

Imperial College London
Department of Electrical and Electronic Engineering
Communication and Signal Processing Group

**From Spline Wavelet to Sampling Theory
on Circulant Graphs and Beyond—
Conceiving Sparsity in Graph Signal Processing**

Madeleine S. Kotzagiannidis

Declaration of Originality

I herewith certify that this thesis, and the research it comprises, are the product of my own work under the guidance of my thesis supervisor Prof. Pier Luigi Dragotti. Any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline. The material of this thesis has not been submitted for any degree at any other academic or professional institution.

Für meine Eltern.

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Abstract

Graph Signal Processing (GSP), as the field concerned with the extension of classical signal processing concepts to the graph domain, is still at the beginning on the path toward providing a generalized theory of signal processing. As such, this thesis aspires to conceive the theory of *sparse representations on graphs* by traversing the cornerstones of wavelet and sampling theory on graphs.

Beginning with the novel topic of *graph spline wavelet theory*, we introduce families of spline and e-spline wavelets, and associated filterbanks on circulant graphs, which leverage an inherent vanishing moment property of circulant graph Laplacian matrices (and their parameterized generalizations), for the reproduction and annihilation of (exponential) polynomial signals. Further, these families are shown to provide a stepping stone to generalized graph wavelet designs with adaptive (annihilation) properties. Circulant graphs, which serve as building blocks, facilitate intuitively equivalent signal processing concepts and operations, such that insights can be leveraged for and extended to more complex scenarios, including arbitrary undirected graphs, time-varying graphs, as well as associated signals with space- and time-variant properties, all the while retaining the focus on inducing sparse representations.

Further, we shift from sparsity-inducing to sparsity-leveraging theory and present a novel *sampling* and *graph coarsening* framework for (wavelet-)sparse graph signals, inspired by Finite Rate of Innovation (FRI) theory and directly building upon (graph) spline wavelet theory. At its core, the introduced Graph-FRI-framework states that any K -sparse signal residing on the vertices of a circulant graph can be sampled and perfectly reconstructed from its dimensionality-reduced graph spectral representation of minimum size $2K$, while the structure of an associated coarsened graph is simultaneously inferred. Extensions to arbitrary graphs can be enforced via suitable approximation schemes.

Eventually, gained insights are unified in a graph-based image approximation framework which further leverages graph partitioning and re-labelling techniques for a maximally sparse graph wavelet representation.

Acknowledgements

I would like to thank my supervisor, Prof. Pier Luigi Dragotti, for his guidance and support throughout my PhD studies. In addition, I would also like to thank the Department of Electrical and Electronic Engineering for providing the scholarship that facilitated this research.

I am grateful to my friends in London and beyond, who made these past four years an enjoyable experience, and at last and most of all, to my mom and dad, for their endless love and support in all of my projects.

Contents

1	Introduction	26
1.1	Motivation and Objectives	26
1.2	Contributions and Outline of Thesis	28
1.3	Publications	30
2	On Graphs and Sparsity: A Brief Review	31
2.1	Sparse Signal Processing	31
2.2	Graph Signal Processing	34
2.2.1	Graph Theory and Linear Algebra	35
2.2.2	The Basics of GSP	37
2.2.3	Wavelets and Sparsity on Graphs	40
2.2.4	Sampling on Graphs	42
2.3	The Class of Circulant Graphs	43
2.3.1	Downsampling and Reconnection on Circulant Graphs	44
3	Wavelets & Filterbanks on Circulant Graphs and Beyond	48
3.1	Motivation and Objectives	48
3.2	Families of Spline Wavelets on Circulant Graphs	49
3.2.1	Vanishing Moments on the Graph	49
3.2.2	The Graph Spline Wavelet	51
3.2.3	The Signless Laplacian	53
3.3	Families of E-Spline Wavelets on Circulant Graphs	54
3.3.1	A Generalized Graph Laplacian Operator	54
3.3.2	Graph E-Spline Wavelets	56
3.3.3	Special Cases and Discussion	57
3.4	Splines on Graphs	60
3.4.1	The Directed Graph Spline	64

3.5	Complementary Graph Wavelets	65
3.5.1	The Bipartite Semi-IIR Graph Filterbank	66
3.5.2	Design and Discussion	67
3.6	Computational Experiments	72
4	Generalized and Adaptive Wavelets on Graphs	78
4.1	Generalized Bandlimiting Graph Wavelet Transforms	79
4.1.1	The GWT in Perspective	79
4.1.2	Generalized Vanishing Moments	82
4.1.3	Generalizations to Random Walk	85
4.2	Space-Variant Graph Wavelets	85
4.3	Time-Variant Graph Wavelets	89
4.3.1	The Cross-Graph Walk	90
4.3.2	Filtering Across Different Graphs	91
4.4	The Condition Number of the GWT	94
4.4.1	Comparison of Different GWTs	96
4.5	Graph Products and Approximations: A Multidimensional Extension	98
4.5.1	Graph Products of Circulants	99
4.5.2	Multi-dimensional Wavelet Analysis on Product Graphs	101
4.5.3	Smoothness and Sparsity on Product Graphs	105
4.6	Computational Experiments	107
4.7	Overview of GWTs	111
5	Sparse Sampling on Graphs	112
5.1	Sampling on Circulant Graphs	113
5.1.1	Related Work	113
5.1.2	Sampling in the Time Domain	113
5.1.3	Wavelets for Sampling	114
5.1.4	The Graph FRI-framework	116
5.1.5	Extensions to Path Graphs	124
5.2	Generalized & Multidimensional Sparse Sampling	125
5.2.1	Exact vs Approximate Graph Product Decomposition	127
5.3	Sampling under Noise: Circulant Graphs with Perturbations	128
5.3.1	Notes on Perturbation Theory	132
5.3.2	Computational Experiments	135

6	Image Processing on Graphs	137
6.1	Graph Wavelets for Non-Linear Image Approximation	138
6.1.1	The General Framework	138
6.1.2	The Matter of the Labelling	141
6.2	Examples	146
6.3	Open Problem: Graph Labelling Under Noise	153
7	Conclusion	156
7.1	Summary	156
7.2	Open Problems and Future Work	158
	Appendices	161
A	Proofs of Chapter 3	161
A.1	Proof of Theorem 3.1	161
A.2	Proof of Theorem 3.2	163
A.3	Proof of Corollary 3.3	168
B	Proofs of Chapter 4	169
B.1	Proof of Corollary 4.3	169
C	Proofs of Chapter 5	175
C.1	Proof of Corollary 5.1	175
C.2	Proof of Theorem 5.1	176
	Bibliography	179

List of Tables

3.1	Continuous e-Spline and Graph e-Spline Definitions in Comparison. ©2017 Elsevier Inc.	63
3.2	Condition Numbers of the GWTs.	74
4.1	Overview of Proposed GWTs.	111

List of Figures

1.1	Sparsity on Graphs: Theory and Applications.	28
2.1	Minnesota Traffic Graph with Graph Signal in the (a) Vertex and (b) Spectral Domain. The color bar in (a) describes the intensity values of the signal on the graph.	39
2.2	Circulant Graphs with generating sets $S = \{1\}$, $S = \{1, 2\}$, and $S = \{1, 3\}$ (from left). ©2017 Elsevier Inc.	44
3.1	Localization of the <i>HGSWT</i> filters for $k = 2$ in the graph vertex domain for circulant G with $S = \{1, 2\}$: shown at vertex $v = 5 \in V$ on G (left), and corresponding graph filter functions at alternate vertices. ©2017 Elsevier Inc.	53
3.2	Classical Discrete Polynomial and Exponential Splines.	61
3.3	The <i>HGSWT</i> filter functions at $k = 1$ for different bipartite circulant graphs, $N = 16$. ©2017 Elsevier Inc.	62
3.4	The <i>HGESWT</i> filter functions ($k = 1$) for different bipartite circulant graphs at $\alpha = \frac{2\pi}{N}$, $N = 16$. ©2017 Elsevier Inc.	63
3.5	Comparison of NLA (lower left) and denoising performance (lower right) for a linear polynomial (top).	73
3.6	Comparison of NLA (lower left) and denoising performance (lower right) for a linear polynomial (top) on a circulant graph with $S = (1, 2, 3, 4)$	74
3.7	Comparison of NLA (lower left) and denoising performance (lower right) for a sinusoidal with ($\alpha = \frac{2\pi 4}{N}$) (top).	75
3.8	Comparison of NLA (lower left) and denoising performance (lower right) for a sum of sinusoidals with ($\alpha_1 = \frac{2\pi 1}{N}$, $\alpha_2 = \frac{2\pi 5}{N}$) (top). ©2017 Elsevier Inc.	76
3.9	Comparison of Graph-filter functions: analysis low-pass (top), synthesis low-and high-pass (from left) at one level for the linear spline constructions on the graph with $S = (1, 2)$	76
3.10	Comparison of Graph-filter functions: analysis low-pass (top), synthesis low-and high-pass (from left) at one level for the linear spline constructions on the graph with $S = (1, 2, 3, 4)$	77

3.11	Comparison of Graph-filter functions: analysis low-pass (top), synthesis low-and high-pass (from left) at one level for the convolved e-spline constructions on the graph with $S = (1, 2)$	77
4.1	Piecewise Smooth Graph Signal on a Circulant Graph.	86
4.2	Illustrative Time-Varying-Graph at $t = 0, 1, 2$	89
4.3	Walk across two circulant graphs.	91
4.4	Graph Cartesian Product of two unweighted circulant graphs. ©2017 Elsevier Inc.	101
4.5	Graph Downsampling and Coarsening of G_\diamond in Fig 4.4 on G_2 w.r.t. $s = 1 \in S_2$ with coarsened \tilde{G}_2 . ©2017 Elsevier Inc.	103
4.6	Non-linear Approximation Performance Comparison for a piecewise smooth signal (top) on a toroidal graph product (left).	108
4.7	Non-linear Approximation Performance Comparison for a noisy piecewise smooth signal (top) on a circulant graph with bandwidth $M = 1, 5, 9$ (from left).	109
4.8	Non-linear Approximation Performance Comparison for a noisy piecewise smooth signal (top) on a circulant graph with bandwidth $M = 1, 5, 9$ (from left).	110
5.1	Traditional Sampling Scheme. ©2017 Elsevier Inc.	113
5.2	Graph Coarsening for a Circulant Graph with $S = \{1, 2, 3\}$. ©2017 Elsevier Inc.	122
5.3	Sampling Scheme with Preceding Sparsification Step and One Level of Coarsening. ©2017 Elsevier Inc.	122
5.4	Graph Cartesian Product of two unweighted path graphs. ©2017 Elsevier Inc.	128
5.5	GFRI-Scheme under Graph Perturbations.	130
5.6	Coefficient Matrix \mathbf{C} under Perturbations (displayed in magnitude, scale-factor 10).	134
5.7	Reconstruction Performance on a Perturbed Simple Cycle ($N = 256$), for 100 randomly generated sparse signals \mathbf{x}_l ($K = 4$, minimum separation of 3 between entries). ©2015 IEEE	135
6.1	Non-Linear Approximation Comparison for the 2D Haar, 2D Linear Spline and proposed Graph Wavelet Transform (GWT) with 5 levels at 5% of non-zero coefficients (f. left).	138
6.2	Random Graph Signal (b) with corresponding Graph Adjacency Matrix (a) before and after ((c) – (d)) applying a simple sort (graph kernel as in Eq. (6.2), with threshold $T = 0.3$).	145

6.3	Original G ($N = 64$) after thresholding of weights (a), after RCM relabelling (b), signal \mathbf{x} before/after relabelling (c), multiscale $HGSWT$ representation (at $k = 1$) of \mathbf{x} (in magnitude) on \tilde{G} for 3 levels (d). ©2017 Elsevier Inc.	146
6.4	Comparison of NLA performance at 5 levels for a 64×64 image patch.	147
6.5	Comparison of NLA performance at 5 levels for a 64×64 image patch.	148
6.6	Localized Basis Functions of the $sparseGWT(bil,RCM)$ depicted on selected (nodes) pixels (b) of the original image patch.	149
6.7	Localized Basis Functions of the $sparseGWT(I,sort)$ depicted on selected (nodes) pixels (b) of the original image patch.	150
6.8	Comparison of NLA performance at 5 levels for a 64×64 image patch.	151
6.9	Comparison of NLA performance at 5 levels for a 64×64 image patch.	151
6.10	(a) Original 64×64 image patch, (b)-(c) Graph Cut Regions for 1 cut & (d) Comparison of NLA performance. ©2017 Elsevier Inc.	152
6.11	(a) Original 64×64 image patch, (b) – (e) Graph Cut Regions for 3 cuts & (f) Comparison of NLA performance.	152
6.12	Denoising performance comparison for a 64×64 image patch.	153
6.13	Denoising performance comparison between BM3D and simple cycle GWT with ideal labelling on the 64×64 cameraman image patch.	155

Notation

Sets and Numbers

\mathbb{N}	Natural Numbers (not including 0)
\mathbb{Z}	Integers
$\mathbb{Z}^{\geq 0}$	Nonnegative Integers
\mathbb{R}	Real Numbers
\mathbb{C}	Complex Numbers
$S = \{s_i\}_i$	Set S with elements s_i
S^c	Complement of set S
\bar{z}, z^*	Complex Conjugate of $z \in \mathbb{C}$
$[a, b]$	Closed Interval: $\{x \in \mathbb{R} : a \leq x \leq b\}$
(a, b)	Open Interval: $\{x \in \mathbb{R} : a < x < b\}$
$[a, b)$	Partially closed Interval: $\{x \in \mathbb{R} : a \leq x < b\}$
$C^\infty([a, b])$	space of infinitely differentiable continuous functions over $[a, b]$
$C^p([a, b])$	space of continuous functions with p continuous derivatives over $[a, b]$
$n \bmod N$	n modulo N
$(i + j)_N$	$i + j$ modulo N
$f(t) \gtrless a$	$f(t) > a$ or $f(t) < a$, for function $f : \mathbb{R} \rightarrow \mathbb{R}$ and $a \in \mathbb{R}$

Vectors and Matrices

$\mathbf{x} \in \mathbb{C}^N$ (or $\vec{x} \in \mathbb{C}^N$)	vector or discrete-time signal of dimension N
$x(i)$ or x_i	i -th entry of vector \mathbf{x}
$\ \mathbf{x}\ _2$	l_2 -norm of vector \mathbf{x} , $\ \mathbf{x}\ _2 = \sqrt{\sum_{i=1}^N x_i ^2}$
$\ \mathbf{x}\ _0$	l_0 -pseudo-norm of vector \mathbf{x} , $\ \mathbf{x}\ _0 = \#\{i : x_i \neq 0\}$
$\mathbf{1}_N$	constant vector of 1's of dimension N
$\mathbf{1}_{\mathcal{S}}$	vector with 1's on set (or interval) \mathcal{S} and 0's otherwise
\mathbf{e}_i	unit vector with $e_i(i) = 1$ and $e_i(j) = 0$, $j \neq i$, of dimension N
δ_{ij}	Kronecker delta (0 if $i \neq j$, 1 if $i = j$)
$\mathbf{A} \in \mathbb{C}^{N \times M}$	matrix of size $N \times M$

$A_{i,j}$ or $A(i,j)$	entry of matrix \mathbf{A} at position (i,j)
\mathbf{A}^T	transpose of matrix \mathbf{A}
\mathbf{A}^H	Hermitian (or conjugate) transpose of matrix \mathbf{A}
\mathbf{A}^{-1}	inverse of matrix \mathbf{A}
$\ \mathbf{A}\ _F$	Frobenius-norm of matrix \mathbf{A} , $\ \mathbf{A}\ _F = \sqrt{\sum_{i=1}^N \sum_{j=1}^N A_{i,j} ^2}$
\mathbf{I}_N	identity matrix of dimension $N \times N$
$\mathbf{0}_{N,N}$	zero matrix of dimension $N \times N$
\circ	Hadamard (entrywise) product
\otimes	Kronecker (or tensor) product

Signal Processing Notation

$x(t)$	continuous-time signal
$f(t) * g(t)$	continuous-time convolution: $\int_{-\infty}^{+\infty} f(\tau)g(t - \tau)d\tau$

Graph Notation

\mathbf{A}	adjacency matrix
\mathbf{D}	degree matrix
\mathbf{L}	graph Laplacian matrix
\mathbf{A}_n	normalized adjacency matrix: $\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$
\mathbf{L}_n	normalized graph Laplacian matrix: $\mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$
d	degree per node
$\tilde{\mathbf{L}}_\alpha$	(exponential) e-graph Laplacian matrix
d_α	e-degree per node

Abbreviations

DFT	Discrete Fourier Transform
FRI	Finite Rate of Innovation
GSP	Graph Signal Processing
GFT	Graph Fourier Transform
GWT	Graph Wavelet Transform
LSI	Linear Shift-Invariant
SP	Signal Processing
wlog	without loss of generality

Proposed Graph Wavelet Transforms

HGSWT	Higher-Order Graph Spline Wavelet Transform	Thm. 3.1 (p. 52)
HGESWT	Higher-Order Graph E-Spline Wavelet Transform	Thm. 3.2 (p. 56)
HCGSWT	Higher-Order Complementary Graph Spline Wavelet Transform	Thm. 3.4 (p. 69)
HCGESWT	Higher-Order Complementary Graph E-Spline Wavelet Transform	Thm. 3.5 (p. 71)
HBGWT	Higher-Order Bandlimiting Graph Wavelet Transform	Cor. 4.1 (p. 82)
HSVGSWT	Higher-Order Space-Variant Graph-Spline Wavelet Transform	Thm. 4.1 (p. 86)

Chapter 1

Introduction

1.1 Motivation and Objectives

Graphs, as high-dimensional (and often sparse) dependency structures, have become an increasingly favorable tool for the representation and processing of large data, as exemplified by i. a. social, transportation and neuronal networks, primarily due to their potential to capture (geometric) complexity.

The appeal of operating with respect to data encapsulated within the higher-dimensional dependency structures of a graph lies not only in the potential for superior data representation and processing for real-world applications, but also emerges in the development of a comprehensive mathematical framework, which seeks to extend conventional signal processing concepts to the graph domain, thus naturally challenging the structural confinement of existing frameworks and posing intriguing new questions.

The resultant field of Graph Signal Processing (GSP) can be characterized as the collective of theoretical and experimental efforts toward a more generalized theory of Signal Processing (SP), which attempts to leverage the complex connectivity of graphs in order to facilitate more sophisticated processing of (high-dimensional) data beyond traditional methods. Nevertheless, the developed underlying theory is still in its infancy and there is a need for a thorough and rigorous theoretical foundation in order to fully comprehend and exploit the capabilities of networks.

A graph consists of a set of vertices and a set of edges, whose associated weights characterize the similarity between the two vertices they respectively connect. In its essence, GSP aspires to use graphs as data representations, whose specific connectivity and edge weights are imposed by or inferred from the problem and/or data at hand, while further instilling the notion of an associated graph signal, which maps the data to a finite sequence of samples such that each sample value corresponds to a vertex in the graph [1].

The inherent challenge of incorporating and interpreting newly arising data dependencies, while maintaining equivalencies to classical cases, has given rise to a variety of different approaches, borrowing notions from i. a. algebraic and spectral graph theory ([2], [3]), algebraic signal processing theory [4], and general matrix theory [5]; here, it has remained an open problem to link structural properties of signals and underlying graphs to properties of graph operators and/or arising transform coefficients [1]. Graph theory and linear algebra in particular create a productive interplay in that any graph can be represented by a matrix and thus be subjected to (and benefit from) purely linear algebraic results, while at the same time, any generic (square) matrix can be interpreted as the connectivity information of a network, facilitating a geometrically richer approach.

Under this theme, the present thesis aspires to conceive *sparsity on graphs* from a theoretical perspective, which i. a. features the problem of identifying an optimal (graph-dependent) basis for the sparse representation of a graph signal. Having evolved from a central issue in the area of classical transform analysis, in light of newly arising data-dependencies that need to be accommodated and a plethora of techniques and interpretations targeted at variable and, usually application-driven, desirable properties, this problem remains largely unanswered from a theoretically rigorous perspective. Further, we delve into the topic of sampling sparse signals on graphs, before eventually tackling the problem of graph-based sparse image approximation. In the course of this conception, we create a bridge from spline wavelet to sampling theory on graphs, which further links to their traditional counterparts in signal processing, on the basis of circulant graphs. Notions of matrix-bandedness and graph-relabelling additionally play a crucial role.

Sparse graph signals, characterized by a small number of non-zero values relative to the dimension of the graph, represent a scarcely studied area of GSP, in particular, it remains unexplored how a sparse representation can be induced via a valid graph operator with respect to the underlying connectivity of the graph, or put differently, what classes of signals can be perfectly annihilated on a graph beyond piecewise constant signals. At the same time, it is of interest to investigate how sparsity on graphs can be leveraged for dimensionality reduction of the signal and coarsening of the associated graph, i. a. for efficient data processing and storage.

The proposed theoretical analysis is primarily conducted in an effort to understand sparse graph signals in the light of the structure and connectivity of graphs and contribute to a more rigorous theoretical foundation of GSP, while simultaneously developing insight for possible applications, such as sparse signal- and image-approximation on graphs.

As will be revealed in subsequent chapters, the inherent polynomial and LSI (Linear Shift-Invariance) quality of circulants gives rise to a range of interesting mathematical properties which can be leveraged for drawing (intuitive) connections and extensions from SP to GSP (see Fig. 1.1). In particular, graph operators defined on circulant graphs can be diagonal-

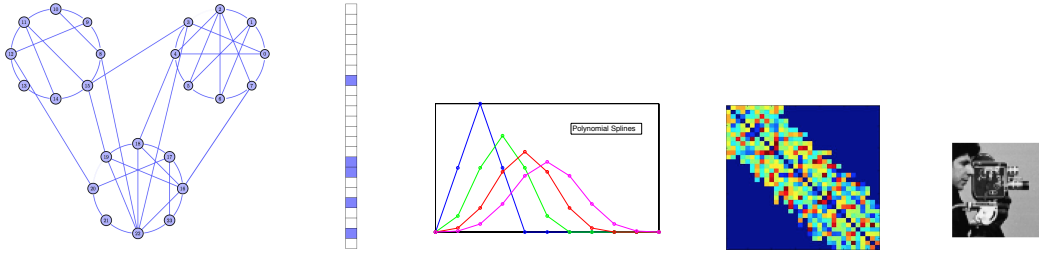


Figure 1.1: Sparsity on Graphs: Theory and Applications.

ized by the DFT-matrix, which is central to traditional SP, while being well characterized in the vertex as well as in the spectral graph domain as a result of the regularity of the graph. Fundamental mathematical properties of the circulant graph Laplacian matrix are detected and incorporated into novel generalized graph differencing operators, which give rise to basis functions that are structurally similar to the classical discrete (e-)splines, thereby inspiring the creation of multilevel wavelet transforms on circulant graphs. The inherent Fourier characterisation in the spectral graph domain is further leveraged in the problem of sampling sparse signals on circulant graphs. Inspired by classical Finite Rate of Innovation theory, the derived framework facilitates the perfect recovery of sparse signals from a dimensionality reduced spectral representation while simultaneously identifying an associated coarsened graph; this not only extends the classical framework to a broader class of signals defined on complex structures beyond the real line, but further characterises the sampled signal through a reduced graph which preserves essential properties of the original.

1.2 Contributions and Outline of Thesis

In the following, we present an outline of the thesis with its major contributions.

Chapter 2 provides an overview of the general landscape of *graph signal processing*, commencing with a brief discussion of the most notable works and contributions on traditional sparse signal processing, followed by basic definitions and concepts from linear algebra and spectral graph theory, which we will be making use of throughout. Subsequently, we proceed to state the general problem context of GSP and underlying notions, along with an overview of recent contributions to wavelet and sampling theory on graphs, and concluding with a specialized review of *circulant graphs*.

In Chapter 3, we introduce *graph spline wavelet theory* on circulant graphs and present a range of properties and results that derive from the discovered vanishing moment property of the circulant graph Laplacian, while also drawing connections to traditional spline the-

ory. We develop novel families of wavelets and associated filterbanks for the analysis and representation of functions defined on circulant graphs, where we distinguish between the vertex domain-based graph spline and e-spline wavelet transforms, and the complementary graph (e-)spline wavelet transforms (derived via spectral factorization), and discuss their properties and special cases. As such, the theory developed in Ch. 3 serves as the foundation for subsequent chapters and their contributions.

Chapter 4 seizes the main insights and derived properties of Chapter 3 and appropriates them for *generalized wavelet design* on arbitrary undirected graphs targeted at the broader class of (piecewise) smooth graph signals. The chapter begins with a review and analysis of a general graph wavelet transform, elucidating how the theory of Ch. 3 constitutes a special case for circulant graphs, and evolves into the derivation of further graph wavelet transforms for undirected graphs with distinct annihilation properties. These include the generalized bandlimiting, space-variant and time-variant graph wavelet transforms as well as the multi-dimensional graph wavelet transform defined on product graphs. Additional analysis of the condition number and sparsifying-level of derived constructions is conducted.

Chapter 5 tackles the problem of *sparse sampling on graphs* by introducing the novel GFRI-framework on circulant graphs and beyond, inspired by traditional FRI theory and directly leveraging the developed graph spline wavelet theory. In particular, the perfect recovery of (wavelet-)sparse signals from a dimensionality-reduced representation is established, while an associated coarsened graph can be identified; properties and special cases are discussed, as well as extensions to multi-dimensional sampling, while generalizations to arbitrary graphs are enforced via suitable approximation schemes. At last, an alternative approach to the latter is explored in form of (noisy) sampling on circulant graphs with perturbations.

Eventually, Chapter 6 presents a novel *graph-based image processing* framework which employs image (graph) segmentation followed by a variable, sparsity-driven graph wavelet analysis step for images featuring distinct discontinuities or patterns. More precisely, circulant graph wavelets with variable localization properties are applied on approximations of the partitioned subgraphs, which represent image regions of homogeneous intensity content. Further relevant performance-enhancing concepts such as the optimal graph labelling are discussed, while the superiority of the method compared to traditional tensor product wavelet bases is illustrated on the basis of real and artificial image patches. In essence, this final chapter is designed to unify and implement certain gained notions on the theory of GSP within the concrete application of image processing, thereby directly leveraging the main strength of GSP to capture, and operate with respect to, the inherent geometry of given (image) data.

1.3 Publications

The following papers in press ([6],[7]) and publications ([8], [9], [10], [11]) were produced over the course of the PhD studies, and are incorporated throughout this thesis, which is accordingly annotated in footnotes.

Journal Papers (in press)

1. M. S. Kotzagiannidis and P. L. Dragotti, "Sampling and Reconstruction of Sparse Signals on Circulant Graphs - An Introduction to Graph-FRI," *Appl. Comput. Harmon. Anal.* (2017), <https://doi.org/10.1016/j.acha.2017.10.003>, *available on arXiv: <http://arxiv.org/abs/1606.08085>*.
2. M. S. Kotzagiannidis and P. L. Dragotti, "Splines and Wavelets on Circulant Graphs," *Appl. Comput. Harmon. Anal.* (2017), <https://doi.org/10.1016/j.acha.2017.10.002>, *available on arXiv: <http://arxiv.org/abs/1603.04917>*.

Conference Papers

3. M. S. Kotzagiannidis and P. L. Dragotti, "The Graph FRI framework-Spline wavelet theory and sampling on circulant graphs," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, March 2016, pp. 6375–6379.
4. M. S. Kotzagiannidis and P. L. Dragotti, "Higher-order graph wavelets and sparsity on circulant graphs," in *SPIE Optical Engineering+ Applications. Wavelets and Sparsity XVI*, vol. 9597. San Diego, USA: International Society for Optics and Photonics, 2015, pp. 95 971E–95 971E-9.
5. M. S. Kotzagiannidis and P. L. Dragotti, "Sparse graph signal reconstruction on circulant graphs with perturbations," in *10th IMA Conference on Mathematics in Signal Processing, Birmingham, UK*, 2014.
6. M. S. Kotzagiannidis and P. L. Dragotti, "Sparse graph signal reconstruction and image processing on circulant graphs," in *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2014, pp. 923–927.

Chapter 2

On Graphs and Sparsity: A Brief Review

2.1 Sparse Signal Processing

The search for a signal representation that is efficient, and hence *sparse*, has been largely thematised in areas such as signal processing and computational harmonic analysis, with recent efforts merging and going beyond established notions. It is not the purpose of this thesis to provide a comprehensive review or detailed analysis of classical definitions and methods in the first place, except in order to specifically motivate a need for and/or draw comparisons to graph-based theory and methods, and as such, only its rough outlines will be retraced, whilst referring to more elaborate review works.

A vector $\mathbf{f} \in \mathbb{R}^N$ is described as sparse if its l_0 -pseudo-norm $\|\mathbf{f}\|_0 = \#\{i : f_i \neq 0\}$ is small relative to the vector-dimension N .

(Sparse) signal approximation in a given basis $\{\mathbf{u}_n\}_{n=0}^{N-1}$ is thus essentially conducted by projecting a signal (or function) $\mathbf{x} \in \mathbb{R}^N$ onto a small number $K \ll N$ of suitable basis elements

$$\mathbf{x} \approx \sum_{n \in I_K(\mathbf{x})} c_n \mathbf{u}_n.$$

The coefficients $c_n = \tilde{\mathbf{u}}_n^T \mathbf{x}$ are obtained via the analysis with a (bi-)orthogonal basis $\{\tilde{\mathbf{u}}_n\}_{n=0}^{N-1}$ and $I_K(\mathbf{x})$ denotes the set of K indices. In the case of linear approximations, the latter are selected to correspond to fixed subspaces of lowest possible dimension, while non-linear approximations consider the best (signal-dependent) spaces, such as the atoms corresponding to the highest magnitude coefficients for the best K -term approximation, which take advantage of and capture i. a. some form of regularity of the signal as well as occurring localized discontinuities [12]. The latter, non-linear approach, in particular,

signifies an increase in sparsity with more flexibility to achieve the best approximation.

Notably, the discrete-time Fourier transform, as the basis of linear time-invariant operators with orthogonal atoms $\{u_n(x) = e^{inx}\}_{n \in \mathbb{Z}}$, emerged as one of the earliest signal transforms and gained popularity due to its particular efficiency in characterizing globally smooth signals for i.a. denoising purposes, as an instance of linear approximation [13]. The Discrete Fourier Transform (DFT) nevertheless falls short of adequately (sparsely) describing signals with arising discontinuities due to a lack of localization, and invites the approaches of bases with compact localized support, beginning with the Short Time Fourier Transform (STFT) or Gabor transform, which consists of windowed waveforms that can be applied locally, and culminating in the construction of multiscale wavelet transform bases, from the translations and dilations of localized low-frequency scaling functions and high-frequency mother wavelets, respectively providing a coarse approximation and detail preservation of the signal [13]. The class of *spline-wavelets*, noted for their symmetry, regularity as well as localization and approximation properties for (piecewise) smooth functions, and characterized by synthesis functions which are higher-order polynomial splines, with variable design choices, including i.a. compact support (B-spline and biorthogonal spline wavelets), in particular emerged as superior [14].

In an effort to refine performance, wavelets have been further developed to incorporate adaptivity to more complex signal properties, such as orientation or translation-invariance at the sacrifice of critical sampling, in form of i.a. steerable [15] and stationary wavelet transforms [16]. Eventually, the sparse signal representation problem evolved into the task of selecting elements from a suitable dictionary, with a plethora of directions for the design and/or learning of overcomplete dictionaries, which may be analytic on the basis of mathematical functions, such as curvelets [17], contourlets [18] or bandelets [19], as well as directly extracted from data based on iterative training algorithms, such as the K-SVD algorithm [13]. Thereby, the former have i.a. tackled the problem of efficiently characterizing higher-dimensional signals, such as images, which have one-dimensional, as opposed to zero-dimensional, smooth discontinuities that cannot be absorbed by traditional wavelet bases [20], via constructions that are more intricate than their predecessors; in particular, Chapter 6 will revisit this aspect through the conceptually simpler lens of graph theory.

Moving from a domain which primarily targeted the creation of sparsity to another which specifically leverages it, we further consider sampling theory, as the bridge between discrete- and continuous-time signal processing, which has experienced a mathematical interplay with wavelet theory, and, specifically, the more recent advent of *sparse sampling* [21]. Sampling traditionally signifies the discretization of a continuous-time signal to a sequence of samples whose subsequent interpolation in turn generates a continuous signal. More generally, it denotes a dimensionality reduction in the discrete domain, such as that of a finite vector, followed by reconstruction or interpolation (a dimensionality increase),

and presents a relevant topic in the general investigation of sparsity on the real line, grids and ultimately graphs.

Traditionally, sampling theory and methods focus on the perfect reconstruction of a continuous-time signal $x(t)$, which is typically filtered by a selected kernel $\varphi(t)$ before being uniformly sampled with a sampling period T , from its samples $y_n = \langle x(t), \varphi(t/T - n) \rangle$. Its foundation is laid by the Nyquist-Shannon sampling theorem for bandlimited signals of the form

$$x(t) = \sum_{k \in \mathbb{Z}} x_k \text{sinc}(Bt - k)$$

with samples $x_k = \langle B \text{sinc}(Bt - k), x(t) \rangle$ and band $[-B/2, B/2]$, which states a sufficient condition for their perfect reconstruction at a sampling rate that is at least twice the maximum frequency of $B/2$ [21]. Yet, it has been established that bandlimitedness is only a sufficient rather than necessary condition for perfect reconstruction.

Within a more comprehensive sampling framework, certain classes of signals have been identified, beyond the bandlimited or those confined to fixed subspaces, which possess a parametric representation with a finite number of degrees of freedom per unit of time, or Finite Rate of Innovation (FRI), of the form

$$x(t) = \sum_{k' \in \mathbb{Z}} \sum_{k=1}^K x_{k',k} g_k(t - t_{k'})$$

for known $\{g_k(t)\}_{k=1}^K$, and free coefficients $x_{k',k}$ and time-shifts $t_{k'}$, and which can be accordingly sampled and perfectly reconstructed, based on a spectral estimation scheme, known as the annihilating filter or Prony's method [22], [23], [24]. One may choose from a range of different sampling kernels that satisfy the so-called Strang Fix conditions and their generalizations [25], i.e. a linear combination of their shifted versions can reproduce polynomials or exponentials. The rate of innovation is then established as $\rho = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} C_x(-\frac{\tau}{2}, \frac{\tau}{2})$, with function $C_x(t_a, t_b)$ counting the number of free parameters over an interval $[t_a, t_b]$ [22].

In particular, the sparsity of a signal consisting of K Diracs $x(t) = \sum_{k=1}^K a_k \delta(t - t_k)$, $t \in \mathbb{R}$ with $t_k \in [0, \tau)$, is encapsulated in the parameter pairs (or innovations) $\{a_k, t_k\}_{k=1}^K$, which completely determine the sampling rate of $\rho = 2K/\tau$ and signal, as for distinct t_k , one can retrieve $x(t)$ from $2K$ consecutive values of its transformed samples $\tau_m = \sum_n c_{m,n} y_n$, for a suitable choice of coefficients $c_{m,n}$ [22]. For instance, consider the samples $y_n = \langle x(t), \varphi(\frac{t}{T} - n) \rangle$, $n = 0, \dots, N-1$, obtained through filtering with a kernel $\varphi(t)$ of compact support that reproduces exponentials; that is $\varphi(t)$ satisfies $\sum_{n \in \mathbb{Z}} c_{m,n} \varphi(t - n) = e^{\alpha_m t}$ for a proper choice of coefficients $c_{m,n}$ $m = 0, \dots, P$, and $\alpha_m \in \mathbb{C}$ ([26], [25]). Then the moments take the form $\tau_m = \sum_{k=1}^K a_k e^{\alpha_m \frac{t_k}{T}}$, which corresponds to the Fourier transformation for $\alpha_m = -2\pi i m/N$.

As will be enlarged upon in Chapter 5, the discrete matrix-based nature of the reconstruction approach, paired with a direct link to circularity in matrices (and hence graphs), facilitates its appropriation for and extension to the graph setting.

At last, due to our focus on sparse signals on graphs, a comparison with compressive sensing (CS) [27] is imperative. According to CS theory, a sparse signal $\mathbf{x} \in \mathbb{R}^N$ can be recovered with high probability from the dimensionality-reduced (sampled) signal $\mathbf{y} = \mathbf{A}\mathbf{x}$ under suitable conditions on the rectangular sampling operator $\mathbf{A} \in \mathbb{R}^{M \times N}$ with $M \ll N$ and sparsity $K = \|\mathbf{x}\|_0$, by solving an l_1 -minimization problem, or alternatively, using greedy reconstruction algorithms [28]. While, in contrast to compressive sensing approaches [29], the recovery of the sparse vector \mathbf{x} in the previous scheme is exact at the critical dimension of $2K$ measurements and based on a direct, spectral estimation technique, known as Prony's method ([30], [22]), it should be noted that neither requires knowledge of the locations of the non-zero entries. Further, CS theory can be extended to the recovery of non-sparse signals $\mathbf{x} = \mathbf{D}\mathbf{c}$ that have a sparse representation \mathbf{c} in properly designed, overcomplete dictionaries \mathbf{D} [31], which has also been addressed in the context of graphs by training a graph-based dictionary [32]. The sampling framework proposed in Chapter 5 envisions a similar approach in that smooth (wavelet-sparse) graph signals \mathbf{x} are filtered with a (circulant) multilevel graph wavelet transform in order to produce sparse signals \mathbf{c} which can subsequently be sampled; nevertheless, the recovery of \mathbf{x} from \mathbf{c} ultimately follows from the invertibility of the wavelet transform.

Contrary to the more recent learning- or optimization-driven approaches, this work primarily focuses on signal representations on graphs that are exactly sparse, following the derivation of suitable annihilating operators and wavelet transforms. Thereby, the notion of sparsity on graphs, and, in particular, how to induce a sparse representation with respect to the graph connectivity, is conceived and developed on a fundamental level.

2.2 Graph Signal Processing

Motivated by the need for efficient and sophisticated data processing and representation, in light of the surge of available information in applications such as social, transportation or biological networks, as well as by the promise of developing a universal mathematical framework that goes beyond conventional signal processing, the field of Graph Signal Processing (GSP) emerged from a wide range of contributions, both novel and established. Some of the key challenges of this field comprise the identification and/or construction of the graph which captures the inherent geometry of a given data set (if not otherwise imposed by the application), the development suitable graph transforms which operate with respect to the graph structure as well as the application of powerful intuitions and techniques from traditional signal processing while simultaneously accommodating newly

arising data dependencies in the irregular graph domain.

At its core, GSP unifies basic concepts from algebraic and spectral graph theory with (computational) harmonic analysis [1], while linear algebra and convex optimization have gained an increasingly sustaining role. Spectral graph theory in particular has been instrumental in extending mathematical concepts and intuitions from Fourier analysis to the graph domain, thereby introducing the notion of graph frequency spectra and graph Fourier transform bases. With the aim of establishing comparable SP properties, operations and concepts in the graph domain, a breadth of intriguing GSP problems ranging from simple filtering operations up to more sophisticated constructions of graph wavelet filterbanks ([33], [34], [35], [36], [37]), graph signal interpolation and recovery ([38], [39], [40]), as well as applications encompassing graph-based image processing ([35], [41]), and semi-supervised learning ([42], [43]), have been derived in the wake of two elementary model assumptions for the central graph operator: the (positive semi-definite) graph Laplacian matrix, and the more generalized graph adjacency matrix. Whereas graph Laplacian-based approaches are focused on undirected graphs and leverage the associated convenient properties of positive semi-definite matrices for spectral graph analysis, alternative avenues have featured both undirected and directed graph scenarios by resorting to the Jordan normal form of the adjacency matrix.

Apart from their suitability for clustering or filtering operations [1], graphs have further been employed for the dimensionality reduction of high-dimensional data sets to allow for localized operations on fewer graph vertices; diffusion maps in particular have been developed as a manifold learning technique with a random walk interpretation [44].

2.2.1 Graph Theory and Linear Algebra

A graph $G = (V, E)$ is characterized by a set of vertices $V = \{0, 1, \dots, N-1\}$ of cardinality $|V| = N$, and a set of edges $E = \{e_0, \dots, e_{M-1}\}$. Its underlying connectivity is captured in an adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ with entries

$$A_{i,j} = \begin{cases} w_{i,j} > 0, & \text{if nodes } i \text{ and } j \text{ are connected by an edge, } (i \neq j) \\ 0, & \text{otherwise} \end{cases}$$

for some non-zero weight $w_{i,j}$ and degree matrix \mathbf{D}

$$D_{i,j} = \begin{cases} d_i, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}$$

with degree $d_i = \sum_j A_{i,j}$ per node i , which give rise to the non-normalized graph Laplacian matrix

$$\mathbf{L} = \mathbf{D} - \mathbf{A}.$$

As a fundamental graph matrix within both algebraic and spectral graph theory, the graph Laplacian has been subject to extensive investigation, with a number of results relating to its spectra [3], among others, and as such provides a key operator in the interpretation as well as implementation of classical signal processing concepts in the graph domain. Notably, when the graph is constructed via a kernel from a point cloud, \mathbf{L} can be interpreted as a second-order differential operator which, under certain conditions, converges to the Laplace-Beltrami differential operator on the underlying manifold [45]. The oriented (vertex-edge) incidence matrix \mathbf{S} of an unweighted G describes the $N \times M$ -matrix whose rows and columns are indexed by V and E respectively, i.e. its (i, j) -th entry is 1 (or -1) if edge e_j originates (or terminates) at i ; specifically, each edge of G is assigned an orientation arbitrarily [46]. It is further related to the graph Laplacian via $\mathbf{L} = \mathbf{S}\mathbf{S}^T$, both of which are of the same rank $N - k$ for k connected components, while for undirected graphs, its unoriented version with $(0, 1)$ -weights also exists [46]. For weighted graphs, equivalently, the incidence matrices are weighted and preceding statements continue to hold. The operation of \mathbf{S}^T on a real-valued function $f : V \rightarrow \mathbb{R}$ with value $(\mathbf{S}^T f)(e)$ takes the difference of f at the end-points of edge e in the manner of a discrete first-order differential operator [47]. The relation between \mathbf{S}^T and \mathbf{L} , and their differential counterparts, can be linked through a discrete version of Greens's formula [47].

A graph whose vertices all have the same degree d is regular (or d -regular) which translates to its graph adjacency and Laplacian matrix sharing the same eigenbasis \mathbf{V} , with $\mathbf{A} = \mathbf{V}\mathbf{\Gamma}\mathbf{V}^H$ and $\mathbf{L} = d\mathbf{I}_N - \mathbf{A} = \mathbf{V}(d\mathbf{I}_N - \mathbf{\Gamma})\mathbf{V}^H$. For irregular graphs, the following relation can be established between the adjacency spectrum $\{\gamma_i\}_i$ and the maximum degree per node: $\bar{d} \leq \gamma_{\max} \leq d_{\max}$, with \bar{d} denoting the average degree per node and equality $\gamma_{\max} = d$ for the d -regular case [48]. In graph theory, it has been of interest to characterize special graph classes, such as paths, cycles or trees [2], which also extends to the probabilistic setting; the Erdős-Rényi model, for instance, describes a popular process to generate random graphs $G(n, p)$, from a fixed number n of vertices and probability p for the existence of an edge [49]. A relevant class is that of bipartite graphs, which are described by a vertex set $V = X \cup Y$ consisting of two disjoint sets X and Y , such that no two vertices within the same set are adjacent. Most prominently, the property that the eigenvalues of a bipartite graph adjacency matrix are symmetric with respect to zero [50], termed as the *spectral folding* phenomenon, has motivated various GSP contributions, including perfect reconstruction graph filterbanks, sampling and approximation schemes ([35], [36]).

Further of interest is the symmetric normalized graph Laplacian $\mathbf{L}_n := \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$ with eigenvalues $\tilde{\lambda} \in [0, 2)$, where $\tilde{\lambda}_{\max} = 2$ if and only if the graph is bipartite. The random-walk matrix $\mathbf{A}^{RW} := \mathbf{D}^{-1}\mathbf{A}$ denotes another well-known graph matrix, whose entries $A_{i,j}^{RW}$ represent the transition probability of going from vertex i to j on G , as one step of a Markov random walk.

In an effort to coin a broader class of graph matrices [47], the *generalized graph Laplacian* (or discrete Schrödinger operator) of G defines a symmetric matrix \mathbf{M} with entries

$$M_{i,j} = \begin{cases} l_{i,j}, & \text{if nodes } i \text{ and } j \text{ are connected by an edge, } (i \neq j) \\ p_{i,i} + l_{i,i}, & (i = j) \\ 0, & \text{otherwise} \end{cases}$$

for weights $l_{i,j}$ and potential $p_{i,i}$; alternatively, it is given by $\mathbf{M} = \mathbf{L} + \mathbf{P}$ for graph Laplacian \mathbf{L} and arbitrary diagonal matrix \mathbf{P} .

Despite the fact that graph theory is generally concerned with the study of (pairwise) relations between objects and arising structures, it provides a substantial interplay with linear algebra in that any graph can be represented as a matrix and thus be subjected to (and benefit from) purely linear algebraic results, while at the same time, any generic (square) matrix can be interpreted as the connectivity information of a network, facilitating a geometrically richer approach.¹ Many insights in algebraic graph theory, encompassing i.a. spectral graph theory, as the field of study focusing on the graph eigenvalues, hail from matrix theory, with the Perron-Frobenius Theorem for (symmetric) nonnegative irreducible matrices [47] playing a particularly crucial role. As will become evident, fundamental results in this thesis employ the linear algebra perspective, notably when dealing with graph operators and transforms, nevertheless, specialized notions such as graph cuts and labelling are leveraged, all the while drawing connections to basic signal processing concepts.

2.2.2 The Basics of GSP

For the ensuing discussion, we mainly consider graphs, which are undirected, connected, (un-)weighted, and do not contain any self-loops. A graph signal \mathbf{x} is traditionally a real-valued scalar function defined on the vertices of a graph G of dimension N , with sample value $x(i)$ at node i , and can be represented as the vector $\mathbf{x} \in \mathbb{R}^N$ [1]; in this work, we extend this definition to include complex-valued graph signals $\mathbf{x} \in \mathbb{C}^N$, for illustration purposes, while maintaining real weights between connections on G . Time-periodic signals have been commonly mapped to circular graphs, due to their sequential structure, but generally, signals can be represented on more complex graph structures, where edges

¹This assumes the admission of negative weights and self-loops.

are not only used to invoke a sense of sequencing, but also i.a. similarity between sample values.

When G is undirected and connected, the graph Laplacian \mathbf{L} is a *positive semi-definite* matrix and has a complete set of orthonormal eigenvectors $\{\mathbf{u}_l\}_{l=0}^{N-1}$, with corresponding nonnegative eigenvalues $0 = \lambda_0 < \lambda_1 \leq \dots \leq \lambda_{N-1}$, constituting a convenient property that has facilitated key definitions and generalizing steps from traditional SP toward the field of GSP.

In particular, while the classical Fourier transform characterizes the expansion of a function $f \in L^2(\mathbb{R})$

$$\hat{f}(\xi) := \langle f, e^{2\pi i \xi t} \rangle = \int_{\mathbb{R}} f(t) e^{-2\pi i \xi t} dt \quad (2.1)$$

in terms of the eigenfunctions of the 1-D Laplace operator $-\Delta(\cdot) = -\frac{\partial^2}{\partial t^2}$, the definition of an equivalent Graph Fourier Transform (GFT) $\hat{\mathbf{f}}$ of a (vectorized) function \mathbf{f} residing on the vertices of G , entails its representation in terms of the graph Laplacian eigenbasis $\mathbf{U} = [\mathbf{u}_0 | \dots | \mathbf{u}_{N-1}]$ such that $\hat{\mathbf{f}} = \mathbf{U}^H \mathbf{f}$, where H denotes the Hermitian transpose, extending the concept of the Fourier transform to the graph domain [1]. Its inverse is given by $f(i) = \sum_{l=0}^{N-1} \hat{f}(\lambda_l) u_l(i)$, with expansion coefficients

$$\hat{f}(\lambda_l) := \langle \mathbf{f}, \mathbf{u}_l \rangle = \sum_{i=0}^{N-1} f(i) u_l^*(i). \quad (2.2)$$

Emanating from the classical eigendecomposition of the Laplacian operator, the graph Laplacian spectrum of a connected graph carries a notion of frequency in that small eigenvalues correspond to eigenvectors that vary smoothly across the graph (thus directly reflecting its connectivity), while larger eigenvalues are associated with rapidly oscillating eigenvectors, justifying its ordering [1]. Here, the eigenvalue $\lambda_0 = 0$ is associated with the all-constant $\mathbf{u}_0 = \frac{1}{\sqrt{N}}$ and of multiplicity $m_0 = 1$, signifying the number of its connected components; alternatively, this is indicated by the so-called Fiedler value (or algebraic connectivity) of the form $\lambda_1 > 0$ [51]. Nevertheless, the instance of a complete graph, whose spectrum consists of only two values, presents a special case where large multiplicities can interfere with this conception of order and a more sophisticated theory is needed. Fig. 2.1 depicts a sample piecewise-constant graph signal on the unweighted Minnesota graph (from [52]), which has served as a common model for GSP contributions, both in the vertex and spectral domain.

A graph (wavelet) filter \mathbf{H} in the vertex domain generally describes a linear transform which takes weighted averages (differences) of components of the input signal \mathbf{x} at a vertex i within its k -hop local neighborhood $N(i, k)$, given by $\tilde{x}(i) = H_{i,i}x(i) + \sum_{j \in N(i,k)} H_{i,j}x(j)$

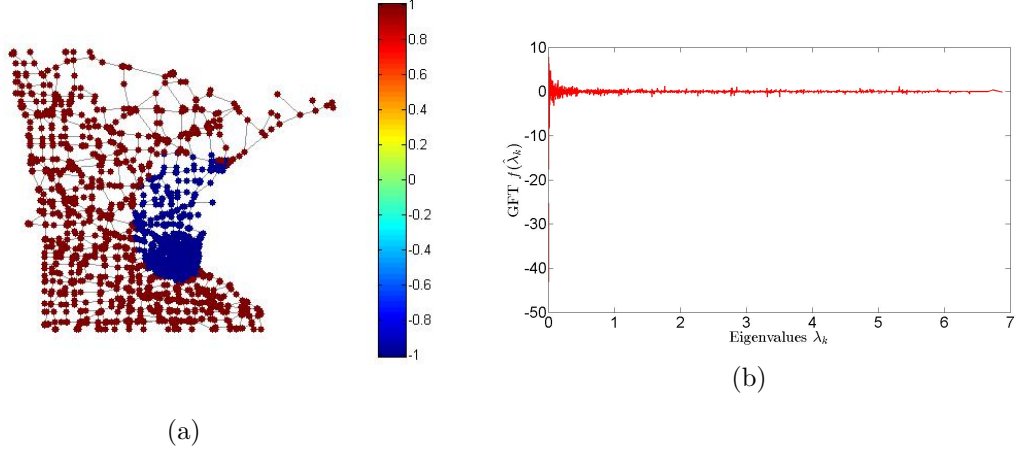


Figure 2.1: Minnesota Traffic Graph with Graph Signal in the (a) Vertex and (b) Spectral Domain. The color bar in (a) describes the intensity values of the signal on the graph.

for some coefficients $\{H_{i,j}\}_{i,j \in V}$; where applicable, it is expressed as a polynomial in the graph Laplacian matrix $\mathbf{H} = h(\mathbf{L}) = \sum_{k=0}^{N-1} h_k \mathbf{L}^k$ for suitable h_k (or alternatively, as a polynomial in the adjacency matrix) [1]. The polynomial form is particularly favored for spectral domain design techniques since it simultaneously gives rise to an interpretation in the vertex domain.

Graph spectral filtering of a signal $\mathbf{x} \in \mathbb{R}^N$ can be represented as $\tilde{\mathbf{x}} = h(\mathbf{L})\mathbf{x}$ with

$$\tilde{x}(i) = \sum_{l=0}^{N-1} \hat{x}(\lambda_l) h(\lambda_l) u_l(i) \quad (2.3)$$

or equivalently through the GFT $\hat{\tilde{\mathbf{x}}} = h(\mathbf{\Lambda})\hat{\mathbf{x}}$ with $\hat{\tilde{x}}(\lambda_l) = \hat{x}(\lambda_l)h(\lambda_l)$, where frequency coefficients $h(\lambda_l)$ are selected to attenuate or amplify certain frequency contributions. Hence, when the spectral graph filter is given by a polynomial $h(\lambda_l) = \sum_{k=0}^K h_k \lambda_l^k$, it is localized with respect to the vertex domain with

$$\tilde{x}(i) = \sum_{j=0}^{N-1} x(j) \sum_{k=0}^K h_k (\mathbf{L}^k)_{i,j}, \quad (2.4)$$

whereby the coefficients of the previous vertex-based transform can now be expressed as $H_{i,j} = \sum_{k=d_G(i,j)}^K h_k (\mathbf{L}^k)_{i,j}$, with $(\mathbf{L}^k)_{i,j} = 0$ when the shortest-path distance $d_G(i,j)$ between vertices i, j (minimum number of hops) is greater than k [37]. More specifically, a graph transform is said to be strictly k -hop localized in the spatial domain of the graph if the filter coefficients $H_{i,j}$ are zero beyond the k -hop neighborhood of each node i . Other classical signal processing notions such as convolution, translation or modulation can be similarly (directly or indirectly) generalized to the graph setting on the basis of the graph Laplacian eigenvectors [1].

Another line of work has sought to broaden the discussion to directed graphs by focusing on the properties of the graph adjacency matrix instead, deriving fundamental notions on frequency and sampling on the basis of its Jordan decomposition ([53], [54]).

Since the discussion in this thesis is largely focused on undirected graphs, most of the results can be easily adapted for different (properly normalized) graph matrices, however, for consistency, the graph Laplacian-based frequency (GFT) interpretation is adopted here.

2.2.3 Wavelets and Sparsity on Graphs

The notion of wavelets on graphs presents a promising avenue for the sophisticated analysis of complex data, which may be captured in form of a graph and underlying graph signal, beyond classical wavelet theory, due to the potential to operate with respect to the inherent geometry of the data in a more localised manner.

A range of designs have been proposed, notably including the diffusion wavelet [33], the biorthogonal and perfect reconstruction filterbank on bipartite graphs ([35], [36]), and the spectral graph wavelet [37], tailored to satisfy a set (or subset) of properties, which have evolved from the traditional domain, such as localization in the vertex or spectral graph domain, critical sampling and invertibility, along with notions of *graph-specific downsampling* and *graph-coarsening* for a multiscale representation, as well as to facilitate generalizations to arbitrary graphs, for applications including image processing [1] and wavelet-regularized semi-supervised learning [55].

More specifically, the diffusion wavelets by Coifman and Maggioni [33] are orthogonalized basis functions based on compressed representations of (powers of) a diffusion operator, within a framework that is applicable to both graphs and smooth manifolds. The spectral graph wavelets by Hammond et al. [37] describe a class of wavelet operators which are constructed in the graph spectral domain of the graph Laplacian via dilations and translations of a bandpass kernel. Further, the graph wavelet transform by Crovella and Kolaczyk (CKWT) [56] constitutes a multiscale design in the vertex domain (of unweighted graphs) based on the shortest-path distance; here, each wavelet is constant across vertices within a certain hop-neighborhood from the given center vertex.

A drawback of the aforementioned designs is their overcompleteness. This is remedied by i.a. lifting- [57] and tree-based [58] designs, as well as the graph wavelet filterbanks on bipartite graphs by Ortega et al. ([35], [36]). The latter in particular rely on convenient properties of bipartite graphs which facilitate intuitive downsampling operations (by simply retaining either disjoint set) as well as a targeted design of spectral filters $h(\lambda)$ for i.a. graph localization and compact support, following similar conditions as regular domain (circulant) z -transformed filters, and can be generalized to arbitrary graphs

through a bipartite subgraph decomposition problem. In order to ensure that transforms are localized in the graph vertex domain, and can thus be expressed as polynomials of the graph Laplacian as well as provide efficient computation, some works employ a Chebyshev approximation for designated kernels, incurring a small reconstruction error ([37], [35]). Further, the spline-like graph wavelet filterbanks on circulant graphs in [59], [34], [55], which are further detailed in the next section, describe vertex-localized transforms which similarly leverage mathematical properties of special graph classes for multiscale processing. For both the bipartite and circulant graph filterbanks, and contrary to traditional signal processing notions, the low-and high-pass filtered content is retained by complementary sets of nodes.

While not necessarily of relevance to all of the above transforms, many of which are redundant, the problem of down-and upsampling a signal on the vertices of a graph is central to GSP, as the evolution of a key component of discrete multilevel signal transformations and filterbanks, and poses a particular challenge due to the complex connectivity of graphs. Along with it arises the problem of identifying a *coarsened* or *reduced graph* for subsequent multilevel operations, requiring a method of assigning suitable edges to a reduced set of vertices, while ensuring that certain properties of the original graph, among other desirable or essential ones, are preserved. Unless the graph at hand is highly structured or special, it remains unclear how to consistently extract a downsampling pattern and solution approaches vary; for instance, in [60], it is proposed to select vertices based on the polarity of the components of the largest graph Laplacian eigenvector. Overall, it is usually desirable to retain a representative half of the vertex set with connections between nodes in the retained as well as in the removed set being of relatively low weight [60]. Moreover, the coarsened graph should preserve essential graph connectivity properties of the original such as structure or sparsity, while being representative of the latter in both the vertex and spectral domain. These questions, among further GSP notions, are reviewed in i.a. [1], [60], while links between the problem of graph coarsening and the more established (approximate) *graph coloring* [61], which describes the search for a partition into vertex subsets of distinct color such that no two adjacent vertices share the same color, as well as to the dual *spectral clustering* [62], as the task of partitioning a (similarity) graph into clusters based on the graph Laplacian eigenvectors, have been discovered.

The topic of sparsity on graphs via wavelet analysis appears as a natural extension to its foundation in the discrete-time domain, and some works have opened its discussion through topics such as the wavelet coefficient decay at small scales of graph-regular signals [63] via the spectral graph wavelet transform, the tight wavelet frame transform on graphs [64], as well as the overcomplete Laplacian pyramid transform with a spline-like interpolation step [60]. Nevertheless, such approaches so far lacked a concrete graph wavelet design methodology which targets the annihilation of graph signals, or, alternatively, the characterization of (classes of) graph signals which can be annihilated by existing con-

structions.

While established graph wavelet constructions, such as the spectral or tight graph wavelet ([37], [64]) may attain sufficiently (approximate) sparse graph wavelet domain representations, i.a. for appropriate design choices of the associated wavelet kernel, there is no concrete (or intuitive) theory on what types of graph signals can be annihilated, beyond the class of piecewise-constant signals, in particular, based on the properties and connectivity of the graph at hand. Sparsity on graphs has been more tangibly addressed through the topic of dictionary learning on graphs [32], which considers the problem of identifying an (overcomplete) basis \mathbf{D} under which a given graph signal \mathbf{y} can be sparsely represented as $\mathbf{y} = \mathbf{D}\mathbf{x}$. Accordingly, the work in this thesis breaks away from previous efforts, in that it examines how the connectivity of a graph can be leveraged to induce ‘exact’ sparsity in data on circulant graphs as well as more complex scenarios, beyond the intuitive example of the former.

2.2.4 Sampling on Graphs

Further, the topics of signal sampling and reconstruction on graphs have gained a growing interest within GSP, and in an effort to complement wavelet theory on graphs and provide context for the proposed bridge between wavelet and sampling theory, a brief overview of the latter is provided.

Signal recovery on graphs, constituting more broadly the empirical study as opposed to the analytical framework, has been tackled i.a. under the premise that the given signal is smooth with respect to the underlying graph, and may be formulated as an optimization problem within different settings ([40], [32]). In several works ([65], [66], [67], [68]), sampling theory for graphs, providing the specialized and more rigorous theorization of the former, is explored with predominant regard to the subspace of bandlimited graph signals under different assumptions; specifically, Anis et al. [67] and Chen et al. [68] provide two alternative interpretations of bandlimitedness in the graph domain, where, in particular, the latter employs matrix algebra to establish a linear reconstruction approach, based on a suitable choice of the retained (and known) sample locations. Moving beyond the traditional domain, sampling theory in the context of graphs has furthermore attempted to address and incorporate graph coarsening, such as by Chen et al. in [68], also a problem in itself ([69], [70]), which bears the challenge of identifying a meaningful underlying graph for the sampled signal and has been generally featured to a lesser extent.

2.3 The Class of Circulant Graphs

Circulant graphs represent a special class of graphs that reveal a set of convenient properties, which, not least of all, can be leveraged for the preservation of traditional signal processing concepts and operations. In particular, a circulant graph G is characterized by a generating set $S = \{s_1, \dots, s_M\}$, with $0 < s_k \leq N/2$, whose elements indicate the existence of an edge between node pairs $(i, (i \pm s_k)_N)$, $\forall s_k \in S$, where $()_N$ is the mod N operation; more intuitively, a graph is circulant if its associated adjacency matrix is a circulant matrix under a particular node labelling [59] (see examples in Fig. 2.2). As part of a sub-class of circulant graphs, the M -connected ring graph G is defined via the generating set $S = \{1, \dots, M\}$, such that there exists an edge between nodes i and j , if $(i - j)_N \leq M$ is satisfied; the associated circulant adjacency matrix is banded of bandwidth M . These graphs are i. a. utilized in the creation of small-world network graphs in the Watts-Strogatz model [71], prior to randomized edge rewiring, and can be embedded as tessellations in high-dimensional flat tori² [72].

The eigenvalues of a circulant graph adjacency matrix of the form

$$\mathbf{C} = \begin{bmatrix} 0 & c_1 & \cdots & c_{N-2} & c_{N-1} \\ c_{N-1} & 0 & \ddots & \ddots & c_{N-2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ c_2 & c_3 & \ddots & \ddots & c_1 \\ c_1 & c_2 & \cdots & \cdots & 0 \end{bmatrix}$$

are given by $\gamma_j = 0 + c_1\omega_j + c_2\omega_j^2 + \cdots + c_{N-1}\omega_j^{N-1}$, for $\omega_j = e^{-\frac{i2\pi j}{N}}$, with a possible choice of corresponding eigenvectors $v_j = \frac{1}{\sqrt{N}}[1, \omega_j, \omega_j^2, \dots, \omega_j^{N-1}]^T$, $j = 0, \dots, N-1$. The fact that the graph Laplacian eigenbasis, or GFT, of a circulant graph can be represented by the DFT matrix reveals a first major link between classical and graph-based signal processing, since Fourier and frequency notions are indirectly upheld.

A circulant (adjacency) matrix is symmetric if its first row has the following structure $C_{0,0:N-1} = [0 \ c_1 \ c_2 \ \dots \ c_{N/2-1} \ c_{N/2} \ c_{N/2-1} \ \dots \ c_2 \ c_1]$; for an $N \times N$ -matrix with $N \in 2\mathbb{N}$, this entails $N/2$ degrees of freedom (with $c_0 = 0$, as we do not allow self-loops). Further noteworthy about general (symmetric) circulant matrices in this context is the occurrence of eigenvalue multiplicities. While there is no general mathematical rule for the eigenvalue multiplicity distribution of a circulant (graph) matrix, with an arbitrary generating set, and it can only be determined by applying an exhaustive search approach, basic results on, for instance, the occurrence of odd and even multiplicities, such as in [73], can be inferred.

²The analysis of data residing on general geometrical shapes, which can be described by meshes, in particular, following the inference of corresponding graphs, can thus be conducted with graph-based methods.

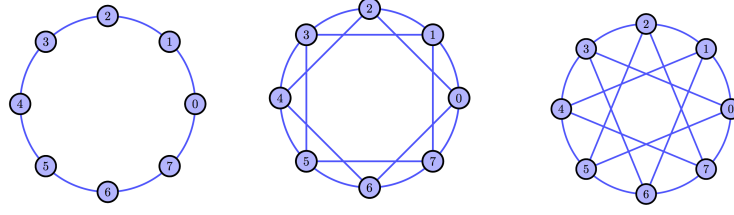


Figure 2.2: Circulant Graphs with generating sets $S = \{1\}$, $S = \{1, 2\}$, and $S = \{1, 3\}$ (from left). ©2017 Elsevier Inc.

Yet, the general intuition remains that a densely connected unweighted circulant graph is usually associated with large eigenvalue multiplicities, as evidenced by the complete graph of size N with adjacency spectrum consisting of $\gamma_0 = N - 1$ and $\gamma_1 = \dots = \gamma_{N-1} = -1$.

The symmetric, circulant graph Laplacian matrix \mathbf{L} , with first row $[l_0 \ \dots \ l_{N-1}]$, can be further defined through its so-called representer polynomial $l(z) = \sum_{i=0}^{N-1} l_i z^i$ with $z^{-j} = z^{N-j}$, which bears a resemblance to the z -transform in SP. For the circulant permutation matrix $\mathbf{\Pi}$ with first row $[0 \ 1 \ 0 \ \dots]$, we obtain $\mathbf{L} = \sum_{i=0}^{N-1} l_i \mathbf{\Pi}^i$. Every circulant matrix is associated with a representer polynomial [74], which not only provides a means to accelerate circulant matrix multiplication, thereby unfolding an inherent polynomial quality of circulants, but also establishes a link to their spectral information. In particular, the polynomial $l(z)$ of a circulant matrix \mathbf{L} gives rise to its (unsorted) eigenvalues³ via $l(e^{\frac{2\pi i k}{N}}) = \lambda_k$, $k = 0, \dots, N - 1$ [74]. This is not to be confused with its characteristic polynomial, given by $p_{\mathbf{L}}(x) = \det(x\mathbf{I}_N - \mathbf{L}) = \prod_{k=0}^{N-1} (x - \lambda_k)$, which as the unique monic polynomial of degree N vanishes at the eigenvalues λ_k , with determinant $\det \mathbf{L} = \prod_{k=0}^{N-1} (\sum_{j=0}^{N-1} \omega_k^{-j} l_j) = \prod_{k=0}^{N-1} \lambda_k$ [74]. In fact, the roots of $p_{\mathbf{L}}(x)$ (eigenvalues of \mathbf{L}) can be easily computed through $l(z)$; this has further inspired a method to solve polynomial equations by first finding a circulant matrix whose characteristic polynomial matches the former and subsequently recovering the solutions through its representer polynomial ([75], [76]). As another consequence of their relation to polynomials, the product of circulant matrices can be computed by simply evaluating the product of their corresponding representer polynomials modulo the matrix dimension [77].

2.3.1 Downsampling and Reconnection on Circulant Graphs

Due to their regularity and structure, circulant graphs further lend themselves for defining meaningful downsampling operations in GSP. As established in [59] by Ekambaram et al., one can downsample a given graph signal by 2 on the vertices of G with respect to any

³Hence, analogously to the z -transform which converts a discrete-time signal into a complex frequency domain representation, the circulant representer polynomial transfers the graph edge weight information to the spectral graph domain.

element $s_k \in S$. In the simplest scenario, which will also be employed in this work, the downsampling operation is conducted with respect to the outmost cycle ($s_1 = 1$) of a given circulant G , i.e. skipping every other labelled node, assuming that the graph at hand is connected such that $s_1 \in S$, and the dimension is $N = 2^n$ for $n \in \mathbb{N}$, the latter of which facilitates a multiresolution analysis.

In addition, the same authors introduced a set of vertex-domain localized filters constituting the ‘spline-like’ graph wavelet filterbank on circulant graphs ([34],[55]), which satisfies critical sampling and perfect reconstruction properties:

Theorem 2.1 ([34]). *The set of low-and high-pass filters, defined on an undirected connected circulant graph with adjacency matrix \mathbf{A} and degree d per node, take (weighted) averages and differences with respect to neighboring nodes at 1-hop distances of a given graph signal, and can be expressed as:*

$$\mathbf{H}_{LP} = \frac{1}{2} \left(\mathbf{I}_N + \frac{\mathbf{A}}{d} \right) \quad (2.5)$$

$$\mathbf{H}_{HP} = \frac{1}{2} \left(\mathbf{I}_N - \frac{\mathbf{A}}{d} \right). \quad (2.6)$$

The filterbank is critically sampled and invertible as long as at least one node retains the low-pass component, while the complementary set of nodes retains the high-pass components.

The structure of the above filterbank motivated the families of graph spline wavelets which will be introduced in Chapter 3, and, as will be demonstrated, bear actual spline properties⁴.

Multiscale analysis is conducted by iterating the result on the respective downsampled low-pass branches, in form of coarsened graphs, that may be obtained through suitable reconnection strategies [55]. In particular, this work will be primarily focused on critically-sampled filterbanks for which the sampled output can be well-defined on (suitably) coarsened graphs so as to control the problem dimensionality for multiscale analysis.

Succeeding the definition of a wavelet transform on a circulant graph, one thus needs to examine the problem of identifying suitable coarsened graph(s) on the vertices of which the downsampled low-(and high-)pass-representations of the original graph signal can be defined, as a means to facilitate the multiresolution decomposition in the graph domain. In general, it is not straightforward to determine if or how to reconnect the reduced set of vertices to form a coarsened graph, and the set of desired properties, comprising closure,

⁴According to [55], the ‘spline-like’ filterbank of Theorem 2.1 derived its name from the linear spline FIR filters which are equivalent to the graph filters for a simple cycle graph, however, comparable properties are not mentioned for other graphs.

preservation of the initial connectivity and spectral characterisation of the graph and/or graph type, among others, is rather difficult to satisfy entirely, such that priorities need to be set in keeping with the overall goal to be achieved. In the traditional domain, the downsampled signal samples are ‘reconnected’ through a simple stacking operation, and its graph-analogy on a simple cycle would entail the reconnection of 2-hop neighbours; yet a straightforward graph generalization is hindered by the overall complex connectivity of a graph, which leaves ambiguity to which extent one generally needs to reconnect downsampled nodes.

Kron-reduction [70] is a commonly used method for both circulant and arbitrary graphs, which employs a sub-matrix approximation scheme that takes into account the entire given graph Laplacian matrix for a well-defined dimensionality-reduced graph-representation. Nevertheless, it often leads to denser (circulant) graphs, and thus an increased matrix bandwidth, due to the maximum reconnection. In particular, given the graph Laplacian matrix \mathbf{L} and set V_α of nodes to retain, Kron-reduction evaluates the graph-Laplacian matrix $\tilde{\mathbf{L}}$ of the coarsened graph via

$$\tilde{\mathbf{L}} = \mathbf{L}(V_\alpha, V_\alpha) - \mathbf{L}(V_\alpha, V_\alpha^c) \mathbf{L}(V_\alpha^c, V_\alpha^c)^{-1} \mathbf{L}(V_\alpha, V_\alpha^c)^T,$$

and is known to preserve circularity under the standard downsampling pattern [55].

Since one of the main objectives in this work is a sparse graph wavelet representation, and, as derived contributions will show, the obtained sparsity K is related to (and increases with) a border effect from the bandwidth M of the graph (adjacency) matrix, we favor a graph reconnection method which reduces or maintains M when conducting multiresolution analysis. For an optimally sparse multilevel graph wavelet representation, one can therefore resort to two variations, both of which preserve circularity with little to no reconnection, to determine the coarse graph \tilde{G} : (1) nodes are not reconnected, and existing edges are preserved (with exception of maintaining $s = 1 \in S$ to ensure that \tilde{G} is connected), (2) a subset of nodes are (re-)connected, such that \tilde{G} is identical in structure to the initial G , i.e. the generating set S is preserved.

While the former approach leads to the sparsest possible option, as given an M -connected (banded) circulant graph, edges resulting from odd elements in the generating set are continuously removed, reducing its band, it fails to preserve the global connectivity of the initial graph, which for the objectives at hand, may be deemed of secondary importance. In addition, for bipartite circulant graphs it produces the trivial simple cycle. The latter approach, while inducing a slightly less sparse representation due to the constant bandwidth, reconnects a subset of nodes, which were initially connected via a path, and preserves connectivity, through an exact replication in lower dimension, as well as certain spectral information, which will also be directly proved in Chapter 5. This discussion will be revisited throughout the course of this work due to its relevance in various aspects of

GSP.

Chapter 3

Wavelets & Filterbanks on Circulant Graphs and Beyond

3.1 Motivation and Objectives

In classical signal processing, one of the central issues within the area of transform analysis revolves around determining the best (often signal-and/or application-dependent) basis in which a given signal can be sparsely approximated with high accuracy [78]. While the standard Fourier basis has been largely preferred for classes of smooth signals, the advent of wavelet theory has revealed that wavelet bases, which can operate locally and across existing signal-changes and discontinuities, are more suitable for signals that are piecewise smooth.

Similarly, one of the fundamental questions in graph signal processing is concerned with how to build the best (graph-dependent) basis for a signal associated with a graph, yet in light of the newly arising data-dependencies that need to be incorporated and plethora of interpretations and techniques tackling variable and, usually data/application-driven, desirable properties, remains largely unanswered from a theoretically rigorous perspective. This chapter is dedicated to the topic of *graph spline wavelet theory*, introducing families of spline and e-spline wavelets, and associated filterbanks on circulant graphs, which leverage an inherent *vanishing moment* property of circulant graph Laplacian matrices, and will eventually provide a crucial stepping stone to more generalized graph wavelet designs, sampling theory as well as applications in the following chapters. With sparsity as the driving concept of this work, the goal of this chapter is to impart intuition behind the proposed wavelets and their properties on the basis of circulant graphs all the while drawing connections to their counterparts in the classical domain, before moving to more complex

scenarios.¹

3.2 Families of Spline Wavelets on Circulant Graphs

3.2.1 Vanishing Moments on the Graph

According to general consensus [1], a signal or function $\mathbf{x} \in \mathbb{R}^N$ on a graph $G = (V, E)$ with respective node and edge sets V and E is considered *smooth* if the samples x_i vary minimally over the graph, i. e. its total variation with respect to neighbouring vertices on the graph is small.

A set of discrete differential operators, which can be considered as discrete versions of continuous differential geometry operators, have been defined in order to instil a mathematical description of the smoothness of a graph signal with respect to the intrinsic structure of the underlying graph; nevertheless, it should be noted these graph operators do not necessarily converge to their continuous versions, e.g. on a manifold, unless additional assumptions are made [1].

In the style of its counterparts in the continuous setting, let $\frac{\partial \mathbf{x}}{\partial e}|_i := \sqrt{A_{i,j}}(x(j) - x(i))$ represent the edge derivative of \mathbf{x} with respect to an edge $e = (i, j)$ at vertex i and $\nabla_i \mathbf{x} := [\{\frac{\partial \mathbf{x}}{\partial e}|_i\}_{e \in E \text{ s.t. } e=(i,j), j \in V}]$ the graph gradient of \mathbf{x} at i [1]. The discrete p -Dirichlet form of \mathbf{x}

$$S_p(\mathbf{x}) := \frac{1}{p} \sum_{i \in V} \|\nabla_i \mathbf{x}\|_2^p = \frac{1}{p} \sum_{i \in V} \left[\sum_{j \in N_i} A_{i,j} (x(j) - x(i))^2 \right]^{p/2}$$

provides a concrete measure of global graph signal smoothness by summing over (powers of) the local variation $\|\nabla_i \mathbf{x}\|_2$; for $p = 1$, it characterizes the total variation of a signal with respect to a graph, and for $p = 2$, the graph Laplacian quadratic form $S_2(\mathbf{x}) = \mathbf{x}^T \mathbf{L} \mathbf{x}$, which in turn gives rise to a semi-norm $\|\mathbf{x}\|_{\mathbf{L}} := \|\mathbf{L}^{1/2} \mathbf{x}\|_2 = \sqrt{\mathbf{x}^T \mathbf{L} \mathbf{x}}$ [63]. Its normalized form

$$\frac{\mathbf{x}^T \mathbf{L} \mathbf{x}}{\|\mathbf{x}\|_2^2} = \frac{1}{\|\mathbf{x}\|_2^2} \sum_{n=1}^N \lambda_n |\hat{x}_n|^2$$

has also been associated with the spectral spread of a signal [79].

For an undirected connected graph, an all-constant signal is also an eigenvector of \mathbf{L} corresponding to $\lambda_0 = 0$, and signals of the form of low-order graph Laplacian eigenvectors thus provide concrete examples that minimize the latter while elucidating that smoothness is understood as a measure within the chosen graph-eigenbasis. In another line of work, total variation on a graph is adapted for directed graph cases, and the p -Dirichlet form is

¹A significant part of the content in this chapter appears in accepted article [7], and in part in publications [9], [8] and accepted article [6].

defined as $S_p(\mathbf{x}) := \|\mathbf{x} - \frac{1}{|\lambda_{max}(\mathbf{A})|} \mathbf{A}\mathbf{x}\|_p^p$ [40].

Nevertheless, since these measures provide little information on the sparsity level of \mathbf{x} on the graph, such as the l_0 -norm $\|\mathbf{L}\mathbf{x}\|_0$, beyond the trivial case, we begin to investigate annihilation on graphs and broaden the term *smooth* to encompass signals with a *sparse representation* in a designated graph-basis. As such, we begin by considering vanishing moments of wavelets which annihilate polynomial signals, and thus induce a sparse representation.

Traditionally, a high-pass filter \mathbf{h} with taps (weights) h_k that is orthogonal with respect to the subspace of polynomials of up to degree $N - 1$ has N *vanishing moments*, i.e. it is characterized by

$$m_n = \sum_{k \in \mathbb{Z}} h_k k^n = 0, \text{ for } n = 0, \dots, N - 1$$

where m_n is the n -th order moment of \mathbf{h} . This is equivalent to its z -transform $H(z)$ having N zeros at $z = 1$. Ideally, one would like to extend a similar notion to graphs. In [33], the number of vanishing moments of a scaling function on a graph or manifold are defined as the number of eigenfunctions of the given diffusion operator \mathbf{T} it is orthogonal to, up to a precision measure. This is further seized on in [80], with graph Laplacian eigenvectors providing the basis for generalized vanishing moments on graphs. Nevertheless, this definition of vanishing moments on graphs does not accommodate equivalencies between the graph and Euclidean domains, in particular, for a discrete periodic (time) signal on a simple cycle, which is considered the transitional base case.

This work mainly adheres to the traditional definition for purposes of illustrating analogies to the developed spline wavelet theory on circulant graphs, yet annihilation is considered in a broader sense as the property of a matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ with a nontrivial *nullspace* $Null(\mathbf{A}) = \{\mathbf{x} \in \mathbb{K}^N | \mathbf{A}\mathbf{x} = \mathbf{0}_N\}$, where \mathbb{K}^N denotes a field of dimension N and the dimension of the nullspace indicating the degree (or order) of annihilation.²

In its original conception, a polynomial is a continuous infinitely differentiable function within a subspace of $C^\infty([a, b])$ [81]. When considered in a discrete and finite domain, the annihilation of the discretized polynomial by a high-pass filter (wavelet) is subject to a boundary effect dependent on the length of the support of the filter. In order to capture the ensuing result, which provides the foundation for spline wavelet theory on graphs, we maintain the standard definition of polynomial (graph) signals as functions which can be annihilated by a suitable graph differential operator and are similarly subject to a boundary effect, which is additionally magnified through the number of connections.

Definition 3.1. A graph signal $\mathbf{p} \in \mathbb{R}^N$ defined on the vertices of a graph G is (piecewise)

²We refrain from a classification into local and non-local annihilation for now since the definitions are transitional, as will be clarified later.

polynomial if its labelled sequence of sample values, with value $p(i)$ at node i , is the discrete, vectorized version of a standard (piecewise) polynomial, such that $\mathbf{p} = \sum_{j=1}^K \mathbf{p}_j \circ \mathbf{1}_{[t_j, t_{j+1})}$, where $t_1 = 0$ and $t_{K+1} = N$, with pieces $p_j(t) = \sum_{d=0}^D a_{d,j} t^d$, $j = 1, \dots, K$, for $t \in \mathbb{Z}^{\geq 0}$, coefficients $a_{d,j} \in \mathbb{R}$, and maximum polynomial degree $D = \deg(p_j(t))$.

Lemma 3.1. *For an undirected, circulant graph $G = (V, E)$ of dimension N , the associated representer polynomial $l(z) = l_0 + \sum_{i=1}^M l_i(z^i + z^{-i})$ of graph Laplacian matrix \mathbf{L} , with first row $[l_0 \ l_1 \ l_2 \ \dots \ l_2 \ l_1]$, has two vanishing moments. Therefore, the operator \mathbf{L} annihilates up to linear polynomial graph signals (i.e. of up to degree $D = 1$), subject to a border effect determined by the bandwidth M of \mathbf{L} , provided $2M \ll N$.*

Proof. The representer polynomial of \mathbf{L} with degree $d = \sum_{i=1}^M 2d_i$ per node and symmetric weights $d_i = A_{j, (j+i)_N}$, can be expressed as:

$$l(z) = (-d_M z^{-M} - \dots - d_1 z^{-1} + d - d_1 z - \dots - d_M z^M) = \sum_{i=1}^M d_i (z^i - 1)(z^{-i} - 1),$$

whose factors are divisible by $(z^{\pm 1} - 1)$ respectively using the equality $z^n - 1 = (z - 1)(1 + z + \dots + z^{n-1})$, thus proving that the matrix \mathbf{L} has two vanishing moments. Therefore, for a sufficiently small M with respect to the dimension N of the graph G , or in other words, if the adjacency matrix of G is a symmetric, banded circulant matrix of bandwidth M , the corresponding \mathbf{L} annihilates linear polynomial graph signals on G up to a boundary effect. \square

Due to the equivalency between polynomial and matrix multiplication, the property from Lemma 3.1 can be generalized to $2k$ vanishing moments for matrix \mathbf{L}^k , $k \in \mathbb{N}$. In general, if the rows of the graph Laplacian of an arbitrary (not necessarily undirected) graph have the above centrosymmetric form (i.e. each node has a symmetric neighborhood), annihilation is maintained, but not carried over to higher order $k > 1$. The (transposed) oriented incidence matrix \mathbf{S}^T of an arbitrary graph has, by default, one vanishing moment, yet it is only for the previously described special cases that the combination $\mathbf{S}\mathbf{S}^T = \mathbf{L}$ can produce two vanishing moments, as a result of neighborhood symmetry.

3.2.2 The Graph Spline Wavelet

Following the interpretation of the graph Laplacian as a high-pass filter, a new range of graph wavelet bases and filterbanks is introduced, with high-pass filters that can annihilate higher-order polynomial graph signals for a sparse representation in a graph wavelet basis. In the first instance, one can generalize the spline-like graph wavelet transform (see Thm. 2.1) to higher order and prove that it remains a valid transform.

Theorem 3.1. *Given the undirected, and connected circulant graph $G = (V, E)$ of dimension N , with adjacency matrix \mathbf{A} and degree d per node, define the higher-order graph-spline wavelet transform (HGSWT), composed of the low-and high-pass filters*

$$\mathbf{H}_{LP} = \frac{1}{2^k} \left(\mathbf{I}_N + \frac{\mathbf{A}}{d} \right)^k \quad (3.1)$$

$$\mathbf{H}_{HP} = \frac{1}{2^k} \left(\mathbf{I}_N - \frac{\mathbf{A}}{d} \right)^k \quad (3.2)$$

whose associated high-pass representer polynomial $H_{HP}(z)$ has $2k$ vanishing moments. This filterbank is invertible for any downsampling pattern, as long as at least one node retains the low-pass component, while the complementary set of nodes retains the high-pass components.

Proof. See Appendix A.1.

For a graph signal $\mathbf{p} \in \mathbb{R}^N$ defined on G , the HGSWT yields

$$\tilde{\mathbf{p}} = \left(\frac{1}{2}(\mathbf{I}_N + \mathbf{K})\mathbf{H}_{LP} + \frac{1}{2}(\mathbf{I}_N - \mathbf{K})\mathbf{H}_{HP} \right) \mathbf{p} = \frac{1}{2}(\mathbf{I}_N + \mathbf{K})\tilde{\mathbf{p}}_{LP} + \frac{1}{2}(\mathbf{I}_N - \mathbf{K})\tilde{\mathbf{p}}_{HP}$$

where \mathbf{K} is a diagonal sampling matrix, with $K_{i,i} = 1$ at $i = 0, 2, \dots, N-2$ and $K_{i,i} = -1$ otherwise, as downsampling is conducted w. r. t. $s = 1 \in S$ and even-numbered nodes are retained. The resulting signals $\tilde{\mathbf{p}}_{LP}, \tilde{\mathbf{p}}_{HP} \in \mathbb{R}^N$ represent (coarsened) low-and high-pass versions of \mathbf{p} on G within a k -hop local neighborhood $N(i, k), \forall i \in V$. With increasing degree of the filterbank, the number of vanishing moments increases, and its filters become less localized in the vertex domain; in particular, a filter of vanishing moment order $2k$ is localized within a k -hop neighborhood. Figure 3.1 plots the low-and high-pass graph filter functions of the HGSWT at $k = 2$ and illustrates the spread in the vertex domain for a sample circulant graph.

A bipartite circulant graph G is characterized by a generating set S which contains only odd elements $s_k \in S$ for even dimension $|V| = N$; the simple cycle $S = \{1\}$ is a clear example. For such graphs, the HGSWT acquires the following interesting property:

Corollary 3.1. *When G is an undirected, circulant, bipartite graph, with adjacency matrix \mathbf{A} of bandwidth M , the polynomial representation $H_{LP}(z)$ of the low-pass filter \mathbf{H}_{LP} in Eq. (3.1) can reproduce polynomial graph signals up to order $2k - 1$, subject to a border effect determined by the bandwidth Mk of \mathbf{H}_{LP} , provided $2Mk \ll N$.*

Proof. Similarly, as in Lemma 3.1, we can express the representer polynomial as

$$H_{LP}(z) = \frac{1}{(2d)^k} (d_M z^{-M} + \dots + d_1 z^{-1} + d + d_1 z + \dots + d_M z^M)^k$$

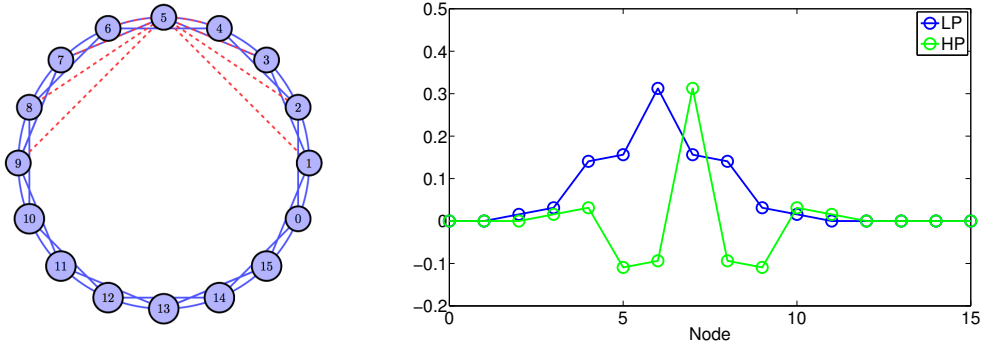


Figure 3.1: Localization of the *HGSWT* filters for $k = 2$ in the graph vertex domain for circulant G with $S = \{1, 2\}$: shown at vertex $v = 5 \in V$ on G (left), and corresponding graph filter functions at alternate vertices. ©2017 Elsevier Inc.

$$= \left(\frac{1}{2d} \sum_{1 \leq i \leq M, i \in 2\mathbb{Z}+1} d_i (z^i + 1)(z^{-i} + 1) \right)^k$$

and note that the RHS factors $(z + 1)^k (z^{-1} + 1)^k$, since $(z^i + 1)$ has a root at $z = -1$ only for $i \in 2\mathbb{Z} + 1$. According to the Strang-Fix condition ([82], [22]), this is necessary and sufficient for ensuring the reproduction of polynomials. \square

3.2.3 The Signless Laplacian

The matrix $\mathbf{Q} = \mathbf{D} + \mathbf{A}$, while primarily chosen as a low-pass filter due to its averaging effect between 1-hop connected neighbors, is also known as the *signless Laplacian*. It has been featured to a lesser extent than the standard graph Laplacian, however, following recent interest, the signless Laplacian has been conceived as more convenient among graph matrices to study graph properties [83], and there exist distinct properties which further reinforce the previously derived results. In particular, for a connected graph, \mathbf{Q} is a positive semi-definite matrix, whose smallest (simple) eigenvalue is 0 if and only if the graph is bipartite; further, its multiplicity is equal to the number of connected bipartite components [84]. A more generalized form of this result has been independently derived in Cor. 3.3 of Sect. 3.3.2. It is further interesting to note that the characteristic polynomials of the signless and regular Laplacian are the same for a bipartite graph, i. e. their eigenvalues are the same; as a result of the spectral folding phenomenon, with the addition of circularity, such that a common eigenbasis exists that reorders the eigenvalues, which are symmetric around the node degree d , this translates into the reproduction property of Cor. 3.1. Here we have frequency parameters $z = e^{\frac{2\pi i k}{N}}$ and $-z = e^{\frac{2\pi i (k+N/2)}{N}}$, $k = 0, \dots, N-1$, in $H_{LP}(-z) = H_{HP}(z)$, which induce graph-filter eigenvalues that are shifted by $N/2$ in their position within the DFT-ordered spectrum, following the folding $A(-z) = -A(z)$ of the representer polynomial $A(z)$ of the adjacency matrix \mathbf{A} . Another notable property of the

signless Laplacian is the interpretation of random walks, i.e. the (i, j) -th entry of \mathbf{Q}^k is equal to the number of semi-edge walks of length k starting at vertex i and terminating at j [83].

3.3 Families of E-Spline Wavelets on Circulant Graphs

3.3.1 A Generalized Graph Laplacian Operator

Following the theory of cardinal exponential splines [85] in the classical domain, we identify an equivalent framework which maintains and extends these properties to the graph domain to broaden the classes of relevant graph signals and graph wavelets. This primarily requires the definition of a generalized graph difference operator with wide-ranging annihilation properties.

Definition 3.2. A complex exponential polynomial graph signal $\mathbf{y} \in \mathbb{C}^N$ with parameter $\alpha \in \mathbb{R}$, is defined such that node j has sample value $y(j) = p(j)e^{i\alpha j}$, for polynomial $\mathbf{p} \in \mathbb{R}^N$ of degree $\deg(p(t))$.

Definition 3.3. Let $G = (V, E)$ be an undirected, circulant graph with adjacency matrix \mathbf{A} and degree $d = \sum_{j=1}^M 2d_j$ per node with symmetric weights $d_j = A_{i, (j+i)_N}$. Then the parameterised e-graph Laplacian of G is given by $\tilde{\mathbf{L}}_\alpha = \tilde{\mathbf{D}}_\alpha - \mathbf{A}$, with exponential degree $\tilde{d}_\alpha = \sum_{j=1}^M 2d_j \cos(\alpha j)$.

Hereby, the e-graph Laplacian matrix $\tilde{\mathbf{L}}_\alpha = \tilde{\mathbf{D}}_\alpha - \mathbf{A}$ is established as a generalized graph Laplacian with a parameterised diagonal degree matrix $\tilde{\mathbf{D}}_\alpha$. The standard graph Laplacian \mathbf{L} can therefore be regarded as a special case of the e-graph Laplacian $\tilde{\mathbf{L}}_\alpha$ for $\alpha = 0$, however, with $\tilde{d}_\alpha \leq d$ the matrix ceases to be positive semi-definite otherwise.

Lemma 3.2. For an undirected, circulant graph $G = (V, E)$ of dimension N , the associated representer polynomial $\tilde{l}(z) = \tilde{l}_0 + \sum_{i=1}^M \tilde{l}_i(z^i + z^{-i})$ of the e-graph Laplacian matrix $\tilde{\mathbf{L}}_\alpha$, with first row $[\tilde{l}_0 \ \tilde{l}_1 \ \tilde{l}_2 \ \dots \ \tilde{l}_2 \ \tilde{l}_1]$, has two vanishing exponential moments, i.e. the operator $\tilde{\mathbf{L}}_\alpha$ annihilates complex exponential polynomial graph signals with exponent $\pm i\alpha$ and $\deg(p(t)) = 0$. Unless $\alpha = \frac{2\pi k}{N}$ for $k \in [0, N-1]$, this is subject to a border effect determined by the bandwidth M of $\tilde{\mathbf{L}}_\alpha$, where $2M \ll N$.

Proof. Consider the representer polynomial $\tilde{l}(z)$ of $\tilde{\mathbf{L}}_\alpha$:

$$\tilde{l}(z) = \sum_{j=1}^M 2d_j \cos(\alpha j) - d_j(z^j + z^{-j}) = \sum_{j=1}^M d_j(1 - e^{i\alpha j} z^j)(1 - e^{-i\alpha j} z^j)(-z^{-j})$$

and note that $(1 - e^{\mp i\alpha} z^{-1})$ is a factor of $(1 - e^{\pm i\alpha j} z^j)$, which corresponds to two exponential vanishing moments [85]. \square

The construction of $\tilde{\mathbf{L}}_\alpha$ with polynomial factors $(1 - e^{i\alpha}z^{-1})(1 - e^{-i\alpha}z^{-1})$ ensures that its edge weights, and, by extension, those of the graph wavelet filter \mathbf{H} , remain symmetric and real-valued. In other words, symmetry is preserved since vanishing (exponential) moments occur in pairwise increments of complex conjugates.

While the class of complex exponential polynomial graph signals of the form $y(t) = p(t)e^{i\alpha t}, \alpha \in \mathbb{R}$, which can be represented by trigonometric splines in the traditional domain [85], is of primary concern, one can transition to further interesting cases e.g. by letting $\alpha = -i\beta, \beta \in \mathbb{R}$. In particular, this results in real exponential polynomials, which can be represented in terms of hyperbolic functions of the form $(\cosh(\beta t), \sinh(\beta t))$, inducing the class of hyperbolic splines [85] on graphs, with redefined e-degree $\tilde{d} = \sum_{k=1}^M 2d_k \cos(-i\beta k) = \sum_{k=1}^M 2d_k \cosh(\beta k)$.

The following remark further elucidates the distinct structure of the e-graph Laplacian:

Remark 3.1. The eigenvalues $\{\lambda_j\}_{j=0}^{N-1}$ of a circulant matrix, in particular those of \mathbf{A} in Lemma 3.2, can be expressed as $\lambda_j = \sum_{k=1}^M 2d_k \cos\left(\frac{2\pi k j}{N}\right)$, $j = 0, \dots, N-1$, ordered by the DFT-matrix as eigenbasis \mathbf{U} . When imposing $\alpha = \frac{2\pi j}{N}$ and $j \in [0, N-1]$, the e-graph Laplacian nullspace $\text{Null}(\tilde{\mathbf{L}}_\alpha)$ contains the j -th eigenvector \mathbf{u}_j , where \mathbf{u}_j represents a complex exponential graph signal with exponent $\alpha = \frac{2\pi j}{N}$ and $\deg(p(t)) = 0$. In particular, we have $\tilde{d}_\alpha = \lambda_j$, and for $\alpha = 0$ this becomes the maximum eigenvalue $\tilde{d}_0 = d = \lambda_{\max}$, whose associated eigenvector is the all-constant $\mathbf{u}_{\max} = \mathbf{1}_N$, i.e. the nullspace of standard graph Laplacian \mathbf{L} . This facilitates the reinterpretation of the e-graph Laplacian as $\tilde{\mathbf{L}}_\alpha = \lambda_j \mathbf{I}_N - \mathbf{A}$, for $\alpha = \frac{2\pi j}{N}$ and $j \in [0, N-1]$, or more generally, $(\lambda_j \mathbf{I}_N - \mathbf{A})\mathbf{u}_j = \mathbf{0}_N$, with $\tilde{\mathbf{L}}_\alpha$ representing the shift of \mathbf{L} by $\lambda_j - d = -\tilde{\lambda}_j$ toward annihilation of \mathbf{u}_j , where $\{\tilde{\lambda}_j\}_{j=0}^{N-1}$ denotes the spectrum of \mathbf{L} . Depending on eigenvalue multiplicities, the nullspace of $\tilde{\mathbf{L}}_\alpha$ is accordingly extended.

In revisiting the discussion on vanishing moments, it becomes evident that the annihilation property of the restricted e-graph Laplacian bears similarities to the definition of vanishing moments on graphs in [33], that is as the nullspace of a (wavelet) operator. Here, however, orthogonality is not extended up to a precision metric, with the selected operator being parametric instead, i.e. the nullspace can be ‘shifted’ to contain (a subset of) its eigenvectors. The significance of the proposed approach lies in the fact that annihilation can also be local, and is not restricted to the nullspace of $\tilde{\mathbf{L}}_\alpha$. This notion will be further realized for arbitrary graphs in the next chapter.

The e-graph Laplacian (and its counterparts for arbitrary graphs) can be seen as a family of generalized graph Laplacian matrices $\mathbf{M} = \mathbf{L} + \mathbf{P}$ (see [47]), with $\mathbf{P} = \tilde{\mathbf{D}}_\alpha - \mathbf{D}$ (or $\mathbf{P} = \lambda_j(\mathbf{A})\mathbf{I}_N - \mathbf{D}$). The graph Laplacian of a connected graph has rank $N-1$, while the e-graph Laplacian has rank $N-m$ if $d_\alpha = \lambda_j(\mathbf{A})$, $j \in [0, N-1]$, where m is the multiplicity of $\lambda_j(\mathbf{A})$, and rank N otherwise for $d_\alpha \neq \lambda_j(\mathbf{A})$, $j \in [0, N-1]$. Since the (ordered) eigenvalues of $\tilde{\mathbf{L}}_\alpha$ are no longer nonnegative, their interpretation as graph frequencies which order the corresponding eigenvectors in terms of the number of their oscillations

(zero crossings) [1] remains valid, with the only difference that the graph frequency $d - \lambda_j$ of \mathbf{L} becomes the new zero or DC-frequency in $\tilde{\mathbf{L}}_\alpha = (\lambda_j \mathbf{I}_N - \mathbf{A})$. Further, the e-graph Laplacian quadratic form, as a measure of smoothness, though no longer a semi-norm, can be expressed as $\tilde{S}_{\alpha 2}(\mathbf{x}) = \mathbf{x}^T \tilde{\mathbf{L}}_\alpha \mathbf{x} = (\tilde{d}_\alpha - d) \|\mathbf{x}\|_2^2 + \mathbf{x}^T \mathbf{L} \mathbf{x} = (\tilde{d}_\alpha - d) \|\mathbf{x}\|_2^2 + S_2(\mathbf{x})$.

3.3.2 Graph E-Spline Wavelets

The graph e-spline wavelet filterbank is designed following a similar line as the (higher-order) graph-spline wavelet filterbank, where the e-graph Laplacian $\tilde{\mathbf{L}}_\alpha$ is employed as the high-pass filter and its signless version, $\tilde{\mathbf{D}}_\alpha + \mathbf{A}$, as the low-pass filter, respectively subject to normalization. The parametric form of the filterbank with respect to $\vec{\alpha} = (\alpha_1, \dots, \alpha_T) \in \mathbb{R}^T$ is achieved via simple convolution of different graph filter functions, resulting in an invertible transform. This facilitates the expansion of the types of graph signals which can be reproduced and/or annihilated by a single GWT.

Theorem 3.2. *The higher-order graph e-spline wavelet transform (HGESWT) on a connected, undirected circulant graph G , is composed of the low-and high-pass filters*

$$\mathbf{H}_{LP\vec{\alpha}} = \prod_{n=1}^T \frac{1}{2^k} \left(\beta_n \mathbf{I}_N + \frac{\mathbf{A}}{d} \right)^k \quad (3.3)$$

$$\mathbf{H}_{HP\vec{\alpha}} = \prod_{n=1}^T \frac{1}{2^k} \left(\beta_n \mathbf{I}_N - \frac{\mathbf{A}}{d} \right)^k \quad (3.4)$$

where \mathbf{A} is the adjacency matrix, d the degree per node and parameter β_n is given by $\beta_n = \frac{\tilde{d}_{\alpha_n}}{d}$ with $\tilde{d}_{\alpha_n} = \sum_{j=1}^M 2d_j \cos(\alpha_n j)$ and $\vec{\alpha} = (\alpha_1, \dots, \alpha_T)$. Then the high-pass filter annihilates complex exponential polynomials (of $\deg(p(t)) \leq k-1$) with exponent $\pm i\alpha_n$ for $n = 1, \dots, T$. The transform is invertible for any downsampling pattern as long as the eigenvalues γ_i of $\frac{\mathbf{A}}{d}$ satisfy $|\beta_n| \neq |\gamma_i|$, $i = 0, \dots, N-1$ for $n = 1, \dots, T$, under either of the sufficient conditions

(i) $k \in 2\mathbb{N}$, or

(ii) $k \in \mathbb{N}$ and β_n, T such that $\forall \gamma_i, f(\gamma_i) = \prod_{n=1}^T (\beta_n^2 - \gamma_i^2)^k > 0$ or $f(\gamma_i) < 0$.

If parameters β_n , are such that $\beta_n = \gamma_i$, for up to T distinct values, the filterbank continues to be invertible under the above as long as $\beta_n \neq 0$ and at least $\sum_{i=1}^T m_i$ low-pass components are retained at nodes in set V_α such that $\{\mathbf{v}_{+i,k}(V_\alpha)\}_{i=1,k=1}^{i=T,k=m_i}$ (and, if eigenvalue $-\gamma_i$ exists, complement $\{\mathbf{v}_{-i,k}(V_\alpha)\}_{i=1,k=1}^{i=T,k=m_i}$) form linearly independent sets, where m_i is the multiplicity of γ_i and $\{\mathbf{v}_{\pm i,k}\}_{k=1}^{m_i}$ are the eigenvectors respectively associated with $\pm \gamma_i$.

Proof. See Appendix A.2.

The essential property of the above transform, as captured in the proof, is that invertibility is governed by the parameters β_n ; Thm. 3.2 also applies to real exponential polynomial signals with e-degree of the form $\tilde{d}_{i\alpha} = \sum_{k=1}^M 2d_k \cosh(\alpha k)$, $\cosh(x) \in [1, \infty) \forall x$, as previously mentioned. In particular, whenever the chosen β_n coincide with the magnitude(s) of certain eigenvalues $\{\gamma_i\}_i$ of normalized adjacency matrix $\frac{\mathbf{A}}{d}$, it is essential to select a (valid) downsampling pattern under which the eigenvectors $\{\mathbf{v}_{i,k}\}_{i,k}$ associated with $\{\gamma_i\}_i$ remain linearly independent.³ Naturally, this can become challenging for certain graph topologies as well as when the number T of such β_n is large. Circulant adjacency matrices, in particular, can be shown to exhibit large eigenvalue multiplicities as a result of regularity and increased graph connectivity.

Example: The normalized adjacency matrix of an unweighted complete (and hence circulant) graph of dimension N has eigenvalues $\gamma_i = -\frac{1}{d}$ of multiplicity $N - 1$ and $\gamma_{max} = 1$, which is simple. As such, when downsampling is conducted w. r. t. $s = 1$, the transform is invertible for $\beta = 1$, but not for $\beta = -\frac{1}{d}$. The introduction of distinct edge weights, however, can remedy this, as the spectrum of a complete weighted (circulant) graph is less concentrated.

A note on multiresolution: In general, one can iterate on the low-pass branch of either of the introduced wavelet transforms to obtain a multilevel graph signal representation on a suitably coarsened graph. However, as a consequence of the non-stationarity (or scale-dependency) of the latter (see [25] for the traditional case), corresponding adjustments need to be made to parameters $\vec{\alpha}$, i. e. we require the parameterization by $2^j \vec{\alpha}$ at level j in order to preserve annihilation properties. Technically, both graph wavelet filterbank constructions in Thms. 3.1 and 3.2 are ‘non-stationary’ in the sense that the representer polynomials of their respective graph filters at different levels are not necessarily dilates of one another, being functions of the graph adjacency matrix. Hence, unless the coarsened graph, on which the downsampled low-pass output is defined, bears identical edge relations to the initial graph (e. g. when the generating sets are identical for bandwidth B with $2B < N$), the representer functions will change with the graph. Nevertheless, the general outer structure of the filters, as polynomials in the adjacency matrix, may only change in Thm. 3.2 due to the parameterization by $\{\beta_i\}_i$.

3.3.3 Special Cases and Discussion

One easily deduces that the *HGESWT* of Thm. 3.2 ‘converges’ to the *HGSWT* of Thm. 3.1 as $\alpha \rightarrow 0$ and $\beta = \frac{\tilde{d}_\alpha}{d} \rightarrow 1$, since we have $\mathbf{H}_{LP_\alpha} = \mathbf{H}_{LP}$ and $\mathbf{H}_{HP_\alpha} = \mathbf{H}_{HP}$ for $\alpha = 0$ resulting from their definition through the e-graph Laplacian, in which case the conditions

³This includes, but is not limited to, the signals which lie in the nullspace of the high-pass filter.

on invertibility are relaxed.⁴

The introduced transforms can be applied on any undirected circulant graph G , yet, there exist some noteworthy property distinctions between the cases of bipartite and non-bipartite circulant graphs, as well as when $|\beta_n| = |\gamma_i|$ is satisfied for some n and $i \in [0 \ N - 1]$. In particular, the low-pass filter $\frac{1}{d}(\mathbf{D}_\alpha + \mathbf{A})$ is of the form of a parameterized signless Laplacian. As a consequence of this structural similarity, one can detect similar reproduction properties for circulant bipartite graphs:

Corollary 3.2. *Let $G = (V, E)$ be an undirected, bipartite circulant graph with adjacency matrix \mathbf{A} of bandwidth M , and e -degree \tilde{d}_α . Then the low-pass filter \mathbf{H}_{LP_α} in Eq. (3.3) reproduces complex exponential polynomial graph signals \mathbf{y} with exponent $\pm i\alpha$, up to a border effect determined by the bandwidth Mk of \mathbf{H}_{LP_α} , provided $2Mk \ll N$.*

Proof. The representer polynomial $H_{LP_\alpha}(z)$ is of the form

$$\begin{aligned} H_{LP_\alpha}(z) &= \frac{1}{(2d)^k} (d_M z^{-M} + \dots + d_1 z^{-1} + \sum_{j=1}^M 2d_j \cos(\alpha j) + d_1 z^1 + \dots + d_M z^M)^k \\ &= \left(\frac{1}{2d} \sum_{j=1}^M d_j (1 + e^{i\alpha j} z^j)(1 + e^{-i\alpha j} z^j)(z^{-j}) \right)^k, \quad j \in 2\mathbb{Z}^+ + 1, \end{aligned}$$

where $(1 + e^{\mp i\alpha} z^{-1})$ is a factor of $(1 + e^{\pm i\alpha j} z^j)$ if j is odd, i.e. the elements in the generating set of G are odd. This filter therefore satisfies the generalized Strang-Fix conditions for the reproduction of exponentials [25]. \square

Further, an interesting property pertaining to the low-pass filter, which will be made use of in further graph wavelet constructions, can be deduced:

Corollary 3.3. *Let $G = (V, E)$ be an undirected, circulant graph with adjacency matrix \mathbf{A} and degree $d = \sum_{j=1}^M 2d_j$ per node with symmetric weights $d_j = A_{i, (j+i)_N}$. Then the low-pass filter \mathbf{H}_{LP_α} in Eq. (3.3) is invertible unless (i) G is bipartite while β_n satisfies $|\beta_n| = |\gamma_i|$ or (ii) $\beta_n = -\gamma_i$, $i \in [0 \ N - 1]$.*

Proof. See Appendix A.3.

Note that for a bipartite graph, when $\{\beta_n\}_n$ are of eigenvalue form, both the spline

⁴We note that invertibility of the wavelet transforms in Thms 3.1 and 3.2 does not require \mathbf{A} to be a circulant matrix. In particular, for the former, it is sufficient that $\frac{\mathbf{A}}{d}$ be replaced by the normalized adjacency matrix $\hat{\mathbf{A}} = \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}$ of an undirected connected graph [3], and for the latter, its eigenbasis \mathbf{V} is subject to similar constraints, if chosen parameters β_n coincide with the eigenvalues of $\hat{\mathbf{A}}$ (see Appendix A.2). Nevertheless, when applying these transforms to non-circulant graphs, (higher-order) vanishing moments are lost and downsampling may become less intuitive and/or accurate.

and e-spline wavelet constructions also carry a ‘collateral’ annihilation property in the low-pass branch as a result of the spectral folding phenomenon, i.e. when \mathbf{H}_{HP} annihilates the eigenvector corresponding to $\beta_n = \gamma_n$, then \mathbf{H}_{LP} annihilates the eigenvector of $-\beta_n = -\gamma_n$. In the case of the simple spline filterbank, this translates to $\mathbf{u} = \mathbf{1}_N$ being annihilated by the high-pass and $\tilde{\mathbf{u}}$, with $\tilde{u}(i) = 1$ for $2i \in V$ and $\tilde{u}(i) = -1$ otherwise, by the low-pass filter.

The following corollaries capture restrictions on the invertibility of the graph e-spline wavelet transform for special cases of α_n :

Corollary 3.4. *The HGESWT ceases to be invertible for any downsampling pattern as well as fails to reproduce certain graph signals when $\exists \alpha_i, \alpha_j$ in $\vec{\alpha}$ such that $\tilde{d}_{\alpha_i} = -\tilde{d}_{\alpha_j}$ for \tilde{d}_{α_l} of the form $\sum_{k=1}^B 2d_k \cos\left(\frac{2\pi kl}{N}\right)$ for $l \in [0, N-1]$ and $2B < N$, including the case $\tilde{d}_{\alpha_i} = 0$.*

Proof. In the case of a general circulant graph, we have, for α_i, α_j as above, $H_{LP_{\alpha_i}}(z) = -H_{HP_{\alpha_j}}(z)$, or $\mathbf{H}_{LP_{\alpha_i}} = -\mathbf{H}_{HP_{\alpha_j}}$ in matrix form and vice versa, which leads to annihilation in the low-pass and reproduction in the high-pass branch.

Further, one cannot demonstrate linear independence of the eigenvectors associated with corresponding eigenvalues γ_i, γ_j , and hence invertibility of the HGESWT, for any downsampling pattern, which follows from Thm 3.2 mutatis mutandis; for brevity we refer to the complete proof in Appendix A.2. When $\tilde{d}_{\alpha_i} = 0$, the filterbank reduces to the normalized adjacency matrix $\frac{\mathbf{A}}{d}$ up to a sign per row (and its powers), which is singular if \mathbf{A} is singular, while its representer polynomial contains the zero root. \square

Corollary 3.5. *Let $\gamma^{DFT} = \{\gamma_i\}_i$ denote the DFT-ordered spectrum of $\frac{\mathbf{A}}{d}$ for $\frac{\mathbf{A}}{d} = \mathbf{V}\mathbf{T}\mathbf{V}^H$ with \mathbf{V} as the $N \times N$ DFT-matrix, and consider the HGESWT, with parameters of the form $\beta_k = \frac{\tilde{d}_{\alpha_k}}{d}$ for α_k in $\vec{\alpha}$. When downsampling is conducted with respect to $s = 1 \in S$, the HGESWT ceases to be invertible if $\exists \alpha_i, \alpha_j$ in $\vec{\alpha}$ for $\frac{\tilde{d}_{\alpha_i}}{d}, \frac{\tilde{d}_{\alpha_j}}{d} \in \gamma^{DFT}$, with respective multiplicities at positions $M_i = \{i_k\}_k$ and $M_j = \{j_k\}_k$ in γ^{DFT} , and such that some $\frac{\tilde{d}_{\alpha_i}}{d} = \gamma_i$ is located at position $(s + N/2)_N, s \in M_j \cup M_i$ of the DFT-ordered spectrum (and vice versa for \tilde{d}_{α_j}). When the graph is bipartite, this condition becomes equivalent to that of Cor. 3.4 for the fixed downsampling pattern.*

Proof. Given parameters of the form $\beta_i = \frac{\tilde{d}_{\alpha_i}}{d}$ which are contained in the spectrum γ^{DFT} of $\frac{\mathbf{A}}{d}$, we distinguish between the eigenvalues γ_i (and if existent, $-\gamma_i$) and their multiplicities, such that $|\beta_i| = |\gamma_i|$, with corresponding eigenvectors $\mathbf{V}_{\pm\gamma_i} = \{\mathbf{v}_{\pm i, l}\}_l$. Then the invertibility of the HGESWT is conditional upon the eigenvector sets $\mathbf{V}_{\pm\gamma_i}$ respectively being linearly independent after downsampling each vector by 2 to give $\mathbf{v}_{+i, l}(V_\alpha)$, with $V_\alpha = \{0 : 2 : N - 2\}$ as the set of retained nodes (see also Appendix A.2). Since

\mathbf{V} is the DFT-matrix and $\mathbf{V}(V_\alpha, 0 : N - 1) = [\tilde{\mathbf{V}} \ \tilde{\mathbf{V}}]$ with $\tilde{\mathbf{V}}$ as the DFT of dimension $N/2$ (up to a normalization constant), we observe that eigenvector pairs $(\mathbf{v}_k, \mathbf{v}_{k+N/2})$, at position $k \in [0 : N - 1]$ become linearly dependent. We therefore need to ensure that the parameters $\{\beta_i\}_i$ with $|\beta_i| = |\gamma_i|$ are chosen such that the corresponding values of $\{\gamma_i\}_i$ (and multiplicities) respectively do not take the aforementioned positions in the DFT-ordered spectrum; for existing $-\gamma_i$, the same relation and deductions hold for complement $\mathbf{V}_{-\gamma_i}(V_\alpha^C) = \{\mathbf{v}_{-i,l}(V_\alpha^C)\}_l$. When the graph is additionally bipartite, we note that given α_i, α_j at respective positions i, j , with $j = (i + N/2)_N$, due to the relation $\cos\left(\frac{2\pi k(i+N/2)}{N}\right) = -\cos\left(\frac{2\pi ki}{N}\right)$ for odd k , we have $\tilde{d}_{\alpha_i} = -\tilde{d}_{\alpha_j}$ and Cor. 3.4 applies. \square

3.4 Splines on Graphs

(Exponential) polynomial splines form a subset of the more generalized variational splines and are commonly characterized as the solutions of certain variational problems ([86], [87]). More specific to the signal processing community, the classical B -spline of degree zero β_+^0 is represented by the box-function

$$\beta_+^0(x) = \begin{cases} 1, & x \in [0, 1) \\ 0, & \text{otherwise} \end{cases}$$

and defined as $\beta_+^0(x) = \Delta_+ x_+^0$, i.e. through the action of a discrete (finite difference) operator $\Delta_+\{\cdot\}$, with z -transform $(1 - z^{-1})$, on the step function x_+^0 . Simultaneously, $x_+^0 = D^{-1}\{\delta(x)\}$ is the Green's function of continuous first-order differential operator $D\{\cdot\}$ [88]. The spline of degree n is then obtained through $(n + 1)$ -fold convolution of the box function $\beta_+^{(0,\dots,0)} = \beta_+^n(x) = \beta_+^0 * \beta_+^0 * \dots * \beta_+^0(x)$, and similarly defined through the higher-order operator $\Delta_+^{n+1}\{\cdot\}$, with z -transform $(1 - z^{-1})^{n+1}$, such that $\beta_+^n(x) = \frac{\Delta_+^{n+1} x_+^n}{n!}$, where x_+^n is the one-sided power function [88].

The relation $\Delta_+^m\{f\} = \beta_+^{m-1} * D^m\{f\}, \forall f \in S'^5$ and, more generally, $\Delta_+^\alpha\{f\} = \beta_+^\alpha * (D - \alpha I)\{f\}$ for operator $\Delta_+^\alpha(z) = (1 - e^\alpha z^{-1})$ and exponential spline β_+^α , with $\alpha \in \mathbb{C}$, derives the spline as the connection between continuous and discrete-time domain operators. In particular, $\Delta_+^\alpha\{\cdot\}$ can be regarded as a discrete approximation of continuous differential operator $(D - \alpha I)\{\cdot\}$ ([89], [88], [85]). The classical (convolved) discrete splines and e-splines are depicted in Fig. 3.2. Biorthogonal spline wavelet transforms in the Euclidean domain are commonly characterized by (dual) scaling functions which represent (combinations of) polynomial splines, with the Cohen-Daubechies-Feauveau wavelet as a prominent example [88].

In order to motivate a framework of splines in the graph domain, one first takes note

⁵Here, S' denotes Schwartz's class of tempered distributions.

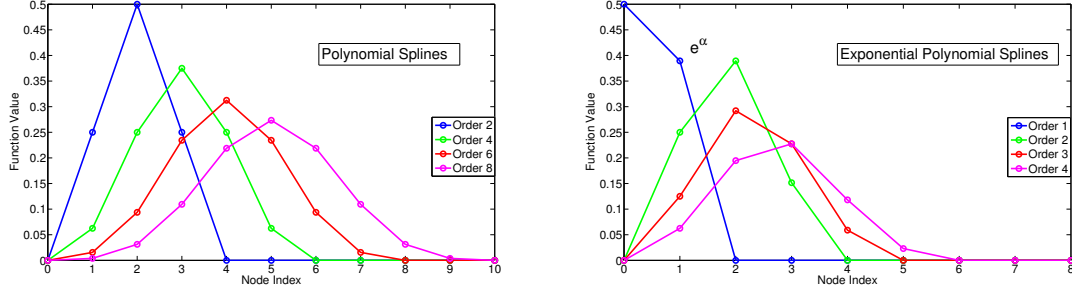


Figure 3.2: Classical Discrete Polynomial and Exponential Splines.

that the graph Laplacian matrix can be viewed as an approximation of the continuous Laplacian operator $-\nabla^2$ via the discrete Laplacian, nevertheless, a brute-force generalization of standard definitions and relations is primarily hindered by the fact that the graph Laplacian is a singular matrix.

In [90], variational splines on graphs, closely adhering to the traditional definition, are determined as the Green's functions of the approximate graph differential operator $(\mathcal{L} + \epsilon \mathbf{I}_N)^t$ for small $\epsilon > 0$ and $t \in \mathbb{R}^+$, which minimize the Sobolev norm, of the form $(\mathcal{L} + \epsilon \mathbf{I}_N)^{-t} \mathbf{e}_i$ for normalized graph Laplacian \mathcal{L} and elementary basis vector $\mathbf{e}_i \in \mathbb{R}^N$. In another line of work, Chung et al. [91] describe the Green's function of a connected graph, without direct reference to splines, as $G = \sum_{\lambda_j > 0} \frac{1}{\lambda_j} \mathbf{u}_j \mathbf{u}_j^H$, for normalized graph Laplacian eigenvectors \mathbf{u}_j and associated eigenvalues λ_j , and propose closed-form expressions for elementary cases, including the simple cycle graph, which can be further extended to Cartesian graph products [92]. Here, instead of resorting to approximations, a variation of the pseudo-inverted graph Laplacian operator is employed.

To substantiate the structural (and intuitive) link between the spline-like properties of the proposed graph wavelet functions and the traditional B-spline, consider the least connected example of a circulant graph, the simple cycle, which is denoted with $G_{S_1=(1)}$. In particular, the graph-representation of a signal (residing) on the vertices of a simple cycle graph can be regarded as an analogy to a periodic signal in the discrete-time domain, with existing edges indicating the sequence of sample values [1].

It appears that the rows and columns of the low-pass filter matrix in Eq. (3.1) of the *HGSWT*, given by

$$\mathbf{H}_{LP} := \begin{bmatrix} 0.5 & 0.25 & 0 & \cdots & 0 & 0.25 \\ 0.25 & 0.5 & 0.25 & 0 & \cdots & 0 \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & \ddots & \ddots \\ 0.25 & 0 & \cdots & 0 & 0.25 & 0.5 \end{bmatrix}^k,$$

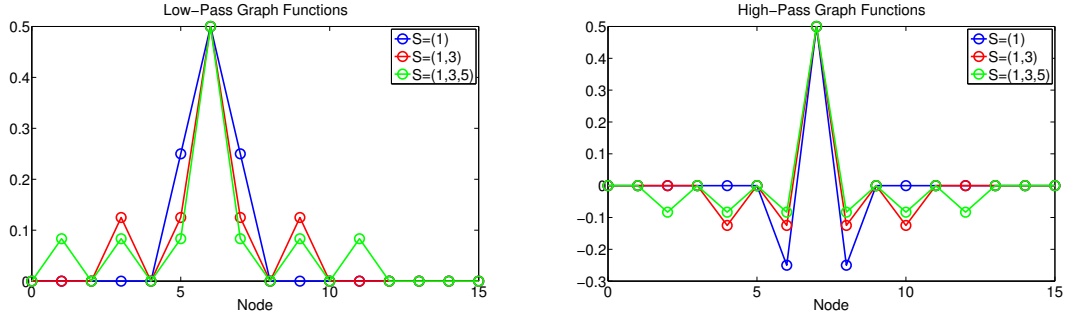


Figure 3.3: The *HGSWT* filter functions at $k = 1$ for different bipartite circulant graphs, $N = 16$. ©2017 Elsevier Inc.

produce the traditional discrete linear spline $\tilde{\beta}_+^{(0,0)}(t)$ for $k = 1$, which extends to the higher-order polynomial spline of order $2k$ as a result of convolution/circulant matrix multiplication. The associated (high-pass) graph Laplacian in Eq. (3.2), not only provides the stencil approximation of the second order differential operator for certain types of graphs such as lattices [55], but in the case of a simple cycle (and symmetric circulant graphs by extension) gains the actual vanishing moment property, thus completing the notion of a spline-wavelet filterbank. Coincidentally, the fact that unweighted lattice graphs can be expressed as the graph product of two path graphs, which are in turn circulant up to a missing edge, implies that the inherent vanishing property of a circulant graph is to some extent preserved via the product operation; this phenomenon will be further addressed in Sect. 4.5 of the next chapter on graph products.

Bipartite circulant graphs preserve the spline-property most distinctly, since their associated *HGSWT* filters retain both the reproduction and annihilation property of traditional spline-wavelets, as noted in Cor. 3.1, while for all other circulant graph cases, it is only the high-pass filter. Similarly, the rows and columns of the low-pass filter matrix \mathbf{H}_{LP_α} in Eq. (3.3) of the *HGESWT* (at $k = 1$) describe a second-order e-spline resulting from the convolution of two first-order complex conjugate e-splines

$$\mathbf{H}_{LP_\alpha} := \begin{bmatrix} 0.5 \cos(\alpha) & 0.25 & 0 & \cdots & 0 & 0.25 \\ 0.25 & 0.5 \cos(\alpha) & 0.25 & 0 & \cdots & 0 \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & \ddots & \ddots \\ 0.25 & 0 & \cdots & 0 & 0.25 & 0.5 \cos(\alpha) \end{bmatrix}.$$

Powers of \mathbf{H}_{LP_α} thus create classical higher-order polynomial e-spline basis functions, while multiplication of low-pass filters with different parameters α_n , as in Thm 3.2, results in convolved heterogeneous e-spline basis functions. One deduces from Cor. 3.2 that the exponential polynomial reproductive properties can be similarly extended to bipartite circulant graphs.

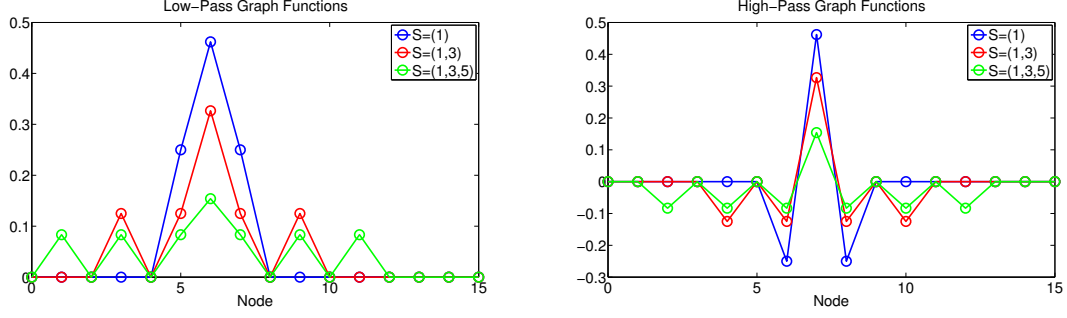


Figure 3.4: The *HGESWT* filter functions ($k = 1$) for different bipartite circulant graphs at $\alpha = \frac{2\pi}{N}$, $N = 16$. ©2017 Elsevier Inc.

e-Spline	Continuous Operator	Order	Graph (e-)Spline	Matrix Operator	Order
$\beta^{(0,0)}(t)$	$D^2\{\} = \frac{d^2}{dt^2}\{\}$	2	$(2d\mathbf{I}_N - \tilde{\mathbf{L}}_0)\mathbf{e}_i$	$\tilde{\mathbf{L}}_0$	2
$\beta^{(0,\dots,0)}(t)$	$D^{2n}\{\} = \frac{d^{2n}}{dt^{2n}}\{\}$	2n	$(2d\mathbf{I}_N - \tilde{\mathbf{L}}_0)^n\mathbf{e}_i$	$\tilde{\mathbf{L}}_0^n$	2n
$\beta^{(i\alpha, -i\alpha)}(t)$	$(D - i\alpha I) * (D + i\alpha I)\{\}$	2	$(2\tilde{d}_\alpha\mathbf{I}_N - \tilde{\mathbf{L}}_\alpha)\mathbf{e}_i$	$\tilde{\mathbf{L}}_\alpha$	2
$\beta^{(i\alpha, -i\alpha, \dots, i\alpha, -i\alpha)}(t)$	$(D - i\alpha I)^n * (D + i\alpha I)^n\{\}$	2n	$(2\tilde{d}_\alpha\mathbf{I}_N - \tilde{\mathbf{L}}_\alpha)^n\mathbf{e}_i$	$\tilde{\mathbf{L}}_\alpha^n$	2n
$\beta^{(i\alpha_1, -i\alpha_1, \dots, i\alpha_m, -i\alpha_m)}(t)$	$\prod_{t=1}^m (D - i\alpha_t I)^n * (D + i\alpha_t I)^n\{\}$	2mn	$\prod_{t=1}^m (2\tilde{d}_{\alpha_t}\mathbf{I}_N - \tilde{\mathbf{L}}_{\alpha_t})^n\mathbf{e}_i$	$\prod_{t=1}^m \tilde{\mathbf{L}}_{\alpha_t}^n$	2mn

Table 3.1: Continuous e-Spline and Graph e-Spline Definitions in Comparison. ©2017 Elsevier Inc.

Hence, for a bipartite circulant graph, the i -column of the generalized low-pass operator $(\tilde{d}_\alpha\mathbf{I}_N + \mathbf{A})\mathbf{e}_i = (2\tilde{d}_\alpha\mathbf{I}_N - \tilde{\mathbf{L}}_\alpha)\mathbf{e}_i$, which is coincidentally a generalized form of the similar signless Laplacian, for e-degree $\tilde{d}_\alpha = \sum_{k=1}^M 2d_k \cos(\alpha k)$, can be interpreted as a convolution of the discrete (e-)spline $\tilde{\beta}_+^{(i\alpha, -i\alpha)}(t)$ of order 2 with a function $\phi_G(t)$, which depends on the connectivity of the graph at hand. Thereby, a link to graphs is established and the notion of a *graph-spline* which converges to the classical discrete spline as $G_S \rightarrow G_{S_1=(1)}$ for an arbitrary circulant bipartite graph G_S with generating set S , becomes more concrete. In Figures 3.3 and 3.4, the low-and high-pass functions of the *HGSWT* and *HGESWT* are respectively compared for different bipartite graph examples which represent second-order graph (e-)splines, with the simple cycle case corresponding to the traditional (e-)spline.

In summary, the functions $(2\tilde{d}_\alpha\mathbf{I}_N - \tilde{\mathbf{L}}_\alpha)\mathbf{e}_i$ reveal spline-characteristics through their reproduction properties of (exponential) polynomials in form of convolutions of discrete (e-)spline basis functions on graphs, and hence structural similarity with classical counterparts, as well as through their description through a suitable differential operator, i. e. the parameterised graph Laplacian $\tilde{\mathbf{L}}_\alpha$ (and more distinctly, its signless counterpart), thus motivating the use of spline-terminology; however, it should be clarified that the analogy is not exact in that these do not constitute Green's functions of $\tilde{\mathbf{L}}_\alpha$. On top of that, the graph spline property can be interpreted more broadly in a GWT as the dual synthesis low-pass filter that is created through orthogonality to (a variation of) the e-graph Laplacian in the analysis high-pass branch (or vice versa), though the resulting function is not necessarily a classical polynomial as will be discussed in Sect. 3.5.

The variational graph splines of the form $(\mathcal{L} + \epsilon \mathbf{I}_N)^{-t} \mathbf{e}_i$, $t > 0$ [90] inherit certain properties for (exponential) polynomial reproduction when \mathcal{L} is circulant, however, contrary to the introduced spline constructions, this is only approximate, while the functions themselves are not well-characterized on the graph nor compactly supported to i. a. facilitate local annihilation, and therefore of lesser interest as basis functions for graph wavelets⁶.

The classes of spline-like functions on bipartite circulant graphs and their classical continuous counterparts are summarized and compared in Table 3.1 in relation to order, with the symmetrization $\beta^n(x) = \beta_+^n(x + \frac{n+1}{2})$.

3.4.1 The Directed Graph Spline

While the emphasis is on undirected graphs with real-valued node degrees and (symmetric) graph filters, we undertake a brief excursion to discuss special properties of *directed* circulant graphs, as they supply crucial insights that complete the previously discussed theory. In particular, the collective of results pertaining to vanishing moments of graph operators can be naturally extended to the case when \mathbf{A} is the adjacency matrix of a directed circulant graph \vec{G}_S , and the associated graph Laplacian is replaced by a first-order (normalized) difference operator of the form $\mathbf{S} = \mathbf{I}_N - \frac{\mathbf{A}}{d}$. Let \vec{G}_S be defined such that edge $(i, (i + s_k)_N)$ is directed from node i to $(i + s_k)_N$, for $s_k \in S$. Due to degree-regularity (i. e. the in- and out-degrees of each node are the same) and circularity, \mathbf{A} maintains the DFT-matrix as its basis.

A simple variation of previous proofs reveals that the representer polynomial of operator \mathbf{S} possesses one vanishing moment, while the degree-parameterised $\tilde{\mathbf{S}}_{\pm\alpha} = \frac{\tilde{d}_{\pm\alpha}}{d} \mathbf{I}_N - \frac{\mathbf{A}}{d}$, featuring the, now complex, e-degree $\tilde{d}_{\pm\alpha} = \sum_{k=1}^M d_k e^{\pm i\alpha s_k}$, per node, possesses one vanishing exponential moment, i. e. $\tilde{\mathbf{S}}_{\pm\alpha}$ respectively annihilate exponential graph signals with exponent $\pm i\alpha$. The same generalizations apply for higher-order k and combinations of basis functions. Equivalently, the low-pass filters $\mathbf{H}_{LP_{\pm\alpha}} = \frac{1}{2^k} \left(\frac{\tilde{d}_{\pm\alpha}}{d} \mathbf{I}_N + \frac{\mathbf{A}}{d} \right)^k$ on directed bipartite circulant graphs retain ‘adapted’ reproduction properties, whereby (linear combinations of) the rows of $\mathbf{H}_{LP_{\pm\alpha}}$ (at $k = 1$) reproduce exponentials with reversed exponent $\mp i\alpha$ (and vice versa for the columns).

It can be further shown that invertibility of the graph spline wavelet filterbank construction in Thm 3.1 remains intact for directed graphs at $k = 1$, as by the Perron Frobenius Thm. [93] for nonnegative matrices, $\frac{\mathbf{A}}{d}$ maintains an (albeit complex) spectrum with $|\gamma_i| < \gamma_{max}$ and $\gamma_{max} = 1$ of multiplicity 1 with eigenvector $\mathbf{1}_N$; here \mathbf{A} is required to be primitive, i. e. $\mathbf{A}^k > 0$ for some $k \in \mathbb{N}$ to ensure $|\gamma_i| < \gamma_{max}$ (Thm. 1.7, [93]). Otherwise, invertibility

⁶An invertible graph wavelet transform may be designed to have analysis filters as a variation of $\mathbf{H}_{LP/HP} = (\mathcal{L} + \epsilon \mathbf{I}_N)^{\pm t}$, as can be easily shown to be consequence of the generalized proof (see next chapter), however, an additional drawback for this type of (dense) construction is instability due to large condition numbers with the filters being close to singular.

of the transform for all remaining graph cases, following proper normalization, depends on the downsampling pattern and requires that the ‘downsampled’ eigenvectors associated with γ_i and $-\gamma_i$ (for $|\gamma_i| = 1$) form linearly independent sets respectively, yet is less evident. The proof of the graph e-spline wavelet transform (see Thm. 3.2) may be extended to accommodate directed graphs under further restrictions.

Directed graphs have been featured to a lesser extent in spectral graph theory. In particular, their treatment is complicated by the fact that the corresponding graph Laplacian matrices are no longer Hermitian, i.e. their eigenvalues (graph frequencies) are complex and the existence of an orthonormal eigenbasis (or general eigendecomposition) is not guaranteed. However in [94], the combinatorial graph Laplacian of a directed graph is defined as a symmetrization via the probability transition matrix \mathbf{P} . When the graph is directed, strongly connected and circulant, this translates into $\tilde{\mathcal{L}} = \mathbf{I}_N - \frac{\mathbf{A} + \mathbf{A}^H}{2d}$, which is equivalent to the normalized graph Laplacian of its undirected graph-counterpart. On the basis of this definition, one can express the undirected normalized e-graph Laplacian via the decomposition

$$\tilde{\mathcal{L}}_\alpha = \frac{\tilde{d}_\alpha}{2d} \mathbf{I}_N - \frac{\mathbf{A} + \mathbf{A}^H}{2d} = \frac{\tilde{\mathbf{S}}_\alpha + \tilde{\mathbf{S}}_\alpha^H}{2} = \frac{\tilde{\mathbf{S}}_{-\alpha} + \tilde{\mathbf{S}}_{-\alpha}^H}{2},$$

with $\tilde{d}_\alpha = \tilde{d}_{+\alpha} + \tilde{d}_{-\alpha}$. Consequentially, the annihilation property of the e-graph Laplacian on a directed circulant graph is preserved through the Hermitian transpose, where $\tilde{d}_{+\alpha}$ serves a dual function as the degree which annihilates complex exponentials with exponent $+i\alpha$ on the graph of \mathbf{A} and $-i\alpha$ on the graph of \mathbf{A}^H (and vice versa for $\tilde{d}_{-\alpha}$). Tying in with the previous discussion on graph spline similarities and analogies, the undirected graph Laplacian operator is a graph extension of a traditional second-order derivative operator, giving rise to graph spline-like functions and associated wavelets in degree steps of 2, which suggests that the directed first-order graph difference operator $\tilde{\mathbf{S}}_\alpha$ provides an extension to the traditional first-order differential operator. For the directed cycle, one therefore ascertains a comprehensive analogy with the traditional spline and e-spline definitions. Nevertheless, it should be distinguished that the graph operators \mathbf{S}^{2k} and \mathbf{L}^k for $\alpha = 0$, albeit possessing the same number of vanishing moments, do not describe the same graph.

3.5 Complementary Graph Wavelets

Since the classical operation of circularly convolving a discrete-time signal with a filter can be defined as the matrix-vector product between a circulant matrix and the given signal-vector, the traditional technique of spectral factorisation in the z -domain is directly extendable to circulant graphs and establishes a convenient analogy to the graph domain.

3.5.1 The Bipartite Semi-IIR Graph Filterbank

Prior to broadening the range of graph wavelet constructions, further analysis of the characteristic structure of current derivations is conducted. The case of a bipartite circulant graph is of particular interest due to its dual vanishing moment property. Contrary to standard biorthogonal wavelet filterbanks, the proposed graph spline wavelet constructions only exhibit well-defined analysis filters, while their corresponding synthesis filters are not concretely characterized. Consider the general analysis matrix of a circulant graph wavelet transform

$$\mathbf{W} = \begin{bmatrix} \Psi_{\downarrow 2} \mathbf{H}_{LP} \\ \Phi_{\downarrow 2} \mathbf{H}_{HP} \end{bmatrix},$$

with downsampling matrices $\Psi_{\downarrow 2}, \Phi_{\downarrow 2} \in \mathbb{R}^{N/2 \times N}$, which respectively retain even and odd numbered nodes such that

$$\Psi_{\downarrow 2}(i, j) = \begin{cases} 1, & j = 2i - 1 \\ 0, & \text{otherwise} \end{cases} \quad \Phi_{\downarrow 2}(i, j) = \begin{cases} 1, & j = 2i \\ 0, & \text{otherwise} \end{cases}$$

and graph-based low-and high-pass filters $\mathbf{H}_{LP}, \mathbf{H}_{HP}$, whose respective representer polynomials are given by $H_{LP}(z), H_{HP}(z)$. Since \mathbf{W} is invertible, and contains two sets of basis functions, their corresponding duals, denoted by $\tilde{H}_{LP}(z), \tilde{H}_{HP}(z)$ are formed in $\mathbf{W}^{-1} = \begin{bmatrix} (\Psi_{\downarrow 2} \tilde{\mathbf{H}}_{LP})^T & (\Phi_{\downarrow 2} \tilde{\mathbf{H}}_{HP})^T \end{bmatrix}$, by design of a biorthogonal system [95]. In general, the following relations hold for $i, j \in \{LP, HP\}$

$$H_i(z) \tilde{H}_i(z) + H_i(-z) \tilde{H}_i(-z) = 2 \quad (3.5)$$

$$H_i(z) \tilde{H}_j(z) + H_i(-z) \tilde{H}_j(-z) = 0, \quad i \neq j \quad (3.6)$$

with $H_{HP}(z) = zH_{LP}(-z)$ emerging from the current constructions when the graph is bipartite and circulant⁷; by substitution into the above, it becomes apparent that $\tilde{H}_{HP}(z) = z^{-1} \tilde{H}_{LP}(-z)$ must equivalently hold for the dual pair.

In particular, the synthesis filters can be explicitly derived from the perfect reconstruction condition [78] on the modulation matrices $\mathbf{H}_i(z) \tilde{\mathbf{H}}_i(z) = 2\mathbf{I}$

$$\begin{pmatrix} H_{LP}(z) & H_{HP}(z) \\ H_{LP}(-z) & H_{HP}(-z) \end{pmatrix} \begin{pmatrix} \tilde{H}_{LP}(z) & \tilde{H}_{LP}(-z) \\ \tilde{H}_{HP}(z) & \tilde{H}_{HP}(-z) \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}. \quad (3.7)$$

Following inversion of the analysis matrix, and substituting $H_{HP}(z) = zH_{LP}(-z)$, we obtain

$$\begin{pmatrix} \tilde{H}_{LP}(z) & \tilde{H}_{LP}(-z) \\ \tilde{H}_{HP}(z) & \tilde{H}_{HP}(-z) \end{pmatrix} = \frac{2}{D(z)} \begin{pmatrix} H_{HP}(-z) & -H_{HP}(z) \\ -H_{LP}(-z) & H_{LP}(z) \end{pmatrix}$$

⁷In an abuse of notation, we incorporate a shift z , equivalently to the traditional z -transform, to signify that odd-numbered rows retain the high-pass component.

with determinant $D(z) = H_{HP}(-z)H_{LP}(z) - H_{HP}(z)H_{LP}(-z) = -z(H_{LP}(z)H_{LP}(z) + H_{LP}(-z)H_{LP}(-z)) \neq 0$ and $D(z) = -D(-z)$. The resulting synthesis filters are thus given by

$$\tilde{H}_{LP}(z) = \frac{H_{HP}(-z)2}{D(z)} = \frac{2H_{LP}(z)}{H_{LP}(z)H_{LP}(z) + H_{LP}(-z)H_{LP}(-z)}$$

and

$$\tilde{H}_{HP}(z) = -\frac{H_{LP}(-z)2}{D(z)} = \frac{2z^{-1}H_{LP}(-z)}{H_{LP}(z)H_{LP}(z) + H_{LP}(-z)H_{LP}(-z)},$$

confirming the relation $\tilde{H}_{HP}(z) = z^{-1}\tilde{H}_{LP}(-z)$.

The derived synthesis filters now describe rational functions whose zeros and poles are respectively the zeros of the numerator and denominator [81], which entails that the vanishing moments of the analysis filters are, in fact, preserved. Here, the order of the polynomial in the denominator clearly exceeds that of the numerator. In classical SP, these type of filters with typically exponentially decaying coefficients are known to induce an infinite impulse response (IIR) and occur i. a. in semi-orthogonal filterbanks [78].

3.5.2 Design and Discussion

At last, a generalized approach is presented for the tailored design of circulant graph wavelet filterbanks, whose vertex-localized (analysis) low-pass filters can acquire reproduction properties, while the high-pass filter is maintained as is.

In previous transforms, both low-and high-pass filters are well-defined in the analysis domain and have compact support of the same length $2Mk + 1$, based on the given graph bandwidth, yet, as a result, lack a concrete representation of their synthesis filters whose support is comparatively larger (exponentially decaying). Further, dual vanishing moments emerge, by design, only for bipartite circulant graphs. In light of this, a new class of graph wavelet filterbanks is developed through traditional spectral factorization techniques, ordinarily employed for the creation of biorthogonal perfect reconstruction filterbanks, and more recently used within a variation in [36] to create bipartite spectral graph filters, to implement the desired properties. These novel transforms are composed of well-defined analysis and synthesis filters of compact support and can be related to the previous filterbanks via a symmetric circulant transformation filter \mathbf{C} in the low-pass branch, depending on the invertibility of the low-pass filters in Eqs. (3.1) and (3.3).

As a result of spectral factorization, it can be ensured that the graph filter matrices in both the analysis and synthesis branch are of finite and balanced bandwidth for $2Mk < N$. Nevertheless, as a limitation of this type of filterbank, only the standard alternating down-sampling pattern for circulant graphs, which samples every other node, can be enforced, contrary to the preceding constructions.

Graph Spline

Consider the case of a simple spline for $\alpha = 0$. Given analysis high-pass filter $H_{HP}(z) = \frac{l(z)^k}{(2d)^k}$ with $2k$ vanishing moments, the synthesis lowpass filter is determined as $\tilde{H}_{LP}(z) = H_{HP}(-z)$ and thus, analysis lowpass filter $H_{LP}(z)$ can be derived from the biorthogonality relations of a traditional filterbank [95], with $P(z) = H_{LP}(z)\tilde{H}_{LP}(z)$ subject to the constraint of the half-band condition⁸ $P(z) + P(-z) = 2$. By additionally imposing symmetry (for an undirected graph) on the resulting filter $H_{LP}(z) = \sum_{i=0}^T r_i(z^i + z^{-i})$, one arrives at the equality

$$P(z) = 1 + \sum_{i=0}^L p_{2i+1}(z^{2i+1} + z^{-(2i+1)}) = \frac{1}{(2d)^k} \left(d - \sum_{i=1}^M (-1)^i d_i(z^i + z^{-i}) \right)^k \left(\sum_{i=0}^T r_i(z^i + z^{-i}) \right) \quad (3.8)$$

where $P(z)$ is a polynomial of odd powers. One may additionally impose that analysis and synthesis filters have an equal number of vanishing moments $2k$, by setting $H_{LP}(z) = (z+1)^k(z^{-1}+1)^k R(z)$, where $R(z)$ is the polynomial to be determined.

At $k = 1$, the highest degree of each side of Eq. (3.8) is $2L + 1 = M + T$, resulting in a linear system with $L + 1$ constraints $p_{2n} = 0$, $n = 1, \dots, L$, and $p_0 = 1$, and $T + 1$ unknowns r_i , $i = 0, \dots, T$. In order to obtain a unique solution, we require $L = T = \frac{M+T-1}{2}$, or $T = M - 1$. At higher-order with $k > 1$, the constraints change as follows: $T = L = \frac{Mk+T-1}{2}$, or $T = Mk - 1$. For both synthesis and analysis filters to have an equal number of vanishing moments, we further need to include the additional factor $(z+1)^k(z^{-1}+1)^k$ for $H_{HP}(z) = \frac{1}{(2d)^k} l(z)^k$, and require $T = Mk + k - 1$.

The necessary existence of a complementary analysis low-pass filter for a given high-pass filter on a graph G , follows from the Bézout theorem:

Theorem 3.3 (Bézout [25]). *Given $C(z) \in \mathbb{R}[z]$, there exists a polynomial $D(z) \in \mathbb{R}[z]$ such that*

$$C(z)D(z) + C(-z)D(-z) = 2$$

if and only if $C(z)$ has neither zero as a root, nor a pair of opposite roots. In this case, there exists a unique polynomial $D_0(z) \in \mathbb{R}[z]$ satisfying the above and such that $\deg D_0(z) \leq C(z) - 1$. The set of all polynomials $D(z) \in \mathbb{R}[z]$ that satisfy the above is

$$\{D_0(z) + z\lambda(z^2)C(-z), \quad \lambda(z) \in \mathbb{R}[z]\}.$$

It can be directly observed that graph Laplacian-based filter $C(z) = H_{HP}(-z)$ cannot

⁸Since the general set up of the proposed graph filterbanks is such that even-numbered nodes retain the low-pass and odd-numbered nodes the high-pass component, one is technically considering orthogonality between the shifted $zH_{HP}(z)$ and its dual $\tilde{H}_{LP}(z) = -z^{-1}((-z)H_{HP}(-z))$, but this notation is omitted for simplicity.

have a zero root since $d = \sum_{i=1}^M 2d_i > 0$ (as graph weights are nonnegative $d_i \geq 0$). If the generating set S of the graph at hand consists of only even elements, $C(z)$ contains pairs of opposing roots such that $H_{HP}(-z) = H_{HP}(z)$; however, G is assumed to be connected with $s = 1 \in S$, so this cannot occur.⁹

One can equivalently resort to spectral factorization on bipartite circulant graphs, however, since both $P(z)$ and $H_{HP}(z)$ are odd degree polynomials, $R(z)$ is required to be of higher degree T than the remaining factor in $P(z)$ in order to produce a non-trivial solution; the resulting underdetermined linear system can be uniquely solved by imposing additional constraints on the coefficients r_i (such as roots at $z = -1$).

The proposed biorthogonal graph spline wavelet constructions for circulant graphs are captured in the following theorem:

Theorem 3.4. *Given the undirected, and connected circulant graph $G = (V, E)$ of dimension N , with adjacency matrix \mathbf{A} and degree d per node, define the higher-order ‘complementary’ graph-spline wavelet transform (HCGSWT) via the set of analysis filters:*

$$\mathbf{H}_{LP,an} \stackrel{(*)}{=} \mathbf{C} \bar{\mathbf{H}}_{LP} = \frac{1}{2^k} \mathbf{C} \left(\mathbf{I}_N + \frac{\mathbf{A}}{d} \right)^k \quad (3.9)$$

$$\mathbf{H}_{HP,an} = \frac{1}{2^k} \left(\mathbf{I}_N - \frac{\mathbf{A}}{d} \right)^k \quad (3.10)$$

and the set of synthesis filters:

$$\mathbf{H}_{LP,syn} = c_1 \mathbf{H}_{HP,an} \circ I_{HP} \quad (3.11)$$

$$\mathbf{H}_{HP,syn} = c_2 \mathbf{H}_{LP,an} \circ I_{LP} \quad (3.12)$$

where $\mathbf{H}_{LP,an}$ is the solution to the system from Eq. (3.8) under specified constraints, with coefficient matrix \mathbf{C} arising from the relation $\mathbf{H}_{LP,an} \bar{\mathbf{H}}_{LP}^{-1}$ when $\bar{\mathbf{H}}_{LP}$ is invertible (see Cor. 3.3). Here, \circ is the Hadamard product, $c_i, i \in \{1, 2\}$ are normalization coefficients, and $I_{LP/HP}$ circulant indicator matrices with first row of the form $[1 \ -1 \ 1 \ -1 \ \dots]$.

Proof. Follows from above discussion.

As a result of spectral factorization, the shape and vertex spread of $\mathbf{H}_{LP,an}$ does not directly coincide with the adjacency matrix of the graph (and its powers), but rather encompasses a subset $S_i \subseteq N(i, \tilde{k})$ of vertices, per node i within its \tilde{k} -hop local neighborhood, for some \tilde{k} dependent on the initial constraints on $H_{LP,an}(z)$.

To establish a structural link to the analysis branch of the HGSWT in Thm 3.1, and in particular, to the (adjacency matrix-based) low-pass graph filter $\bar{\mathbf{H}}_{LP}$ of Eq. (3.1), a

⁹This is a phenomenon one would however observe for the adjacency matrix of a bipartite circulant graph with power $k = 2N$, since the diagonals are zero.

transitional graph filter is formed with the symmetric circulant coefficient matrix \mathbf{C} in (*), which is determined via matrix inversion of $\bar{\mathbf{H}}_{LP}$ (when G is non-bipartite, see Cor. 3.3 for $\alpha = 0$).

Given initial graph signal $\mathbf{p} \in \mathbb{R}^N$ and its GWT-representation $\tilde{\mathbf{p}} \in \mathbb{R}^N$ from Sect. 3.2, the synthesis stage can thus be explicitly expressed as follows:

$$\left(\frac{1}{2} ((\mathbf{I}_N + \mathbf{K})\mathbf{H}_{LP, syn})^T + \frac{1}{2} ((\mathbf{I}_N - \mathbf{K})\mathbf{H}_{HP, syn})^T \right) \tilde{\mathbf{p}} = \mathbf{p},$$

for $K_{i,i} = 1$ at even-numbered positions and $K_{i,i} = -1$ otherwise.

Graph E-spline

The process is analogous for the graph e-spline case and we further proceed to create complementary graph e-spline filterbanks with filters that can reproduce and annihilate signals within the generalized class of complex exponential polynomials.

According to (Thm. 1, [25]), a scaling function $H_j(z)$ at level j can reproduce a function of the form $P(t)e^{\gamma_m t}$, where $\deg P(t) \leq (L_m - 1)$ for multiplicity L_m of γ_m , if and only if the former is divisible by the term $R_{2^j \vec{\gamma}}(z)$, $\forall j \leq j_0 - 1$, where $R_{\vec{\gamma}}(z) = \prod_{m=1}^M (1 + e^{\gamma_m} z^{-1})$, with $\vec{\gamma} = (\gamma_1, \dots, \gamma_M)^T \in \mathbb{C}^M$, and $H_j(z)$ has no roots of opposite sign, i. e. $H_j(z)$ satisfies the generalized Strang-Fix conditions for suitable $\vec{\gamma}$.

Mirroring the constructions of Thm. 3.4, analysis lowpass filter $H_{LP_\alpha}(z)$ is determined from analysis high-pass filter $H_{HP_\alpha}(z) = \frac{\tilde{l}_\alpha(z)}{2d}$ with 2 vanishing exponential moments, and can be expressed as an extension of Eq. (3.3) via coefficient matrix \mathbf{C} (subject to constraints, see Cor. 3.3). By imposing the constraints of Bézout's Thm. [25], and setting $P(z) = H_{LP_\alpha}(z)H_{HP_\alpha}(-z)$, the previous equality takes the form

$$P(z) = 1 + \sum_{i=0}^L p_{2i+1}(z^{2i+1} + z^{-(2i+1)}) = \frac{1}{2d} \left(\tilde{d}_\alpha - \sum_{i=1}^M (-1)^i d_i (z^i + z^{-i}) \right) \left(\sum_{i=0}^T r_i (z^i + z^{-i}) \right) \quad (3.13)$$

whose emerging linear system can be solved for unknown symmetric coefficients r_i of $H_{LP_\alpha}(z)$, in a similar fashion as discussed for Eq. (3.8).

Moreover, one may impose $H_{LP_\alpha}(z) = (z + e^{i\alpha})(1 + e^{-i\alpha} z^{-1})R(z)$, with unknown polynomial $R(z)$, for the analysis and synthesis filters to have (an equal number of) vanishing moments. Generalizations to multiple parameters $\vec{\alpha} = (\alpha_1, \dots, \alpha_T)$ and higher order $k > 1$ are realized by defining the high-pass filter $H_{HP_{\vec{\alpha}}}(z) = \prod_{n=1}^T \frac{1}{(2d)^k} \tilde{l}_{\alpha_n}(z)^k$ (as in Thm. 3.2), and $H_{LP_{\vec{\alpha}}}(z) = \prod_{n=1}^T (z + 2\cos(\alpha_n) + z^{-1})^k R(z)$, however, a solution $R(z)$ exists only if the remainder term in $P(z)$ does not contain zero and/or opposing roots [25].

The existence of a suitable low-pass filter $\mathbf{H}_{LP_{\vec{\alpha}, an}}$ which completes the above filterbank is conditional upon the representer polynomial $H_{HP_{\vec{\alpha}, an}}(z)$ of $\mathbf{H}_{HP_{\vec{\alpha}, an}}$ yielding no opposing

or zero roots for given $\vec{\alpha}$, thereby satisfying Bézout's Theorem ([7], [25]). For instance, opposing roots occur for the case $e^{-i\alpha} = -e^{i\alpha}$, at $\alpha = \pi/2, 3\pi/2$, and one cannot create filterbanks for these parameters.

Similarly, a multiresolution representation of the filterbank can be realized if real-valued filters exist, which maintain their reproduction/annihilation properties up to a certain level $j \leq J - 1$. In the Euclidean domain, the filters of a non-stationary biorthogonal exponential wavelet filterbank with exponents $\vec{\alpha} = (\alpha_1, \dots, \alpha_T)$ do not contain roots of opposite sign nor the zero root at level j , as long as there are no distinct α, α' in $\vec{\alpha}$ that satisfy $2^j(\alpha - \alpha') = i(2k + 1)\pi$, for some $j \leq J - 1$ and $k \in \mathbb{Z}$ [25]; otherwise, a multilevel representation is only possible up to a finite level $J - 1$, when this condition ceases to be fulfilled. This result becomes particularly relevant when considering multi-parameter graph e-spline wavelet filterbanks. As an extension of the latter, one can further deduce from Cor. 3.5 on the *HGESWT* conditions on the existence of a complementary filterbank, given a parameterized high-pass filter. The discussed approach gives rise to the following filterbank:

Theorem 3.5. *Given the undirected, and connected circulant graph $G = (V, E)$ of dimension N , with adjacency matrix \mathbf{A} and degree d per node, we define the higher-order ‘complementary’ graph e-spline wavelet transform (*HCGESWT*) via the set of analysis filters:*

$$\mathbf{H}_{LP_{\vec{\alpha}}, an} \stackrel{(*)}{=} \mathbf{C} \bar{\mathbf{H}}_{LP_{\vec{\alpha}}} = \mathbf{C} \prod_{n=1}^T \frac{1}{2^k} \left(\beta_n \mathbf{I}_N + \frac{\mathbf{A}}{d} \right)^k \quad (3.14)$$

$$\mathbf{H}_{HP_{\vec{\alpha}}, an} = \prod_{n=1}^T \frac{1}{2^k} \left(\beta_n \mathbf{I}_N - \frac{\mathbf{A}}{d} \right)^k \quad (3.15)$$

and the set of synthesis filters:

$$\mathbf{H}_{LP_{\vec{\alpha}}, syn} = c_1 \mathbf{H}_{HP_{\vec{\alpha}}, an} \circ I_{HP} \quad (3.16)$$

$$\mathbf{H}_{HP_{\vec{\alpha}}, syn} = c_2 \mathbf{H}_{LP_{\vec{\alpha}}, an} \circ I_{LP} \quad (3.17)$$

where $\mathbf{H}_{LP_{\vec{\alpha}}, an}$ is the solution to the system from Eq. (3.13) for $\vec{\alpha}$ under specified constraints, with coefficient matrix \mathbf{C} arising from the relation $\mathbf{H}_{LP_{\vec{\alpha}}, an} \bar{\mathbf{H}}_{LP_{\vec{\alpha}}}^{-1}$ where applicable (see Cor. 3.3). Here, $c_i, i \in \{1, 2\}$ are normalization coefficients, and $I_{LP/HP}$ are circulant indicator matrices with first row of the form $[1 \ -1 \ 1 \ -1 \ \dots]$.

At $\vec{\alpha} = \mathbf{0}$, this coincides with the previous graph spline wavelet filterbank.

While the proposed design is applicable to bipartite circulant graphs, it is less relevant, as the transform given in Thm. 3.2 already provides the desired reproduction properties, except when compactly supported synthesis filters are desired. As an interesting aside,

however, for a bipartite circulant graph, the special case from Cor. 3.4 in Sect. 3.3.3 can be further extended to provide scenarios which violate Bézout's Thm. and hence the existence of a complementary filterbank construction when downsampling is conducted with respect to $s = 1 \in S$. Specifically, when $\exists \alpha_i, \alpha_j$ in $\vec{\alpha}$ such that $\tilde{d}_{\alpha_i} = -\tilde{d}_{\alpha_j}$ for \tilde{d}_{α} , the representer polynomial of the high-pass filter product $\mathbf{H}_{HP_{\alpha_i}} \mathbf{H}_{HP_{\alpha_j}}$ contains opposing roots due to $H_{HP_{\alpha_j}}(z) = -H_{HP_{\alpha_i}}(-z)$. We will further revisit this discussion in Ch. 5 on sampling.

3.6 Computational Experiments

The proposed graph wavelet designs are exemplified and assessed in the following experiments, which compare their non-linear approximation performance for smooth signals in both clean and noisy states; these primarily serve an illustrative purpose as opposed to providing a comprehensive numerical study, in light of the variation present for different choices of graphs.

Let signals reside on the circulant graph of dimension $N = |V| = 2^{10} = 1024$ and generating set $S = (1, 2)$, and consider 5 levels of the graph (e-)spline wavelet transform (see Thms. 3.1 and 3.2) in comparison with complementary GWT constructions (see Thms. 3.4 and 3.5), which feature the same analysis high-pass filter, suitably parameterized to annihilate the signal \mathbf{x} at hand, and a variable low-pass filter. The *HCGESWT* is presented in two variations with either dual or unilateral (on the analysis side) vanishing moments, denoted with the suffixes *M.M* and *M.0* respectively for number of vanishing moments M . The wavelet atoms (rows) have been normalized to length 1 and reconnection is conducted by retaining the same generating set.

The non-linear approximation of a graph signal $\mathbf{x} \in \mathbb{R}^N$ within a given graph basis is defined as $\hat{\mathbf{x}} = \sum_{k \in I_K} w_k \hat{\mathbf{w}}_k$, for graph wavelet coefficients $\{w_k\}_k$ and basis functions $\{\hat{\mathbf{w}}_k\}_k$ of the inverse GWT, where I_K is the index set of the K -largest magnitude coefficients w_k . In the presence of white Gaussian noise \mathbf{n} , \mathbf{x} is recovered from the graph wavelet representation $\tilde{\mathbf{w}}$ of $\tilde{\mathbf{x}} = \mathbf{x} + \mathbf{n}$ via hard thresholding by setting a prescribed number $T = N - K$ of the smallest graph wavelet coefficients to zero and subsequently applying the inverse transformation. In the treated examples, this is done at several different levels of noise, for a fixed percentage close (or equal) to the sparsifying threshold, i.e. minimum number of non-zero graph wavelet coefficients required for perfect reconstruction, of either of the transforms.¹⁰ The reconstruction error is measured in dB via $SNR = 10 \log_{10} \frac{\|\mathbf{x}\|_2^2}{\|\tilde{\mathbf{x}} - \hat{\mathbf{x}}\|_2^2}$ and,

¹⁰This assumes that the type and order of the clean signal is known, however, the exact threshold level can also be directly computed from the parameters of the transform.

in the noisy case, averaged over 10 trials per noise level; perfect reconstruction is indicated via a large plateaued value $SNR > 200$ dB, as per MATLAB's numerical precision. Performance of the transforms is then assessed with respect to the fraction of retained non-zero wavelet coefficients K/N , and, in the noiseless case, sparsity K can be explicitly stated through a formula, given in Cor. 5.1 of subsequent Sect. 5.1.3.

Figure 3.5 shows the performance for a linear polynomial graph signal, and it becomes apparent that best performance under noise is achieved by the complementary construction with dual vanishing moments, while all transforms exhibit the same sparsity (efficiency) level in the clean state.

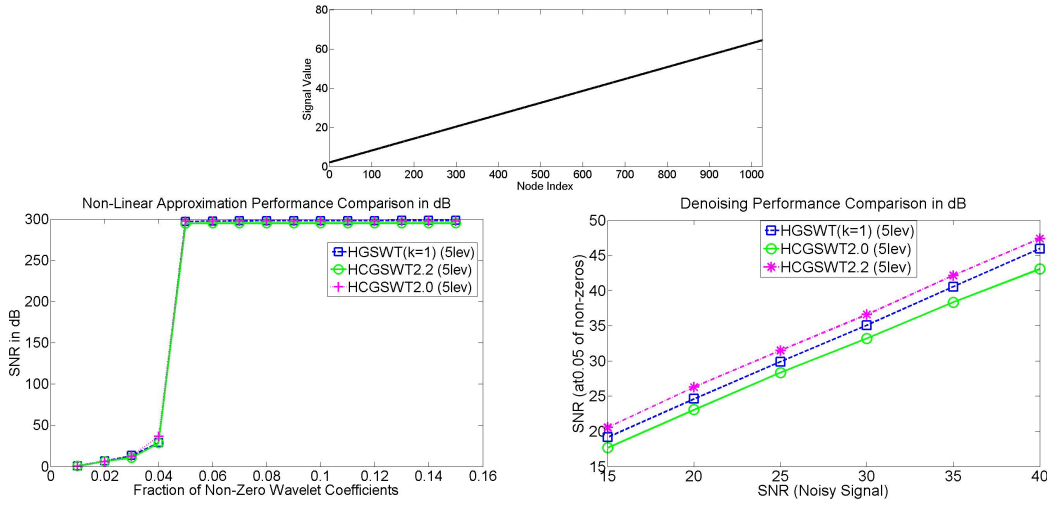


Figure 3.5: Comparison of NLA (lower left) and denoising performance (lower right) for a linear polynomial (top).

In addition, we consider the same linear polynomial graph signal on a slightly more connected circulant graph with $S = (1, 2, 3, 4)$, which reveals a shift in performance, particularly noting that the matrix-based transform outperforms the complementary constructions under noise, which in turn prove destructive as a result of an increased bandwidth (see Fig. 3.6).

In Figure 3.7, we consider the graph e-spline based constructions for a sinusoidal graph signal, which due to lack of a border effect, are equally efficient when there is no noise, while the dual complementary construction again prevails in the noisy scenario. At last, Fig. 3.8 presents a sum of sinusoidals, which is best represented under noise by the dual complementary e-spline transform, while the matrix-based construction suffers a distinct drop in performance, indicating a higher susceptibility to noise at higher order.

Table 4.1 lists the condition numbers of the specified multilevel graph wavelet matrices, which provide an indication, although not a universal explanation, for the observed performances under noise. While the tailored complementary designs possess a consistently

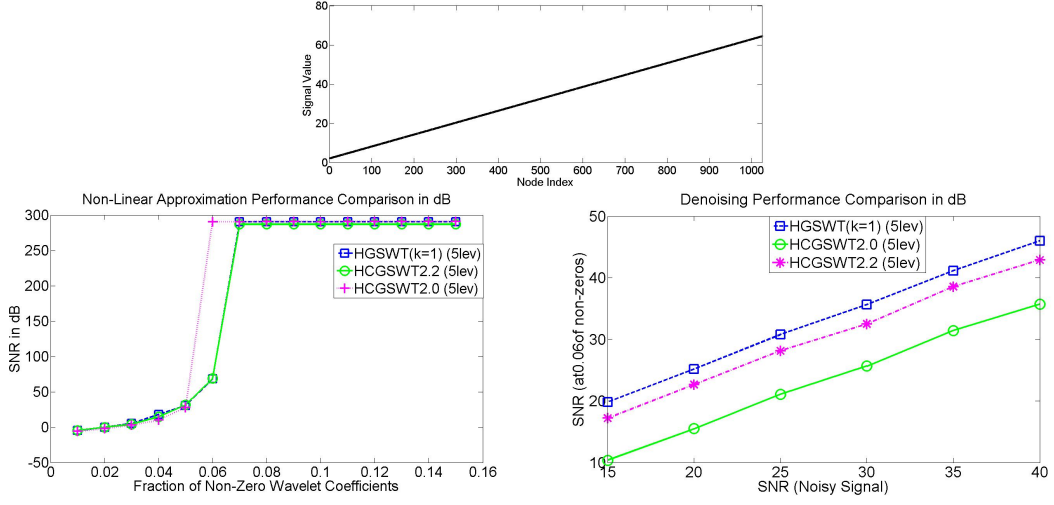


Figure 3.6: Comparison of NLA (lower left) and denoising performance (lower right) for a linear polynomial (top) on a circulant graph with $S = (1, 2, 3, 4)$.

lower condition number, in keeping with their mostly superior/solid performance, the purely graph matrix-based designs gain large condition numbers as the number of parameters increase, which proves destructive. This downside will be further investigated (and partially remedied) in the next chapter.

Graph Generating Set	Type	Order	Parameters $\vec{\alpha}$	Cond. No
{1, 2}	HGSWT	$k = 1$	0	14.98
	HCGSWT	2.0		13.13
	HCGSWT	2.2		13.00
	HGESWT	$k = 1$	$\frac{2\pi 4}{N}$	19.60
	HCGESWT	2.0		12.93
	HCGESWT	2.2		13.58
	HGESWT	$k = 1$	$(\frac{2\pi 1}{N}, \frac{2\pi 5}{N})$	331.69
	HCGESWT	4.0		41.47
	HCGESWT	4.4		49.37
{1, 2, 3, 4}	HGSWT	$k = 1$	0	16.07
	HCGSWT	2.0		35.16
	HCGSWT	2.2		16.03

Table 3.2: Condition Numbers of the GWTs.

Nevertheless, performance also depends on the graph topology, whereby an increase in the number of connections (or graph band) may similarly contribute to decrease in performance (at any order); it would appear that among the graph matrix-based constructions, the condition number of the simple spline is less dramatically affected by an increase in connections (which will also be proved in Ch. 4 for bipartite graphs).

The individual (normalized) graph filter-functions for selected cases (at level 1) are depicted in Figs. 3.9 – 3.11. One expectedly observes that with an increase in the number of parameters (or order) and graph connectivity, the graph filter functions (especially in

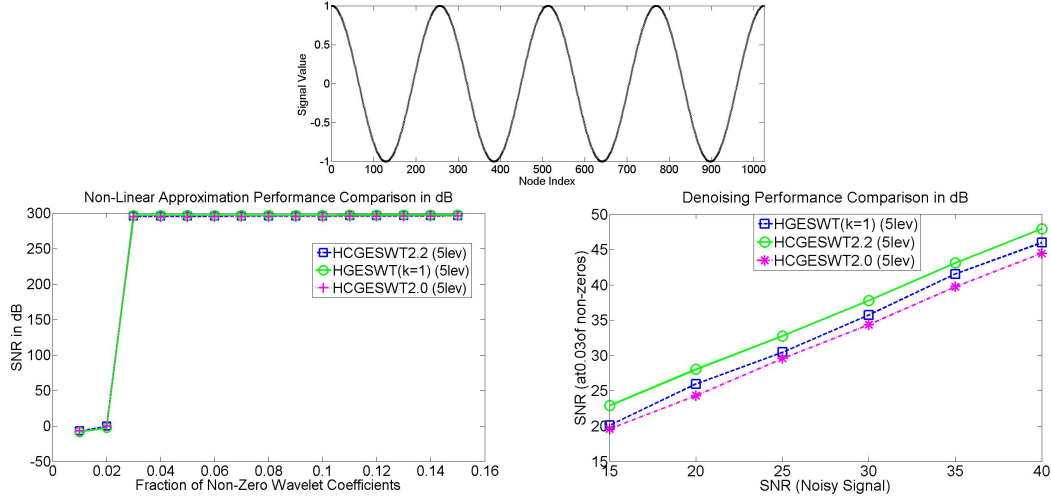


Figure 3.7: Comparison of NLA (lower left) and denoising performance (lower right) for a sinusoidal with $(\alpha = \frac{2\pi^4}{N})$ (top).

the synthesis domain) become less smooth.

To summarize, this chapter has introduced a range of novel graph (e-)spline wavelet transforms on circulant graphs, and provided a comprehensive analysis of their properties and implications for both spectral graph theory and GSP as well as established their link to traditional signal processing on the basis of the theory of circulants.

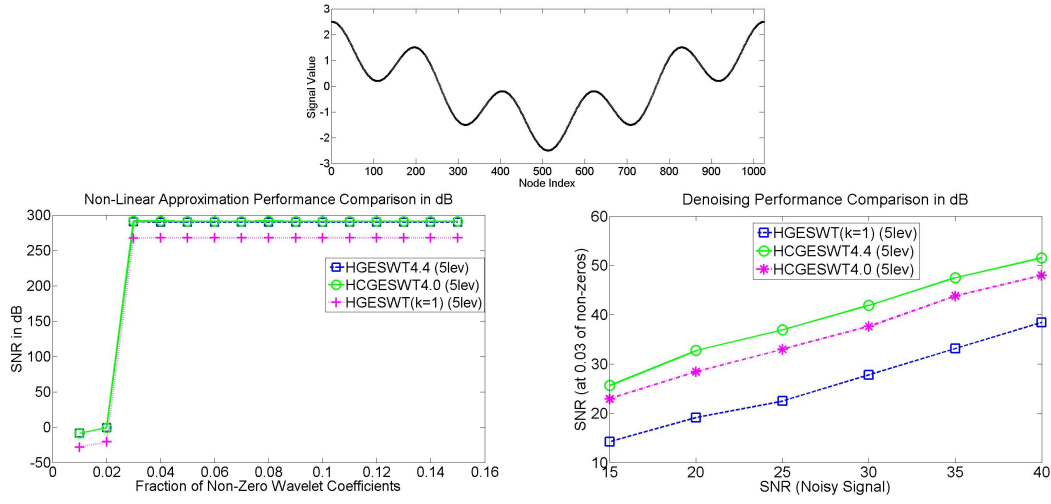


Figure 3.8: Comparison of NLA (lower left) and denoising performance (lower right) for a sum of sinusoids with $(\alpha_1 = \frac{2\pi 1}{N}, \alpha_2 = \frac{2\pi 5}{N})$ (top). ©2017 Elsevier Inc.

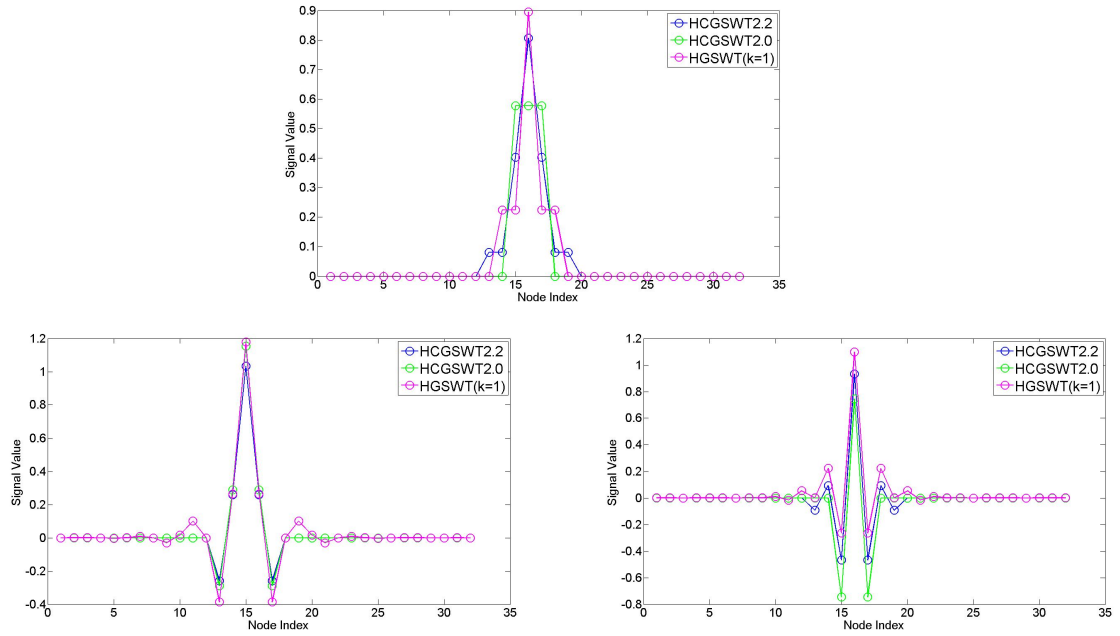


Figure 3.9: Comparison of Graph-filter functions: analysis low-pass (top), synthesis low- and high-pass (from left) at one level for the linear spline constructions on the graph with $S = (1, 2)$.

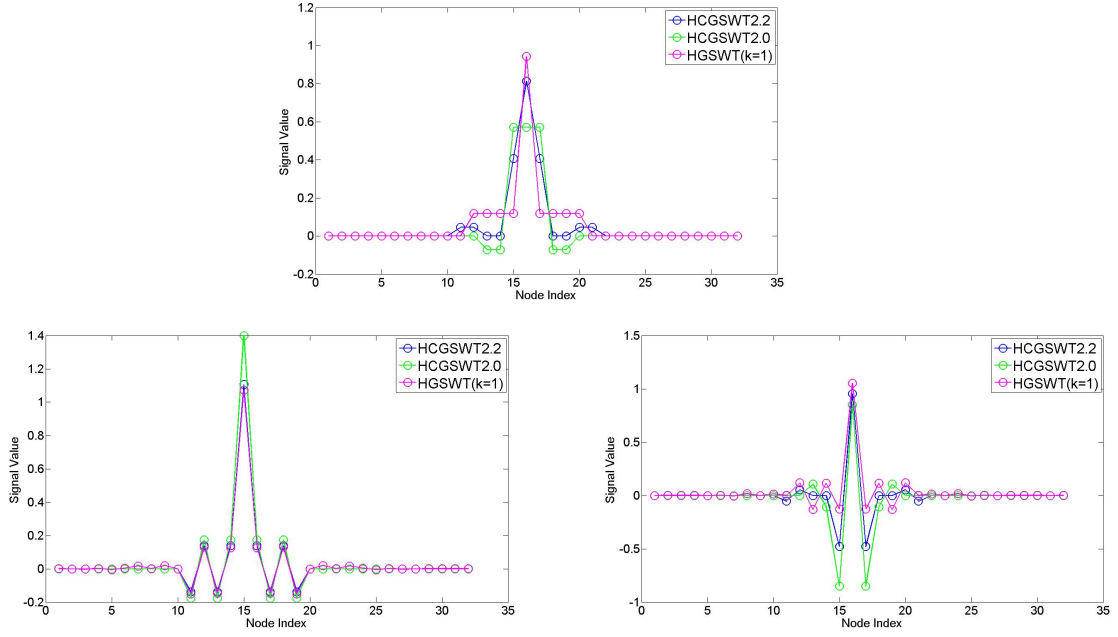


Figure 3.10: Comparison of Graph-filter functions: analysis low-pass (top), synthesis low- and high-pass (from left) at one level for the linear spline constructions on the graph with $S = (1, 2, 3, 4)$.

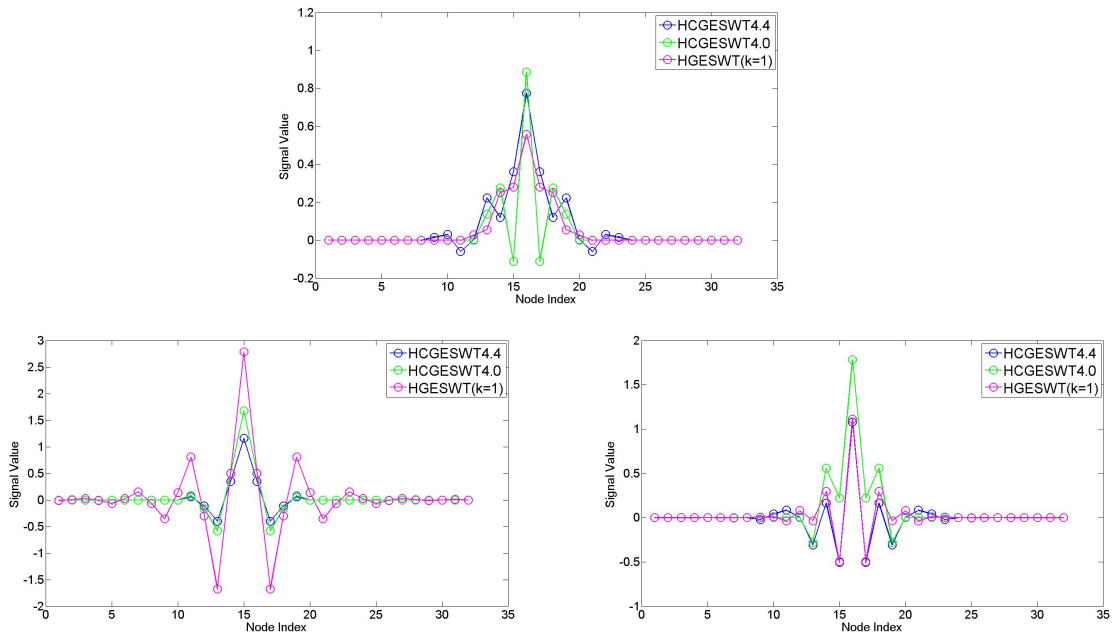


Figure 3.11: Comparison of Graph-filter functions: analysis low-pass (top), synthesis low- and high-pass (from left) at one level for the convolved e-spline constructions on the graph with $S = (1, 2)$.

Chapter 4

Generalized and Adaptive Wavelets on Graphs

Following a discussion which was largely focused on circulant graphs, what has transpired is that a number of key ideas and properties actually have widespread relevance, and as such the present chapter is designed to expand upon how the established structure of the graph wavelet transform (GWT) can be leveraged for more complex scenarios, including non-circulant *arbitrary undirected* graphs, a *set (or system) of graphs*, as well as associated signals with *space-and time-variant* properties, all the while retaining the focus on inducing *sparse* representations.

We will begin with an analysis and review of the graph wavelet transform in order to demonstrate its implications, and summarize how the proposed designs constitute special cases of a more general matrix-based filterbank notion when the graph at hand is of a regular structure. Subsequently, the discussion shifts to a range of novel designs, encompassing *(i)* generalized bandlimiting, *(ii)* space-variant, and *(iii)* time-variant graph wavelet transforms. In the latter instance, this may entail the cases of a fixed signal over a graph that changes over time, where topological information is fused through the sum or product of graph-filters at different time instants in order to capture the evolution of the graph structure, as well as a time-varying signal over a fixed (multi-dimensional) graph, realized through graph product operations within a multi-dimensional scheme.

4.1 Generalized Bandlimiting Graph Wavelet Transforms

4.1.1 The GWT in Perspective

In the following, we review the graph wavelet transform from a linear algebraic perspective and make connections to existing and proposed variations. This a priori necessitates the specification of a *graph shift-operator*:

Note on the shift operator: *For circulant graphs, which are LSI, the graph shift-operator is given by the powers of the circulant permutation matrix $\mathbf{\Pi}$ with first row $[0 \ 1 \ 0 \ \dots]$, equivalent to the classical time-shift [55]. In order to determine the translations of a given signal with respect to the structural irregularity of an arbitrary graph, one requires the definition of an appropriate shift-operator underlying all graph filtering operations. Within several GSP works, a generalized graph-shift operator is described as a linear transformation with local (topological) significance, which may take the form of any graph matrix; filters that are polynomials in the shift-operator are thus deemed to be shift-invariant with respect to the graph. In [54], graph operations derive from the adjacency matrix and its Jordan decomposition $\mathbf{A} = \mathbf{V}\mathbf{J}\mathbf{V}^{-1}$, admitting extensions to directed graphs, and as such filters are deemed shift-invariant on the graph if and only if they constitute polynomials in \mathbf{A} ; here it is required that the minimal and characteristic polynomial of \mathbf{A} are equal, which entails that each distinct eigenvalue is associated with a single Jordan block \mathbf{J}_i . In another line of work, shift operators are defined as having the same eigenbasis as the graph matrix at hand, yet their eigenvalues are complex exponential phase shifts, having been created to facilitate the preservation of the energy of the signal [96]. All of these operators are based in some form on a common graph-eigenbasis which ensures that resulting (filter-) matrices commute, as a defining property for many relevant transforms.*

Hermitian matrices, along with real skew-symmetric matrices, form a subset of normal matrices, which are unitarily diagonalizable. As such, normal matrices commute if they are simultaneously unitarily diagonalizable (Thm. 2.5.5, [97]), while their product is also normal. Therefore, a shift-invariant operator (filter) on an undirected graph may take any form, beyond being polynomial in a designated graph matrix¹, provided it is diagonalizable by the same graph eigenbasis, and hence commutes with the graph matrix.

Let \mathbf{H}_{LP} and \mathbf{H}_{HP} denote generic graph filters on an undirected graph $G = (V, E)$ with $|V| = N$, whose sole restriction is their shift-invariance with respect to the normalized graph Laplacian matrix wlog² $\mathbf{L}_n = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, i.e. they share the same eigenbasis \mathbf{U} , and

¹Note that the polynomial formulation may be advantageous, though not always necessary, for other properties such as localization in the vertex domain [1].

²While transforms are further on expressed in terms of the normalized adjacency matrix as the central operator, the graph Laplacian-based perspective is adopted here to draw connections to existing work.

\mathbf{K} a suitable downsampling matrix, with $K_{i,i} = 1$ for retained nodes and $K_{i,i} = -1$ otherwise. In a generalization of Thms. 3.1 and 3.2, the graph wavelet transform

$$\mathbf{W} = \frac{1}{2}(\mathbf{I}_N + \mathbf{K})\mathbf{H}_{LP} + \frac{1}{2}(\mathbf{I}_N - \mathbf{K})\mathbf{H}_{HP}$$

is invertible for any \mathbf{K} as long as the sufficient condition $\lambda_{LP,i}\lambda_{HP,i} \geq 0, \forall i \in [0, N-1]$ on the graph filter eigenvalues is satisfied, where $\lambda_{k,i}$ denotes the i -th eigenvalue of the filter decomposition $\mathbf{H}_k = \mathbf{U}\mathbf{\Lambda}_k\mathbf{U}^H$. In particular, by assuming that the nullspace of \mathbf{W} , represented by $\mathbf{z} = \mathbf{U}\mathbf{r}$ for some $\mathbf{r} \in \mathbb{C}^N$, is non-empty and setting $\mathbf{W}\mathbf{z} = \mathbf{0}_N$, one arrives at the equality

$$\mathbf{U}(\mathbf{\Lambda}_{LP} + \mathbf{\Lambda}_{HP})\mathbf{r} = -\mathbf{K}\mathbf{U}(\mathbf{\Lambda}_{LP} - \mathbf{\Lambda}_{HP})\mathbf{r},$$

which, following l_2 -normalization and subsequent squaring of both sides, results in $\sum_{i=0}^{N-1} |r(i)|^2 \lambda_{LP,i} \lambda_{HP,i} = 0$ and, hence, $\mathbf{r} = \mathbf{0}_N$ under suitable restrictions on the eigenvalues. If the spectral product becomes zero for certain i , invertibility can be further guaranteed as long as essentially the corresponding eigenvector(s) remain linearly independent after downsampling by a suitably chosen pattern \mathbf{K} .

As such, analysis filter-functions of the form $H_{LP}(1-\lambda) = H_{HP}(-(1-\lambda)) = \prod_n (\beta_n + (1-\lambda))^k$ introduced within previous circulant constructions and expressed with respect to the normalized graph Laplacian, present only a subset of valid possibilities; in particular, as long as the constant-sign requirement is satisfied for $\{H_{LP}(1-\lambda_i)H_{HP}(1-\lambda_i)\}_i$, one may i. a. combine analysis low-and high-pass filters of different order k or parameter β , beyond circulant graphs, for a desired frequency response. More generally, filters may adopt the form $\mathbf{H}_{LP} = \sum_{i=0}^{N-1} s_i \mathbf{L}_n^i$ and $\mathbf{H}_{HP} = \sum_{i=0}^{N-1} \tilde{s}_i \mathbf{L}_n^i$ subject to

$$\left(\sum_{i=0}^{N-1} s_i \lambda_j^i \right) \left(\sum_{i=0}^{N-1} \tilde{s}_i \lambda_j^i \right) \geq 0, \forall j \in [0, N-1],$$

while the coefficients $\{s_i, \tilde{s}_i\}_{i=0}^{N-1}$ need not be inter-related; alternatively, one may define the filters with respect to the adjacency matrix, and, in case of regularity and/or normalization, these graph matrices are interchangeable due to a common eigenbasis.

For synthesis filters $\tilde{\mathbf{H}}_{LP}$ and $\tilde{\mathbf{H}}_{HP}$, the overall transfer function \mathbf{T} of a critically-sampled perfect reconstruction filterbank then yields

$$\mathbf{T} = \underbrace{\frac{1}{2}(\tilde{\mathbf{H}}_{LP}\mathbf{H}_{LP} + \tilde{\mathbf{H}}_{HP}\mathbf{H}_{HP})}_{\mathbf{T}_1} + \underbrace{\frac{1}{2}(\tilde{\mathbf{H}}_{LP}\mathbf{K}\mathbf{H}_{LP} - \tilde{\mathbf{H}}_{HP}\mathbf{K}\mathbf{H}_{HP})}_{\mathbf{T}_2}. \quad (4.1)$$

One can achieve perfect reconstruction (PR) with $\mathbf{T} = \mathbf{I}_N$ by e. g. setting $\mathbf{T}_1 = c\mathbf{I}_N$, $c \in \mathbb{R}$, and $\mathbf{T}_2 = \mathbf{0}_{N,N}$, and accordingly deriving conditions on the individual filters. When the graph filters are restricted to a function $h(\cdot)$ of graph Laplacian \mathbf{L} , with representation

$\mathbf{H} = h(\mathbf{L}) = \sum_{\lambda_i \in \lambda(\mathbf{L})} h(\lambda_i) P_{\lambda_i}$, for spectrum $\lambda(\mathbf{L})$ of \mathbf{L} and projection-matrix $\mathbf{P}_{\lambda_i} = \mathbf{u}_i \mathbf{u}_i^H$ corresponding to λ_i , their design can be tailored to satisfy the PR condition directly in the spectral domain. The previous relations thus simplify to $\mathbf{T}_1 = \frac{1}{2} \sum_{j=0}^{N-1} (\tilde{\lambda}_{LP,j} \lambda_{LP,j} + \tilde{\lambda}_{HP,j} \lambda_{HP,j}) \mathbf{P}_{\lambda_j}$ and $\mathbf{T}_2 = \frac{1}{2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (\tilde{\lambda}_{HP,i} \lambda_{HP,j} - \tilde{\lambda}_{LP,i} \lambda_{LP,j}) \mathbf{P}_{\lambda_i} \mathbf{K} \mathbf{P}_{\lambda_j}$. In the case of a bipartite graph, the spectral folding property of the graph-eigenvalues, along with the distinct bipartite downsampling pattern \mathbf{K} , which retains vertices of one disjoint set, and the relation $\mathbf{P}_\lambda \mathbf{P}_\gamma = \delta_{\lambda\gamma} \mathbf{P}_\lambda$ for Kronecker delta δ_{ij} , facilitate the further simplification of $\mathbf{T}_2 = \frac{1}{2} \sum_{i=0}^{N-1} (\tilde{\lambda}_{HP,i} \lambda_{HP,N-i} - \tilde{\lambda}_{LP,i} \lambda_{LP,N-i}) \mathbf{K} \mathbf{P}_{\lambda_{N-i}}$, which translates into the set of PR conditions for $i \in [0 \ N-1]$

$$\tilde{\lambda}_{LP,i} \lambda_{LP,i} + \tilde{\lambda}_{HP,i} \lambda_{HP,i} = c^2, \quad c \in \mathbb{R}$$

$$\tilde{\lambda}_{LP,i} \lambda_{LP,N-i} - \tilde{\lambda}_{HP,i} \lambda_{HP,N-i} = 0$$

as is demonstrated as part of the bipartite graph filterbank by Ortega et al. [35]. The filter eigenvalues may be tailored to satisfy the above relations, i. a. through the spectral design of a QMF or biorthogonal filterbank, under the assumption that these are given by (polynomial) functions in a graph matrix with $\lambda_h = h(\lambda)$ ([35], [36]).

The above system resembles the form of classical perfect reconstruction and anti-aliasing conditions, as seen in Eq. (3.7) of Sect. 3.5.1, however, only when the graph at hand is *circulant* and the filter eigenvalues invoked in the *z-domain* of representer polynomials $H_{LP/HP}(z)$ (as well as the downsampling pattern fixed with respect to $s = 1 \in S$), there is an exact analogy, as elucidated in Sect. 3.5.

According to circulant matrix theory [75], the latter are defined as functions of the roots of unity, and give rise to the corresponding filter-eigenvalues at position k of the DFT-ordered spectrum for $z = e^{\frac{2\pi i k}{N}}$; the decisive difference with bipartite filter constructions is that the representer polynomial of circulants is directly defined in the vertex domain, as opposed to the spectral domain as a function of a graph matrix, and by extension, its eigenvalues. Hence, in case of circularity, one may resort to the complementary PR design proposed in Sect. 3.5.2 by fixing one filter and imposing desired properties on its dual.

If the graph is both bipartite and circulant, special scenarios such as the introduced semi-IIR graph-filterbank arise, and, more significantly, properties from either design strategy may be leveraged. Moreover, the bipartite downsampling pattern coincides with that of a circulant graph, if taken with respect to the outmost cycle. While in the bipartite graph case, the filter-eigenvalues $h(\lambda)$ are given as a function $h(\cdot)$ of the graph Laplacian eigenvalues λ , for a circulant graph, the eigenvalues are induced via the representer-polynomial $H(z)$ as a function of the roots of unity; hence for a circulant bipartite graph, one can easily switch between the two via transformation $\lambda(z) = \sum_{j \in 2\mathbb{Z}+1} d_j (-z^j + 2 - z^{-j})$ for weights $d_j = A_{i,(j+i)_N}$ in $h(\lambda(z))$.

Overall, it is for the aforementioned special cases that synthesis filters can be explicitly stated and/or derived from a given set of analysis filters. When the graph at hand has neither of the preceding properties, the spectral system resulting from Eq. (4.1) remains valid, but it is less straightforward to tailor its design and determine explicit synthesis filters, particularly considering the absence of a definitive downsampling scheme. The general filterbank structure is preserved through inverse $\mathbf{W}^{-1} = \frac{1}{2}\tilde{\mathbf{H}}_{LP}^T(\mathbf{I}_N + \mathbf{K}) + \frac{1}{2}\tilde{\mathbf{H}}_{HP}^T(\mathbf{I}_N - \mathbf{K})$, which consists of two sets of basis functions, yet it is unclear how they are related to the graph, beyond the fact that they may be interpreted to preserve shift-invariance in certain cases. The synthesis filters $\tilde{\mathbf{H}}_{LP/HP}$ before downsampling are not uniquely defined from the inverse, unless there is a distinct underlying structure and sampling pattern, such as for the circulant case.

4.1.2 Generalized Vanishing Moments

In an effort to challenge the annihilation property of previously developed transforms beyond (exponential) polynomials, and accommodate more general types of signals as well as graphs, one can make the general observation that any filter of the form $\mathbf{H} = (\gamma(\mathbf{A})\mathbf{I}_N - \mathbf{A})$ for symmetric \mathbf{A} and eigenvalue $\gamma(\mathbf{A})$ is *bandlimiting*, i.e. it annihilates selected signals of banded frequency support by attenuating their content outside that support. The inherent technique of successive nulling of the eigenvalues of a linear graph operator has also been employed for average-consensus in multi-agent systems [98], albeit with the objective to restrict an initial signal to a fixed (eigen-)subspace.

In particular, any generalized graph filter that is defined as a polynomial in the adjacency matrix $\mathbf{H} = \sum_{i=0}^{N-1} h_i \mathbf{A}_i$, and is thus *shift-invariant* with respect to \mathbf{A} , with $\mathbf{H}\mathbf{A} = \mathbf{A}\mathbf{H}$, can alternatively be expressed in the factorized form $\mathbf{H} = \tilde{h}_0 \prod_{i=1}^{N-1} (\tilde{h}_i \mathbf{I}_N - \mathbf{A})$ for $h_i, \tilde{h}_i \in \mathbb{C}$ according to the Fundamental Thm. of Algebra, and hence tailored to attenuate certain frequencies by setting its coefficients $\tilde{h}_i = \gamma_k$ equal to the eigenvalues $\{\gamma_k\}_k$ of \mathbf{A} . If \mathbf{A} is circulant and banded of bandwidth M , annihilation can prominently be further expanded to localized form for generic e-degree $\tilde{h}_i = \sum_{k=1}^M 2d_k \cos(\alpha_n k)$ and some $\alpha_n \in \mathbb{R}$.

As already noted in Ch. 3, both the spline and e-spline graph wavelet transforms do not require circularity (albeit they do require symmetry) for invertibility, which prompts the creation of a generalized (higher-order) *bandlimiting* graph wavelet transform, following the notion of a nullspace-shifted high-pass operator à la Rmk 3.1:

Corollary 4.1. *Given the undirected graph G with symmetric normalized adjacency matrix \mathbf{A}_n , we define the higher-order bandlimiting graph wavelet transform (HBGWT), composed of the low-and high-pass filters:*

$$\mathbf{H}_{LP} = \prod_{n=1}^T \frac{1}{2^k} (\gamma_n \mathbf{I}_N + \mathbf{A}_n)^k \quad (4.2)$$

$$\mathbf{H}_{HP} = \prod_{n=1}^T \frac{1}{2^k} (\gamma_n \mathbf{I}_N - \mathbf{A}_n)^k \quad (4.3)$$

where \mathbf{H}_{HP} annihilates bandlimited graph signals associated with eigenvalues $\{\gamma_n\}_{n=1}^T$. This filterbank is invertible, subject to equivalent restrictions on the downsampling pattern and parameters, as in Thms 3.1 and 3.2.

The above transform may further be expressed in terms of any equivalent (symmetric) graph shift-operator (or generalized graph Laplacian) that need not be the adjacency matrix; in the case when normalized graph adjacency matrix eigenvalues γ_n are replaced by other constants, invertibility conditions may be accordingly relaxed.

According to the Perron-Frobenius Thm. $\gamma_{\max}(\mathbf{A}_n) = 1$ has multiplicity 1 for a connected undirected graph (Thm. 2.22, [47]), which simplifies the preceding Corollary (and equivalent Theorems) when $\gamma_n = 1, \forall n$, since in that case only at least one low-pass component needs to be retained to guarantee invertibility. However, connectivity is generally not a requirement, provided linear independence of the downsampled eigenvectors (corresponding to γ_n) is upheld for larger multiplicities (and multiple eigenvalues) for a suitable downsampling pattern.

While the higher order invoked by powers $k > 1$ is in this case less meaningful than the higher order vanishing moments of a circulant graph, it nevertheless grants a related interpretation, according to which an increase in order k , leads to an increase in the space of functions which can be approximately annihilated, in the spirit of the vanishing moments of a diffusion operator by Coifman et al. [33]. In particular, for those $\gamma_j(\mathbf{A}_n)$ that are close to selected graph frequencies $\{\gamma_n\}_n$, or rather, for certain eigenvalues of \mathbf{H}_{HP} that are close to zero, the corresponding eigenvector \mathbf{u}_j is accordingly smooth with respect to the graph, i.e. closer to annihilation. For $k \gg 1$, the eigenvalues of \mathbf{H}_{HP} become smaller and concentrate more densely around zero (i.e. as a result of normalization), such that $(\gamma_n(\mathbf{A}_n)\mathbf{I}_N - \mathbf{A}_n)^k \mathbf{u}_j \approx \mathbf{0}_N$ for some $\gamma_j \approx \gamma_n$.

Further, when the graph at hand is bipartite, the filter-structure of the *HBGWT* is affected by the spectral folding phenomenon, as previously noted for bipartite circulant transforms in Sect. 3.3.3, which entails that graph signals bandlimited to frequency $-\gamma_n$ are annihilated in the low-pass branch and maintained in the high-pass branch.

We adopt the following definition for the class of bandlimited graph signals, while noting that the characterisation of a ‘band’ has been interpreted differently in other work, such as through a cut-off value [67] as opposed to a fixed number of frequencies [68].

Definition 4.1. A graph signal $\mathbf{x} \in \mathbb{C}^N$ is bandlimited with respect to graph basis \mathbf{U} with bandwidth K if it has the form $\mathbf{x} = \sum_{i=1}^K a_i \mathbf{u}_i$, where \mathbf{u}_i is the i -th (ordered) eigenvector of an appropriately chosen (un-)normalized graph matrix, for coefficients $a_i \in \mathbb{C}$.

While conventionally the graph Laplacian eigenbasis is chosen as a basis \mathbf{U} , one may also consider the eigenvectors of the (normalized) adjacency matrix, as done in [68]. One can extend this definition to comprise arbitrary (sparse) linear combinations of basis vectors, up to a permutation, without the restriction to a single ordered graph frequency band, which we term *generalized bandlimited*.

At last, in an effort to allow for directionality in edges, consider an extension of the GWT to classes of graphs with normal adjacency matrices, which may include asymmetry, and hence *direction*:

Remark 4.1. As previously noted, normal matrices, admitting the relation $\mathbf{A}\mathbf{A}^H = \mathbf{A}^H\mathbf{A}$, form a generalized class of matrices, comprising i. a. unitary, Hermitian and skew-Hermitian matrices, which are diagonalizable by a unitary basis \mathbf{U} [97]. It can be shown that Cor. 4.1 with $\gamma_n = 1$ (or Thm. 3.1) for $k = 1$ is extendable to nonnegative normal matrices under further restrictions; leveraging Perron-Frobenius theory, here, \mathbf{A}_n is required to be primitive so that $|\gamma_i| < \gamma_{\max}$, with $\gamma_{\max} = 1$ under proper normalization, and the proof carries over, otherwise, invertibility depends on the downsampling pattern. In particular, a directed circulant matrix is normal, and the already discussed conditions apply equivalently (see Sect. 3.4.1).

Real skew-symmetric matrices on the other hand, defined through $\mathbf{A}^T = -\mathbf{A}$, possess exclusively purely imaginary eigenvalues (in conjugate pairs) [97], which, if imposed as transform parameters $\{\gamma_n\}_n$, uphold Cor. 4.1, leaving the simplifications due to symmetry in the proof(s) unchanged, as imaginary terms are cancelled out; this further extends to arbitrary purely imaginary parameters.

Following Rmk. 4.1, note that any square matrix \mathbf{A} can be expressed as the sum of symmetric and antisymmetric parts in $\mathbf{A} = \mathbf{A}_{\text{sym}} + \mathbf{A}_{\text{asym}} = \frac{1}{2}(\mathbf{A} + \mathbf{A}^T) + \frac{1}{2}(\mathbf{A} - \mathbf{A}^T)$ [97]. Given an arbitrary directed graph with adjacency matrix \mathbf{A} , one may conduct operations with respect to both a symmetrized \mathbf{A}_{sym} and anti-symmetrized version \mathbf{A}_{asym} , which possibly contains negative weights, of \mathbf{A} separately and subsequently average (or merge) the results. In particular, for normalized $\mathbf{A}_{\text{asym},n} = \frac{1}{|\gamma_{\max}|}\mathbf{A}_{\text{asym}}$, where γ_{\max} is the largest magnitude eigenvalue of \mathbf{A}_{asym} ³, the proposed bandlimiting graph wavelet transform is provably invertible for either matrix class, without incurring further restrictions due to the complex eigenvalues of arbitrary unsymmetric matrices or resorting to the less flexible Jordan decomposition; here, Cor. 4.1 applies directly to the latter for purely imaginary parameters $\{\gamma_i\}_i$.⁴

³This normalization ensures that the spectrum $\tilde{\gamma}(\mathbf{A}_{\text{asym},n})$ of $\mathbf{A}_{\text{asym},n}$ is bounded with $|\tilde{\gamma}| \leq 1$, which in the case of nonnegative matrices, would be otherwise facilitated by the symmetric normalized $\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$ or random-walk normalized form $\mathbf{D}^{-1}\mathbf{A}$, as per the Gerschgorin Circle Theorem [5].

⁴In particular, the symmetry of the spectrum of a skew-symmetric matrix, containing purely imaginary pairs of complex conjugates (including 0), is similar to the (real) spectrum symmetry of a bipartite graph with respect to zero. As such, for special cases $|\beta_n| = |\gamma_i(\mathbf{A}_{\text{asym},n})|$, the proof of Thm. 3.2 (and Cor. 4.1) continues to hold.

The case of $\gamma = 1$ (as per Thm. 3.1), continues to hold for properly normalized skew-symmetric matrices when $k = 1$, however, one incurs the restriction that invertibility depends on the linear independence of downsampled eigenvectors corresponding to $|\gamma_i(\mathbf{A}_{asym,n})| = 1$ since Perron Frobenius theory is not applicable.

4.1.3 Generalizations to Random Walk

When dealing with non-regular connected undirected graphs, it may be desirable for the (spline-)GWT to retain orthogonality with respect to constant signals, however, as a consequence of the normalization of the adjacency matrix to the form $\mathbf{A}_n = \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}$, the high-pass filter $\mathbf{L}_n = \mathbf{I}_N - \mathbf{A}_n$ instead only annihilates signals of the form $\mathbf{D}^{1/2} \mathbf{1}_N$, as the eigenvector in its nullspace. In this case, \mathbf{A}_n may be replaced by the random-walk normalized version $\mathbf{D}^{-1} \mathbf{A}$; while the latter is consequentially not symmetric in general and hence does not necessarily possess an orthonormal eigenbasis, it is similar to the former, i. e. they share the same eigenvalues.

If a transform in \mathbf{A}_n is invertible, it must also be invertible under the same conditions for $\mathbf{D}^{-1} \mathbf{A}$, which can be shown by applying a simple transformation to the filters of the form $\mathbf{H}_{LP/HP} = \prod_n (\beta_n \mathbf{I}_N \pm \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2})^k$ for some coefficients β_n :

$$\begin{aligned} \mathbf{H}_{LP/HP,RW} &= \mathbf{D}^{-1/2} \mathbf{H}_{LP/HP} \mathbf{D}^{1/2} \\ &= \mathbf{D}^{-1/2} \prod_n (\beta_n \mathbf{I}_N \pm \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}) \dots \mathbf{D}^{1/2} \mathbf{D}^{-1/2} \dots \prod_n (\beta_n \mathbf{I}_N \pm \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}) \mathbf{D}^{1/2} \\ &= \prod_n (\beta_n \mathbf{I}_N \pm \mathbf{D}^{-1} \mathbf{A})^k. \end{aligned}$$

Thus, the random-walk-normalized transform

$$\begin{aligned} \mathbf{W}_{RW} &= \frac{1}{2} \mathbf{D}^{-1/2} (\mathbf{H}_{LP} + \mathbf{H}_{HP} + \mathbf{K}(\mathbf{H}_{LP} - \mathbf{H}_{HP})) \mathbf{D}^{1/2} \\ &= \frac{1}{2} (\mathbf{H}_{LP,RW} + \mathbf{H}_{HP,RW} + \mathbf{K}(\mathbf{H}_{LP,RW} - \mathbf{H}_{HP,RW})), \end{aligned}$$

continues to be invertible for non-zero node degrees and subject to the same restrictions.

4.2 Space-Variant Graph Wavelets

For previously derived transforms, the signals of interest have mainly been confined to being globally and locally (piecewise) smooth, with pieces of the same function type. A more flexible version of the graph (spline) wavelet transform, which is termed *space-* or

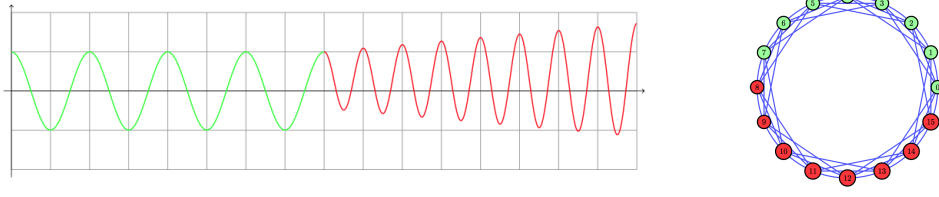


Figure 4.1: Piecewise Smooth Graph Signal on a Circulant Graph.

node-variant, is therefore introduced carrying localized annihilation properties for piecewise smooth (or bandlimited) graph signals, with pieces that do not necessarily belong to the same class of functions.

Let $\mathbf{1}_C \in \mathbb{R}^N$ denote the graph signal on $G = (V, E)$, with $(\mathbf{1}_C)_i = 1$ if node $i \in C$ and $(\mathbf{1}_C)_j = 0$ otherwise, for local vertex set $C \in V$. Further, consider the series of vertex sets $\{C_i\}_{i=0}^T$ such that $C_i := [j, j + |C_i| - 1]$, with $j - 1 \in C_{i-1}$, describe consecutively labelled partitions of V ; without loss of generality, (piecewise) smooth signals are thus defined on local vertex sets with ordered labelling (see Fig. 4.1).

Definition 4.2. A graph signal $\mathbf{x} \in \mathbb{R}^N$ on $G = (V, E)$ is piecewise smooth if it can be written as $\mathbf{x} = \sum_{j=0}^T \mathbf{x}^j \circ \mathbf{1}_{C_j}$ for node sets $\{C_j\}_{j=0}^T$, with pieces \mathbf{x}^j belonging to the class of smooth graph signals, including bandlimited and (exponential) polynomial signals.

As this necessitates the design of graph filters that behave differently in different neighborhoods of the graph, one may adopt the form $\tilde{\mathbf{D}} \pm \mathbf{A}$, parameterized by the diagonal degree matrix $\tilde{\mathbf{D}} = \sum_{i=0}^T \mathbf{D}_i$, which consists of pieces such that $\mathbf{D}_i(C_i, C_i) = \beta_i \mathbf{I}_{|C_i|}$ and $\mathbf{D}_i(C_i^c, C_i^c) = \mathbf{0}_{N_i, N_i}$, with $N_i = |C_i^c|$, for selected (degree) parameters $\{\beta_i\}_i$. In the case of circulant graphs, one may set the latter to be of the form of the e-degree $d_{\alpha_i}, \alpha_i \in \mathbb{R}$, or in general, the eigenvalues of (normalized) \mathbf{A} .

This ultimately gives rise to a node-variant transform of the following form:

Theorem 4.1. Given the undirected graph $G = (V, E)$ of dimension N , with symmetric normalized adjacency matrix \mathbf{A}_n , we define the higher-order space-variant graph-spline wavelet transform (HSVGSWT), composed of the low-and high-pass filters

$$\mathbf{H}_{LP} = \frac{1}{2^k} (\mathbf{I}_N + \tilde{\mathbf{D}}^{-1/2} \mathbf{A}_n \tilde{\mathbf{D}}^{-1/2})^k \quad (4.4)$$

$$\mathbf{H}_{HP} = \frac{1}{2^k} (\mathbf{I}_N - \tilde{\mathbf{D}}^{-1/2} \mathbf{A}_n \tilde{\mathbf{D}}^{-1/2})^k \quad (4.5)$$

for $\tilde{\mathbf{D}} = \sum_{i=0}^T \mathbf{D}_i$, such that $\mathbf{D}_i(C_i, C_i) = \beta_i \mathbf{I}_{|C_i|}$ and $\mathbf{D}_i(C_i^c, C_i^c) = \mathbf{0}_{N_i, N_i}$, with $N_i = |C_i^c|$. This filterbank is invertible subject to $\forall i, \beta_i \geq 0$ and $k \in 2\mathbb{Z}^+$, for any downsampling pattern, as long as $\tilde{\gamma}(\tilde{\mathbf{A}}) \neq 1$, for eigenvalues of $\tilde{\mathbf{A}} = \tilde{\mathbf{D}}^{-1/2} \mathbf{A}_n \tilde{\mathbf{D}}^{-1/2}$. If $\exists \tilde{\gamma}(\tilde{\mathbf{A}}) = \pm 1$, the

transform continues to be invertible for a suitable downsampling pattern such that the respective associated downsampled eigenvectors $\{\mathbf{v}_{1,j}(V_\alpha)\}_j$ (and complement $\{\mathbf{v}_{-1,j}(V_\alpha^c)\}_j$) form linearly independent sets, for retained nodes in set V_α and multiplicities j .

Proof. Since the matrix $\tilde{\mathbf{A}} = \tilde{\mathbf{D}}^{-1/2} \mathbf{A}_n \tilde{\mathbf{D}}^{-1/2}$ continues to be Hermitian (though not normalized) with an orthonormal basis, the proof from Thm. 3.1 in Appendix A.1 carries over, up to Eq. (A.3)

$$\sum_{i=0}^{N-1} r(i)^2 (1 - \tilde{\gamma}_i^2)^k = 0$$

for eigenvalues $\{\tilde{\gamma}_i\}_{i=0}^{N-1}$ of $\tilde{\mathbf{A}}$. By the submultiplicativity of the spectral norm [97]

$$\|\tilde{\mathbf{D}}^{-1/2} \mathbf{A}_n \tilde{\mathbf{D}}^{-1/2}\| \leq \|\tilde{\mathbf{D}}^{-1/2}\| \|\mathbf{A}_n\| \|\tilde{\mathbf{D}}^{-1/2}\| \leq \frac{1}{\tilde{d}_{\min}} \gamma_{\max}(\mathbf{A}_n)$$

and noting the spectral radius $\|\mathbf{A}_n\|_2 = \gamma_{\max}(\mathbf{A}_n) = 1$, we observe for the spectrum $\tilde{\gamma}(\tilde{\mathbf{A}})$ that $1 \leq \tilde{\gamma}_{\max} \leq \frac{1}{\tilde{d}_{\min}}$, so we require $k \in 2\mathbb{Z}^+$ to ensure all-positive summands, and thus $r(i) = 0$, $i = 0, \dots, N-1$. Here, $\beta_i = \tilde{d}_{\min}$ is the smallest magnitude entry of $\tilde{\mathbf{D}}$, where $|\beta_i| \leq 1$, $\forall i$ generally holds for normalized eigenvalues (e-degrees). One merely needs to exclude the special case of $\tilde{\gamma} = \pm 1$, for the construction to be invertible under any downsampling pattern, which due to the re-normalization is not an anticipated case; otherwise, one may further need to show that for $\tilde{\gamma} = \pm 1$, the corresponding eigenvectors of $\tilde{\mathbf{A}}$ remain respectively linearly independent after (suitable) downsampling, following the proof of Thm. 3.2 in Appendix A.2.

In addition, to ensure that $\tilde{\mathbf{A}}$ is real symmetric, we require all parameters $\{\beta_i\}_{i=0}^T$ to have the same sign; for negative β_i , $\tilde{\mathbf{D}}^{-1/2}$ gains purely imaginary entries, which cancel out only if all β_i are negative. Hence, for parameters of different sign, $\tilde{\mathbf{A}}$ becomes complex symmetric with $\tilde{\mathbf{A}} = \tilde{\mathbf{A}}^T$; according to (Cor. 4.4.4, [97]), a complex symmetric matrix admits the so-called Takagi factorization $\tilde{\mathbf{A}} = \mathbf{V} \bar{\mathbf{D}} \mathbf{V}^T$, with unitary \mathbf{V} and nonnegative diagonal $\bar{\mathbf{D}}$, however, it is not diagonalizable per se and the proof does not apply.

If the graph is bipartite, the property that the spectrum $\tilde{\Gamma}$ is symmetric with respect to 0 is maintained so $\tilde{\gamma}_{\min} = -\tilde{\gamma}_{\max}$, since the characteristic structure of its un-normalized adjacency matrix, with

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{A}_1 \\ \mathbf{A}_1^T & \mathbf{0} \end{bmatrix}$$

is upheld beyond multiplication with arbitrary (non-zero) diagonals

$$\mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2} = \begin{bmatrix} \mathbf{D}_1^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2^{-1/2} \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{A}_1 \\ \mathbf{A}_1^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{D}_1^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2^{-1/2} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{0} & \mathbf{D}_1^{-1/2} \mathbf{A}_1 \mathbf{D}_2^{-1/2} \\ (\mathbf{D}_1^{-1/2} \mathbf{A}_1 \mathbf{D}_2^{-1/2})^T & \mathbf{0} \end{bmatrix}.$$

As such, in case of $|\tilde{\gamma}| = 1$, one needs to consider the linear dependency of the corresponding eigenvectors of $\tilde{\mathbf{A}}$ (after downsampling) for both opposing eigenvalues (and their multiplicities). \square

Hence, for a sufficiently banded, arbitrary undirected graph, the above transform gives rise to sparse representations of signals that are piecewise bandlimited with respect to the graph, subject to border effects dependent on the bandwidth of the graph; in the case of a circulant graph with e-degree parameterization, higher-order vanishing moments can be invoked for $k > 1$.

As a result of their structure, the proposed node-variant filters are invertible, which in the case of the high-pass filter gives rise to a promising independent sparsification method.

Corollary 4.2. *The filters of Thm. 4.1 are invertible for $\beta_i \geq 0, \forall i$, and in general, if $\tilde{\mathbf{A}}$ is normal, provided $\tilde{\gamma}(\tilde{\mathbf{A}}) \neq 1$.*

Proof. We need to show that the nullspace of the filters with eigenbasis \mathbf{U} , represented by vectors satisfying $\mathbf{z} = \mathbf{U}\mathbf{r}$, is empty and jump to the step

$$\mathbf{U}(\mathbf{I}_N \pm \tilde{\mathbf{\Gamma}})^k \mathbf{r} = \mathbf{0}_N,$$

followed by taking the l_2 -norm and power 2,

$$\mathbf{r}^H (\mathbf{I}_N \pm \tilde{\mathbf{\Gamma}})^{2k} \mathbf{r} = \sum_{i=0}^{N-1} |r(i)|^2 (1 \pm \tilde{\gamma}_i)^{2k} = 0.$$

It becomes evident that this is always satisfied for any k , as the summands are always positive and thus $\mathbf{r} = \mathbf{0}_N$.

If $\exists \tilde{\gamma} = 1$, this ceases to be valid for the high-pass version (and vice versa for the low-pass filter with $\tilde{\gamma} = -1$), as well as for either filter in the bipartite graph case, when $\exists \tilde{\gamma} = \pm 1$, by similar reasoning as in Cor. 3.3. When $\tilde{\mathbf{A}}$ is normal with complex $\tilde{\mathbf{\Gamma}}$, one may further show that

$$\mathbf{r}^H (\mathbf{I}_N \pm \tilde{\mathbf{\Gamma}}^H)^k (\mathbf{I}_N \pm \tilde{\mathbf{\Gamma}})^k \mathbf{r} = \mathbf{r}^H (\mathbf{I}_N \pm \tilde{\mathbf{\Gamma}}^H \pm \tilde{\mathbf{\Gamma}} + \tilde{\mathbf{\Gamma}} \tilde{\mathbf{\Gamma}}^H)^k \mathbf{r} = \sum_{i=0}^{N-1} |r(i)|^2 (1 + |\tilde{\gamma}_i|^2 \pm (\tilde{\gamma}_i + \tilde{\gamma}_i^*))^k,$$

and letting $\tilde{\gamma}_i = x_i + iy_i$, $x_i, y_i \in \mathbb{R}$, one has $(1 + |\tilde{\gamma}_i|^2 \pm (\tilde{\gamma}_i + \tilde{\gamma}_i^*))^k = (1 + x_i^2 + y_i^2 \pm 2x_i)^k = ((1 \pm x_i)^2 + y_i^2)^k > 0$ and hence $\mathbf{r} = \mathbf{0}_N$. In case of $\tilde{\gamma} = \pm 1$, prior reasoning applies. \square

Further note that this is an interesting feature of the space-variant high-pass filter which the (circulant) e-spline graph Laplacian does not possess (i.e. when parameter β is an eigenvalue, it is not invertible, see Rmk. 3.1).

A different realization of a *node-variant* filter was introduced in [99] by Segarra et al. of the form $\mathbf{H} = \sum_{i=0}^{N-1} \mathbf{D}_i \mathbf{S}^i$ for diagonal matrices $\{\mathbf{D}_i\}_i$ and graph shift-operator \mathbf{S} , which re-weights the contributions for each node, in an effort to design more general graph operators. This, however, is distinct from the proposed construction in that it does not have any desired and/or proven annihilation effect.

4.3 Time-Variant Graph Wavelets

When the graph at hand changes over time or a series of graphs are constructed from given data to capture information from multiple (geometric) views, it is desirable to design wavelets that operate across and merge different connectivities or topologies as opposed to conducting a separate graph wavelet analysis on each individual graph. The existence of a (time-)varying graph transform on a set of data-driven graphs, in particular, facilitates the merging of information from different views or modalities, given in form of (time-)varying edges, for a *joint* multimodal (geometric) analysis of the data set involving multiple data spaces.⁵

In the following, we present various families of graph-variant transforms and discuss their specific interpretation. We distinguish between (i) graph-fusing and (ii) graph-extending transforms. The former, which is the focus of this section, is built on the assumption that a set or sequence of given graphs $\{G_k\}_k$ share the same vertex set V while each graph has a different edge set E_k , with the goal to merge or unify information from the different graphs within the same dimension $|V| = N$. The latter, to be treated in Sect. 4.5, additionally considers the inter-connectivity between individual graphs and, by contrast, adds another dimension through the Kronecker operator for multi-dimensional processing.

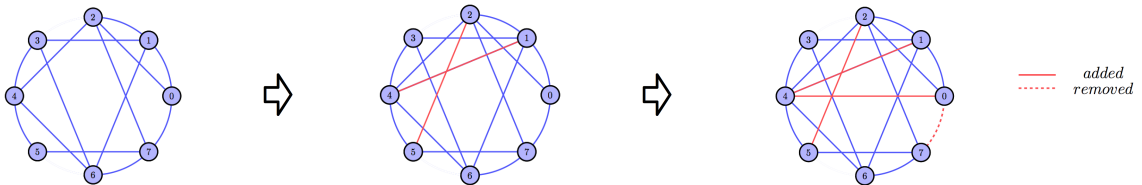


Figure 4.2: Illustrative Time-Varying-Graph at $t = 0, 1, 2$.

⁵This notion of multimodal manifold analysis has in particular been realized for approaches such as spectral clustering or diffusion maps [100].

4.3.1 The Cross-Graph Walk

We restrict our focus at first to families or sequences of undirected graphs which share the same (graph matrix-)eigenbasis. The primary motivation for this choice is to maintain shift-invariance with respect to the same basis across graphs of different connectivity, such that essential GSP operations, including downsampling, are preserved over time, and suitable (symmetric) transforms of provable invertibility can be identified. This may, for instance, encompass series of graphs which are subject to specific perturbations over time (see Fig. 4.2) and are modelled to possess steady eigenvectors with changing eigenvalues.

If for a given sequence of graphs this property is not fulfilled, one may resort to approximation schemes such as joint diagonalization [100] to find an approximate common eigenbasis. A further requirement is that a common downsampling pattern is established; note that while graphs may share the same eigenbasis, the vectors associated with certain graph eigenvalues, which can serve as a possible indicator function in generalized downsampling schemes [60], may not always be the same⁶. As a next step, a joint coarsening strategy needs to be established.

In an effort to create *global* filtering operations across graphs, we primarily distinguish between the addition and multiplication of time-varying graph matrices, which, wlog and for consistency of discussion, are constrained to be of the form of adjacency matrices here; as such, generalized polynomial graph filters on a set of adjacency matrices $\{\mathbf{A}_j\}_j$ may take the form of $\mathbf{H} = \sum_{j,k} a_{j,k} \mathbf{A}_j^k$ or $\mathbf{H} = \sum_k \prod_j \tilde{a}_{j,k} \mathbf{A}_j^k$ for coefficients $a_{j,k}, \tilde{a}_{j,k} \in \mathbb{C}$. Similarly as for stationary graphs, coefficients may be chosen to induce a desired frequency response, for the targeted annihilation of smooth (generalized bandlimited) graph signals as well as for ranking the contribution of individual graph topologies. The summation of two (or more) adjacency matrices presents the simplest way of fusing connectivity information from different graphs by simply adding the edge weights of equivalent nodes in a form of (weighted) averaging; multiplication, however, appears to be more intricate and we proceed to analyse its significance in mathematical terms.

It is known from graph theory that the (i, j) -th entry of \mathbf{A}^k , for binarized version \mathbf{A} of a given adjacency matrix, indicates the number of k -hop walks between the node pair (i, j) [101]; we wish to further generalize this interpretation for products of different graph adjacency matrices. Consider the binary adjacency matrices \mathbf{A} and \mathbf{B} of graphs $G_{\mathbf{A}}$ and $G_{\mathbf{B}}$ respectively, with product $\mathbf{C} = \mathbf{AB}$; then $C_{i,j} \neq 0$ iff there exists a walk from i to j in set $(E_{\mathbf{A}}, E_{\mathbf{B}})$, which indicates ‘double-hop’ walks whose first edge is in $G_{\mathbf{A}}$ and second in $G_{\mathbf{B}}$, and $C_{i,j} = 0$ otherwise. In particular, from the relation $C_{i,j} = \sum_{k=0}^{N-1} A_{i,k} B_{k,j}$, we observe that $(C^s)_{i,j}$ counts the number of $2s$ -hop walks from node i on graph $G_{\mathbf{A}}$ to node

⁶Conditions which ensure the ‘matching’ of eigenvectors on different graphs are in particular addressed in the proof of Cor. 4.3 in Appendix B.1.

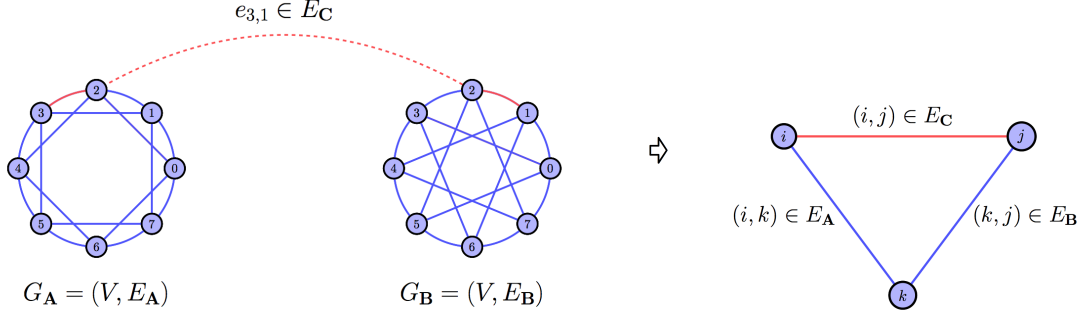


Figure 4.3: Walk across two circulant graphs.

j on graph $G_{\mathbf{B}}$, with the first edge on $G_{\mathbf{A}}$ and second on $G_{\mathbf{B}}$, while traversing a common node k as an *anchor* or central point (see Fig. 4.3). Provided that \mathbf{A} and \mathbf{B} are symmetric and share the same eigenbasis, \mathbf{C} is also symmetric.

Further, as a consequence of ([37], Lemma 5.2), when its diagonals are set to have unit weight, $(C^s)_{i,j}$ represents the number of $2r$ -hop walks of length $r \leq s$, thereby alluding to the fact that products of graph Laplacians give rise to localized operations within $2s$ -hop neighborhoods across graphs. This interpretation assumes no further ‘transitional’ probability between nodes on different graphs (as e.g. done in [102] for graph layers), in the sense of conducting operations with respect to past and future neighbours of a node, all of which are equally weighted.

In [101], the idea of graph multiplication is realized through König digraphs (as a special kind of bipartite graphs) which are shown to preserve essential graph properties under multiplication. In this case, while not a proper graph operation which preserves desirable properties, the multiplication of adjacency (or graph Laplacian) matrices rather provides insight into the graph evolution and connectivity over time in form of the intermediate or transitional graph $G_{\mathbf{C}}$.

One should further take note that the above operations are affected by the individual labelling of each graph component; as such, we assume a fixed node labelling over time and apply permutations, if necessary, on all graphs, prior to processing.

4.3.2 Filtering Across Different Graphs

Consider the set of graphs $\bar{G} = \{G_0, \dots, G_T\}$ with corresponding adjacency matrices $\bar{A} = \{\mathbf{A}_0, \dots, \mathbf{A}_T\}$, captured at time instant $t = 0, \dots, T$, and let \mathbf{x} denote a graph signal on G_t , which may be stationary or time-varying⁷. In the former case, the signal is assumed to reside on the (fixed) nodes of each graph layer.

⁷In the latter case, a variation of the proposed transforms can be applied on each individual signal.

If a common basis \mathbf{U} does not exist, one may pose an optimization problem and conduct joint diagonalization of existing graph (Laplacian) matrices \mathbf{L}_k , as done in [100]

$$\min_{\hat{\mathbf{U}} \in \mathbb{R}^{N \times N}} \sum_{k=0}^T \text{off}(\hat{\mathbf{U}} \mathbf{L}_k \hat{\mathbf{U}}) \text{ s.t. } \hat{\mathbf{U}}^T \hat{\mathbf{U}} = \mathbf{I}_N, \text{ with } \text{off}(\mathbf{A}) = \sum_{i \neq j} A_{i,j}^2$$

among other variations (see [103], [100] for a review), while a common downsampling pattern may be found i. a. via partitioning of a suitably chosen, common eigenvector (e. g. using λ_{max} [60] of a suitable transition graph). For a multilevel scheme, the individual graphs may be coarsened through Kron-reduction, with subsequent repetition of joint diagonalization (or potential coarsening with respect to common eigenbasis \mathbf{U}) at further levels to find a new common basis, where applicable.

The filtering of \mathbf{x} with respect to varying graphs can then be realized through the successive application of filters $\{\mathbf{H}_i\}_i$ on the individual graphs $\{G_i\}_i$

$$\mathbf{x}^t = \mathbf{H}_t \dots \mathbf{H}_0 \mathbf{x}$$

as a generalization of the iteration $\mathbf{x}^t = \mathbf{H}^t \mathbf{x}$; under the assumption of a common graph shift-operator, the former matrices commute and their order is interchangeable.

A time-varying graph transform (or filterbank), as the fusion of different graphs within a single transform, facilitates critical sampling and may capture and incorporate graph similarities and/or inter-relations into the signal analysis, whereas the fragmentation into separate GWTs for each graph is more costly and information on graph-connectivity evolution is lost. While the simplest realization would be the transform product $\mathbf{x}^t = \mathbf{W}_t \dots \mathbf{W}_0 \mathbf{x}$ for transform \mathbf{W}_i on graph G_i ⁸, other possible structures may feature i. a. the successive low-pass filtering of the signal on each graph (up to a certain time instant $t - 1$), with a proper downsampling on low-and high-pass branches imposed on the t -th graph, or, more consistently, the successive application of low-and high-pass filters on each graph and eventually, given a common downsampling pattern, unification in a transform.

Following the style of previous generalized graph transforms, for the sequence of undirected graphs $\{G_t = (V, E_t)\}_{t=0}^T$, $|V| = N$, with symmetric normalized (not necessarily circulant) adjacency matrices $\{\tilde{\mathbf{A}}_t\}_{t=0}^T$, which possess the same eigenbasis, i. e. $\tilde{\mathbf{A}}_t = \mathbf{V} \mathbf{\Gamma}_t \mathbf{V}^H$, we propose a set of time-varying graph wavelet *averaging* transforms, consisting of the filters

$$\mathbf{H}_{LP} = \sum_{t=0}^T h_t \mathbf{H}_{LPt} = \sum_{t=0}^T h_t \prod_{n=1}^M \left(\beta_{t,n} \mathbf{I}_N + \tilde{\mathbf{A}}_t \right)^k \quad (4.6)$$

⁸Instead of separate graphs, a given set may also characterize a partition (decomposition) of a single graph into subgraphs with $E = \cup_k E_k$ (that are ideally connected, e. g. through cycles) as done in [35].

$$\mathbf{H}_{HP} = \sum_{t=0}^T \tilde{h}_t \mathbf{H}_{HP_t} = \sum_{t=0}^T \tilde{h}_t \prod_{n=1}^M (\beta_{t,n} \mathbf{I}_N - \tilde{\mathbf{A}}_t)^k \quad (4.7)$$

and graph wavelet *product* transforms, with filters

$$\mathbf{H}_{LP} = \prod_{t=0}^T h_t \mathbf{H}_{LP_t} = \prod_{t=0}^T h_t \prod_{n=1}^S (\beta_{t,n} \mathbf{I}_N + \tilde{\mathbf{A}}_t)^k \quad (4.8)$$

$$\mathbf{H}_{HP} = \prod_{t=0}^T \tilde{h}_t \mathbf{H}_{HP_t} = \prod_{t=0}^T \tilde{h}_t \prod_{n=1}^S (\beta_{t,n} \mathbf{I}_N - \tilde{\mathbf{A}}_t)^k, \quad (4.9)$$

for respectively non-zero coefficients $h_t, \tilde{h}_t \in \mathbb{C}$.

Corollary 4.3. *The time-varying graph wavelet transform, consisting respectively of the filters*

(i) *in Eqs. (4.8) and (4.9), with parameters $\beta_{t,n} = 1$ and $S = 1 \forall t$, is invertible for any downsampling pattern, as long as at least one node retains the low-pass component for connected graphs with non-negative weights, and otherwise, for a suitable set of retained nodes V_α such that the partitioned eigenvectors $\{\mathbf{v}_{t,j}(V_\alpha)\}_{t,j}$ corresponding to $\gamma_{t,i} = 1$ (and for bipartite graphs, also $\{\tilde{\mathbf{v}}_{t,j}(V_\alpha^{\mathbb{C}})\}_{t,j}$ for $\gamma_{t,i} = -1$) of multiplicity j are linearly independent. If the graphs are bipartite, non-negativity is sufficient.*

(ii) *in Eqs. (4.8) and (4.9), is invertible for any downsampling pattern as long as $|\beta_{t,n}| \neq |\gamma_{t,i}|$ and subject to restrictions (a) the eigenvalues of $\mathbf{H}_{LP}\mathbf{H}_{HP}$ are non-zero and of the same sign or (b) $k \in 2\mathbb{N}$; otherwise, if for some $\beta_{t,n} = \gamma_{t,i}$, and the graph is non-bipartite, invertibility follows, under the previous conditions, from linear independence of the corresponding partitions $\{\mathbf{v}_{t,j}(V_\alpha)\}_{t,j}$ for suitable V_α .*

(iii) *in Eqs. (4.6) and (4.7), is invertible under the same conditions as (ii), where in (b) it is additionally required that $h_t, \tilde{h}_t \leq 0 \forall t$; otherwise, if $\beta_{t,n} = \gamma_{t,i}$ holds on each graph G_t such that all \mathbf{H}_{HP_t} share the same nullspace \mathbf{v} , invertibility follows under the preceding provided the corresponding partitions $\mathbf{v}(V_\alpha)$ (and $\tilde{\mathbf{v}}(V_\alpha^{\mathbb{C}})$ in the bipartite case, for $\tilde{\mathbf{v}}$ inducing $-\gamma_{t,i}$ on each graph G_t) are linearly independent for suitable V_α .*

Proof: See Appendix B.1.

For the special case of circulant \mathbf{A}_t , with $\{\beta_n\}_n = \{\frac{\tilde{d}_{\alpha_n}}{d}\}_n$ and suitable coefficients $\{h_t, \tilde{h}_t\}_t$, commutativity and the identification of a common downsampling pattern are ensured, while a wider range of annihilation possibilities can be incorporated.

For more generalized designs, in the sense of Sect. 4.1, one may consider a filterbank of the form

$$\mathbf{H}_{LP} = \sum_k a_k \mathbf{H}_{LP_k}$$

$$\mathbf{H}_{HP} = \sum_k \tilde{a}_k \mathbf{H}_{HP_k}$$

with equivalent restrictions on the frequency response. Alternatively, instead of aligning pre-constructed individual graph filters, a more generalized option is to formulate the transform directly in terms of the graph adjacency matrices:

$$\mathbf{H}_{LP} = \sum_{k=0}^K h_k \tilde{\mathbf{A}}^k$$

$$\mathbf{H}_{HP} = \sum_{k=0}^K \tilde{h}_k \tilde{\mathbf{A}}^k$$

with symmetric $\tilde{\mathbf{A}} = \prod_{t=0}^T a_t \mathbf{A}_t$ or $\tilde{\mathbf{A}} = \sum_{t=0}^T a_t \mathbf{A}_t$ subject to suitable restrictions on \mathbf{A}_t and $\{h_t, \tilde{h}_t\}_t$.

When different graph filters on a given set of (circulant) graphs have the same annihilation property (i.e. annihilate the same type of signal), their sum preserves that, while their product may generalize it to higher (or multiple) order, where applicable; in either case when annihilation is localized, graphs need to be sufficiently banded due to the border effect. For instance, the product of (banded) circulant graph Laplacians $\mathbf{L}_1 \dots \mathbf{L}_k$ has the same number of vanishing moments as the single (banded) circulant \mathbf{L}^k , but on different graphs combined. In certain cases, the sum of graph e-spline filters of different e-degrees (eigenvalues) can lead to the creation of new vanishing moments, or more generally, create a shift in the nullspace. The case of the symmetrized circulant graph Laplacian $\tilde{\mathcal{L}}_\alpha$ for directed circulant graphs, treated in Sect. 3.4.1 is a direct example for when the addition of two operators on different graphs with opposing vanishing moments can induce new vanishing moments.

While the present constructions can also be interpreted as formations of a unifying graph, the main objective is to provide a notion of a *time-varying* graph filterbank, whose filters can be adapted to rank and merge the contribution of individual graphs, while the latter are still separately preserved i. a. for subsequent (multiscale) operations.

4.4 The Condition Number of the GWT

In order to assess the ability of the derived graph wavelet transforms \mathbf{W} to sparsely represent smooth signals \mathbf{x} under perturbations or noise, as well as evaluate their proximity to orthogonal bases, one may compute the frame bounds of the analysis operator \mathbf{W} , given

by the extremal eigenvalues of $\mathbf{W}^T \mathbf{W}$

$$\lambda_{\min} \|\mathbf{x}\|_2^2 \leq \|\mathbf{W}\mathbf{x}\|_2^2 \leq \lambda_{\max} \|\mathbf{x}\|_2^2$$

with $0 < \lambda_{\min} \leq \lambda_{\max} < \infty$. The ratio $C = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}}$ is known as the condition number of \mathbf{W} , as a measure of its sensitivity with respect to errors or noise, with $C = 1$ characterizing a unitary matrix and $C \rightarrow \infty$ a singular (ill-conditioned) one [97].

It can be demonstrated that in the case of a general bipartite graph, there exists an explicit expression for the condition number of the GWT as a result of the spectral folding phenomenon. Hence, the set of imposed filter functions, which satisfy the conditions of the generalized perfect reconstruction filterbank on a bipartite graph, as discussed earlier, facilitate the direct computation of C .

Consider the symmetric operator

$$\mathbf{W}^T \mathbf{W} = \frac{1}{2}(\mathbf{H}_{LP}^2 + \mathbf{H}_{HP}^2 + \mathbf{H}_{LP} \mathbf{K} \mathbf{H}_{LP} - \mathbf{H}_{HP} \mathbf{K} \mathbf{H}_{HP}) = \frac{1}{2}(\mathbf{H}_{LP}^2 + \mathbf{H}_{HP}^2)$$

where \mathbf{K} is the downsampling matrix on a bipartite graph. The last simplification follows from $\mathbf{K} \mathbf{H}_{LP} = \mathbf{H}_{HP} \mathbf{K}$, which, for appropriately chosen filter functions, is specific to a bipartite graph. In particular, assuming wlog the filter form $\mathbf{H}_{LP/HP} = \varphi_{LP/HP}(\mathbf{A}_n)$, where $\varphi(\cdot)$ is a function applied on the symmetric normalized graph adjacency matrix⁹, and letting $\gamma(\mathbf{A}_n)$ denote the spectrum of \mathbf{A}_n and $\mathbf{P}_{\gamma_i} = \mathbf{u}_i \mathbf{u}_i^T$ the projection matrix of eigenvector \mathbf{u}_i of \mathbf{A}_n , corresponding to γ_i , such that $\mathbf{H}_{LP/HP} = \sum_{\gamma_i \in \gamma(\mathbf{A}_n)} \varphi_{LP/HP}(\gamma_i) \mathbf{P}_{\gamma_i}$, it can be shown that

$$\mathbf{K} \mathbf{H}_{LP} = \sum_{\gamma_i \in \gamma(\mathbf{A}_n)} \varphi_{LP}(\gamma_i) \mathbf{K} \mathbf{P}_{\gamma_i} = \sum_{\gamma_i \in \gamma(\mathbf{A}_n)} \varphi_{LP}(-\gamma_i) \mathbf{P}_{\gamma_i} \mathbf{K}$$

and further, in an equivalent fashion to the derivation in ([35], Eq. (33))

$$\mathbf{H}_{LP} \mathbf{K} \mathbf{H}_{LP} - \mathbf{H}_{HP} \mathbf{K} \mathbf{H}_{HP} = \sum_{\gamma_i \in \gamma(\mathbf{A}_n)} (\varphi_{LP}(\gamma_i) \varphi_{LP}(-\gamma_i) - \varphi_{HP}(\gamma_i) \varphi_{HP}(-\gamma_i)) \mathbf{K} \mathbf{P}_{\gamma_i}$$

due to $\mathbf{K} \mathbf{P}_{\gamma_i} = \mathbf{P}_{-\gamma_i} \mathbf{K}$, $\forall \gamma_i \in \gamma(\mathbf{A}_n)$. Thus, for $\varphi_{LP}(-\gamma) = \varphi_{HP}(\gamma)$, which is the case for most proposed constructions, the eigenvalues of $\mathbf{W}^T \mathbf{W}$ can be directly expressed as $\lambda = \frac{1}{2}(\varphi_{LP}(\gamma)^2 + \varphi_{HP}(\gamma)^2)$, resulting in the explicit condition number $C = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}}$ of \mathbf{W} . In the special case of a circulant bipartite graph, one can additionally show that the representer polynomials satisfy $H_{LP}(-z) = H_{HP}(z)$, which is equivalent to $\mathbf{K} \mathbf{H}_{LP} = \mathbf{H}_{HP} \mathbf{K}$, while the matrix $\mathbf{W}^T \mathbf{W}$ is also circulant.

⁹In the case of node-variant filters, this may be replaced by $\tilde{\mathbf{A}}_n$, or for multiple graphs G_i , by $\mathbf{A}_{n,i}$.

Note that the expression for the condition number coincides with that derived in [35], yet the latter uses an approximation to reduce to that, while here specific filter functions $\varphi(\cdot)$ are considered that facilitate exact simplifications.

4.4.1 Comparison of Different GWTs

We proceed to evaluate and compare the condition numbers for a selection of proposed graph transforms, in order to further comprehend their variation in performance, when the graph is undirected and bipartite, and downsampling is fixed as above. Here we assume for simplicity that the adjacency matrix, and by extension any symmetric (parameter-)normalized version, always has the eigenvalue $\gamma = 0$ and is thus singular. While this is a frequent case, there also exist bipartite graphs with non-singular adjacency matrices, however, an intuitive or direct classification into either set is not known (see [104] for an algorithmic proposition); in the latter case, one may thus simply adjust derived formulae by replacing $\gamma = 0$ with the eigenvalue of smallest magnitude, where applicable.

Corollary 4.4. *The condition number of the HGSWT of Thm. 3.1 for undirected bipartite graphs is $C = \sqrt{2^{2k-1}}$.*

Proof. In the case of the simple spline filterbank, i. e. when $\beta = 1$, we have $\varphi_{LP/HP}(\gamma) = (1 \pm \gamma)^k$ and define $f(\gamma) = \frac{1}{2}(1 + \gamma)^{2k} + \frac{1}{2}(1 - \gamma)^{2k}$, which is assumed to be continuous; the filter coefficients are omitted for simplicity as they eventually cancel out within the fraction of eigenvalues in C . With $\gamma_{max} = 1$ and $\gamma_{min} = -1$, we prove that $f(\gamma)$ is strictly increasing for $\gamma \in [0, 1]$, and decreasing for $\gamma \in [-1, 0]$, by considering derivative $f'(\gamma)$: as a consequence of the Mean Value Theorem, we have $f'(\gamma) = k((1 + \gamma)^{2k-1} - (1 - \gamma)^{2k-1}) > 0, \forall \gamma \in (0, 1]$ and $f'(\gamma) < 0, \forall \gamma \in [-1, 0)$. Thus, $f_{min} = f(0) = 1$ and $f_{max} = f(\gamma_{max}) = f(\gamma_{min}) = 2^{2k-1}$, given $f(-\gamma) = f(\gamma)$, which proves $C = \sqrt{2^{2k-1}}$. \square

In another interesting property of bipartite graphs, it can be shown that the above condition number stays the same regardless of the graph connectivity for (non-parameterized) strictly monotonic simple spline filters $\varphi(\gamma) = (1 \pm \gamma)^k$, since the extremal eigenvalues $\gamma_{max} = -\gamma_{min} = 1$ stay the same; this ceases to be the case for parameterized constructions, where the e-degree (or eigenvalue) parameter β changes with the graph, or non-singular graphs.

Corollary 4.5. *The condition number of the HGESWT of Thm. 3.2 (by extension, Cor. 4.1) is $C_B = \sqrt{\frac{\prod_{n=1}^T (\beta_n + 1)^{2k} + \prod_{n=1}^T (\beta_n - 1)^{2k}}{2 \prod_{n=1}^T \beta_n^{2k}}}$, for parameters of the same sign.*

Proof. When $\beta_n \leq 0, \forall n$, monotonicity is upheld, with $f(\gamma) = \frac{1}{2} \prod_{n=1}^T (\beta_n + \gamma)^{2k} + \frac{1}{2} \prod_{n=1}^T (\beta_n - \gamma)^{2k}$ and $f'(\gamma) \leq 0$ since the resulting series $f(\gamma) = \sum_{i=0}^{2kT} \frac{1}{2} (s_i + (-1)^i s_i) \gamma^i$ is strictly monotonic for $s_i \leq 0$, increasing for $\gamma \in [0, 1]$ and decreasing for $\gamma \in [-1, 0]$, with $f'(\gamma) = \frac{1}{2} \sum_{i=1}^{2kT} i s_i \gamma^{i-1} + \frac{1}{2} \sum_{i=1}^{2kT} (-1)^i i s_i \gamma^{i-1} = \sum_{i=1}^{kT} 2i s_{2i} \gamma^{2i-1}$. Note that the elementary symmetric polynomials s_i are of the form $s_0 = \beta_1 \dots \beta_{2Tk}$, ..., $s_{2Tk-1} = \beta_1 + \dots + \beta_{2Tk}$, $s_{2Tk} = 1$ for redundant $\{\beta_i\}_{i=1}^{2Tk}$, i.e. at even numbered positions we always have $s_{2i} > 0$ provided all β_n are of the same sign (as a sufficient condition). Thus $C = \sqrt{\frac{f(1)}{f(0)}} = \sqrt{\frac{\prod_{n=1}^T (\beta_n + 1)^{2k} + \prod_{n=1}^T (\beta_n - 1)^{2k}}{2 \prod_{n=1}^T \beta_n^{2k}}}$. \square

Further, we compare the condition numbers of the node-variant and parameterized bandlimiting wavelet transform for bipartite (not necessarily circulant) graphs; this is of relevance when the given signal to analyze is piecewise smooth (generalized bandlimited) and noisy, considering that both transforms may be tailored to annihilate the clean version.

Corollary 4.6. *The condition number C_S of the space-variant HSVGSWT transform in Thm. 4.2 is bounded as follows $\sqrt{2^{2k-1}} \leq C_S \leq \sqrt{\frac{1}{2}(1 + \frac{1}{\tilde{d}_{min}})^{2k} + \frac{1}{2}(1 - \frac{1}{\tilde{d}_{min}})^{2k}}$.*

Proof. Based on the function $f(\tilde{\gamma}) = \frac{1}{2}(1 + \tilde{\gamma})^{2k} + \frac{1}{2}(1 - \tilde{\gamma})^{2k}$, where $\tilde{\Gamma}$ are the eigenvalues of $\tilde{\mathbf{A}} = \tilde{\mathbf{D}}^{-1/2} \mathbf{A}_n \tilde{\mathbf{D}}^{-1/2}$, monotonicity follows from Cor. 4.4. Hence, we have $f_{max} = f(\tilde{\gamma}_{max}) = f(-\tilde{\gamma}_{min}) \leq f(\frac{1}{\tilde{d}_{min}}) = \frac{1}{2}(1 + \frac{1}{\tilde{d}_{min}})^{2k} + \frac{1}{2}(1 - \frac{1}{\tilde{d}_{min}})^{2k}$ and $f_{min} = f(0) = 1$, where \tilde{d}_{min} and \tilde{d}_{max} denote the extremal magnitude values of diagonal $\tilde{\mathbf{D}}$, giving $C_S \leq \sqrt{\frac{1}{2}(1 + \frac{1}{\tilde{d}_{min}})^{2k} + \frac{1}{2}(1 - \frac{1}{\tilde{d}_{min}})^{2k}}$. Since $f(1) \leq f(\tilde{\gamma}_{max})$, we also have the condition number of the HGSWT as a lower bound $C = \sqrt{2^{2k-1}} \leq C_S$. \square

Corollary 4.7. *The condition number C_S of the space-variant HSVGSWT is consistently smaller than that of the bandlimiting HBGWT C_B for equivalent parameterizations (of the same sign).*

Proof. Consider for simplicity an e-spline/bandlimiting GWT with two parameters β_1 and β_2 and space-variant GWT with two diagonal elements $\beta_1 = \tilde{d}_{min}$ and $\beta_2 = \tilde{d}_{max}$ of the same sign; we know that $|\tilde{d}| \leq 1$, as these constitute normalized adjacency eigenvalues (or e-degrees), and so $\frac{1}{|\tilde{d}|} \geq 1$, provided $\tilde{d} \neq 0$. Then the function corresponding to the former, $f(\gamma) = \frac{1}{2}(\beta_1 + \gamma)^{2k}(\beta_2 + \gamma)^{2k} + \frac{1}{2}(\beta_1 - \gamma)^{2k}(\beta_2 - \gamma)^{2k}$ gives $f_{max} = f(1) = \frac{1}{2}(\beta_1 + 1)^{2k}(\beta_2 + 1)^{2k} + \frac{1}{2}(\beta_1 - 1)^{2k}(\beta_2 - 1)^{2k}$ and $f_{min} = (\beta_1 \beta_2)^{2k}$, and hence

$$C_B = \sqrt{\frac{1}{2} \left(1 + \frac{1}{\beta_1}\right)^{2k} \left(1 + \frac{1}{\beta_2}\right)^{2k} + \frac{1}{2} \left(1 - \frac{1}{\beta_1}\right)^{2k} \left(1 - \frac{1}{\beta_2}\right)^{2k}}.$$

Assume, wlog that $\beta_1, \beta_2 > 0$; thus, for condition numbers C_S and C_B , we have

$$\begin{aligned} C_S^2 &\leq \frac{1}{2} \left(1 + \frac{1}{\beta_1}\right)^{2k} + \frac{1}{2} \left(1 - \frac{1}{\beta_1}\right)^{2k} \leq \left(1 + \frac{1}{\beta_1}\right)^{2k} \\ &\leq \frac{1}{2} \left(1 + \frac{1}{\beta_1}\right)^{2k} (1+1)^{2k} \leq \frac{1}{2} \left(1 + \frac{1}{\beta_1}\right)^{2k} \left(1 + \frac{1}{\beta_2}\right)^{2k} \\ &\leq \frac{1}{2} \left(\left(1 + \frac{1}{\beta_1}\right)^{2k} \left(1 + \frac{1}{\beta_2}\right)^{2k} + \left(1 - \frac{1}{\beta_1}\right)^{2k} \left(1 - \frac{1}{\beta_2}\right)^{2k} \right) = C_B^2 \end{aligned}$$

This can then be generalized for multiple parameters, since the upper bound for C_S always depends on the smallest single parameter \tilde{d}_{min} , while C_B grows rapidly for an increasing number of parameters. \square

While the previous explicit results and comparisons are reserved for bipartite graphs, one can nevertheless make similar observations about the condition numbers of general undirected graphs, in that transforms consisting of filters with multiple matrix factors (of the same sign) gain rather large condition numbers rapidly, while an increase in the graph connectivity or complexity (at a consistent transform order) leads to a comparatively slower increase in condition number, as was already evidenced in the experiments of Chapter 3; nevertheless, these generalizations remain observational.

4.5 Graph Products and Approximations: A Multidimensional Extension

As part of a more generalized motivation which facilitates the multi- and lower-dimensional processing and representation of signals on graphs, we at last explore graph product approximations as a promising avenue. In particular, given an arbitrary undirected graph, one may consider its approximation as the graph product of factor graphs of certain desirable structures, such as circulants, which may be further leveraged to implement transformations and induce sparsity.¹⁰ Multi-dimensional wavelet analysis has been considered i. a. for bipartite graphs as the operation with respect to separate edge sets on the same vertex set within a bipartite subgraph decomposition [35]. In the proposed framework for graph approximation and wavelet analysis via graph product decomposition, however, the notion of multiple dimensions arises from the graph product operation itself, with each factor constituting a separate dimension.

Graph products [105] have been studied and applied in a variety of contexts for purposes such as modelling realistic networks and/or rendering matrix operations computationally

¹⁰The content of this section appears in [7].

efficient ([106], [107], [108]). Their relevance for GSP was first considered in [109] as a means of modelling and representation of complex (high-dimensional) data as graph signals defined on product graphs, with the potential of promoting more efficient implementations of graph operations, such as graph filtering. The motivation for considering graph products here is twofold: (1) as a scheme which can decompose arbitrary graphs into circulant graphs, to facilitate the processing of graph signals with respect to the circulant approximations, and (2) as a means to conduct operations in lower dimensional settings in order to increase efficiency. Beyond that graph products can be used to model time-varying processes, and, in contrast to previously considered fusion-based transformations, operate on an extended set of nodes.

4.5.1 Graph Products of Circulants

The product, \diamond , of two (undirected) graphs $G_1 = (V(G_1), E(G_1))$ and $G_2 = (V(G_2), E(G_2))$, also referred to as factors, with respective adjacency matrices $\mathbf{A}_1 \in \mathbb{R}^{N_1 \times N_1}$ and $\mathbf{A}_2 \in \mathbb{R}^{N_2 \times N_2}$, is formed by denoting the new vertex set of the resulting graph G as the Cartesian product $V(G) = V(G_1) \times V(G_2)$, and defining the new edge relations $E(G)$ according to the characteristic adjacency rules of the product operation, resulting in adjacency matrix $\mathbf{A}_\diamond \in \mathbb{R}^{N_1 N_2 \times N_1 N_2}$. We identify four main graph products of interest:

- Kronecker product $G_1 \otimes G_2$: $\mathbf{A}_\otimes = \mathbf{A}_1 \otimes \mathbf{A}_2$
- Cartesian product $G_1 \times G_2$: $\mathbf{A}_\times = \mathbf{A}_1 \times \mathbf{A}_2 = \mathbf{A}_1 \otimes \mathbf{I}_{N_2} + \mathbf{I}_{N_1} \otimes \mathbf{A}_2$
- Strong product $G_1 \boxtimes G_2$: $\mathbf{A}_\boxtimes = \mathbf{A}_1 \boxtimes \mathbf{A}_2 = \mathbf{A}_1 \otimes \mathbf{A}_2 + \mathbf{A}_1 \otimes \mathbf{I}_{N_2} + \mathbf{I}_{N_1} \otimes \mathbf{A}_2$
- Lexicographic product $G_1[G_2]$: $\mathbf{A}_\lceil \rfloor = \mathbf{A}_1[\mathbf{A}_2] = \mathbf{A}_1 \otimes \mathbf{J}_{N_2} + \mathbf{I}_{N_1} \otimes \mathbf{A}_2$

where $\mathbf{J}_{N_2} = \mathbf{1}_{N_2} \mathbf{1}_{N_2}^T$. The lexicographic product in particular can be regarded as a variation of the Cartesian product, yet contrary to the others, it is not commutative for unlabelled graphs [105]. Furthermore, the adjacency matrices \mathbf{A}_\diamond of the first three products possess the same (orthogonal) eigenbasis $\mathbf{V} = \mathbf{V}_1 \otimes \mathbf{V}_2$, for decompositions $\mathbf{A}_1 = \mathbf{V}_1 \mathbf{\Gamma}_1 \mathbf{V}_1^H$ and $\mathbf{A}_2 = \mathbf{V}_2 \mathbf{\Gamma}_2 \mathbf{V}_2^H$, and eigenvalues of the form $\mathbf{\Gamma}_\diamond = \mathbf{\Gamma}_1 \diamond \mathbf{\Gamma}_2$ ([105]).

Graph products have been employed to model realistic networks due to their ability to capture present regularities, such as patterns and recursive community growth [107], and may therefore serve as suitable approximations to networks with inherent substructures, such as social networks consisting of similarly structured communities or time-evolving sensor networks [109]. For the ensuing analysis, we exploit an existing matrix scheme

for arbitrary graphs which imposes desired constraints, such as that of circularity, on the individual factors¹¹.

The Kronecker Product Approximation

Given an arbitrary undirected graph G with adjacency matrix \mathbf{A} , one may resort to a result from matrix theory by Pitsianis et al. [110] which facilitates the approximate Kronecker product decomposition $\mathbf{A} \approx \mathbf{A}_1 \otimes \mathbf{A}_2$ into circulant (adjacency) matrices \mathbf{A}_i of suitably chosen dimension N_i , by solving the convex optimization problem

$$\min_{\mathbf{C}_1^T \text{vec}(\mathbf{A}_1)=0, \mathbf{C}_2^T \text{vec}(\mathbf{A}_2)=0} \|\mathbf{A} - \mathbf{A}_1 \otimes \mathbf{A}_2\|_F$$

subject to linear constraints in the form of structured, rectangular matrices \mathbf{C}_i with entries $\{0, 1, -1\}$, which impose circularity (or other pattern) on \mathbf{A}_i via column-stacking operator vec . Closed-form solutions $\text{vec}(\mathbf{A}_i), i = 1, 2$ are obtained by solving a reduced unconstrained problem, after expressing the above as a rank-1 approximation problem (see [110] for details). In addition, desirable properties such as symmetry and bandedness may be imposed, among others ([110], [108]).

Exact Graph Products

In reverse action, we observe that the general graph product of circulants gives rise to block-circulant structures (or sums thereof), as the example in Fig. 4.4 illustrates. Here, the Cartesian product of two circulant graphs G_1 and G_2 of respective dimension N_1 and N_2 is formed by connecting N_1 blocks of G_2 according to the connectivity of G_1 .

However, a subset of circulant graphs can be represented as the graph products of circulant factors; while such cases are marginal, they motivate decompositions for lower-dimensional processing.

Circulant graphs are not generally closed under the graph product operation, with the exception of the lexicographic product [111]. In particular, the product $G_1[G_2]$ of two circulant graphs $G_1 = C_{N_1, S_1}$ and $G_2 = C_{N_2, S_2}$ of respective dimensions N_1 and N_2 and with generating sets S_1 and S_2 , is isomorphic to the circulant graph $C_{N_1 N_2, S}$ with generating set $S = \left(\bigcup_{t=0}^{\lfloor \frac{N_2-1}{2} \rfloor} tN_1 + S_1 \right) \cup \left(\bigcup_{t=1}^{\lfloor \frac{N_2}{2} \rfloor} tN_1 - S_1 \right) \cup N_1 S_2$ [111]. The graph $C_{N_1 N_2, S}$ is connected with $1 \in S$ only if G_1 is connected with $1 \in S_1$. The characteristic adjacency matrix $\mathbf{A}_{[\]}$ is not circulant per se, but its isomorphism $\tilde{\mathbf{A}}_{[\]} = \mathbf{P} \mathbf{A}_{[\]} \mathbf{P}^T$ is, where permutation matrix \mathbf{P} performs the relabelling $\{0, \dots, N_1 N_2 - 1\} \rightarrow \{0 : N_2 :$

¹¹On a related note, in [106] Kronecker product approximation is identified as a means to increase computational efficiency for large structured least-squares problems in image restoration, which coincidentally marks an area where circulant approximations are used extensively.

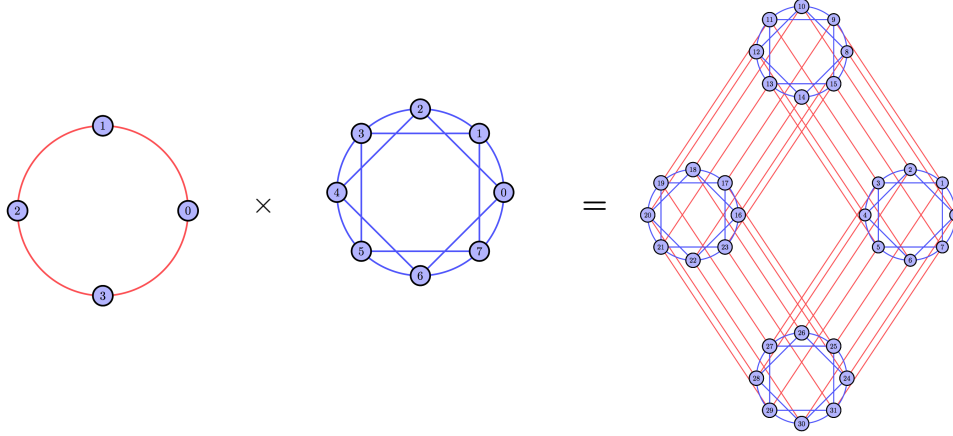


Figure 4.4: Graph Cartesian Product of two unweighted circulant graphs. ©2017 Elsevier Inc.

$N_1 N_2 - 1, 1 : N_2 : N_1 N_2 - 1, \dots, N_2 - 1 : N_2 : N_1 N_2 - 1\}$ such that each product node $(g_{1,j}, g_{2,k}) \in V(G)$ is labelled as $g_{1,j} + N_1 g_{2,k}$, for $g_{i,j} \in V(G_i)$.

Further special cases of graph products that remain circulant with circulant factors are discussed in [112].

4.5.2 Multi-dimensional Wavelet Analysis on Product Graphs

The following analysis explores how the developed theory on graph wavelets, and circulant graphs in particular, can be extended to product graphs¹². Here, we operate under the assumption that the decomposition (and decomposition type) of an arbitrary graph into circulants (or alternative graph structures) is either known (exactly or approximately), or unknown, in which case one can resort to a Kronecker product approximation. This primarily requires the identification of the graph Laplacian of product graphs \mathbf{L}_\diamond as a relevant high-pass filter; its specific interpretation as an extension of the circulant graph Laplacian high-pass filter to higher dimensions, with associated property preservations, will be revisited in Sect. 4.5.3. In particular, the formation of \mathbf{L}_\diamond does not directly reflect the adjacency matrix relations, except in the case of the Cartesian product ([113], [114]):

- Kronecker product: $\mathbf{L}_\otimes = \mathbf{D}_1 \otimes \mathbf{D}_2 - \mathbf{A}_1 \otimes \mathbf{A}_2 = \mathbf{L}_1 \otimes \mathbf{D}_2 + \mathbf{D}_1 \otimes \mathbf{L}_2 - \mathbf{L}_1 \otimes \mathbf{L}_2$
- Cartesian product: $\mathbf{L}_\times = \mathbf{D}_1 \times \mathbf{D}_2 - \mathbf{A}_1 \times \mathbf{A}_2 = \mathbf{L}_1 \otimes \mathbf{I}_{N_2} + \mathbf{I}_{N_1} \otimes \mathbf{L}_2$
- Strong product: $\mathbf{L}_\boxtimes = \mathbf{D}_1 \boxtimes \mathbf{D}_2 - \mathbf{A}_1 \boxtimes \mathbf{A}_2 = \mathbf{L}_1 \otimes \mathbf{D}_2 + \mathbf{D}_1 \otimes \mathbf{L}_2 - \mathbf{L}_1 \otimes \mathbf{L}_2 + \mathbf{L}_1 \otimes \mathbf{I}_{N_2} + \mathbf{I}_{N_1} \otimes \mathbf{L}_2$
- Lexicographic product: $\mathbf{L}_{[\]} = \mathbf{D}_1[\mathbf{D}_2] - \mathbf{A}_1[\mathbf{A}_2] = \mathbf{I}_{N_1} \otimes \mathbf{L}_2 + \mathbf{L}_1 \otimes \mathbf{J}_{N_2} + \mathbf{D}_1 \otimes (N_2 \mathbf{I}_{N_2} - \mathbf{J}_{N_2})$

¹²Note that this can be naturally extended to graph products between more than two factors.

For regular connected graph factors G_i , equivalent relations between the eigenbases \mathbf{U} of $\mathbf{L}_\diamond = \mathbf{U}\mathbf{\Lambda}_\diamond\mathbf{U}^H$ and \mathbf{U}_i of $\mathbf{L}_i = \mathbf{U}_i\mathbf{\Lambda}_i\mathbf{U}_i^H$ as for the adjacency matrices hold, with $\mathbf{U} = \mathbf{U}_1 \otimes \mathbf{U}_2$ and eigenvalues $\mathbf{\Lambda}_\diamond = \mathbf{D}_\diamond - \mathbf{\Gamma}_\diamond$ [113]; these relations are further preserved for the Cartesian product when the factors G_i are generic, while a nearer characterization for the remaining cases is subject to investigation ([114], [113]). The eigenvectors of $\mathbf{L}_{[\]}$ are defined as $\mathbf{U}_{[\]} = [\{\mathbf{u}_{1,i} \otimes \mathbf{1}_{N_2}\}_{i=1}^{N_1} | \{\mathbf{e}_i \otimes \mathbf{u}_{2,j}\}_{i=1, j=2}^{i=N_1, j=N_2}]$, when G_1 is connected [114].

Therefore, under the first three products, the special case of circulant graph factors, for which $\mathbf{U} = \mathbf{V}$ can be represented as the 2D DFT matrix, and arbitrary regular graph factors by extension, reveals that each graph Laplacian eigenvector \mathbf{u}_j of \mathbf{L}_\diamond is the Kronecker product of the graph Laplacian eigenvectors of its factor graphs. This insight motivates a more generalized view of a graph signal for GSP:

Definition 4.3. Any graph signal $\mathbf{x} \in \mathbb{R}^N$, with $N = N_1N_2$, can be decomposed as $\mathbf{x} = \sum_{s=1}^k \mathbf{x}_{s,1} \otimes \mathbf{x}_{s,2} = \text{vec}_r\{\sum_{s=1}^k \mathbf{x}_{s,1}\mathbf{x}_{s,2}^T\}$, where $\text{vec}_r\{\}$ indicates the row-stacking operation, or, equivalently, $\sum_{s=1}^k \mathbf{x}_{s,1}\mathbf{x}_{s,2}^T$ has rank k with $\mathbf{x}_{s,i} \in \mathbb{R}^{N_i}$. For \mathbf{x} residing on the vertices of an arbitrary undirected graph G , which admits the graph product decomposition of type \diamond , such that $G_\diamond = G_1 \diamond G_2$ and $|V(G_i)| = N_i$, one can redefine and process \mathbf{x} as the graph signal tensor factors $\mathbf{x}_{s,i}$ on G_i .

While \mathbf{x} does not generally lie in the graph (Laplacian) eigenspace of the underlying graph, and alternative decompositions are possible, the above perspective is adopted as a promising interpretation of component-wise processing of graph signals defined on product graphs. For an undirected arbitrary graph product with eigenbasis $\mathbf{U} = \mathbf{U}_1 \otimes \mathbf{U}_2$, a signal of the form $\mathbf{x} = \sum_{s=1}^k \mathbf{u}_{s,1} \otimes \mathbf{u}_{s,2}$ is multi-dimensional (generalized) bandlimited as both \mathbf{x} and its components are respectively bandlimited on G and $\{G_i\}_{i=1}^2$. The special case $\mathbf{x} = \mathbf{x}_1 \otimes \mathbf{x}_2$ (for rank $k = 1$) is inspected more closely in Sect. 4.5.3, as it facilitates concrete claims on the smoothness and sparsity relations between a signal and its tensor factors on a graph.

Separable vs Non-separable Wavelet Analysis

Let graph signal \mathbf{x} reside on the vertices of an arbitrary, undirected graph G with graph product decomposition $G_\diamond = G_1 \diamond G_2$, which can be exact or approximate, such that the factors G_i are circulant with adjacency matrices $\mathbf{A}_i \in \mathbb{R}^{N_i \times N_i}$, $i = 1, 2$ and connected with $s = 1 \in S_i$, $i = 1, 2$ in particular (or undirected regular and connected in general).

We propose a non-separable and a separable wavelet transform on G_\diamond : the former operates on the product graph directly, while the latter acts on each factor graph independently, thereby omitting the inter-connections arising through the graph product operation between the two factors.

Define the *non-separable* graph wavelet transform on G_\diamond with adjacency matrix \mathbf{A}_\diamond as

$$\mathbf{H}_\diamond = \frac{1}{2}(\mathbf{I}_N + \mathbf{K}) \prod_{n=1}^T \frac{1}{2^k} \left(\beta_{\diamond,n} \mathbf{I}_N + \frac{\mathbf{A}_\diamond}{d} \right)^k + \frac{1}{2}(\mathbf{I}_N - \mathbf{K}) \prod_{n=1}^T \frac{1}{2^k} \left(\beta_{\diamond,n} \mathbf{I}_N - \frac{\mathbf{A}_\diamond}{d} \right)^k, \quad k \in \mathbb{N}.$$

As has transpired from earlier discussions, this is verifiably invertible for any downsampling pattern \mathbf{K} as long as at least one low-pass component is retained when $\beta_{\diamond,n} = 1, \forall n$ and subject to further conditions otherwise, following Cor. 4.1 and others. The fundamental properties which ensure this extension are that G_\diamond is regular and connected [50], i. e. the spectrum of $\frac{\mathbf{A}_\diamond}{d}$ is such that $|\gamma_{\diamond,i}| < \gamma_{\diamond,max} = 1$, with $\frac{\mathbf{A}_\diamond}{d} \mathbf{1}_{N_1 N_2} = \mathbf{1}_{N_1 N_2}$ and corresponding $\gamma_{\diamond,max}$ of multiplicity 1. Here, G_\diamond is connected under the Cartesian product for connected G_i and under the Kronecker product, if in addition at least one G_i is non-bipartite [115]. Wavelet constructions with exponential degree (or eigenvalue) parameters à la Thm 3.2 and Cor. 4.1 can be similarly extended to product graphs under equivalent restrictions on cases $|\beta_{\diamond,n}| = |\gamma_{\diamond,i}|$.

While the graph product G_\diamond of circulants is not LSI per se, it is invariant with respect to circular shifts on its factors, i. e. the matrix $\mathbf{P}_\otimes = \mathbf{P}_{N_1} \otimes \mathbf{P}_{N_2}$, for circulant permutation matrices \mathbf{P}_{N_i} , commutes with filters defined on \mathbf{A}_\diamond .

As a result, one can conduct multiresolution analysis with respect to product graphs by performing downsampling and graph coarsening operations on its individual factors, where one level corresponds to operating on either G_i . For instance, given downsampling matrix \mathbf{K}_2 (with respect to $s = 1 \in S_2$) for circulant G_2 in Fig. 4.4, one may apply the downsampling pattern $\mathbf{K} = \mathbf{I}_{N_1} \otimes \mathbf{K}_2$ on G_\diamond and redefine the sampled low-pass output on $G_1 \diamond \tilde{G}_2$, where \tilde{G}_2 represents the coarsened version of G_2 (see Fig. 4.5). This notion similarly extends to undirected (regular) product graphs of the first three types as filters of the form $\mathbf{H}_\diamond = \varphi(\mathbf{A}_\diamond)$ (for some function $\varphi(\cdot)$) are shift-invariant with respect to a designated graph-operator with common eigenbasis $\mathbf{U} = \mathbf{U}_1 \otimes \mathbf{U}_2$.

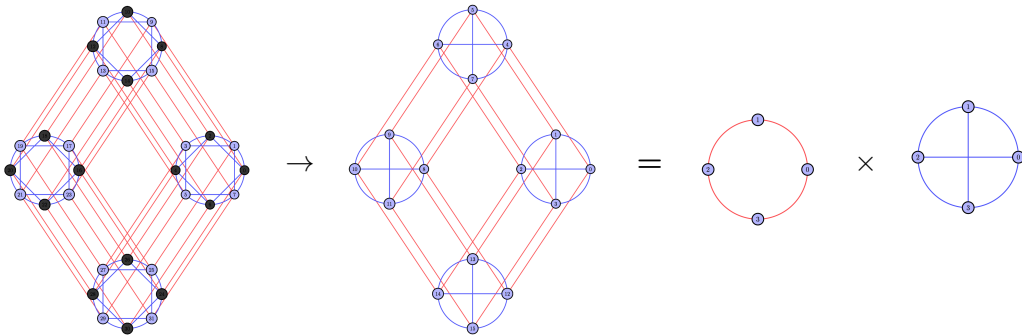


Figure 4.5: Graph Downsampling and Coarsening of G_\diamond in Fig 4.4 on G_2 w.r.t. $s = 1 \in S_2$ with coarsened \tilde{G}_2 . ©2017 Elsevier Inc.

The *separable* graph wavelet transform on G_\diamond is proposed as an alternative construction, which is applied with respect to the individual graph factors.

Let \mathbf{W}_i denote the generalized graph wavelet transform of (circulant) graph factor G_i

$$\mathbf{W}_i = \begin{bmatrix} \Psi_{\downarrow 2} \mathbf{H}_{LP_{\bar{\alpha}}} \\ \Phi_{\downarrow 2} \mathbf{H}_{HP_{\bar{\alpha}}} \end{bmatrix},$$

with downsampling matrices $\Psi_{\downarrow 2}, \Phi_{\downarrow 2}$ and filters $\mathbf{H}_{LP_{\bar{\alpha}}/HP_{\bar{\alpha}}}$ as defined in Sect. 3 (or more generally).¹³ Then $(\mathbf{W}_1 \otimes \mathbf{W}_2)$ represents the separable transform processing \mathbf{x} with respect to G_1 and G_2 , which entails the analysis of N_1 graph signal partitions $\{x((0 : N_2 - 1) + (t - 1) * N_2)\}_{t=1}^{N_1}$ on G_2 , and subsequent N_2 partitions $\{w(t : N_2 : N_1 N_2 - 1)\}_{t=0}^{N_2-1}$ on G_1 , with $\mathbf{w} = (\mathbf{I}_{N_1} \otimes \mathbf{W}_2)\mathbf{x}$. For partition $\bar{\mathbf{x}}_i \in \mathbb{R}^{N_i}$ on G_i , let $\bar{\mathbf{w}}_i = \mathbf{P}_{N_i} \mathbf{W}_i \bar{\mathbf{x}}_i$ denote the graph wavelet domain representation on the same graph, subject to a node relabelling in form of permutation matrix \mathbf{P}_{N_i} ; the respective low-and high-pass values of $\bar{\mathbf{w}}_i$ may be subsequently assigned to suitably coarsened versions of G_i . The recombination $\bar{\mathbf{w}} = \bar{\mathbf{w}}_1 \otimes \bar{\mathbf{w}}_2$ ultimately gives rise to the graph signal $\bar{\mathbf{w}}$ on G . The proposed scheme can be generalized to accommodate iterations on the low-pass branch, by defining the multilevel transform

$$\mathbf{W}_i^{(j)} = \begin{bmatrix} \mathbf{W}_i^j & \\ & \mathbf{I}_{N_i - \frac{N_i}{2^j}} \end{bmatrix} \dots \mathbf{W}_i^0$$

and iterative permutation matrix

$$\mathbf{P}_{N_i}^{(j)} = \mathbf{P}_i^0 \dots \begin{bmatrix} \mathbf{P}_i^j & \\ & \mathbf{I}_{N_i - \frac{N_i}{2^j}} \end{bmatrix}$$

at levels $j \leq J - 1$, for a *multiscale representation*

$$\mathbf{w} = (\mathbf{P}_{N_1}^{(J-1)} \otimes \mathbf{P}_{N_2}^{(J-1)}) (\mathbf{W}_1^{(J-1)} \otimes \mathbf{W}_2^{(J-1)}) \mathbf{x} = \mathbf{P}_{N_1 N_2}^{(J-1)} (\mathbf{W}_1^{(J-1)} \otimes \mathbf{W}_2^{(J-1)}) \mathbf{x},$$

where $(\tilde{\mathbf{W}}_1 \otimes \tilde{\mathbf{W}}_2) = (\mathbf{W}_1^{(j)} \otimes \mathbf{W}_2^{(j)})$ represents the introduced graph product transform at level j . This transform is invertible with inverse $(\tilde{\mathbf{W}}_1^{-1} \otimes \tilde{\mathbf{W}}_2^{-1})$, as per invertibility of its (circulant) sub-wavelet transforms \mathbf{W}_i^j .

Remark 4.2. The application of a 2D discrete tensor-product wavelet transform on an (image) matrix \mathbf{X} can be expressed as $\mathbf{w}_{2D} = \mathbf{W}_1 \mathbf{X} \mathbf{W}_2^T$, whose row-vectorized form is given by $(\mathbf{W}_1 \otimes \mathbf{W}_2)\mathbf{x}$. It becomes evident that this constitutes an analogy to one level of the proposed transform, where in the traditional domain one has $\mathbf{W}_2 = \mathbf{W}_1$, while in the graph domain the \mathbf{W}_i 's generally differ, as they are not defined on the same graph. This elucidates that our derived scheme can be regarded as the equivalent of operating on a graph signal (or vectorized image) with respect to confined direction (rows and columns), which is dictated by the factors in the chosen graph decomposition. One may therefore

¹³Assume for simplicity that the general graph (factor) in each case admits downsampling into two equal-sized vertex sets.

regard the analysis of partitions $\bar{\mathbf{x}}_i$ on G_i , and subsequent re-assignment to the vertices of G_\diamond , as a two-dimensional extension of the graph wavelet analysis of \mathbf{x} on G_\diamond via a suitable transform.

A significant difference in notion between graph product-based and the graph-fused or time-variant graph filters of previous sections is that the former does not accommodate changes in the graph connectivity, since the components (or dimensions) of the graph product are fixed, but is more suitable to analyze signals changing over time. In particular, this can be realized with one graph factor representing the time axis in form of the simple cycle and the other as the underlying graph topology of the signal in question.

4.5.3 Smoothness and Sparsity on Product Graphs

For the remainder of this discussion, following Def. 4.3, the focus is on the analysis of graph signals which admit the rank-1 decomposition $\mathbf{x} = \mathbf{x}_1 \otimes \mathbf{x}_2$ into *smooth* signal (tensor) factors \mathbf{x}_i for a maximally sparse representation.

We begin by investigating how the smoothness of graph signal \mathbf{x} with respect to G_\diamond is related to the smoothness of the subgraph signals \mathbf{x}_i with respect to G_i , as measured by the classical graph Laplacian quadratic form $S_2(\mathbf{x}) = \mathbf{x}^T \mathbf{L} \mathbf{x}$ [1]. Let the individual smoothness coefficients be denoted by $S_\diamond = \mathbf{x}^T \mathbf{L} \mathbf{x}$ for G_\diamond , and $S_i = \mathbf{x}_i^T \mathbf{L}_i \mathbf{x}_i$ for factors G_i , and assuming degree regularity, simplify $\mathbf{D}_i = d_i \mathbf{I}_{N_i}$, which gives rise to the following relations:

- $S_\otimes = d_2 S_1 \|\mathbf{x}_2\|_2^2 + d_1 S_2 \|\mathbf{x}_1\|_2^2 - S_1 S_2$
- $S_\times = S_1 \|\mathbf{x}_2\|_2^2 + S_2 \|\mathbf{x}_1\|_2^2$
- $S_\boxtimes = (1 + d_2) S_1 \|\mathbf{x}_2\|_2^2 + (1 + d_1) S_2 \|\mathbf{x}_1\|_2^2 - S_1 S_2$
- $S_{[\]} = \|\mathbf{x}_1\|_2^2 S_2 + S_1 c_2^2 + d_1 \|\mathbf{x}_1\|_2^2 (N_2 \|\mathbf{x}_2\|_2^2 - c_2^2),$

with constant $c_2 = \sum_{i=0}^{N_2-1} x_2(i)$. The total smoothness S_\diamond is composed of weighted sub-measures S_i , whose individual contribution is scaled by parameters pertaining to the energy of the corresponding subgraph signal and node degree of the opposing factor graph. One deduces that if for a designated decomposition $G_\diamond = G_1 \diamond G_2$ with factors G_i , the measures S_i are small, i.e. the sub-signal factors \mathbf{x}_i are smooth with respect to graph factors G_i , then \mathbf{x} is also relatively smooth on G_\diamond with small S_\diamond (subject to a scaling). Following proper normalization by $\|\mathbf{x}\|_2^2 = \|\mathbf{x}_1\|_2^2 \|\mathbf{x}_2\|_2^2$, it can be inferred that when \mathbf{x} is an eigenvector, the smoothness measure gives rise to an eigenvalue of \mathbf{L}_\diamond . For the symmetric normalized graph Laplacian matrices of non-regular graphs, the above relations continue to hold.

Furthermore, following the interpretation of \mathbf{L}_\diamond as a high-pass filter, consider the signal $\mathbf{L}_\diamond \mathbf{x}$ and analyze its sparsity $\|\mathbf{L}_\diamond \mathbf{x}\|_0$:

- $\mathbf{L}_\otimes \mathbf{x} = (\mathbf{L}_1 \mathbf{x}_1) \otimes d_2 \mathbf{x}_2 + d_1 \mathbf{x}_1 \otimes (\mathbf{L}_2 \mathbf{x}_2) - (\mathbf{L}_1 \mathbf{x}_1) \otimes (\mathbf{L}_2 \mathbf{x}_2)$
- $\mathbf{L}_\times \mathbf{x} = (\mathbf{L}_1 \mathbf{x}_1) \otimes \mathbf{x}_2 + \mathbf{x}_1 \otimes (\mathbf{L}_2 \mathbf{x}_2)$
- $\mathbf{L}_{\boxtimes} \mathbf{x} = (\mathbf{L}_1 \mathbf{x}_1) \otimes d_2 \mathbf{x}_2 + d_1 \mathbf{x}_1 \otimes (\mathbf{L}_2 \mathbf{x}_2) - (\mathbf{L}_1 \mathbf{x}_1) \otimes (\mathbf{L}_2 \mathbf{x}_2) + (\mathbf{L}_1 \mathbf{x}_1) \otimes \mathbf{x}_2 + \mathbf{x}_1 \otimes (\mathbf{L}_2 \mathbf{x}_2)$
- $\mathbf{L}_{[\]} \mathbf{x} = \mathbf{x}_1 \otimes (\mathbf{L}_2 \mathbf{x}_2) + (\mathbf{L}_1 \mathbf{x}_1) \otimes c_2 \mathbf{1}_{N_2} + d_1 \mathbf{x}_1 \otimes (N_2 \mathbf{x}_2 - c_2 \mathbf{1}_{N_2})$

It becomes evident that for constant \mathbf{x}_i (and hence \mathbf{x}) such that $\mathbf{L}_i \mathbf{x}_i = \mathbf{0}_{N_i}$, we have $\mathbf{L} \mathbf{x} = \mathbf{0}$, naturally preserving the nullspace. For linear polynomial \mathbf{x}_i and banded circulant graph factors G_i , signals $\mathbf{L}_i \mathbf{x}_i$ are sparse, which is not necessarily true for $\mathbf{L}_\diamond \mathbf{x}$ under any product operation.

Replacing \mathbf{D}_i by e-degree matrices $\tilde{\mathbf{D}}_{i,\alpha_k}$ for exponential parameter α_k (as defined in Ch. 3, Def. 3.3 on p. 54), gives rise to equivalent relations for parameterized circulant $\tilde{\mathbf{L}}_{i,\alpha_k}$; the same principle can be extended to generalized graph Laplacians of the form $\mathbf{L}_{\diamond,\lambda_j} = \lambda_j(\mathbf{A}_\diamond) - \mathbf{A}_\diamond$. Hence, for periodic complex exponential graph signals \mathbf{x}_i parameterised by $\alpha_k = \frac{2\pi k}{N_i}$, $k \in [0, N_i - 1]$ (or accordingly, bandlimited signals), we obtain $\mathbf{L}_{\alpha_1 \diamond \alpha_2} \mathbf{x} = \mathbf{0}$, with exception of the lexicographic product, for which this holds only if \mathbf{x}_2 is an all-constant vector, as evidenced by its eigenspace property.

Overall, this demonstrates that the vanishing moment properties of circulant (e-)graph Laplacians are to an extent preserved within the graph product, yet sparsity is reduced as a result of the newly arising interconnections between the factors and, hence, growing (graph) border effect. Due to the fact that the above relations cannot be generalized to powers of the graph Laplacian matrix, comparable property preservations can only conditionally be extended to higher order (exponential) polynomial graph signals, which have a sparse representation with respect to factors \mathbf{L}_i^k . This suggests that a sparser representation can be gained by performing a separable signal (wavelet) analysis with respect to inherent circulant substructures, as measured via the graph Laplacian and its powers. In contrast, the annihilation property of generalized or e-graph Laplacians for bandlimited signals continues to be preserved at higher powers, as the eigenbasis does not change.

In light of this, we at last compare the sparsity of representation attained via the proposed graph wavelet transforms on the basis of an example, and discover that the separable approach, apart from preserving higher-order annihilation properties, can induce more sparsity, for certain cases such as non-periodic smooth signals on circulants. It should be noted that two levels of the 1-D non-separable graph wavelet transform (or alternatively, downsampling on both factors, with respect to $\mathbf{K} = \mathbf{I}_{N_1} \otimes \mathbf{K}_2$ followed by $\mathbf{K} = \mathbf{K}_1 \otimes \mathbf{I}_{N_2/2}$), are considered comparable to one level of the 2-D separable transform, yet similarly as in the traditional domain, there is no direct equivalence between the two.

Example 1: Given graph signal $\mathbf{x} = \mathbf{x}_1 \otimes \mathbf{x}_2$ on $G = G_1 \otimes G_2$, where G_i are circulant and banded of bandwidth M_i and $\mathbf{x}_i \in \mathbb{R}^{N_i}$ are linear polynomial, let $\mathbf{W}_i \in \mathbb{R}^{N_i \times N_i}$ and $\mathbf{W}_\otimes \in \mathbb{R}^{N_1 N_2 \times N_1 N_2}$ represent first-order graph-spline wavelet transforms on factors G_i and G respectively. Here, we downsample w. r. t. $s = 1 \in S_i$ on each G_i and reconnect nodes such that generating sets S_i are preserved. Hence, separable representation $\mathbf{w} = \mathbf{w}_1 \otimes \mathbf{w}_2$ has $K = \frac{3}{4}N_1 N_2 - \frac{1}{2}(2M_1 M_2 + M_1 N_2 + M_2 N_1)$ zero entries, whereas non-separable $\mathbf{w}_\otimes = \mathbf{W}_\otimes \mathbf{x}$ has a total of $K_\otimes = K_{\otimes,1} + K_{\otimes,2}$ zeros, with $K_{\otimes,1} = \frac{1}{2}N_1 N_2 - (M_1 N_2 + M_2 N_1 - 2M_2 M_1)$ and $K_{\otimes,2} = \frac{1}{4}N_1 N_2 - (\frac{3}{2}N_1 M_2 + M_1 N_2 - 6M_1 M_2)$ zeros at levels 1 and 2 respectively. We have $K > K_\otimes$ for $4M_i < N_i$, and $K > K_\otimes = K_{\otimes,1}$ with $K_{\otimes,2} = 0$ for $N_i/4 \leq M_i < N_i/2$, which implies that the separable approach induces a sparser representation at any bandwidth $M_i < N_i/2$. Note that the sparsity of \mathbf{w}_\diamond is the same under any of the first three graph products.

Example 2: For circulant lexicographic product graph $G = C_{N_1 N_2, S}$ and decomposition $G = G_1[G_2]$, it can be deduced from the above relations, that we can gain sparsity by conducting the 2-D graph wavelet analysis of \mathbf{x}_i on G_i as opposed to the 1-D analysis of \mathbf{x} on G , for any choice of smooth \mathbf{x}_i as long as they do not lie in the eigenspace of the graph Laplacian. Here, the product-related relabelling $\mathbf{P}\mathbf{x} = \mathbf{x}_2 \otimes \mathbf{x}_1$, which renders a circulant matrix $\tilde{\mathbf{A}}_{[\]} = \mathbf{P}\mathbf{A}_{[\]}\mathbf{P}^T$, corresponds to a simple stacking of columns instead of rows, thus preserving the tensor product, and associated smoothness properties with respect to the subgraphs.

4.6 Computational Experiments

In the following, we consider selected transforms in order to illustrate their sparsification properties.

Let a multi-dimensional piecewise smooth signal reside on the toroidal graph, which results from the Cartesian graph product of two circulants of respective dimension $|V_i| = 32$ with generating sets $S_1 = \{1\}$ and $S_2 = \{1, 2\}$, also illustrated in Fig. 4.4 of Sect. 4.5. We compare the non-linear approximation performance between the suitably parameterized space-variant (*HSVGSWT*) and regular graph e-spline wavelet transform (*HGESWT*) with initial parameters $\beta_1 = \tilde{d}_{max} = \frac{2\cos(\alpha_1) + 2\cos(\alpha_1) + 2\cos(2\alpha_1)}{6}$, $\beta_2 = \tilde{d}_{min} = \frac{2\cos(\alpha_2) + 2\cos(\alpha_2) + 2\cos(2\alpha_2)}{6}$ and order $k = 2$ at 3 levels as well as with the simple graph spline wavelet transform (*HGSWT*). Further, reconnection is conducted such that generating sets are preserved separately on each graph factor and all transform factors are set to 1. The condition numbers of the constructions are respectively $C_{HSVGSWT} = 655.04$, $C_{HGESWT} = 719.538 \times 10^3$ and $C_{HGSWT} = 399.61$, and the signal is given by $\mathbf{x} = \sum_{i=1}^2 \mathbf{x}^i \circ \mathbf{1}_{[t_i \ t_i + N/2 - 1]}$, $t_1 = 0$, $t_2 = N/2$, with pieces $\mathbf{x}^i = (\cos(\alpha_i \mathbf{k}) \circ \mathbf{p}) \otimes \cos(\alpha_i \mathbf{k})$, for linear polynomial $\mathbf{p} \in \mathbb{R}^{N_i}$,

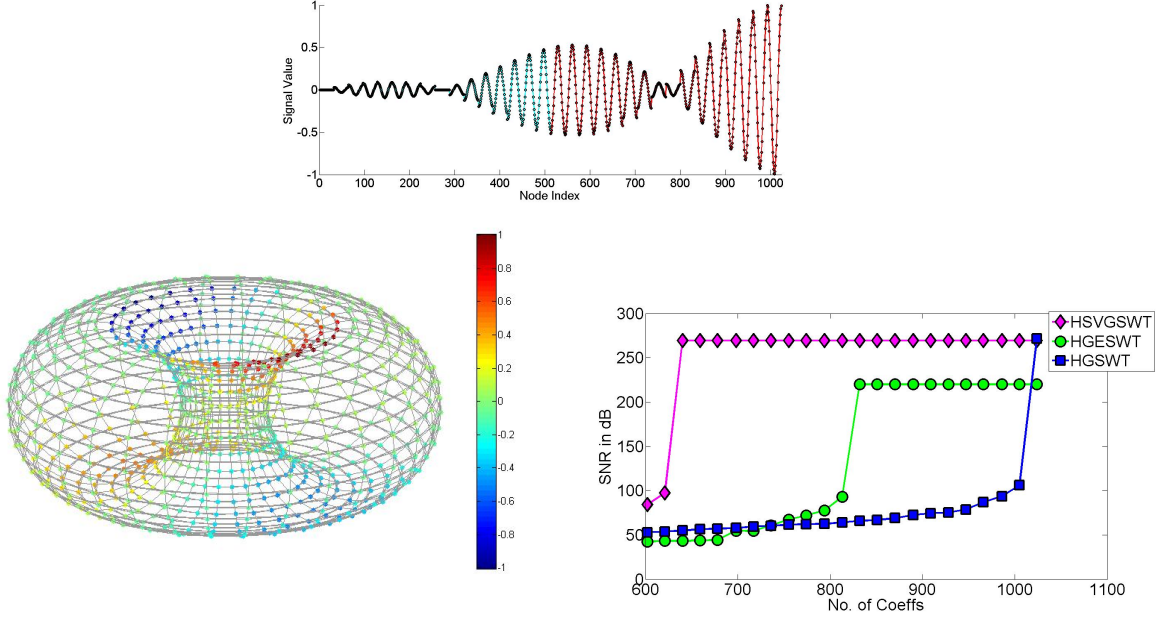


Figure 4.6: Non-linear Approximation Performance Comparison for a piecewise smooth signal (top) on a toroidal graph product (left).

sequence $\mathbf{k} = 0 : N_i - 1$, and with $\alpha_1 = \frac{2\pi}{32}$ and $\alpha_2 = 0.2$ respectively per piece. It becomes evident that the node-variant transform achieves perfect reconstruction at a significantly lower number of retained coefficients than the multiplicative graph e-spline transform (see Fig. 4.6).¹⁴ This can be traced back to a smaller matrix bandwidth, and hence smaller border effect, of the node-variant transform compared to the graph e-spline transform, which includes a larger number of matrix factors to achieve the same annihilation effect. The latter is further outperformed within the region prior to perfect reconstruction by the simple graph spline construction, which albeit not having the required annihilation properties, has a lower condition number, in line with the previous derivations of Sect. 4.4.

Further, we compare the performance of the previous set of transforms (for order $k = 1$) under noise for a piecewise smooth signal on a circulant graph of dimension $|V| = 1024$ with generating set $S = \{1, 2, \dots, M\}$ of varying bandwidth $M = 1, 5, 9$. Given the noisy signal $\mathbf{y} = \mathbf{x} + \mathbf{n}$, for random Gaussian noise vector \mathbf{n} , we recover \mathbf{x} via the hard-thresholding of a fraction of graph wavelets coefficients \mathbf{w} , for a chosen value close to the sparsifying threshold of the annihilating transforms in the clean state, and subsequent application of the inverse GWT, similar to the analysis in Sect. 3.6. Here, 3 levels are considered, with reconnection as before, and standard transform factors omitted, while results are averaged over 10 trials of noise and plotted against the standard deviation of the noise. In Fig. 4.7, a piecewise sinusoidal signal, with frequencies $\alpha_1 = \frac{2\pi}{N}$ and $\alpha_2 = \frac{2\pi 5}{N}$ respectively per

¹⁴The graph in Fig. 4.6 (lower left) was prepared using [52].

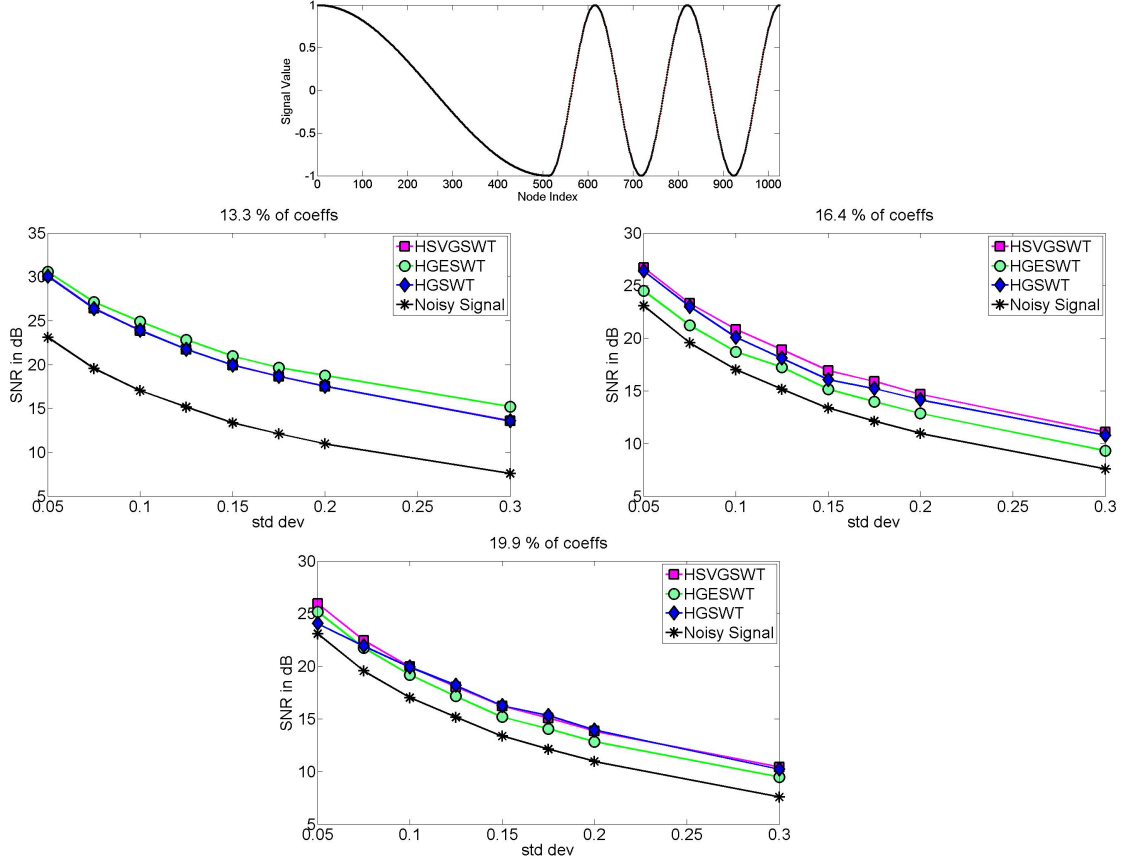


Figure 4.7: Non-linear Approximation Performance Comparison for a noisy piecewise smooth signal (top) on a circulant graph with bandwidth $M = 1, 5, 9$ (from left).

piece, is recovered on the vertices of a circulant graph. Performance generally decreases with increasing bandwidth, in line with the observations for universally smooth signals from Sect. 3.6, however, as is apparent, all three considered (first-order) graph wavelet transforms provide an improvement with respect to the noisy version.

In the case when the first piece of the signal is a polynomial sinusoidal with the same frequency $\alpha_1 = \frac{2\pi}{N}$, while the latter is a sinusoidal with $\alpha_2 = 0.05$, an additional graph border effect is incurred following transform analysis, and we observe in Fig. 4.8, that the matrix-product-based *HGESWT* construction diminishes rapidly in performance as a result of its increased condition number and sensitivity to noise. Here, we have compared the space-variant and simple spline GWTs of order $k = 2$ with the *HGESWT* comprising filter factors of order $k_1 = 2$ and $k_2 = 1$, respectively corresponding to the band-pass filters for α_1 and α_2 .

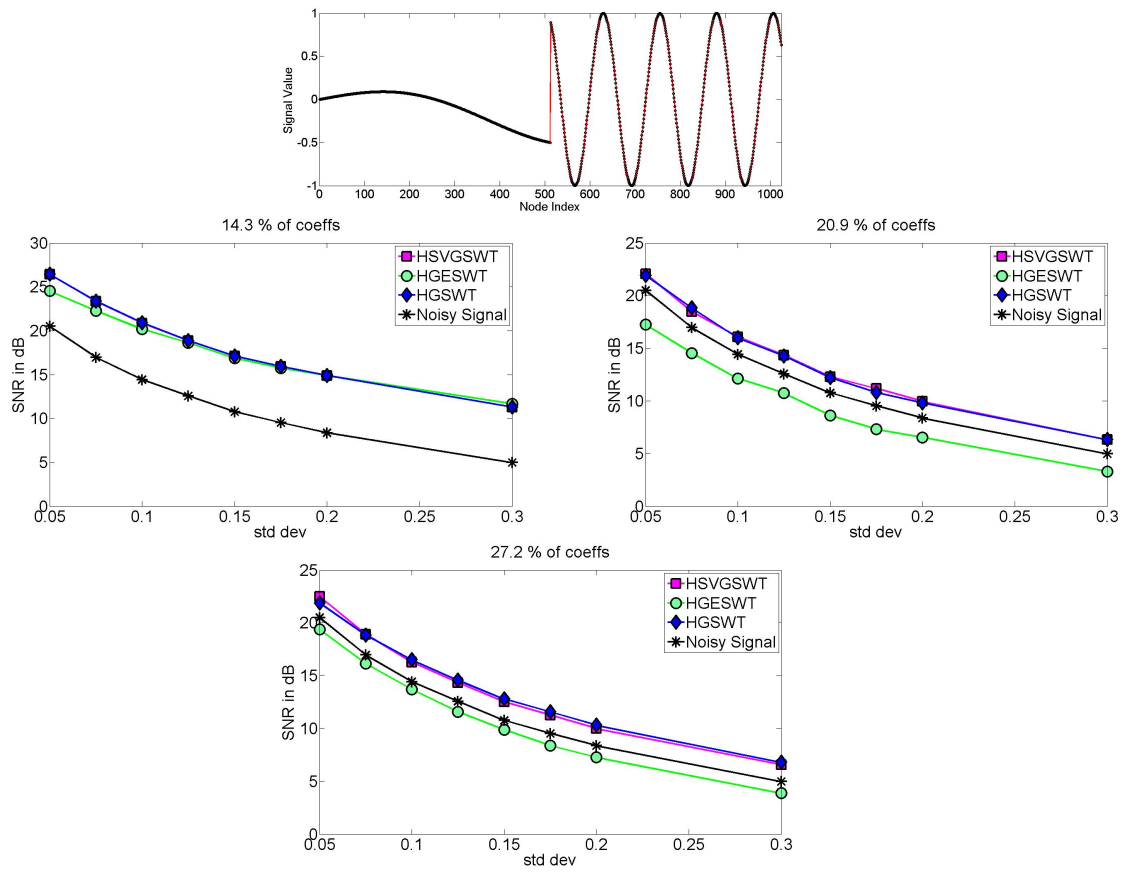


Figure 4.8: Non-linear Approximation Performance Comparison for a noisy piecewise smooth signal (top) on a circulant graph with bandwidth $M = 1, 5, 9$ (from left).

4.7 Overview of GWTs

The following table summarizes the various proposed GWTs from Chapters 3 and 4, along with the corresponding classes of graph signals that can be annihilated (or reproduced, where applicable) on the specified undirected graph type.

We list the analysis filter type as a function of the normalized graph (adjacency) matrix eigenvalues γ , the order of annihilation as either the number k of vanishing moments (local) or the dimension of the nullspace of the high-pass filter $Null(\cdot)$ (or $Null_P(\cdot)$ for piecewise bandlimited signals), which is variable, and the parameterization in form of the e-degree d_α or generalized eigenvalue γ_i .

Name	Filters	Order	Parameters	Graph Type	Signal Class
<i>HGSWT</i>	$(1 \pm \gamma)^k$	$2k$	$d_0 = 1$	Circulant	Polynomial: \mathbf{p}
<i>HGESWT</i>	$\prod_{n=1}^T (\beta_n \pm \gamma)^k$	$2Tk/Null(\cdot)$ $2Tk$	$\beta_n = d_{\alpha_n} = \sum_{k=1}^M 2d_k \cos(\alpha_n k)$ $\beta_n = d_{i\alpha_n} = \sum_{k=1}^M 2d_k \cosh(\alpha_n k)$		Complex Exp. Pol. $e^{\pm i k \alpha_n} \mathbf{p}$ Real Exp. Pol. $e^{\pm k \alpha_n} \mathbf{p}$
<i>HCGESWT</i>	$\prod_{n=1}^T (\beta_n - \gamma)^k$	$2Tk/Null(\cdot)$	$\beta_n = d_{\alpha_n} = \sum_{k=1}^M 2d_k \cos(\alpha_n k)$		Complex Exp. Pol. $e^{\pm i k \alpha_n} \mathbf{p}$
<i>HSVGSWT</i>	$(1 \pm \tilde{\gamma})^k$	$2k/Null_P(\cdot)$	d_i		Piecewise Smooth
<i>HBGWT</i>	$\prod_n (\beta_n \pm \gamma)^k$	$Null(\cdot)$	$\beta_n = \gamma_i$	Arbitrary	(Piecewise) Bandlimited
<i>HSVGSWT</i>	$(1 \pm \tilde{\gamma})^k$	$Null_P(\cdot)$	d_i		Piecewise Bandlimited

Table 4.1: Overview of Proposed GWTs.

We further make the following remarks in relation to Table 4.1:

- The term *piecewise* here refers to pieces of different function type; note that all circulant constructions annihilate piecewise smooth signals of the same function type by default.
- Further $\tilde{\gamma}$ denotes the spectrum of $\tilde{\mathbf{A}} = \tilde{\mathbf{D}}^{-1/2} \mathbf{A}_n \tilde{\mathbf{D}}^{-1/2}$ with parameters \tilde{d}_i and $\mathbf{k} = 0 : N - 1$.
- The complementary construction *HCGESWT* is an alternative to the *HGESWT* for complex exponential polynomials with more reproduction properties, within a variable low-pass filter, and well-defined synthesis filters.
- The proposed time-varying and graph-product-based transforms represent generalizations of the above respectively to sets of multiple graphs and higher-dimensional graphs.

Chapter 5

Sparse Sampling on Graphs

With the focus of previous chapters laid on the problem of inducing sparsity in graph signal representations by identifying suitable graph wavelet bases, the present discussion is intended to further the exploration of *sparsity on graphs* by considering its wide-ranging implications. Given a sparse signal residing on the vertices of an arbitrary graph, it is desirable to exploit the sparsity property for sampling or dimensionality reduction, as conducted in the classical frame of signal processing or compressed sensing.

In particular, by leveraging the core theory of Finite Rate of Innovation (FRI) sampling in the classical domain, paired with insights from classical and, by extension, the derived (circulant) graph spline wavelet theory, we introduce a *graph signal sampling* and *graph coarsening* framework, termed the Graph-FRI (GFRI) framework for sparse graph signals.¹ At its core, the introduced GFRI-framework states that any K -sparse signal residing on the vertices of a circulant graph can be sampled and perfectly reconstructed from its dimensionality-reduced representation in the graph spectral domain of minimum size $2K$, while the structure of an associated coarsened graph can be simultaneously inferred. Equipped with a broad range of sparsifying graph wavelet transforms, the developed framework can thus be naturally extended to signals that are (piecewise) smooth, while the context of noisy recovery under (graph-)perturbations is also explored. Extensions to arbitrary graphs can be further enforced via suitable approximation schemes.

¹A significant part of the content of this chapter appears in accepted article [6], and further led to publications [8] and in part [11].

5.1 Sampling on Circulant Graphs

5.1.1 Related Work

The spectral domain-based coarsening approach in the developed GFRI framework is comparable to the graph coarsening scheme introduced in [68] by Chen et al. up to the choice of the sampling set and resulting property preservations, and with the further distinction that we additionally (iteratively) filter the given graph signal with a suitable graph e-spline filter. While the former requires K entries of suitably chosen sample locations (for bandlimited signals of bandwidth K) for perfect recovery, our downsampling pattern is fixed and independent of the reconstruction scheme, which solely requires the input of the dimensionality-reduced spectral graph signal, as well as primarily used to extract the coarsened graph corresponding to the sampled graph signal, under preservation of certain graph properties. In addition, the present scheme considers sparse and graph wavelet-sparse, as opposed to bandlimited, graph signals, encompassing a wider variety of graph signal classes, which do not necessarily belong to a fixed subspace, as facilitated through suitable graph (e-)spline wavelet analysis. It has already been established by Ekambaram et al. [55] that a K -bandlimited graph signal \mathbf{x} on a circulant graph, whose GFT is non-zero at known locations, can be exactly recovered from K consecutive entries of \mathbf{x} via inversion of its GFT-matrix, following that any $K \times K$ -submatrix of either K consecutive rows or columns of the DFT-matrix is invertible.

5.1.2 Sampling in the Time Domain

The process of sampling a continuous-time signal $x(t)$ in the Euclidean domain traditionally comprises a filtering (convolution) operation with a given filter function $h(t)$, followed by a (uniform) sampling step, which creates the samples $y_n = (x * h)(t)|_{t=nT}$ at sampling rate $f_s = \frac{1}{T}$ for period T ([22], see Fig. 5.1). At its core, sampling theory creates a bridge between continuous-time and discrete-time signals by seeking to identify ideal methods as well as conditions for the perfect reconstruction of $x(t)$ from the given y_n ; this further extends to characterizing distinct classes of signals $x(t)$ and suitable filters $h(t)$ which guarantee perfect recovery. In a broader sense, sampling in discrete-time may be



Figure 5.1: Traditional Sampling Scheme. ©2017 Elsevier Inc.

interpreted as a dimensionality reduction process, which is followed by a dimensionality increase (or interpolation) step to recover the original signal [81]. Nevertheless, in order

to formulate a proper sampling theory in the graph domain, additional questions need to be addressed, in particular, on what kind of graph structure the dimensionality-reduced (sampled) signal resides and how the former relates to the original graph, which challenges the classical problem and invites a more sophisticated take on sampling. Sampling theory on graphs can therefore be collectively described as the study of conditions and methods which facilitate the perfect recovery of a graph signal $\mathbf{x} \in \mathbb{R}^N$ on the vertices of a graph $G = (V, E)$ with $|V| = N$, from a dimensionality-reduced, possibly graph-filtered, signal $\mathbf{y} \in \mathbb{R}^M$, further extending to the identification of a coarsened graph \tilde{G} , with $|\tilde{V}| = M$ and $M < N$, on which \mathbf{y} is defined, as well as to providing a generalized frame for the classical problem. Further intriguing dimensions of the framework may involve the accompanying recovery of G from \tilde{G} , however, in this work we restrict our focus on the former.

5.1.3 Wavelets for Sampling

Equipped with a range of novel families of sparsifying circulant (and non-circulant) graph wavelet transforms, the focus now shifts to the resulting sparse graph signals and, as such, the following conventions are applied.

Let $\mathbf{x}_W \in \mathbb{C}^N$ denote a signal defined on a (circulant) graph G and $\mathbf{W}_j \in \mathbb{R}^{N/2^j \times N/2^j}$ represent a general graph wavelet transform (GWT) of the form

$$\mathbf{W}_j = \begin{bmatrix} \Psi_{j\downarrow 2} \mathbf{H}_{LP_j} \\ \Phi_{j\downarrow 2} \mathbf{H}_{HP_j} \end{bmatrix},$$

composed of low-and high-pass filters $\mathbf{H}_{LP_j}, \mathbf{H}_{HP_j} \in \mathbb{R}^{N/2^j \times N/2^j}$, at level j , where the binary downsampling matrices $\Psi_{j\downarrow 2}, \Phi_{j\downarrow 2} \in \mathbb{R}^{N/2^{j+1} \times N/2^j}$ sample complementary sets of nodes in the standard alternating pattern with respect to $s = 1 \in S$ for circulants (or as appropriate); here, even-numbered nodes are retained in the low-pass branch and subsequently redefined on a suitably coarsened graph. The multiresolution representation of \mathbf{x}_W , following iteration in the low-pass branch, then yields

$$\tilde{\mathbf{x}} = \mathbf{W} \mathbf{x}_W = \begin{bmatrix} \mathbf{W}_j & \\ & \mathbf{I}_{\frac{N(2^j-1)}{2^j}} \end{bmatrix} \cdots \begin{bmatrix} \mathbf{W}_1 & \\ & \mathbf{I}_{\frac{N}{2}} \end{bmatrix} \mathbf{W}_0 \mathbf{x}_W,$$

for multilevel graph wavelet transform matrix \mathbf{W} . In order to re-label $\tilde{\mathbf{x}}$, whose individual partitions reside on a collective of coarsened graphs, with respect to the original G , we introduce the permutation matrix \mathbf{P} , so that, for an appropriate re-assignment, $\mathbf{x} = \mathbf{P} \tilde{\mathbf{x}}$ resides on G .

Prior analysis is thus captured in the definition of a class W of *wavelet- K -sparse* graph signals, with $\mathbf{x}_W \in W$ of dimension N , whose elements possess a K -sparse multiresolution

representation $\mathbf{x} \in \mathbb{C}^N$, $\|\mathbf{x}\|_0 = K$, via a suitable GWT. Given a smooth graph signal \mathbf{x}_W , one can describe, and hence tailor, the sparsity K of \mathbf{x} as a function of decomposition level j as well as of the bandwidth B_j of the graph filter matrix at each j , where applicable. In particular, this analysis model is primarily relevant for (piecewise) smooth signals on circulant graphs due to their breadth of annihilation capabilities, ranging from localized to global (bandlimited) form on multiple scales, whereas annihilation on arbitrary graphs has so far been confined to signals of (piecewise) bandlimited form and is not extendable to multiple levels in general due to the lack of structure and consistency in graph coarsening. For illustration, results assessing the number of non-zero entries of $\mathbf{x} \in \mathbb{C}^N$ are presented, for polynomial \mathbf{x}_W (generalizable to complex exponential polynomials) on circulant graphs. Node reconnection is conducted such that the coarse graphs retain their original generating set after downsampling at each level, thereby maintaining constant filter support; generalizations to other, less regular, graph coarsening schemes such as Kron-reduction, are omitted for simplicity.

Corollary 5.1. *Consider an undirected, circulant graph G of dimension N and bandwidth \tilde{B} , and let \mathbf{x} be the multiresolution decomposition of graph signal \mathbf{x}_W , which is a 1-piece polynomial of maximum degree $D \leq 2k - 1$, on G via the j -level GWT matrix \mathbf{W} .*

(i) *Let \mathbf{W} be the HGSWT of order $2k$, with corresponding low-and high-pass graph filter matrices each of bandwidth $B = k\tilde{B}$, and assume that B is sufficiently small such that $\sum_{n=0}^l \frac{B}{2^n} \leq \frac{N}{2^{l+1}}$ at each level $l \leq j - 1$. The resulting $\mathbf{x} = \mathbf{P}\mathbf{W}\mathbf{x}_W$ is K -sparse, where $K = \frac{N}{2^j} + B(2(j-1) + 2^{1-j})$, when $B = 2^{j-1}r$, $r \in \mathbb{Z}^+$.*

(ii) *Let \mathbf{W} be the HCGSWT of order $2k$, with corresponding low-and high-pass graph filter matrices of bandwidth T and $B = k\tilde{B}$ respectively, such that $B + \sum_{n=1}^l \frac{T}{2^n} \leq \frac{N}{2^{l+1}}$ at each level $l \leq j - 1$. The resulting \mathbf{x} is K -sparse, where $K = \frac{N}{2^j} + Bj + T(j + 2^{1-j} - 2)$, when $T = 2^{j-1}r$, $r \in \mathbb{Z}^+$.*

(iii) *Let \mathbf{W} be the HGSWT at $j = 0$, with the alternative ‘minimum’ downsampling pattern, which retains only one low-pass component. Then \mathbf{x} is K -sparse with $K = 2B$.*

Proof. See Appendix C.1.

When $B \in \mathbb{Z}^+$, the results of (i) & (ii) in Cor. 5.1 continue to apply up to a small correction term, which increases with the number of levels j . For the multiresolution decomposition of periodic (complex exponential) graph signals with parameter $\alpha = \frac{2\pi k}{N}$, $k \in \mathbb{N}$, the maximum sparsity of $K = \frac{N}{2^j}$ can be achieved after j levels, or selectively, up to $K = 1$ for the ‘minimum’ downsampling pattern in (iii) under a suitable GWT.

Further to Cor. 5.1 (iii), since invertibility of the generalized GWT on an arbitrary graph depends on i.a. the downsampling pattern, one can identify a set of distinctly parameterized transforms on circulant and non-circulant graphs, for which this is satisfied when only one (or few) nodes are low-pass filtered; the circulant HGSWT being a prominent example. Here, the downsampling operation is less meaningful from a graph and sampling

theoretical (or filterbank) perspective but rather leveraged as an intermediate processing step to produce a sparse signal. Therefore, one may apply any type of sparsifying graph transform, that does not necessarily capture a graph filterbank, such as a form of invertible generalized (space-variant) graph Laplacian filter (see Cor. 4.2), where applicable.

5.1.4 The Graph FRI-framework

As previously mentioned in the review of Ch. 2, the traditional FRI-framework is built on the central result that certain classes of non-bandlimited signals with a finite rate of innovation can be sampled and perfectly reconstructed using kernels of compact support, which satisfy the Strang-Fix conditions ([116], [22]). In the discrete domain, this prominently entails that a K -sparse signal vector $\mathbf{x} \in \mathbb{R}^N$ with known K can be perfectly reconstructed from $M \geq 2K$ consecutive sample values y_n of the measurement vector $\mathbf{y} = \mathbf{F}\mathbf{x}$, where $\mathbf{F} \in \mathbb{C}^{N \times N}$ is the DFT matrix, of the form

$$y_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{K-1} x_{c_k} e^{-i2\pi c_k n/N} = \sum_{k=0}^{K-1} \alpha_k u_k^n \quad (5.1)$$

with weights x_{c_k} of \mathbf{x} at positions c_k ; here, the locations $u_k = e^{-i2\pi c_k n/N}$ and amplitudes $\alpha_k = x_{c_k}/\sqrt{N}$ are successively recovered using a local reconstruction algorithm known as Prony's method [22]. In other words, the filtering (or acquisition) of a sparse signal with the DFT \mathbf{F} facilitates its exact reconstruction from a dimensionality-reduced version in the Fourier domain.

The insight that the graph frequency-ordered (graph Laplacian) GFT basis of any circulant graph G can be expressed as the DFT-matrix, subject to a graph-dependent permutation σ of columns, motivates an intuitive extension of sparse sampling to the graph-domain, which is termed the Graph FRI-framework (GFRI):

Theorem 5.1. (*Graph-FRI*) Define the permuted GFT basis \mathbf{U} of undirected circulant graph G such that \mathbf{U}^H is the DFT-matrix. We can sample and perfectly reconstruct a (wavelet-) K -sparse graph signal (with multiresolution) $\mathbf{x} \in \mathbb{C}^N$, on the vertices of circulant G using the dimensionality-reduced GFT representation $\mathbf{y} = \mathbf{U}_M^H \mathbf{x}$, $\mathbf{y} \in \mathbb{C}^M$, where \mathbf{U}_M^H are the first M rows of \mathbf{U}^H , as long as $M \geq 2K$.

Proof. See Appendix C.2.

Similarly as in the traditional case, the proof of this theorem follows from the simple application of Prony's method. In reference to the previous sparsity analysis, one therefore requires at least $K < \frac{N}{2}$ for a given K -(wavelet-)sparse graph signal \mathbf{x} , since $M \geq 2K$, for dimension $N = 2^n, n \in \mathbb{Z}^+$ by initial assumption.

While the sampling basis is the same for all circulant graphs, the reduced spectral signal \mathbf{y} carries a variable frequency interpretation that depends on the graph topology and thus permutation σ , encapsulated within frequency coefficients $y(\lambda_i^\sigma)$ for DFT-permuted (partial) graph Laplacian spectrum $\{\lambda_i^\sigma\}_{i=1}^M$; nevertheless, apart from possible inference through \mathbf{x} , the sole representation through \mathbf{y} does not a priori capture (or reveal) the specific topology of the underlying graph, which will be inspected more closely in what follows.

Graph Coarsening for GFRI

The problem of (down-)sampling a signal on a graph $G = (V, E)$ along with graph coarsening, as the task of determining the reduced set of vertices and edges of the coarsened graph $\tilde{G} = (\tilde{V}, \tilde{E})$, are central to GSP theory and represent one of the main challenges that the complex data dependencies of graphs impose on traditional signal processing, as has been remarked in previous chapters on graph wavelet analysis. In an effort to further characterize the spectral graph representation \mathbf{y} in terms of the principal underlying graph G as well as complete the current graph sampling framework, it is thus desirable to identify the coarsened graph \tilde{G} complementing the dimensionality-reduced signal.

Formulated approaches range from spectral graph partitioning, where the largest graph Laplacian eigenvector is used to determine a downsampling pattern [60], up to graph-characteristic operations such as for bipartite graphs ([1], [35]), which naturally comprise a partitioning into two disjoint sets of nodes and can be considered a special case of the former. Graph reconnection may be conducted to satisfy various properties, and is an accompanying problem in itself. In the context of multilevel graph wavelet analysis, the properties of most interest have been the preservation of circularity and sparsity of the GWT representation, and as implied by the foregoing discussion, the latter is achieved when the bandwidth of the graph Laplacian is small, i.e. minimal reconnection is conducted.

Since in the present scenario, one needs to identify the graph structure associated with the dimensionality reduced GFT-representation \mathbf{y} , conversely to prior graph coarsening approaches, both an appropriate downsampling pattern as well as a reconnection strategy need to be extracted from the information given by the spectral coefficients at hand, rather than imposed to fulfill a set of desired properties in the first instance. The main difficulty is posed by the fact that \mathbf{y} resides in the spectral graph domain and does not directly give rise to a downsampling pattern in the vertex domain.

In the traditional FRI-framework [22], a sparse signal can be sampled with a general *exponential reproducing* kernel $\varphi(t)$, beyond the complex exponentials of the Fourier domain,

where the function $\varphi(t)$ in continuous-time and its shifted versions reproduce exponentials for a proper choice of coefficients $c_{m,n}$

$$\sum_{n \in \mathbb{Z}} c_{m,n} \varphi(t - n) = e^{\alpha_m t} \quad \text{for } \alpha_m \in \mathbb{C}, m = 0, \dots, P. \quad (5.2)$$

We note that the coefficients $c_{m,n}$ in Eq. (5.2) can be expressed as $c_{m,n} = c_{m,0} e^{\alpha_m n}$, where $c_{m,0} = \int_{-\infty}^{\infty} e^{\alpha_m x} \tilde{\varphi}(x) dx$ [22], and the functions $\varphi(t)$ and $\tilde{\varphi}(t)$ form a quasi-biorthonormal set, with biorthonormality as a special case ([117], [118]).

Inspired by this notion of sampling a sparse signal within a multi-layered scheme, the Graph FRI-framework is further extended by expressing the reduced GFT-basis \mathbf{U}_M^H as the product between a fat coefficient matrix \mathbf{C} and a row-reduced low-pass GWT filter, which reproduces complex exponential graph signals as per Thms. 3.2 and 3.5.

In the following, we proceed to demonstrate the feasibility of this scheme by first proving the existence of such a coefficient matrix \mathbf{C} and its relation to a row-reduced DFT-matrix.

Lemma 5.1. *Let \mathbf{U}_M^H be the reduced GFT-basis of undirected circulant graph G , as defined in Thm. 5.1, and $\mathbf{E}_{\vec{\alpha}} \in \mathbb{R}^{N \times N}$ a low-pass graph filter matrix in the e-spline GWT family (see Thms. 3.2, 3.5), which can reproduce complex exponential graph signals with parameter $\vec{\alpha} = (\alpha_0, \dots, \alpha_{M-1}) = \left(0, \dots, \frac{2\pi k}{N}, \dots, \frac{2\pi(M-1)}{N}\right)$. We thus have $\mathbf{U}_M^H = \mathbf{C} \Psi_{\downarrow 2} \mathbf{E}_{\vec{\alpha}}$, where $\Psi_{\downarrow 2} \in \mathbb{R}^{N/2 \times N}$ is a binary sampling matrix which retains even-numbered nodes, and $\mathbf{C} \in \mathbb{C}^{M \times N/2}$ is a coefficient matrix. Further, $\mathbf{C} = \hat{\mathbf{C}} \tilde{\mathbf{U}}_M^H$, where $\hat{\mathbf{C}} \in \mathbb{C}^{M \times M}$ is diagonal and $\tilde{\mathbf{U}}^H$ is the DFT matrix of dimension $N/2$.*

Proof. Consider the general complementary graph e-spline wavelet filterbank (Thm. 3.5) with respective analysis and synthesis matrices

$$\mathbf{W} = \begin{bmatrix} \Psi_{\downarrow 2} \mathbf{H}_{LP_{\alpha}} \\ \Phi_{\downarrow 2} \mathbf{H}_{HP_{\alpha}} \end{bmatrix}, \quad \tilde{\mathbf{W}} = \begin{bmatrix} \Psi_{\downarrow 2} \tilde{\mathbf{H}}_{LP_{\alpha}} \\ \Phi_{\downarrow 2} \tilde{\mathbf{H}}_{HP_{\alpha}} \end{bmatrix}$$

such that $\tilde{\mathbf{W}}^T \mathbf{W} = \mathbf{I}_N$, where the high-pass representer polynomials at both branches possess the same number of exponential vanishing moments, i.e. roots at $z = e^{\pm i\alpha}$ for some $\alpha \in \mathbb{R}$. Let $\mathbf{x} \in \mathbb{C}^N$ be a complex exponential graph signal of the form

$$\mathbf{x} = \begin{bmatrix} e^{i\alpha 0} & e^{i\alpha 1} & e^{i\alpha 2} & \dots & e^{i\alpha(N-1)} \end{bmatrix}^T$$

where $\alpha = -\frac{2\pi k}{N}$, i.e. \mathbf{x}^T is the $(k+1)$ -th row of the (unnormalized) DFT-matrix, and define $\mathbf{y} = \mathbf{H}_{LP_{\alpha}} \mathbf{x} = c\mathbf{x}$, for $c \in \mathbb{R}$ (also an eigenvalue of $\mathbf{H}_{LP_{\alpha}}$), such that

$$\Psi_{\downarrow 2} \mathbf{H}_{LP_{\alpha}} \mathbf{x} = c \Psi_{\downarrow 2} \mathbf{x} = c \begin{bmatrix} e^{i\alpha 0} & e^{i\alpha 2} & e^{i\alpha 4} & \dots & e^{i\alpha(N-2)} \end{bmatrix}^T = \mathbf{y}(0:2:N-2) = \mathbf{y}_{\downarrow 2}$$

which denotes a scalar multiple of the $(k+1)$ -th row of the DFT of dimension $N/2$, since

$$\begin{bmatrix} e^{i\alpha 0} & e^{i\alpha 2} & e^{i\alpha 4} & \dots & e^{i\alpha(N-2)} \end{bmatrix}^T = \begin{bmatrix} e^{i(2\alpha)0} & e^{i(2\alpha)1} & e^{i(2\alpha)2} & \dots & e^{i(2\alpha)(N/2-1)} \end{bmatrix}^T$$

with $2\alpha = -\frac{2\pi k}{N/2}$. We obtain

$$\begin{bmatrix} (\Psi_{\downarrow 2} \tilde{\mathbf{H}}_{LP_\alpha})^T & (\Phi_{\downarrow 2} \tilde{\mathbf{H}}_{HP_\alpha})^T \end{bmatrix} \begin{bmatrix} \Psi_{\downarrow 2} \mathbf{H}_{LP_\alpha} \\ \Phi_{\downarrow 2} \mathbf{H}_{HP_\alpha} \end{bmatrix} \mathbf{x} = \begin{bmatrix} (\Psi_{\downarrow 2} \tilde{\mathbf{H}}_{LP_\alpha})^T & (\Phi_{\downarrow 2} \tilde{\mathbf{H}}_{HP_\alpha})^T \end{bmatrix} \begin{bmatrix} \mathbf{y}_{\downarrow 2} \\ \mathbf{0}_{N/2} \end{bmatrix},$$

but since $(\Phi_{\downarrow 2} \tilde{\mathbf{H}}_{HP_\alpha})^T \mathbf{0}_{N/2} = \mathbf{0}_N$, neither $\mathbf{0}_{N/2}$ nor $(\Phi_{\downarrow 2} \tilde{\mathbf{H}}_{HP_\alpha})^T$ contribute, and we can thus write

$$\begin{bmatrix} (\Psi_{\downarrow 2} \tilde{\mathbf{H}}_{LP_\alpha})^T & (\Phi_{\downarrow 2} \tilde{\mathbf{H}}_{HP_\alpha})^T \end{bmatrix} \begin{bmatrix} \mathbf{y}_{\downarrow 2} \\ \mathbf{0}_{N/2} \end{bmatrix} = \begin{bmatrix} (\Psi_{\downarrow 2} \tilde{\mathbf{H}}_{LP_\alpha})^T \end{bmatrix} \begin{bmatrix} \mathbf{y}_{\downarrow 2} \end{bmatrix} = \mathbf{x}$$

i. e. linear combinations of the columns of $(\Psi_{\downarrow 2} \tilde{\mathbf{H}}_{LP_\alpha})^T$ reproduce \mathbf{x} . Rewriting the former, we obtain $\mathbf{y}_{\downarrow 2}^T \Psi_{\downarrow 2} \tilde{\mathbf{H}}_{LP_\alpha} = \mathbf{x}^T$, and reversing the sequence of \mathbf{W} and $\tilde{\mathbf{W}}$, and letting $\mathbf{E}_\alpha = \mathbf{H}_{LP_\alpha}$, we arrive at

$$\begin{bmatrix} \mathbf{c}^T \end{bmatrix} \Psi_{\downarrow 2} \mathbf{E}_\alpha = \mathbf{x}^T$$

with $\mathbf{c} \in \mathbb{C}^{N/2}$ ($\mathbf{c} = \hat{\mathbf{c}} \mathbf{x}_{\downarrow 2}$ for eigenvalue $\hat{\mathbf{c}}$ of $\tilde{\mathbf{H}}_{LP_\alpha}$), $\Psi_{\downarrow 2} \mathbf{E}_\alpha \in \mathbb{R}^{N/2 \times N}$, and $\mathbf{x}^T \in \mathbb{C}^N$. By generalizing the RHS to incorporate M stacked complex exponential vectors \mathbf{x} to form the transposed DFT-matrix $(\mathbf{U}_M^H)^T$, we can similarly show

$$\mathbf{C} \Psi_{\downarrow 2} \mathbf{E}_{\vec{\alpha}} = \mathbf{U}_M^H,$$

with $\mathbf{C} = \hat{\mathbf{C}} \tilde{\mathbf{U}}_M^H$ and $\vec{\alpha} = (\alpha_1, \dots, \alpha_M)$. In particular, the matrix $\hat{\mathbf{C}}$ is diagonal, while $\tilde{\mathbf{U}}_M^H$ represents the first M rows of the DFT-matrix of dimension $N/2$.

In the case of a bipartite graph, using Thm. 3.2, we can proceed similarly, and obtain $\begin{bmatrix} (\Psi_{\downarrow 2} \tilde{\mathbf{H}}_{LP_\alpha})^T \end{bmatrix} \begin{bmatrix} \mathbf{y}_{\downarrow 2} \end{bmatrix} = \mathbf{x}$ with $\mathbf{y}_{\downarrow 2} = c \mathbf{x}_{\downarrow 2}$ for some $c \in \mathbb{R}$, for synthesis low-pass filter $\tilde{\mathbf{H}}_{LP_\alpha}$, which confirms that it reproduces complex exponentials with parameter $\pm\alpha$, just as the analysis low-pass filter, despite not being of the same support (see also the semi-IIR graph filterbank in Sect. 3.5.1). In particular, as previously shown, the inherent biorthogonality constraints of the wavelet transform impose that the representer polynomials of the bipartite *HGESWT* contain opposing roots respectively for analysis and synthesis, i. e. we have $H_{LP_\alpha}(z) = -z^{-1} H_{HP_\alpha}(-z)$ and $\tilde{H}_{LP_\alpha}(z) = -z \tilde{H}_{HP_\alpha}(-z)$. Thus we may interchange the order of synthesis and analysis branch, and obtain $\begin{bmatrix} (\Psi_{\downarrow 2} \mathbf{H}_{LP_\alpha})^T \end{bmatrix} \begin{bmatrix} \hat{\mathbf{c}} \mathbf{x}_{\downarrow 2} \end{bmatrix} = \mathbf{x}$, confirming our previous result that the columns of the adjacency-matrix based, analysis low-pass filter $(\Psi_{\downarrow 2} \mathbf{H}_{LP_\alpha})^T$ reproduce \mathbf{x} as a consequence of the generalized Strang-Fix conditions ([25], [82]). \square

A sensible vertex selection for the coarsened graph can be extracted from the previous

result, in that the factored low-pass graph filter $\Psi_{\downarrow 2} \mathbf{E}_{\vec{\alpha}}$ samples every other node in keeping with the standard circulant downsampling pattern with respect to $s = 1 \in S$.

The popular graph coarsening scheme of Kron-reduction [70], which employs Schur complementation based on the given node sampling pattern, presents a straight-forward realization of reconnection by evaluating the graph-Laplacian matrix $\tilde{\mathbf{L}}$ of the coarsened graph from the graph Laplacian matrix \mathbf{L} of initial graph G and given set $V_{\alpha} = \{0 : 2 : N - 2\}$ of retained nodes: $\tilde{\mathbf{L}} = \mathbf{L}(V_{\alpha}, V_{\alpha}) - \mathbf{L}(V_{\alpha}, V_{\alpha}^c) \mathbf{L}(V_{\alpha}^c, V_{\alpha}^c)^{-1} \mathbf{L}(V_{\alpha}^c, V_{\alpha})^T$. It can be shown that for the given V_{α} and symmetric circulant \mathbf{L} , the scheme preserves these properties for the resulting coarsened graph ([70], [55]), yet faces the drawback of generally inducing denser graphs. In particular, for a banded circulant matrix, the resulting lower-dimensional Schur complement will be of equal or larger bandwidth which may prove destructive in a sparsity-driven filterbank construction and analysis as well as possibly for matrix conditioning.

Alternatively, by leveraging the fact that any graph Laplacian eigenvector \mathbf{u}_k of G can be interpreted as a graph signal on its vertices with sample value $u_k(i)$ at node i , we propose a graph coarsening approach in the spectral domain of the graph at hand, which applies the extracted downsampling pattern on its eigenbasis, and is captured by the following:

Lemma 5.2. *Consider an undirected circulant graph G with generating set S , and adjacency matrix $\mathbf{A} = \frac{1}{N} \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H \in \mathbb{R}^{N \times N}$ with bandwidth B , where $\frac{1}{\sqrt{N}} \mathbf{U}^H$ is the DFT matrix. We downsample by 2 via the binary matrix $\Psi_{\downarrow 2} \in \mathbb{R}^{N/2 \times N}$ on the first $N/2$ rows in \mathbf{U}^H and eigenvalues $\mathbf{\Lambda}$, such that $\tilde{\mathbf{U}}^H = \mathbf{U}_{0:N/2-1}^H \Psi_{\downarrow 2}^T$, and $\tilde{\mathbf{\Lambda}} = \Psi_{\downarrow 2} \mathbf{\Lambda} \Psi_{\downarrow 2}^T$. The resulting adjacency matrix $\tilde{\mathbf{A}} = \frac{1}{N/2} \tilde{\mathbf{U}} \tilde{\mathbf{\Lambda}} \tilde{\mathbf{U}}^H \in \mathbb{R}^{N/2 \times N/2}$ is circulant with the same generating set S as G , provided $2B < N/2$.*

Proof. The eigenvalues of \mathbf{A} with first row $[0 \ a_1 \dots a_1]$ are $\lambda_j = \sum_{k=1}^B 2a_k \cos\left(\frac{2\pi jk}{N}\right)$, $j = 0, \dots, N-1$. Thus the eigenvalues of $\tilde{\mathbf{A}}$ with the same entries a_i and bandwidth $B < N/4$, are $\tilde{\lambda}_j = \sum_{k=1}^B 2a_k \cos\left(\frac{2\pi(2j)k}{N}\right) = \lambda_{2j}$, $j = 0, \dots, N/2-1$. We can similarly show the preservation of the downsampled DFT-eigenbasis. Let

$$\mathbf{x} = \begin{bmatrix} e^{i\alpha 0} & e^{i\alpha 1} & e^{i\alpha 2} & \dots & e^{i\alpha(N-1)} \end{bmatrix}^T$$

with $\alpha = -\frac{2\pi k}{N}$ denote the $(k+1)$ -th row of the non-normalized DFT-matrix. If we discard all entries at odd-numbered positions, we obtain the $(k+1)$ -th row of the DFT of dimension $N/2$, since

$$\begin{bmatrix} e^{i\alpha 0} & e^{i\alpha 2} & e^{i\alpha 4} & \dots & e^{i\alpha(N-2)} \end{bmatrix}^T = \begin{bmatrix} e^{i(2\alpha)0} & e^{i(2\alpha)1} & e^{i(2\alpha)2} & \dots & e^{i(2\alpha)(N/2-1)} \end{bmatrix}^T$$

with $2\alpha = -\frac{2\pi k}{N/2}$. Thus, if we apply the above sampling pattern on the first $N/2$ rows of the DFT of dimension N , we obtain the DFT of dimension $N/2$. In particular, at $k = N/2$,

we have $\alpha = \pi$ and thus corresponding, downsampled row $\mathbf{x} = \mathbf{1}_{N/2}$, and proceeding similarly, we observe that the sampled lower half of the $N \times N$ DFT, equivalently gives the DFT of dimension $N/2$. \square

The above is further reinforced by the fact that adjacency and graph Laplacian matrices of regular graphs possess the same basis, which as will be seen in a later discussion in Sect. 5.1.5, is not upheld for e.g. a path graph. In particular, Lemma 5.2 gives rise to a meaningful coarsening strategy for circulant graphs as it preserves both the original connectivity of the graph by retaining the same generating set, as well as its spectral information given that its eigenvalues and eigenbasis are respectively composed of a subset and subpartition of the original.

Following a generalization of Lemma 5.1, and the preceding discussion, the graph coarsening scheme complementing the proposed Graph-FRI framework is formulated as follows:

Theorem 5.2. *Given GFT $\mathbf{y} \in \mathbb{C}^M$, from Thm 5.1, we determine the coarsened graph $\tilde{G} = (\tilde{V}, \tilde{E})$ associated with the dimensionality-reduced graph signal $\tilde{\mathbf{y}} \in \mathbb{C}^{\tilde{M}}$, via the decomposition*

$$\mathbf{y} = \mathbf{U}_M^H \mathbf{x} = \mathbf{C} \prod_{j=0}^{J-1} (\Psi_{j\downarrow 2} \mathbf{E}_{2^j \tilde{\alpha}}) \mathbf{x} = \mathbf{C} \tilde{\mathbf{y}}$$

where $\mathbf{U}_M^H \in \mathbb{C}^{M \times N}$ is the row-reduced permuted GFT basis (DFT-matrix), $\mathbf{C} \in \mathbb{C}^{M \times \tilde{M}}$ is a coefficient matrix with $\tilde{M} = \frac{N}{2^J}$ given M , $\Psi_{j\downarrow 2} \in \mathbb{R}^{N/2^{j+1} \times N/2^j}$ is a binary sampling matrix which retains even-numbered nodes, and $\mathbf{E}_{2^j \tilde{\alpha}} \in \mathbb{R}^{N/2^j \times N/2^j}$ is a (higher-order) graph e-spline low-pass filter on \tilde{G}_j , which reproduces complex exponentials at level j , with parameter $\tilde{\alpha} = (\alpha_0, \dots, \alpha_{M-1}) = \left(0, \dots, \frac{2\pi(M-1)}{N}\right)$. The associated coarsened graphs \tilde{G}_j at levels $j \leq J$ can be determined following two different schemes:

- (i) Perform Kron-reduction at each level $j \leq J$ using the pattern \mathbf{V}_α in $\Psi_{j\downarrow 2}$ to obtain \mathbf{L}_j
- (ii) Define eigenbasis $(\tilde{\mathbf{U}}_j, \tilde{\mathbf{\Lambda}}_j) \in \mathbb{C}^{N/2^j \times N/2^j}$ at each level $j \leq J$ through the application of $\Psi_{j-1\downarrow 2}$ on $(\tilde{\mathbf{U}}_{j-1}, \tilde{\mathbf{\Lambda}}_{j-1})$ (see Lemma 5.2). The coarse graph \tilde{G}_j for graph signal $\tilde{\mathbf{y}}_j = \prod_{k=0}^{j-1} (\Psi_{k\downarrow 2} \mathbf{E}_{2^k \tilde{\alpha}}) \mathbf{x}$, has adjacency matrix

$$\mathbf{A}_j = (2^j/N) \tilde{\mathbf{U}}_j \tilde{\mathbf{\Lambda}}_j \tilde{\mathbf{U}}_j^H$$

which preserves the generating set S of G for a sufficiently small bandwidth.

Consequentially, the edge set of the coarsened graph associated with the GFRI-framework is not unique, and two possible approaches have been explored which satisfy the main connectivity constraints of symmetry and circularity. Kron-reduction on the one hand preserves basic graph characteristics, yet, while using the entire graph adjacency relations in the computation of the coarsened version, it provides little general intuition on the

topology of the latter. The spectral reduction technique on the other hand, is shown to additionally preserve the original graph connectivity by retaining its generating set (see Fig. 5.2), thereby simultaneously alleviating the issue of an increasing bandwidth. It is therefore also relevant as a coarsening scheme for (sparse) multilevel graph wavelet analysis.

The graph sampling framework, further illustrated in Fig. 5.3, may conclusively be summarized as the filtering of a sparse graph signal \mathbf{x} on G with a graph e-spline (low-pass) filter followed by dimensionality reduction and giving rise to signal $\tilde{\mathbf{y}}$ on coarsened \tilde{G} , which is subsequently transformed into the further dimensionality-reduced, scaled spectral graph domain, resulting in the representation \mathbf{y} . Following an initial sparsification step, this extends to graph signals \mathbf{x}_W with a sparse (multiresolution) representation via a GWT, which can then be similarly sampled. The derived decomposition notably achieves the distinct spectral characterization of \mathbf{y} in that the coefficient matrix \mathbf{C} , which incorporates graph filter eigenvalues, as well as the graph filter depend directly on (and are unique for) the graph.

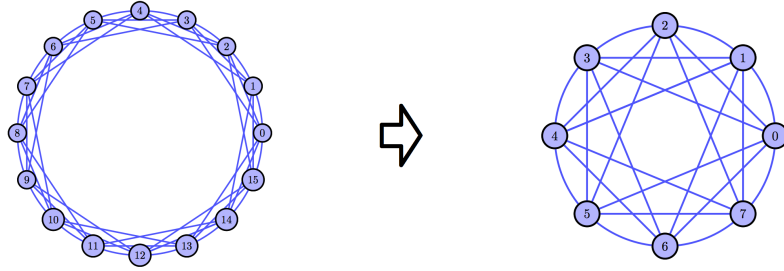


Figure 5.2: Graph Coarsening for a Circulant Graph with $S = \{1, 2, 3\}$. ©2017 Elsevier Inc.

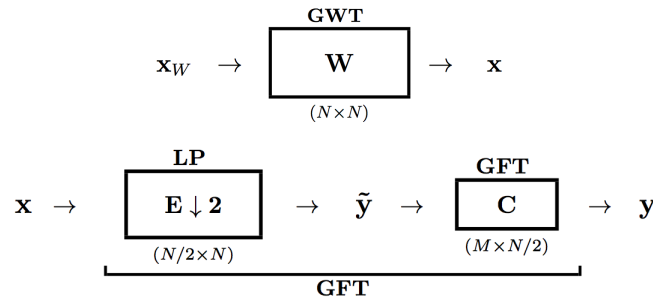


Figure 5.3: Sampling Scheme with Preceding Sparsification Step and One Level of Coarsening. ©2017 Elsevier Inc.

The matrix $\mathbf{E}_{2^j \vec{\alpha}}$ in Thm. 5.2, representing a higher-order, vertex-localized graph e-spline wavelet low-pass filter parameterised by $\vec{\alpha} 2^j$ at level j , is implicitly assumed to reproduce complex exponential graph signals (with $\vec{\alpha}$ as specified), and thus of the form of

a low-pass filter designed in Eqs. (3.3) (of Thm. 3.2) or (3.14) (of Thm. 3.5), depending on whether the graph at hand is also bipartite or not. Since the filter construction in either case is based on the combination (convolution) of different graph e-spline basis functions, for certain scenarios the resulting higher-order function may contain opposing roots, which i. a. would violate Bézout's Thm in the non-bipartite case [25], and thus the necessary biorthogonality condition for filterbank construction. While for complementary graph wavelet filterbanks, it is more intricate to formulate generalized conditions on when exactly this occurs, one can further specify conditions under which the *HGESWT* for bipartite graphs loses reproduction properties and/or ceases to be invertible, as formulated in Cor. 3.4 ².

Further to Thm 5.2, we proceed to specify restrictions on the number of samples M and levels J that ensure the reproduction via low-pass graph filter $\mathbf{E}_{\vec{\alpha}}$. As alluded to in Ch. 3, certain rows of the DFT-matrix cannot be reproduced using a real-valued symmetric graph filter, as, for instance, parameter $\alpha = \pm\pi/2$ induces the opposing (complex conjugate) factors $(1 - iz)(1 + iz)$ in the representer polynomial of a parameterized graph Laplacian filter, which violates Bézouts equality for complementary filterbank construction as well as prevents the reproduction of the corresponding complex exponential in the DFT via a *HGESWT*-based low-pass graph filter.

It follows that one cannot reproduce consecutive rows of the DFT-matrix beyond its $N/4$ -th row, and therefore one needs to ensure within a multiresolution analysis that

$$\alpha 2^{j-1} = \frac{2\pi k 2^{j-1}}{N} < \frac{2\pi(N/4)}{N}, \quad \forall j \leq J$$

such that $k < \frac{N}{2^{J+1}}$ or $J < \log_2\left(\frac{N}{k}\right) - 1$, where J is the total number of levels. In other words, one can approximate the DFT-matrix up to its $M = k + 1$ -th row for a certain number of levels, with parameters $\tilde{M} = \frac{N}{2^J}$ and $M = \frac{N}{2^{J+1}} = \frac{\tilde{M}}{2}$ of Thm 5.2. This condition further coincides with the biorthogonality constraint for traditional e-spline wavelets outlined in [85], [25] which ensures that the corresponding filters do not contain opposing roots, i. e. given distinct $\gamma, \gamma' \in \vec{\gamma}$, $2^j(\gamma - \gamma') = i\pi(2k + 1)$ must not be satisfied for some $k \in \mathbb{Z}$ at level $j \leq j_0 - 1$. Depending on the given signals, one may not always be able to satisfy the sparsity level $K = \frac{N}{2^{J+2}}$ exactly, in which case one may simply require $2(2K) \leq 2M \leq \tilde{M}$ and generalize the formulae accordingly.

Since the graph e-spline filter functions of Thm. 3.2 give rise to traditional e-splines when the graph at hand is a simple cycle, the aforementioned condition with $\gamma = \pm i\alpha$ is sufficient in that case, nevertheless for all other cases, as a result of the complex connectivity of circulant graphs and arising eigenvalue multiplicities, further restrictions on $\vec{\alpha}$ need to

²In particular, this amounts to showing when $\tilde{d}_{\alpha_j} = \sum_{k \in 2\mathbb{Z}+1} 2d_k \cos\left(\frac{2\pi jk}{N}\right) = -\tilde{d}_{\alpha_t} = \sum_{k \in 2\mathbb{Z}+1} 2d_k \cos\left(\frac{2\pi k}{N}\left(t \pm \frac{N}{2}\right)\right)$ can occur, i. e. for which $\alpha_j = \frac{2\pi j}{N}$, $\alpha_t = \frac{2\pi t}{N}$, weights d_k and graph connectivity $k \in 2\mathbb{Z} + 1$, the scheme ceases to be valid.

be imposed.

For an unweighted bipartite circulant graph with consecutive, odd elements $s_k \in S$ and even generating set cardinality $|S|$, and as a special case of Cor. 3.4, one additionally observes at $\alpha = \pm\pi/4$ the e-degree $\tilde{d}_\alpha = \sum_{k \in \mathbb{Z}} 2 \cos(\alpha(2k+1)) = 0$, since $\cos(\alpha(2k+1)) = \frac{\sqrt{2}}{2}$, $k = 0, 3, 4, 7, 8, \dots$, and $\cos(\alpha(2k+1)) = -\frac{\sqrt{2}}{2}$, $k = 1, 2, 5, 6, \dots$. Thus similarly as before, the associated graph filter polynomials contain opposing roots and one can approximate consecutive rows of the DFT-matrix up to at most the $N/8$ -th row, which translates into the constraints $k < \frac{N}{2^{J+2}}$ or $J < \log_2 \left(\frac{N}{k} \right) - 2$, and $M = \frac{N}{2^{J+2}} = \frac{\tilde{M}}{4}$. The parameter M may require further reduction, as a consequence of increasing eigenvalue multiplicities for $\tilde{d}_\alpha = \gamma_i = 0$ for different graph-connectivities.

More generally, if G is bipartite circulant, one needs to ensure that no parameters $\alpha_i, \alpha_j \in \vec{\alpha}$ satisfy $\tilde{d}_{\alpha_i} = -\tilde{d}_{\alpha_j}$ at any level in order to preserve the invertibility property of Thm 3.2, and thus consecutive frequencies (or consecutive rows of the DFT) are considered only up to some cut-off frequency with $\alpha_k = \frac{2\pi k}{N}$ at position $k+1$, such that for $i, j \leq k$, $\tilde{d}_{\alpha_i} \neq -\tilde{d}_{\alpha_j}$. As per Cor. 3.5, $\tilde{d}_{\alpha_i} = -\tilde{d}_{\alpha_j}$ is satisfied when the location parameters fulfill $j = (i + N/2)_N$, which can occur, despite the previously derived constraint with $i, j < N/4$, as a result of large eigenvalue multiplicities (associated with higher graph connectivity) at 0, for some i, j . As an example, we note the normalized adjacency matrix of the unweighted (circulant) complete bipartite graph with bipartite sets of equal size $N/2$, whose eigenvalues are $\gamma_{\max/\min} = \pm 1$ of respective multiplicity $m = 1$, and $\gamma_i = 0$ of multiplicity $N - 2$. Further, the condition $j \neq (i + N/2)_N$ also needs to be satisfied for non-bipartite circulant graphs in the *HCGESWT*, which follows equivalently from the traditional biorthogonality constraints as well as from a special case of the presented graph spline wavelet transform designs.³ We note that these are necessary conditions for the existence of a suitable low-pass filter via the *HGESWT*, however, the set of special cases presented here is not exhaustive.

5.1.5 Extensions to Path Graphs

The path graph, as a simple cycle without the periodic extension, bears similar properties to its circulant counterpart; it is bipartite and its graph Laplacian eigenvectors can be represented as the basis vectors of the DCT-II matrix [119] such that $\mathbf{U}^H = \mathbf{Q}$ is the DCT-III matrix, with entries $Q_{m,n} = c(m) \sqrt{\frac{2}{N}} \cos\left(\frac{\pi m(n+0.5)}{N}\right)$, for $0 \leq m, n \leq N-1$, and constants $c(0) = \frac{1}{\sqrt{2}}$ and $c(m) = 1$ for $m \geq 1$, with corresponding distinct eigenvalues $\lambda_m = 2 - 2 \cos\left(\frac{\pi m}{N}\right)$, $m \in \{0, 1, \dots, N-1\}$. In addition, a signal residing on the vertices of a path graph can be considered analogous to a non-periodic discrete-time signal

³This can be easily demonstrated in a generalization of the proof of Thm. 3.2 (in Appendix A.2), where high-pass filter \mathbf{H}_{HP_α} is maintained and low-pass filter \mathbf{H}_{LP_α} is generalized to the form of Eq. (3.14), with fixed downsampling pattern with respect to $s = 1 \in S$.

in the classical domain. A K -sparse signal sampled with this DCT matrix can be perfectly reconstructed via a variation of Prony's method using at least $4K$ of its consecutive sample values, according to [120], which inspires a specialized extension of the Graph-FRI framework⁴:

Theorem 5.3. (*Graph-FRI for paths*) *Let $\mathbf{x} \in \mathbb{C}^N$ be a K -sparse graph signal defined on the vertices of an undirected and unweighted path graph G , whose GFT basis is expressed such that \mathbf{U}^H is the DCT-matrix \mathbf{Q} . We can sample and perfectly reconstruct \mathbf{x} on G using the dimensionality-reduced GFT-representation $\mathbf{y} = \mathbf{U}_M^H \mathbf{x} \in \mathbb{C}^M$, where \mathbf{U}_M^H corresponds to the first M rows of \mathbf{U}^H , provided $M \geq 4K$.*

Furthermore, as the graph Laplacian matrix of a path is circulant up to its first and last row, it incorporates 2 vanishing moments with a graph-inherent border effect of length $T = 2$, while its powers similarly inherit $2k$ vanishing moments with a border effect of $T = 2k$. Thus, for symmetric normalized adjacency matrix $\mathbf{A}_n = \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}$, the graph spline wavelet construction of Thm. 3.1 applies to the path graph, with the minor restriction that the high-pass filter $\mathbf{L}_{norm} = \mathbf{I}_N - \mathbf{A}_n$ gains an increased border effect of $T = 2(k + 1)$, as the nullspace of \mathbf{L}_{norm} now contains $\mathbf{D}^{1/2} \mathbf{1}_N$, as a consequence of the non-regularity of the graph.

Similar extensions pertain to the e-graph-spline transform in Thm. 3.2 (or bandlimiting transform, Cor. 4.1), with filters of the form $\mathbf{H} = \frac{1}{2^k} (\mathbf{I}_N \lambda_j \pm \mathbf{A}_n)^k$ for eigenvalues λ_j of \mathbf{A}_n , however, these results are of lesser relevance (except for sparse multiresolution analysis), given that the DCT does not give rise to an equivalently intuitive decomposition scheme as the DFT for sampling-based graph coarsening.

At last, it should be noted that, for a multilevel graph wavelet analysis, the path graph can be coarsened via the Kron-reduction of the graph Laplacian matrix, if every other node is sampled, resulting in a weighted path graph with universal weight $1/2$ [60].

5.2 Generalized & Multidimensional Sparse Sampling

In order to extend the derived sampling framework to sparse signals defined on arbitrary graphs, one may leverage a variety of approximation schemes, which facilitate the interpretation of circulant graphs as building blocks for the former. Given the adjacency matrix \mathbf{A} of a general undirected graph G , we therefore propose to conduct (i) nearest circulant

⁴In particular, the DCT matrix is an example of a larger class of invertible sampling bases of the form $\mathbf{Q} = \mathbf{A} \mathbf{V} \mathbf{S}$, with diagonal $\mathbf{A} \in \mathbb{C}^{N \times N}$, Vandermonde matrix $\mathbf{V} \in \mathbb{C}^{N \times M}$ with $[\mathbf{V}]_{n,m} = p_m^n$ and distinct p_m , and $\mathbf{S} \in \mathbb{C}^{M \times N}$, whose columns are at most D -sparse [120]. A K -sparse signal \mathbf{x} can be perfectly recovered from $2DK$ consecutive entries of $\mathbf{y} = \mathbf{Q} \mathbf{x}$ using Prony's method (Prop. 4, [120]), which also facilitates a generalization to graphs whose GFT basis is of that form.

approximation of \mathbf{A} by $\tilde{\mathbf{A}}$, or alternatively (ii) (approximate) graph product decomposition $\mathbf{A} \approx \mathbf{A}_1 \diamond \mathbf{A}_2$ into (banded) circulants \mathbf{A}_i .

The former case entails the projection of \mathbf{A} onto the subspace of circulant matrices \mathbf{C}_N , spanned by circulant permutation matrices $\mathbf{\Pi}^i$, $i = 0, \dots, N-1$, where $\mathbf{\Pi}$ has first row $[0 \ 1 \ 0 \ \dots]$, via the Frobenius inner product $\tilde{\mathbf{A}} = \sum_{i=0}^{N-1} \frac{1}{N} \langle \mathbf{A}, \mathbf{\Pi}^i \rangle_F \mathbf{\Pi}^i$ [121]. Here, \mathbf{A} may be subjected to a prior node relabelling for bandwidth minimization (using for instance the RCM-algorithm [122]) so as to reduce the crude approximation effect of averaging over the diagonals of \mathbf{A} , which introduces additional complementary edges in $\tilde{\mathbf{A}}$, and hence significantly alters the graph, as well as prior partitioning, for when the graph at hand features distinct communities. The set of (wavelet-)sparse signals defined on G can be subsequently sampled with respect to the graph approximation \tilde{G} . In light of the rich sparsifying graph wavelet analysis, the sampling framework may be further extended to accommodate an arbitrary graph and *smooth* graph signal with an initial sparsification step on the original graph, provided that a suitable graph basis can induce a sufficiently sparse representation, while the actual sampling step may be continued with the sparsified signal and the nearest circulant approximation of the given graph.⁵

In the latter case, an additional degree of dimensionality reduction is introduced through the graph product operation, which can be successfully leveraged for suitably defined multi-dimensional sparse signals consisting of sparse tensor factors. Analogously to the investigation of multidimensional graph wavelet analysis, conducted in Sect. 4.5 of the previous chapter, we focus on the generalization of the GFRI sampling framework to arbitrary graphs through graph product decomposition (refer to Sect. 4.5 for definitions).

Given the graph product decomposition $G = G_1 \diamond G_2$, which can be exact or approximate such that factors G_i are undirected, circulant and connected with $s = 1 \in S_i, i = 1, 2$, the tensor factors of a given multi-dimensional graph signal \mathbf{x} on G (as defined in Def. 4.3) can be analyzed with respect to its inherent circulant substructures; we consider $\mathbf{x} = \mathbf{x}_1 \otimes \mathbf{x}_2$ (at rank $k = 1$) for the remainder of this discussion, for simplicity.

A multi-dimensional $K = K_1 K_2$ -sparse graph signal \mathbf{x} , with K_i -sparse tensor factors \mathbf{x}_i , defined on an arbitrary undirected graph G , can be sampled and perfectly reconstructed from dimensionality-reduced GFT-representations of \mathbf{x}_i on the (approximate or exact) graph product decomposition of G into circulant factors G_i . Here, the GFRI framework (Thms. 5.1 & 5.2) can be applied on each component individually, thus perfectly recovering \mathbf{x}_i (using Prony's method) from spectral representations $\mathbf{y}_i = \mathbf{U}_{M_i}^H \mathbf{x}_i$, of dimension $M_i \geq 2K_i$, where $\mathbf{U}_{M_i}^H$ denote the first M_i rows of the permuted GFT (DFT) matrix

⁵This approach nevertheless does not take into account or quantify the loss of accuracy (in i. a. data representation) incurred by working with a graph approximation as opposed to the original graph, which in itself poses an open problem.

of dimension $N_i \times N_i$, $i = 1, 2$. In particular, for all but the lexicographic product, the following holds

$$(\Phi_{M_1} \otimes \Phi_{M_2}) \mathbf{U}^H \mathbf{x} = (\mathbf{U}_{M_1}^H \otimes \mathbf{U}_{M_2}^H)(\mathbf{x}_1 \otimes \mathbf{x}_2) = \mathbf{y}_1 \otimes \mathbf{y}_2 = \mathbf{y},$$

where $\mathbf{y} \in \mathbb{C}^M$ with $M \geq 4K$ and $\Phi_{M_i} \in \mathbb{R}^{M_i \times N_i}$ sample the first M_i rows, or expressed alternatively as $\mathbf{y} = (\mathbf{C}_1 \tilde{\mathbf{y}}_1) \otimes (\mathbf{C}_2 \tilde{\mathbf{y}}_2)$ for graph-filtered representations $\tilde{\mathbf{y}}_i$ and graph spectral transformation matrices \mathbf{C}_i ; otherwise, the \mathbf{x}_i need to be processed separately on the individual graph Laplacian eigenbases of G_i . Ultimately, one may recombine the coarsened circulant graphs \tilde{G}_i associated with representations $\tilde{\mathbf{y}}_i$ under the same graph product operation to form the coarse graph $\tilde{G} = \tilde{G}_1 \diamond \tilde{G}_2$ and redefine the signal $\tilde{\mathbf{y}}_1 \otimes \tilde{\mathbf{y}}_2$ on its vertices.

5.2.1 Exact vs Approximate Graph Product Decomposition

In the case when a graph product decomposition $G = G_1 \diamond G_2$ into circulants G_i is exact and known, one can perfectly recover the tensor-sparse signal \mathbf{x} on G by performing graph operations in lower dimensions, which requires the storage of further lower dimensional spectral representations \mathbf{y}_i . This advantage is particularly evident in the case of the lexicographic product, which is closed under circulant graphs [111], as the GFRI framework may be applied directly on the original graph G , with known lexicographic decomposition and assuming that \mathbf{x} is sufficiently sparse (or smooth on G), yet the decomposition into lower-dimensional circulants can increase efficiency, while preserving the scheme.

The Cartesian product of two unweighted path graphs is known to produce an unweighted lattice graph [105], which facilitates the generalization of both the sparse sampling and wavelet analysis framework to lattice graphs (see Fig. 5.4) as well as to more general graph products of path and circulant graphs. Leveraging the vanishing moment property of the graph Laplacian of a path graph, one can apply a multidimensional wavelet analysis of (piecewise) smooth signals on lattice graphs. In particular, this has revealed an interesting relation to the interpretation of the graph Laplacian as a differential operator: while the graph Laplacian provides the stencil approximation of the second order differential operator for lattices up to a sign, coincidentally, unweighted lattice graphs, as the graph product of two path graphs, which are circulant up to an edge, preserve (to an extent) the inherent vanishing property of the graph Laplacian of a circulant graph via the product operation. This phenomenon was further investigated for general circulants in Sect. 4.5.3.

Hence, following the GFRI Thm. for paths, signals defined on the graph product of path graphs are equivalently sampled and reconstructed in a multidimensional scheme, where sparse subensors \mathbf{x}_i on G_i can be perfectly recovered based on at least $4K_i$ consecutive samples of their dimensionality-reduced GFT (DCT) representation.

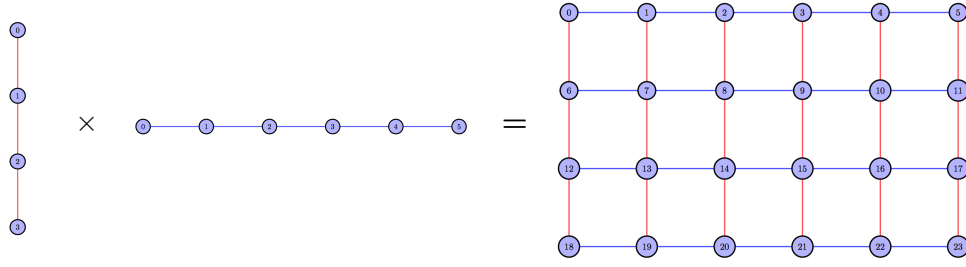


Figure 5.4: Graph Cartesian Product of two unweighted path graphs. ©2017 Elsevier Inc.

In order to extend the GFRI framework beyond circulant and path graphs to arbitrary graphs, which can be decomposed as an approximate graph product of the former, one may resort to Kronecker product decomposition under the constraint of symmetric circularity (and bandedness), as laid out previously in Sect. 4.5; here, the specific case of a 1-banded Toeplitz structure can similarly be imposed on either factor \mathbf{A}_i to obtain a decomposition into path graphs, where applicable.

As a means to complete the foregoing discussion on sampling, one may at last revisit the *separable* graph wavelet transform of Sect. 4.5.2, which is defined on the individual circulant factors G_i of product graphs; while doing so, we preserve previous notation.

Let \mathbf{W}_i denote the graph (e-)spline wavelet transform in the vertex domain of circulant graph factor G_i , and $\mathbf{w}_i = \mathbf{P}_{N_i} \mathbf{W}_i \mathbf{x}_i$ the graph wavelet domain representation of signal factor \mathbf{x}_i on G_i , subject to the node relabelling \mathbf{P}_{N_i} ; the representation $\mathbf{w} = \mathbf{w}_1 \otimes \mathbf{w}_2$ then is the result of a separable, two-dimensional graph spline wavelet transform redefined on G . If \mathbf{x} consists of smooth graph signal tensor factors \mathbf{x}_i such that 2-D multiresolution graph wavelet representation $\mathbf{w} = \mathbf{w}_1 \otimes \mathbf{w}_2 = \mathbf{P}_{N_1 N_2}^{(j)} (\mathbf{W}_1^{(j)} \otimes \mathbf{W}_2^{(j)}) \mathbf{x}$ is K -sparse with $\|\mathbf{w}_i\|_0 = K_i$ and $K = K_1 K_2$, for suitable graph wavelet transforms $\mathbf{W}_i^{(j)}$ and permutation matrices $\mathbf{P}_{N_1 N_2}^{(j)}$ at level $j \leq J - 1$, one can apply the multidimensional framework of sampling and perfect reconstruction of \mathbf{w}_i on the vertices of circulant G_i . Eventually the original \mathbf{x} can be reconstructed from \mathbf{w} via the (invertible) 2-D graph wavelet transform.

5.3 Sampling under Noise: Circulant Graphs with Perturbations

Preceding sections have addressed the challenge of sampling with respect to arbitrary graphs by resorting to (multidimensional) circulant graph approximations, and while such methods are simple in execution, they may fail to preserve basic graph characteristics when the given graph is far from being circulant, as determined via a designated error norm. This final section is therefore devoted to the discussion of sampling on graphs in the presence of (connectivity) noise, which serves to illustrate an alternative avenue of pursuing the above

problem, featuring i. a. matrix perturbation theory, with its accompanying difficulties.

Consider the problem of K -sparse graph signal recovery on perturbed circulant graphs, simulating network clusters, as groups of strongly connected entities within a larger network.⁶ In particular, given such a network, we propose to model the individual network clusters as (un-)weighted undirected circulant subgraphs, which are linked via inter-connecting edges on a main graph $G = (V, E)$, and are subject to perturbations in form of the addition and/or removal of randomly chosen edges; this is distinct from the construction of small-world networks [71], where individual edges of circulants are rewired with a certain probability, yet their total number in the graph is preserved. In light of the aforementioned background and findings, we present a novel model of (blockwise) reconstruction operations with dimensionality reduction; following graph partitioning or clustering, we operate on each subgraph individually using a set of approximation and denoising schemes.⁷

It has been established that the (graph Laplacian) GFT-basis $\mathbf{U} = [\mathbf{u}_0 | \dots | \mathbf{u}_{N-1}] \in \mathbb{R}^{N \times N}$ on circulant graphs can be expressed as a permutation of the columns of the DFT-matrix $\mathbf{F} \in \mathbb{C}^{N \times N}$, however, as a result of the occurring eigenvalue multiplicities, it is not unique. Due to the difference in frequency interpretation, the k -th column of \mathbf{F} , corresponding to the frequency $\omega_k = 2\pi k/N$ is not the same as basis vector \mathbf{u}_k , corresponding to the k -th smallest graph-frequency λ_k , requiring a permutation (and eigensubspace re-combination) in form of coefficient matrix $\mathbf{C} \in \mathbb{C}^{N \times N}$ such that $\mathbf{C}\mathbf{U}^H = \mathbf{F}$.

Let $\mathbf{x} \in \mathbb{R}^N$, with $\|\mathbf{x}\|_0 = K$, denote a K -sparse graph signal defined on the vertices of a circulant graph G , with GFT $\hat{\mathbf{y}} = \mathbf{U}^H \mathbf{x}$, and, following a change of basis, new measurement vector $\mathbf{y} = \mathbf{C}\hat{\mathbf{y}} = \mathbf{C}\mathbf{U}^H \mathbf{x}$. Then \mathbf{x} can be perfectly reconstructed from at least $2K$ consecutive entries of \mathbf{y} using Prony's method. If the samples contain noise, in form of additive Gaussian noise \mathbf{n} , giving $\tilde{\mathbf{y}} = \mathbf{y} + \mathbf{n}$, one needs to further apply denoising schemes, such as Cadzow's algorithm [123], while a larger number of samples $M \geq 2K$ is required to achieve a sufficiently good reconstruction, as has been conducted in the classical domain. The latter is a robust algorithm which performs an SVD-decomposition of $\hat{\mathbf{Y}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ as the Toeplitz-realization of $\hat{\mathbf{y}}$ (see Appendix C.2), zeroing the $L + 1 - K$ smallest diagonal coefficients of $\mathbf{\Sigma}$ giving $\mathbf{\Sigma}'$, and finding the closest Toeplitz-matrix approximation⁸ to the resulting $\hat{\mathbf{Y}}' = \mathbf{U}\mathbf{\Sigma}'\mathbf{V}^T$; a few iterations of this process are conducted to reduce the noise. The shape of $\hat{\mathbf{Y}} \in \mathbb{C}^{L \times R}$ is chosen to be as close to square as possible, with suitable parameters L and R , for optimal performance ([123], [23]).

The devised GFRI-framework is now further challenged through two additional factors:

⁶The content of this section appears in part in publications [11] and [10].

⁷One may alternatively operate on the entire graph at once using a blockwise matrix operation scheme, however, we resort to operating on each subgraph individually to limit the perturbation effect.

⁸Incidentally, this is equivalent to nearest circulant approximation, considering that a circulant matrix is in fact a special Toeplitz matrix.

only a subset of the basis-vectors \mathbf{U} is retained and G is subject to perturbations. In light of this, we develop a reconstruction approach, which employs a variation of Prony's method, along with further iterative denoising, to recover the sparse signal residing on G .

In particular, given a subset of P selected GFT-eigenvectors \mathbf{U}_P and dimensionality-reduced approximation of the GFT $\mathbf{y} = \mathbf{C}\mathbf{U}_P^H \mathbf{x} \in \mathbb{C}^M$, where $\mathbf{C} \in \mathbb{C}^{M \times P}$ and $M < P < N$, we reconstruct the K -sparse graph signal $\mathbf{x} \in \mathbb{R}^N$ on the vertices of a perturbed circulant graph $G = (V, E)$ for suitably chosen P and $M \geq 2K$, with $\mathbf{F}_M \approx \mathbf{C}\mathbf{U}_P^H$ and \mathbf{F}_M denoting the first M rows of the DFT-matrix (see Fig. 5.5).

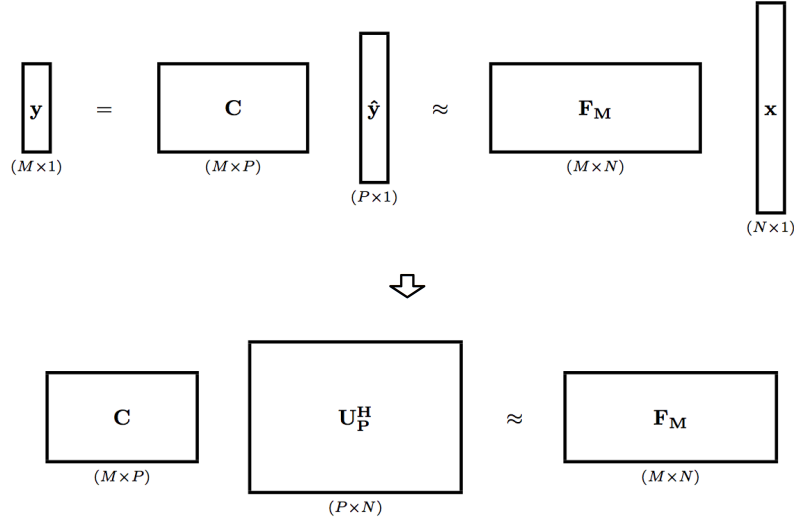


Figure 5.5: GFRI-Scheme under Graph Perturbations.

Here, one imposes a further permutation σ on \mathbf{U} according to the eigenvalue sequence $\mathbf{\Lambda}$ obtained by taking the DFT of the first row of graph Laplacian \mathbf{L} of the circulant G (or its nearest circulant approximation) and arranging repeated eigenvalues of the same subspace j together, as per the location of their first appearance. Depending on the eigenvalue multiplicity distribution function $f(\cdot)$ of the graph at hand, one thus needs to impose the dimensionality offset $P = f(M) \geq f(2K)$, which results in a rectangular coefficient matrix \mathbf{C} . In the case of dimensionality reduction, this allows to obtain a better least-squares approximation of the matrix $\mathbf{C} \in \mathbb{C}^{M \times P}$, i. e. only the relevant eigensubspaces of \mathbf{U} , corresponding to the rows of \mathbf{F}_M , are incorporated in \mathbf{U}_P . In particular, when G is an unperturbed circulant and $P = f(2K)$, one achieves perfect reconstruction.

For instance, in the case of a simple cycle, one can perfectly recover \mathbf{x} with only $P = 2M - 1 = 4K - 1$ GFT-sample values, since the eigenvalues $\lambda_k = 2 - 2\cos(2\pi k/N)$ occur with a maximum multiplicity of $m = 2$ (except at $\lambda_0 = 0$, and $\lambda_N = 4$ with $m = 1$, for even N). While the order of eigenvalue multiplicities may differ by dimension N , it would generally appear for circulant graphs that the higher the connectivity and similarity of edge weights, the less distinct are the eigenvalues. Large eigenvalue multiplicities (along

with proximities) can lead to instabilities in the corresponding eigenspaces [124] as well as require a greater number of GFT-eigenvectors P for the approximation of individual DFT-rows, and thus greater storage. The inclusion of (random) symmetric weights in the graph model generally facilitates a spectrum with smaller multiplicities at higher connectivity, however, as mentioned in earlier discussions, there is no systematic way to detect the particular circulant generating sets inducing lowest multiplicities.

When the circulant graph at hand is perturbed, we apply an iterative denoising scheme to the given samples prior to reconstruction, beginning with Cadzow's algorithm⁹ and followed by further denoising based on removal of the perturbation matrix $\mathbf{E} = \mathbf{C}\mathbf{U}_P^H - \mathbf{F}_M$ at the current estimate $\hat{\mathbf{x}}$ from the measurement vector \mathbf{y} at iteration i : $\mathbf{y}^{i+1} = \mathbf{y} - \mathbf{E}\hat{\mathbf{x}}^i$. In particular, this reduces the error caused by initial estimates $\hat{\mathbf{x}}$ in a setting of highly irregular and/or localized noise, so that for a sufficiently large number of given samples M , and number of iterations i , one can achieve good to nearly perfect reconstruction.

The proposed approaches are summarized in Algorithm 5.1, of which Option 2 only applies to a subset of circulant graphs with certain generating sets, i. a. the simple cycle with $S = \{1\}$; it has been established that for a sufficiently large number of given samples M and iteration number $i \geq 5$, one can usually achieve good reconstruction results. Otherwise, circulant graphs with complementary generating sets have exhibited destructive behavior in the reconstruction process due to highly localized, topologically-dependent, perturbation noise, caused by the permutation/approximation steps, which requires a more refined approach.

Algorithm 5.1 Sparse Graph Signal Recovery on Network Clusters

- 1: **Input:** Adjacency matrix \mathbf{A} of clustered graph $G = (V, E)$, and corresponding signal \mathbf{x} , with $\|\mathbf{x}\|_0 = K$
 - 2: Decompose G into T disconnected subgraphs $\{G_l\}_{l=1}^T$ via graph partitioning (e. g. graph cut). Re-assign \mathbf{x} appropriately on the disconnected subgraphs, resulting in T signals $\{\mathbf{x}_l\}_{l=1}^T$ with sparsity $\{K_l\}_{l=1}^T$. Apply the following scheme(s) on each subgraph and -signal individually
 - 3: **Option 1:** Compute the nearest circulant $\{\tilde{\mathbf{A}}_l\}_{l=1}^T$ to $\{\mathbf{A}_l\}_{l=1}^T$, or alternative. Define \mathbf{x}_l on the vertices of \tilde{G}_l , and, if required, impose a suitable permutation $\sigma_{\tilde{\Lambda}_j}$ on $\tilde{\mathbf{U}}$ of \tilde{G}_l . Compute \mathbf{C} via LS: $\mathbf{C}^T = (\tilde{\mathbf{U}}_P^H)^T \mathbf{F}_M^T$. Only $P = 4K_l - 1$ consecutive samples are required for perfect reconstruction, at best. Ensure that \tilde{G}_l has the required minimum of multiplicities (or adjust P accordingly). Store GFT-vector $\hat{\mathbf{y}} = \tilde{\mathbf{U}}_P^H \mathbf{x}_l \in \mathbb{C}^P$
 - 4: **Option 2:** Model the graph as circulant with a perturbation, by first computing the nearest circulant $\tilde{\mathbf{A}}_l$ to the given \mathbf{A}_l , and imposing the permutation $\sigma_{\tilde{\Lambda}_j}$ on \mathbf{U} of G_l , if required. Compute \mathbf{C} as above, and create $\mathbf{y} = \mathbf{C}\mathbf{U}_P^H \mathbf{x}_l$ for suitable M , and $P = f(M)$ (where mapping $f(\cdot)$ depends on the multiplicities of \tilde{G}_l). Apply the proposed scheme at $P = f(M) \geq 4K_l - 1$ samples:
 - 5: Denoise \mathbf{y} with Cadzow's algorithm, and recover $\hat{\mathbf{x}}_l$ through Prony's method
 - 6: Do further iterative denoising, $\mathbf{y}^{i+1} = \mathbf{y} - \mathbf{E}\hat{\mathbf{x}}_l^i$, as required, while repeating 5. Store GFT-vector $\hat{\mathbf{y}} = \mathbf{U}_P^H \mathbf{x}_l \in \mathbb{C}^P$.
-

⁹Note that this algorithm has been deemed mainly suitable for FRI scenarios where the noise is additive Gaussian, however, in the current scenario, the noise is irregular and multiplicative.

Further, note that the nearest circulant approximation may produce different multiplicity distributions as an otherwise determined circulant graph approximation, so further analysis is needed to investigate which model gives the best eigensubspace-permutation. In this context, consider the following:

Remark 5.1. If the *graph-invasiveness* (or the degree of alteration of an existing graph) of a scheme were to be measured in terms of the number of added or removed links, the nearest circulant (graph) approximation (via subspace-projection) does not necessarily provide the least invasive option, despite minimizing the Frobenius-norm error with respect to the original. In particular, consider the simple cycle graph with an additional (arbitrary) edge; the nearest circulant approximation of the corresponding adjacency matrix yields four additional diagonals with uniform weights $1/N$, which represent $N - 1$ new connections. In contrast, the simple cycle differs only by a link of unit weight from the given perturbed version, and may be interpreted as a structurally closer approximation, despite resulting in a (slightly) larger Frobenius-norm error.

Both the original circulant connectivity (multiplicity distribution) and perturbation location contribute to a setting of highly irregular and localized noise; the performance of Prony's and Cadzow's algorithms is further affected by increasing dimensions, marginal sparsity levels as well as the distribution and strength of perturbation. Moreover, Prony's method is known to be unstable when the sparse vector entries are too close together. For high perturbation noise, the multiplicity offset becomes less critical and/or effective, in which case it is not excluded that more adaptive approaches would lead to better results. As such, it would appear that the variability present due to different graph connectivities and the effect of the perturbation location within the graph largely hinders generalized conclusions on the effectiveness of the approach. However, since a comprehensive numerical analysis is outside the scope of this section and the problem itself, we further conduct a theoretical analysis of simple scenarios in the following section in order to better comprehend the underlying effect of perturbations on graphs.

5.3.1 Notes on Perturbation Theory

In the course of the numerical analysis of perturbed (weighted) circulant graphs, we discover an interesting phenomenon. When the perturbation is particularly strong or localized, due to either a high number of added edges or their specific location, a certain boundary effect causes shifts in the eigen-subspaces, which are usually associated with repeated eigenvalues. The following results from matrix perturbation theory serve to further comprehend this perturbation effect.

Let \mathbf{L} represent the graph Laplacian matrix of a circulant graph G , and \mathbf{E} the graph Laplacian of a single (unweighted) edge not on G ; then $\mathbf{L} + \mathbf{E}$ gives a perturbed graph

caused by a rank-1 matrix perturbation \mathbf{E} .¹⁰ Given that \mathbf{L} and $\mathbf{L} + \mathbf{E}$ are real, symmetric matrices, we can infer certain properties on the eigenvalues of the perturbed graph Laplacian, including the bound $|\lambda_k(\mathbf{L} + \mathbf{E}) - \lambda_k(\mathbf{L})| \leq \|\mathbf{E}\|_2 = 2$, $k = 0, \dots, N-1$ (Cor. 8.1.6, [5]) and the fact that there exist nonnegative m_i such that $\lambda_i(\mathbf{L} + \mathbf{E}) = \lambda_i(\mathbf{L}) + m_i\alpha$, $i = 0, \dots, N-1$, and $\sum_{i=1}^N m_i = 1$ (Thm. 8.1.8, [5]), since for $\mathbf{E} = \alpha \mathbf{c}\mathbf{c}^T$ with some $\alpha \in \mathbb{R}$, $\alpha \geq 0$ and $\mathbf{c} \in \mathbb{R}^N$ of unit l_2 -norm,

$$\lambda_i(\mathbf{L} + \mathbf{E}) \in [\lambda_i(\mathbf{L}), \lambda_{i+1}(\mathbf{L})], \quad i = 0, \dots, N-2$$

whereas for $\alpha \leq 0$

$$\lambda_i(\mathbf{L} + \mathbf{E}) \in [\lambda_{i-1}(\mathbf{L}), \lambda_i(\mathbf{L})], \quad i = 1, \dots, N-1.$$

Matrix perturbation theory provides tighter bounds as well as series expansions for the perturbed eigenvalues of Hermitian matrices, but they cannot be determined explicitly [124]. Nevertheless, there exists an explicit formula for the corresponding perturbed eigenbasis (in the case of distinct λ 's) as well as scheme for the case of repeated λ 's, yet requiring the knowledge of the perturbed spectrum, which was developed by Bunch, Sorensen et al. (1978) [125]. We generally note that the perturbation of an eigenbasis with repeated eigenvalues (which are additionally close to each other, instead of lying in well-separated clusters) leads to instabilities [124].

In particular, according to the previous results, the perturbed eigenvalue $\tilde{\lambda}_i$ lies in the interval $[\lambda_i, \lambda_{i\pm 1}]$ of unperturbed eigenvalues (for which respectively an edge has been added or removed), yet when the outmost perturbation gives $\tilde{\lambda}_i = \lambda_{i\pm 1}$, i.e. the eigenvalue is perturbed to equal the next-highest/lowest unperturbed eigenvalue, we run the risk of a swap, as (equal-valued) perturbed eigenvalues are ordered in a random manner making it impossible to identify which value corresponds to which eigensubspace; this perturbation effect is directly visible when only one eigenvalue per subspace is perturbed. However, this information is particularly crucial for the reordering of the given perturbed GFT matrix according to its unperturbed eigenvalues. The phenomenon of a subspace swap was further studied in [126] in the context of SVD-based methods.

Another boundary perturbation effect occurs for the approximation $\tilde{\lambda}_i \approx \lambda_{i\pm 1}$, which, while not leading to a swap, if sufficiently close, reveals (from the coefficient distribution in the matrix \mathbf{C}) that the perturbed GFT-eigenspace at $\tilde{\lambda}_i$ may provide a closer approximation, within the linear combination of eigenspaces, to the $i \pm 1$ -th DFT-row at hand, than the corresponding eigenspace at $\tilde{\lambda}_{i\pm 1}$. However, these occurrences need to be subjected to further theoretical study. In case of several edge perturbations, the perturbed spectrum is subject to shifts beyond $\tilde{\lambda}_i \geq \lambda_{i\pm 1}$, which renders a coordinated permutation

¹⁰In the case of high perturbation noise, i.e. \mathbf{E} has larger rank, such analysis becomes less definite.

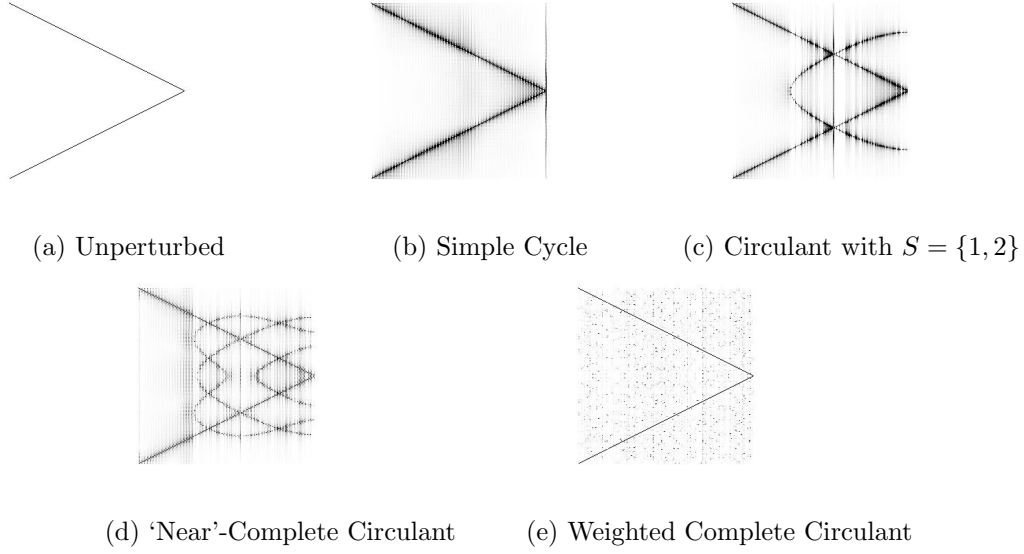


Figure 5.6: Coefficient Matrix \mathbf{C} under Perturbations (displayed in magnitude, scale-factor 10).

approach of the eigensubspaces less effective.

The nature of the perturbation/approximation noise present in different scenarios for circulant graphs is further monitored via the structure of the coefficient matrix \mathbf{C} .

In particular, Fig. 5.6 displays the full coefficient matrix \mathbf{C} (given \mathbf{U}) of size 256×256 , which consists of characteristic diagonals in the general unperturbed state (Fig. 5.6 (a)), for (b) the simple cycle, (c) the circulant graph with $S = \{1, 2\}$, (d) the near-complete unweighted graph with its only zero weights for edges $\{i, (i + N/2 - 2 : i + N/2 + 2)_N\}$ and (e) a randomly weighted complete circulant graph with weights in $(0, 1)$. Each graph is subject to perturbations between the same node pairs $(1, 4)$ and $(5, N/2)$, where for (b) and (c) these are connected with unit weight, while for (d) both are disconnected and for (e) the former is disconnected and the latter re-weighted by a random weight; in the unperturbed state all graphs have a comparable eigenvalue multiplicity distribution with maximum multiplicity $m = 2$.

For a simple cycle (b), we observe the formation of noisy entries (for every perturbed eigenvalue) closely around the diagonals, indicating that perturbed eigenvectors require additional approximation information from preceding and succeeding eigenspaces. The magnitude of the respective coefficients decreases the further away an eigenspace lies (as ordered by eigenvalue magnitude from smallest to largest). This type of noise distribution can be detrimental to performance when certain circulant graphs are considered. While the permutation σ changes the order of the perturbed eigenvectors, the previously noted approximation effect, where nearby eigenspaces, as ordered by magnitude of the corresponding eigenvalues, contribute to the approximation of a given eigenvector via noise, remains in place, and thereby may differ in shape from the diagonal structure of the

coefficient matrix \mathbf{C} , as can be seen for cases (c) and (d). This difference in location impacts the goodness of approximations, when considering P GFT-rows for the reduced M DFT-rows; as the coefficient matrix is of size $M \times P$, it will not contain approximation information from eigenvectors, which occur much later in the new sequence, yet are actually closer (as per eigenvalue magnitude ordering) to the eigenvectors at hand. In (e), however, noise appears to be evenly distributed across \mathbf{C} .

5.3.2 Computational Experiments

Consider the simple cycle graph subject to perturbations in form of two additional, randomly distributed edges $(1, 4)$ and $(5, N/2)$, and let a set of 100 randomly generated sparse graph signals $\{\mathbf{x}_i\}_i$ with i.i.d. entries reside on its vertices; we apply the proposed scheme on the corresponding measurement vectors $\{\mathbf{y}_i\}_i$ to recover the signals (see Option 2 in Algorithm 5.1). Fig. 5.7 illustrates the average reconstruction performance per iteration in form of the average location error between the estimated and true entry locations of the randomly generated sparse vector \mathbf{x} for 100 generated trials, based on the dimensionality reduced measurement vector $\mathbf{y} \in \mathbb{C}^M$, and $P = 2M - 1$ GFT-samples. In the course of the

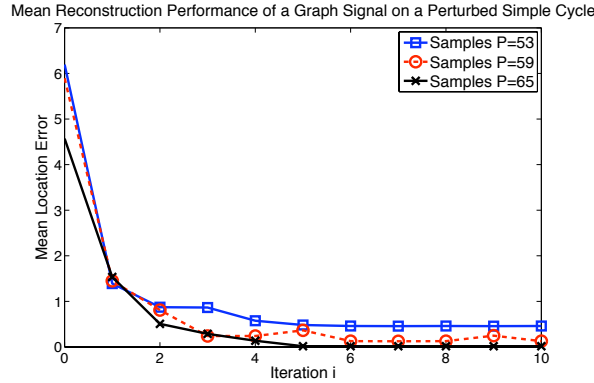


Figure 5.7: Reconstruction Performance on a Perturbed Simple Cycle ($N = 256$), for 100 randomly generated sparse signals \mathbf{x}_l ($K = 4$, minimum separation of 3 between entries). ©2015 IEEE

trials the location of the K -sparse entries as well as the corresponding amplitudes (i.i.d. Gaussian numbers) are re-distributed, and the average location error $\mathbf{e} = \sum_{k=1}^K \frac{|u_k - \tilde{u}_k|}{K}$ between the true locations u_k and the estimates \tilde{u}_k recorded. In this case, the simple cycle and the nearest circulant approximation to the perturbed graph exhibit the same multiplicity distribution and hence ordering scheme.

Overall, while the above analysis has revealed interesting properties of (perturbed) circulant graphs, it appears that it is only feasible for the addition/removal of (individual) links for selected circulant graphs of low multiplicities. Therefore, for an arbitrary graph,

it remains more promising to conduct operations directly on a nearest circulant graph approximation and apply the previously developed GFRI-framework (which further facilitates graph coarsening), at the possible sacrifice of connectivity-preservation. The latter may be regarded as a special case of the presented perturbation scheme for which \mathbf{F}_M is directly assumed as the permuted partial eigenspace of the circulant graph (approximation) at hand, circumventing the need for coefficient matrix \mathbf{C} with dimensionality offset.¹¹

¹¹Here, the coefficient matrix \mathbf{C} used in the GFRI framework is unrelated to the present construction.

Chapter 6

Image Processing on Graphs

Graph-based image processing, in the form of graph wavelet analysis or regularization ([127], [41], [128], [36]), presents a promising application venue for theoretical contributions to harmonic analysis on graphs, that is garnering appeal within the growing field of GSP. The projection of (patches from) images onto higher-dimensional network structures has been employed for i. a. segmentation, compression and/or denoising purposes, as it particularly lends itself for capturing and processing the inherent geometry of the data on a sophisticated level.

While in the one-dimensional case, traditional wavelet bases are most suited for the representation of piecewise smooth functions, higher-dimensionality encompasses the additional challenge of incorporating smoothness across multiple directions and discontinuities i. a. along curves, which prove difficult to capture by simple tensor product wavelet bases and require more sophisticated methods. Classical image processing, in particular, has experienced the advent of curvelets [17] and shearlets [129], which represent highly redundant frames with strong anisotropic directional selectivity, however, are non-adaptive and thus mainly suited for images which are piecewise C^2 with discontinuities along C^2 curves [20]. For less regular images, bases and frames such as the orthogonal bandelets were introduced, which represent anisotropic wavelets that can adapt to the geometric regularity of the image [19]. Nevertheless, the construction of such bases is elaborate, while conceptually non-trivial, which all the more invites the conceptual clarity of graphs and GSP.

This final chapter aims to unify gained notions on wavelets and sparsity on graphs in an application-driven framework for *graph-based non-linear image approximation*, whose discussion is intertwined and furthered i. a. through the substantial, yet largely unexplored (and unexploited), concept of the optimum node labelling. In its essence, the framework facilitates the analysis of homogeneous image regions with respect to data-driven graph wavelet basis functions for increasingly sparse representations. The superiority of the approach, compared to traditional image transforms, is eventually demonstrated on the basis of real and artificial image patches. At last, the setting of noisy (image) data is considered

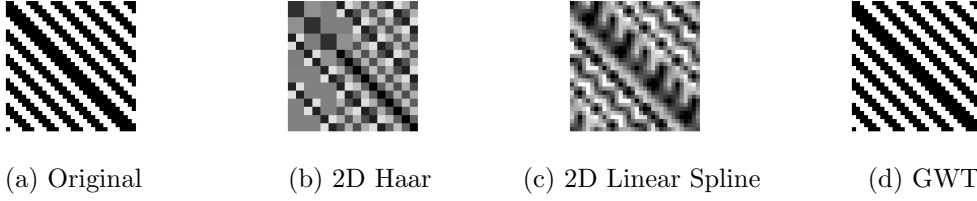


Figure 6.1: Non-Linear Approximation Comparison for the 2D Haar, 2D Linear Spline and proposed Graph Wavelet Transform (GWT) with 5 levels at 5% of non-zero coefficients (f. left).

and its discussion concluded with an open problem.

6.1 Graph Wavelets for Non-Linear Image Approximation

A (gray-scale) image, traditionally represented by a matrix, can be modelled by a graph in a variety of ways. Popular graph models include grid graphs which represent the image structure as a grid of pixels (nodes) of rectangular and/or diagonal connectivity, and nearest neighbor (k - NN) graphs, in which each vertex, representing a vector (object) in a metric space, has exactly k connections to its nearest neighbors, as determined by a designated distance function [130].

Of particular interest in this context are images with distinct geometrical properties, as opposed to natural images, incorporating i. a. irregularly shaped, sharp discontinuities between homogeneous regions or alternating patterns, which challenge the ‘stiff’ row-column-wise operations of conventional tensor product wavelet bases, generally unsuitable to represent geometric structures, and invite more flexible, graph-based operations across irregular (not necessarily spatially confined) image segments.

Consider the example of an image with a simple (binary) stripe pattern, pictured in Fig. 6.1, that is approximated with traditional and graph wavelet basis elements corresponding to 5% of the largest magnitude wavelet coefficients. It becomes apparent that while both of the traditional tensor wavelet transforms produce significantly deformed results, the proposed ‘segmentation-based’ graph wavelet transform achieves perfect reconstruction in this case.

6.1.1 The General Framework

In the following, an approach is devised to conduct non-linear image approximation on graphs which facilitates the analysis of (not necessarily spatially connected) homogeneous image regions through circulant graph wavelets for a maximally sparse representation.

Image Segmentation

A (data-driven) graph $G = (V, E)$ is built from an $N \times N$ image by letting each node represent a pixel and establishing bilaterally weighted connections between node pairs (i, j) , of the form

$$w_{i,j} = e^{-\frac{\|p_i - p_j\|_2^2}{\sigma_p^2}} e^{-\frac{|I(i) - I(j)|^2}{\sigma_I^2}}, \quad i, j \in \{0, \dots, N^2 - 1\} \quad (6.1)$$

based on the distance between their spatial location p and intensity I ; here, σ_p and σ_I denote their respective calibration. Further, the image is converted to a graph signal $\mathbf{x} \in \mathbb{R}^{N^2}$ for processing on G by assigning an intensity value to each node i and adopting the vector-stacked form $x(i) = I(i), \forall i \in V$.

The transformed image data can be subsequently segmented according to graph-based image segmentation methods; here, a variation of the graph cut is employed, which seeks an optimal partition of the vertex set V of G into two sets A and B , while minimizing the number and/or total weight of removed edges between them.

The normalized graph cut problem on a graph with degree matrix \mathbf{D} and weighted adjacency matrix \mathbf{W} is formulated as

$$\min_{\mathbf{y}} Ncut(\mathbf{x}) = \min_{\mathbf{y}} \frac{\mathbf{y}^T (\mathbf{D} - \mathbf{W}) \mathbf{y}}{\mathbf{y}^T \mathbf{D} \mathbf{y}}, \quad \text{subject to } y(i) \in [1, -b] \text{ and } \mathbf{y}^T \mathbf{D} \mathbf{1} = 0$$

for some specified constant b , and can be minimized by solving the generalized eigenvalue problem $(\mathbf{D} - \mathbf{W})\mathbf{y} = \lambda \mathbf{D} \mathbf{y}$ [131]. Its approximate solution, given by the second smallest eigenvector of the previous system, serves as an indicator vector for partitioning, whose splitting point can be selected through a search across evenly spaced points to minimize the so-called *Ncut*-measure, defined as

$$Ncut(A, B) = \frac{cut(A, B)}{assoc(A, V)} + \frac{cut(A, B)}{assoc(B, V)}$$

for resulting partitions. Here, $cut(A, B) = \sum_{u \in A, v \in B} w(u, v)$ is the traditional cut-measure as the total weight of removed edges between partitions A and B , while $assoc(A, V) = \sum_{u \in A, t \in V} w(u, t)$ measures the connections from vertices in one set to all graph vertices. The significance of this approach lies in the fact that the cut cost is computed as a fraction of total connections, which circumvents the bias of cutting isolated points [131].

The selection of suitable weights and calibrations σ in the initial graph model, with the latter described as typically 10-20% of the total range of the distance function $d(x)$ for weight $w(x) = e^{-\left(\frac{d(x)}{\sigma}\right)^2}$ [131], is therefore crucial in order to ensure the goodness of the cut.¹

¹Contrary to the model in [131], the present image graph does not feature connectivity restrictions for spatially distant pixels, since the primary objective is to separate highly homogeneous as opposed to

Performance of the proposed algorithm is in the first instance governed by the realization of an adequate segmentation, and, where applicable, several iterative graph cuts may be performed.

Graph Wavelet Analysis

Eventually, circulant first-order graph spline wavelet transforms are constructed and applied on the individual regions of homogeneous (or constant) content, as captured in form of partitioned sub-graph signals.

To obtain circulant graphs, one resorts to computing circulant graph approximations \tilde{G}_i to the partitioned sub-graphs G_i using the nearest circulant approximation scheme, which entails the minimization of error norm $\min_{\tilde{\mathbf{W}}_i \in C_{N_i}} \|\mathbf{W}_i^P - \tilde{\mathbf{W}}_i\|_F$ over the space C_{N_i} of all $N_i \times N_i$ circulant matrices, for given weighted adjacency matrices $\mathbf{W}_i \in \mathbb{R}^{N_i \times N_i}$, $i = 1, 2$. If G is sparse or a posteriori sparsified (e.g. through the thresholding of edges with small weight), this approximation can be refined through a prior node relabelling P based on the Reverse Cuthill-McKee (RCM) algorithm [122], so as to minimize the bandwidth of \mathbf{W}_i . In particular, this facilitates a restructuring such that \mathbf{W}_i^P is (locally) closer to circulant (sub-)structures, and, hence, reduces the number of complementary edges in $\tilde{\mathbf{W}}_i$ that are introduced as a result of the diagonal averaging of the circulant approximation. The closed-form solution is therefore captured as

$$\tilde{\mathbf{W}}_i = \sum_{j=0}^{N_i-1} \frac{1}{N_i} \langle \mathbf{W}_i^P, \mathbf{\Pi}^j \rangle_F \mathbf{\Pi}^j, \quad i = 1, 2$$

for circulant permutation matrix $\mathbf{\Pi}$, and, graph signals residing on G_i are analyzed with respect to circulant approximation \tilde{G}_i .

Hence, in order to construct a more localized form of the graph wavelet with respect to the image segment, as opposed to the entire image, among other objectives, one can explore various graph sparsification schemes; we concentrate on sparsification with respect to the Euclidean distance by only retaining connections within a square neighborhood of the original pixel grid $\|p_i - p_j\| \leq \sqrt{2}$, as well as sparsification with respect to a (data-dependent) intensity threshold. In the latter case, the graph weights are additionally transformed from bilateral to intensity-based for the application of the GWT, and the RCM algorithm is replaced by a simple sorting of the corresponding subgraph signal to serve as the relabelling step; this simultaneously minimizes the bandwidth, prior to the

spatially confined content, however, this may be tailored.

approximation, and the (graph-unrelated) total variation

$$\|\mathbf{x}^P\|_{TV} = \sum_{j=2}^{N_i} |x^P(j) - x^P(j-1)|$$

of the signal to be annihilated, for maximum sparsity in the graph wavelet domain. In particular, when the similarity measure of \mathbf{W}_i is based solely on the (intensity) samples of the graph signal \mathbf{x}_i , the relabelling from the proposed bandwidth minimization algorithm approximates the one obtained by performing a simple sort operation on \mathbf{x}_i , due to the breadth-first traversal of the former. In an effort to further minimize the border effect of the GWT, one may approximate the given subgraph(s) by the simple cycle, as the sparsest circulant graph, or alternatively, the ‘smoothest cycle’ of G_i , and define the sorted graph signal \mathbf{x}_i on its vertices. This phenomenon and the significance of incorporating a relabelling step is revisited and further elucidated in the next section.

The transform is extended to multiple levels by iterating on the respective lowpass-branches, where downsampling is conducted with respect to the outmost cycle ($s = 1$) and no reconnection applied for graph coarsening, so as to ensure maximum sparsity; if either subgraph dimension is odd, one may alternatively employ nearest circulant approximation for reconnection to preserve circularity.

While, as has emerged from the discussion in Ch. 4, the *HGSWT* remains valid on arbitrary connected undirected graphs, albeit the high-pass filter loses vanishing moments, suitable downsampling and reconnection strategies need to be identified in order to continue to preserve invertibility and sparsity for a multiscale representation, and are thus more difficult to tailor for the chosen graph model and objectives than for circulant graphs.

Overall, the proposed non-linear approximation scheme consists of two main components: a *graph-based image segmentation* step followed by the construction of suitable *circulant graph wavelets* on the segmented regions. Here, the selection of parameters and graph weights may be updated in each step so as to optimize the respective task, i. e. the initial (complete) graph is built to ensure an optimal cut, while for graph wavelet analysis a sparser graph may be preferred for refined processing and enhanced performance. A complete description is outlined in Algorithm 6.1.

6.1.2 The Matter of the Labelling

The GWT for image approximation may be regarded as an instance of transforming a 2D into a 1D problem, which is directed to benefit from graph-specific operations, such as *optimization of labelling* and *localization of basis functions*.

While the primary purpose for introducing the RCM-algorithm in the second stage of the

Algorithm 6.1 Non-linear Image Approximation on Graphs

- 1: **Input:** Grayscale Image \mathbf{I}
 - 2: Construct undirected, weighted graph G from \mathbf{I} with weights $w_{i,j} = e^{-\frac{|I_i - I_j|^2}{\sigma_I^2}} e^{-\frac{\|p_i - p_j\|_2^2}{\sigma_p^2}}$ between node pair (i, j) . Let graph signal $\mathbf{x} \in \mathbb{R}^{N^2}$ be the vector-stacked form of \mathbf{I} .
 - 3: Perform one (or more) normalized graph cuts on G , and partition the graph signal accordingly s.t. signals $\{\mathbf{x}(V_k)\}_k$ respectively reside on graph-partitions $G_k = (V_k, E_k)$ for $k = 1, 2, \dots$
 - 4: Sparsify G_k by a suitable threshold and relabel using the RCM
or
 - 5: Reweight G_k with $w_{i,j} = e^{-\frac{|I_i - I_j|^2}{\sigma_I^2}}$, then sparsify by a suitable threshold and relabel using the simple sort on $\{\mathbf{x}(V_k)\}_k$
 - 6: Compute the nearest circulant graph approximations \tilde{G}_k to G_k
 - 7: Construct a circulant GWT on \tilde{G}_k to analyze sub-signals $\{\mathbf{x}(V_k)\}_k$
 - 8: Coarsen \tilde{G}_k with no reconnection and iterate on the respective LP-branches for a multiresolution representation.
-

proposed framework, has been to obtain a reordered matrix with more similar neighborhoods, and, whose structure is closer to that of a circulant matrix, it appears that in light of our treatment of maximum annihilation, the obtained reduced bandwidth can be additionally leveraged for increased sparsity in the graph wavelet domain, provided the corresponding re-ordered graph signal is smooth. In point of fact, the problem of finding the optimal band form of a symmetric matrix has an equivalence in the graph domain, as captured by the following theorem:

Theorem 6.1. ([51]): *Let \mathbf{A} be a symmetric square matrix of order N which is (p, p) -banded. Let the vertices of its graph $G(\mathbf{A})$ be labelled by the indices $1, 2, \dots, N$ of the corresponding rows (and columns) of \mathbf{A} . If each edge (i, k) of $G(\mathbf{A})$ is assigned the number $|i - k|$, then p equals the maximum over all edges in $G(\mathbf{A})$ of the numbers assigned in this way.*

Namely, the problem of finding a labelling, consisting of distinct integers 0 to $N - 1$, for the vertices of a graph of size N , such that the maximum value of the distances $|i - k|$ over all edges (i, k) is minimized, belongs to the class of NP -complete problems, comprising i. a. the travelling salesman problem, yet several (heuristic) algorithms such as the presently employed RCM or GPS algorithm can find adequate (suboptimal) solutions [51].

Both algorithms are designed to find a so-called level structure $R = [V_1, \dots, V_r]$ of the given graph $G = (V, E)$ which partitions the vertex set into smaller sets of connected components, such that the condensation of G with respect to R is connected and the width $w(R) = \max_{i=1, \dots, r} |V_i|$ of R is minimized [51]. The condensation of G describes the path of size r through vertices V_1, \dots, V_r , resulting from the contraction of each partition to a single vertex, which has an edge from V_i to V_j , $i \neq j$, if and only if there is an edge from x to y in E , for elements $x \in V_i$ and $y \in V_j$ [51]. The employed RCM algorithm, which comprises a slight modification [132] of the original by Cuthill and McKee [122], is

summarized in Algorithm 6.2.

Algorithm 6.2 The RCM algorithm [51]

- 1: **Input:** Graph G with $|V| = N$ and individual vertex degrees $\{s_i\}_{i=1}^N$
 - 2: Define $S = \max_{i=1,\dots,N} s_i$, $s = \min_{i=1,\dots,N} s_i$ and $\sigma = \frac{1}{N} \sum_{i=1}^N s_i$, and set threshold $T = \max\{\min[(S + s)/2, \sigma - 1], s\}$
 - 3: for $v = 1 : N$, if $s_v \leq T$
 - Generate R_v at vertex v : put $V_1 = \{v\}$ and define $V_k, k > 1$ as the set of all vertices with distance $k - 1$ from v
 - Compute $w(R_v)$
 - end
 - Consider the R_v with minimum $w(R_v)$ and label as follows:
 - 4: Let $V_1 = v$ be 1, for $k = 1 : \dots$,
 - Given the labelling in V_k , for V_{k+1} label first the vertices adjacent to the lowest numbered vertex in V_k , in order of increasing degree
 - Move to the next lowest numbered vertex in V_k and repeat until all vertices in V_{k+1} are exhausted
 - Repeat above until all vertices in G are numbered. If the next vertex cannot be chosen uniquely, it is done arbitrarily
 - end
 - 5: Compute bandwidth for each labelling and choose the (or a) minimizing one
 - 6: Number vertices in reverse order of final labelling
-

As laid out, the final labelling is based on the creation of a level structure to form a smooth multi-dimensional path. In addition, the algorithm assumes an unweighted graph, as it only takes the number of connections, not the weights, into account. It is therefore beneficial to consider graphs that are sparse or have been reduced to their most significant edges via thresholding, since the removal of redundant or less meaningful edges may improve the search for smooth labelling paths. A *threshold* may be chosen heuristically and depends on the graph at hand, i. e. if chosen too small it may decrease differentiation between regions and lead to similar adverse effects as if chosen too large as well as possible disconnectedness; accordingly, the labelling within individual neighborhoods (level components) of less nuanced connectivity is subject to increased inaccuracy.

The bandwidth minimization problem presents one of several graph labelling problems, which have been incorporated into a comprehensive framework in [133]. For a connected undirected graph $G = (V, E)$, with $|V| = N$, an optimum one-to-one mapping (i. e. labelling) P from V to the vertex set $V_H = \{0, \dots, N - 1\}$ of a host graph $G_H = (V_H, E_H)$ is sought, as per minimization (or maximization) of a suitable objective function; this specifically entails that distances between labels correspond to pairs of adjacent vertices in G . The labelling P that minimizes the graph bandwidth, as has been of interest, can be expressed as

$$\hat{P} = \arg \min_P \max_{\{u,v\} \in E} d_H(P(u), P(v))$$

where d_H is the shortest path distance between two vertices (as simply the absolute value of the label difference) and $\max_{\{u,v\} \in E} d_H(P(u), P(v))$ denotes the labelling bandwidth.

The related bandwidth (or minimum) sum problem has a different objective function

$$\hat{P} = \arg \min_P \sum_{\{u,v\} \in E} d_H(P(u), P(v)),$$

and is set to minimize the total length sum of the edges. To date, the most widely used host graph model in graph labelling problems is the simple path, which has provided the foundation for a variety of approaches, while alternatives, such as the simple cycle or (higher-dimensional) grid graph have been studied to a lesser extent [133]. A theoretical investigation with respect to weighted graphs has not yet been realized, to the best of our knowledge, albeit heuristic approaches exist.

While operations on graphs with respect to graph signals, as considered, are usually invariant to labelling permutations, the latter nonetheless provide essential information on the general structure of the network which can be leveraged for various objectives, including *bandedness* (sparsity) and *visualization*. In [134], a transformation technique is devised that converts networks to time series using multidimensional scaling, with results uncovering i. a. that the level of periodicity of resulting time series is an indicator for the similarity of networks to ring lattices (i. e. circulants). For instance, the time series resulting from circulant graphs coincide with its periodic sinusoidal eigenvectors that reflect the underlying ‘smooth’, circular labelling, while those of the small-world network are ‘noisy’ periodic. This has been further studied as an inverse operation [135]. Nevertheless, the smoothness of eigenvectors is generally not a sufficient criterion for our purposes, considering that every circulant graph has the same periodic eigenvectors, but not necessarily a small bandwidth. In [136], a new heuristic is introduced for the approximate solution of the cyclic bandwidth sum problem, consisting of the identification of a set of paths by a depth-first search for vertices with similar neighborhoods, which are subsequently merged; further, the algorithm adapts its search to the structure of the graph as it traverses it. It is further argued that a labelling that reflects or is consistent with the topology of the network, i. e. takes into account the adjacency relations, facilitates spectral analysis [136]. The related problems of *community detection*, *clustering*, and *image segmentation*, which can be posed as graph partitioning problems, thus present popular examples that consider vertices in a consistent order with respect to the graph topology. It becomes evident on the basis of a simple form of execution, the graph cut, which determines the most economical (with respect to some cost measure) partition of a graph into disjoint sets of high intra- and low inter-similarity via the Fiedler vector² [47], how the sample-distribution (or smoothness) of the corresponding graph eigenvectors, which represent a form of graph signal in their own right, is crucial in determining vertex adjacency. As such, the use of the graph cut for segmentation, in the first step of the proposed framework, presents a

²The line of work originating in spectral graph theory employs (graph) eigenvectors for bisection, and, in early heuristics the Fiedler vector, which belongs to the second smallest eigenvalue of the graph Laplacian.

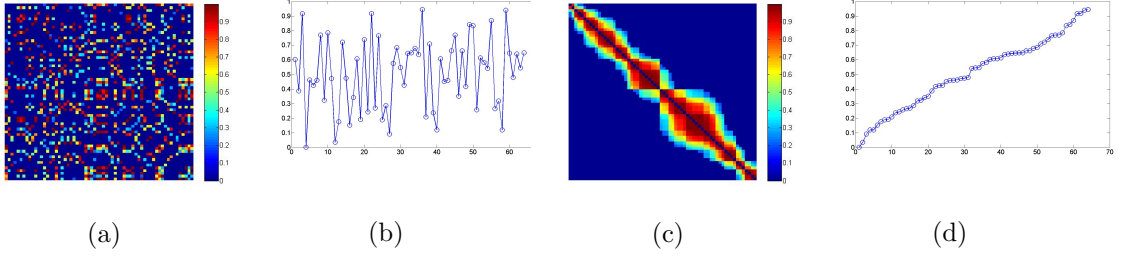


Figure 6.2: Random Graph Signal (b) with corresponding Graph Adjacency Matrix (a) before and after ((c) – (d)) applying a simple sort (graph kernel as in Eq. (6.2), with threshold $T = 0.3$).

preliminary means of label categorization in the pursuit of an optimal graph labelling.

Signals that vary smoothly over the network, or in other words, whose distribution is reflected in the network structure through similar values in neighborhoods of higher connectivity, can therefore be leveraged in a superior relabelling scheme as indicator-functions, whose simple ordering is related to and may be used to reduce the bandwidth of the graph. In fact, while originally an exclusive graph problem, the addition of the graph signal has simplified the bandwidth-minimization problem, as an optimal solution can now be found in form of the simple sort operation, which minimizes the total variation of the graph signal. The sorted signal is (ideally) piecewise smooth, with pieces conceived as approximate (or noisy) polynomials, and can be (near-)annihilated by a suitable graph Laplacian filter with a sufficient number of vanishing moments, while the minimum bandwidth ensures a reduced graph border effect (see Fig. 6.2).

In the proposed image processing framework, this is realized by confining the similarity kernel of the partitioned subgraphs to the intensity of corresponding graph signals which subsequently facilitates the use of signal-sorting as opposed to sub-optimal algorithms for relabelling; the projection of the sorted signal onto a simple cycle is the sparsest option and directly reflects the, now, simplified task of finding the smoothest possible path (or cycle) within a graph. Nevertheless, when the underlying similarity kernel of the graph goes beyond reflecting the sample value similarity, this becomes less reliable; when additional measures, such as the bilateral weight, which includes spatial (Euclidean) proximity, are incorporated, while desirable for other purposes such as image segmentation, compromise the previously stated relation between sorting and bandwidth. A path of spatially close vertices in an image-graph is not necessarily smooth, with the simple-sort relabelling revealing that similar pixels (nodes) are often spread out, hence, the inclusion of the Euclidean distance in the weight measure is limiting to the bandwidth-reduction model.

By way of illustration, Fig. 6.3 depicts a data-driven graph with weights

$$w_{i,j} = e^{-\frac{d(x_i, x_j)^2}{\sigma^2}}, \quad d(a, b) = |a - b|, \quad i, j = 0, \dots, N - 1 \quad (6.2)$$

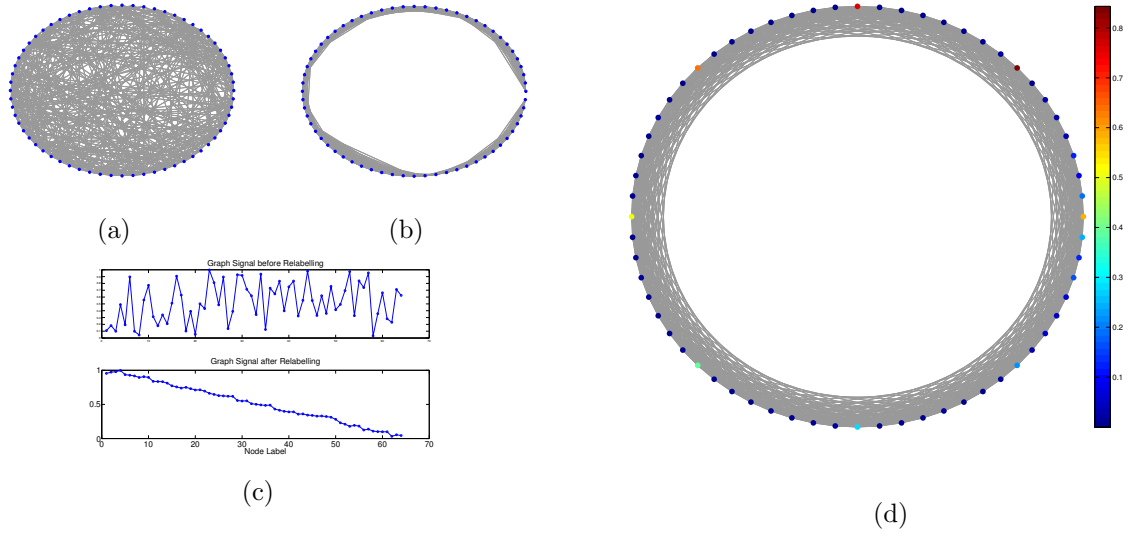


Figure 6.3: Original G ($N = 64$) after thresholding of weights (a), after RCM relabelling (b), signal \mathbf{x} before/after relabelling (c), multiscale $HGSWT$ representation (at $k = 1$) of \mathbf{x} (in magnitude) on \tilde{G} for 3 levels (d). ©2017 Elsevier Inc.

with σ as 10% of the total range of $d(a, b)$, and random graph signal \mathbf{x} with uniformly distributed entries $x_i \in (0, 1)$, along with the corresponding (sparse) graph approximation and multiscale representation via the $HGSWT$ (at $k = 1$, no reconnection), prepared using [52]. The obtained representation is highly sparse as a consequence of the breadth-first traversal of the RCM algorithm, where \mathbf{x}^P has reduced total variation $\|\mathbf{x}^P\|_{TV}$ and simultaneously \mathbf{W}^P (and by extension $\tilde{\mathbf{W}}$) is of small bandwidth. When a simple sort operation is applied instead, both functions are minimized.

6.2 Examples

We consider and compare the non-linear approximation performance for variations of the introduced algorithm with standard 2D wavelet transforms for a variety of artificial and natural images; here, we restrict our attention to the 2D Haar and linear spline (CDF 5/3 wavelet transform), with respectively 1 and 2 (dual) vanishing moments. While the employed graph constructions are all of first order with 2 vanishing moments, the 2D Haar is considered due to its smaller support and proven superiority for certain natural images.

Succeeding a partitioning of the same starting graph with weights as in Eq. (6.1), the following variations of the proposed algorithm are employed (where applicable), for each of which the final GWT(s) are constructed on circulant approximations of the specified graph type: (i) **nGWT**: the original complete weighted graph, (ii) **sparseGWT(bil, RCM)**: the graph is sparsified by Euclidean distance $\|p_i - p_j\|_2 \leq \sqrt{2}$ and subsequently relabelled

by the RCM, (iii) **sparseGWT(I, sort)**: the graph is re-weighted as intensity-only, sparsified by a data-dependent intensity threshold I , and reordered using signal sorting and (iv) **GWT(S=(1), sort)**: the graph is reduced to the smoothest simple cycle.

In certain cases, we consider multiple graph cuts and extend the described approaches to all segments, however, for simplicity, the chosen thresholds per GWT remain the same. The chosen GWT-type is the *HGSWT* ($k = 1$) of Thm. 3.1. It should be noted that while we consider the same number of levels for all transforms, the comparison is technically not favourable to the graph-based methods, since it can be demonstrated that the traditional tensor product wavelets can be represented as two levels of a 1D transform (see also Sect. 4.5 on graph products). In addition, all GWT rows have been normalized to 1. The performance is measured as $PSNR = 20 \log_{10} \left(\frac{I_{\max}}{\sqrt{MSE}} \right)$, with maximal intensity value I_{\max} and

$$MSE = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |I(i, j) - \tilde{I}(i, j)|^2$$

between original image \mathbf{I} and approximation $\tilde{\mathbf{I}}$; for $I(i, j) \in [0, 1]$, $\tilde{\mathbf{I}}$ is subject to post-processing by adjusting outliers $\tilde{I}(i, j) \notin [0, 1]$ to the original extrema.

We begin by comparing two different stripe patterns, a binary and a sinusoidal one. The application of the proposed scheme on the former produces constant sub-graph signals on two distinct graph partitions, which can be perfectly annihilated (irrespective of graph bandwidth), and hence ensures perfect reconstruction from a small number of basis elements, while the traditional 2D wavelet transforms struggle (Fig. 6.4). In the latter case, the two distinct regions are no longer constant, which hinders complete annihilation, yet invites further variations. As can be seen in Fig. 6.5, all proposed graph wavelets outperform the traditional, with the smooth cycle (iv) being the best by a margin; here the *sparseGWT(bil,RCM)* marks a slight improvement over the un-processed *nGWT*, while all intensity-only based transforms that employ the simple sort fare better in order of decreasing support.

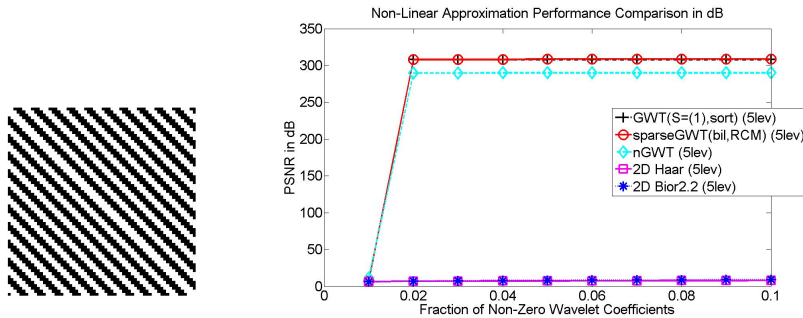


Figure 6.4: Comparison of NLA performance at 5 levels for a 64×64 image patch.

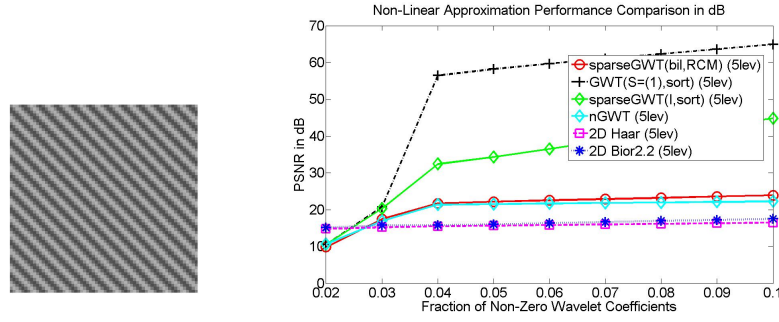


Figure 6.5: Comparison of NLA performance at 5 levels for a 64×64 image patch.

Figs. 6.6-6.7 respectively illustrate individual (multi-level) analysis low-pass basis elements of the euclidean-and intensity-thresholded transforms on the sinusoidal stripe image.

Further, we consider images which feature regions with respectively the same stripe patterns as well as homogeneous regions, which are created to be linear polynomial of the form $w_{i,j} = \alpha_1 x(i) + \alpha_2 y(j)$, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$ and $\alpha_1, \alpha_2 \in \mathbb{R}$. In both cases, we apply a total of two graph cuts and compare the NLA performance in Fig. 6.8. When the stripe pattern is homogeneous, the best performance is again achieved by the smoothest cycle, where performance is nearly identical for both one and two cuts, as a result of the minimum border effect. This is followed by the *sparseGWT(bil, RCM)*, while the *sparseGWT(I,sort)* fares less well, yet still outperforms traditional methods. The performance of the *nGWT* and *sparseGWT(I,sort)* is almost comparable since their annihilation potential is identical for the stripe region, with the homogeneous region constituting only a small gain for the latter due to it being non-constant. When the pattern features non-binary stripes, this trend is again reversed, and the performance of intensity-based GWTs superior, for both one and two cuts (see Fig. 6.9). Eventually, consider a real image patch extracted from the 256×256 ‘cameraman’, which features distinct regions with sharp discontinuities and details (see Figs. 6.10, 6.11). When one graph cut is applied, previous observations are upheld with the intensity-based variations outperforming the remaining, while all graph-based methods outperform the traditional tensor wavelets. The effect of the former is further exemplified beyond the simple cycle through the use of two different intensity thresholds, with the larger one achieving better results (*sparseGWT(I,sort)-1* in Fig. 6.10). When three graph cuts are applied, all GWTs gain in performance.

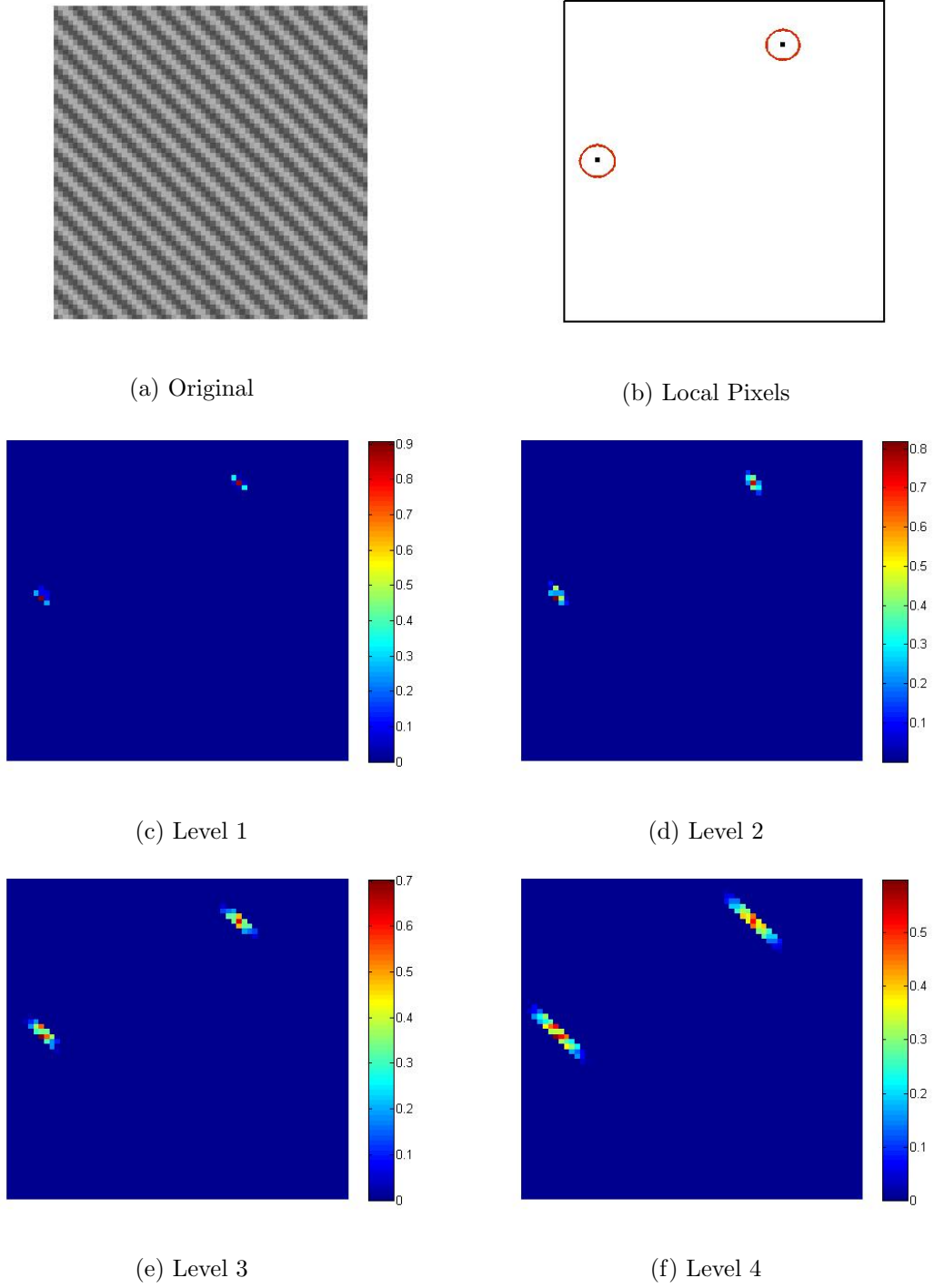


Figure 6.6: Localized Basis Functions of the $\text{sparseGWT}(\text{bil}, \text{RCM})$ depicted on selected (nodes) pixels (b) of the original image patch.

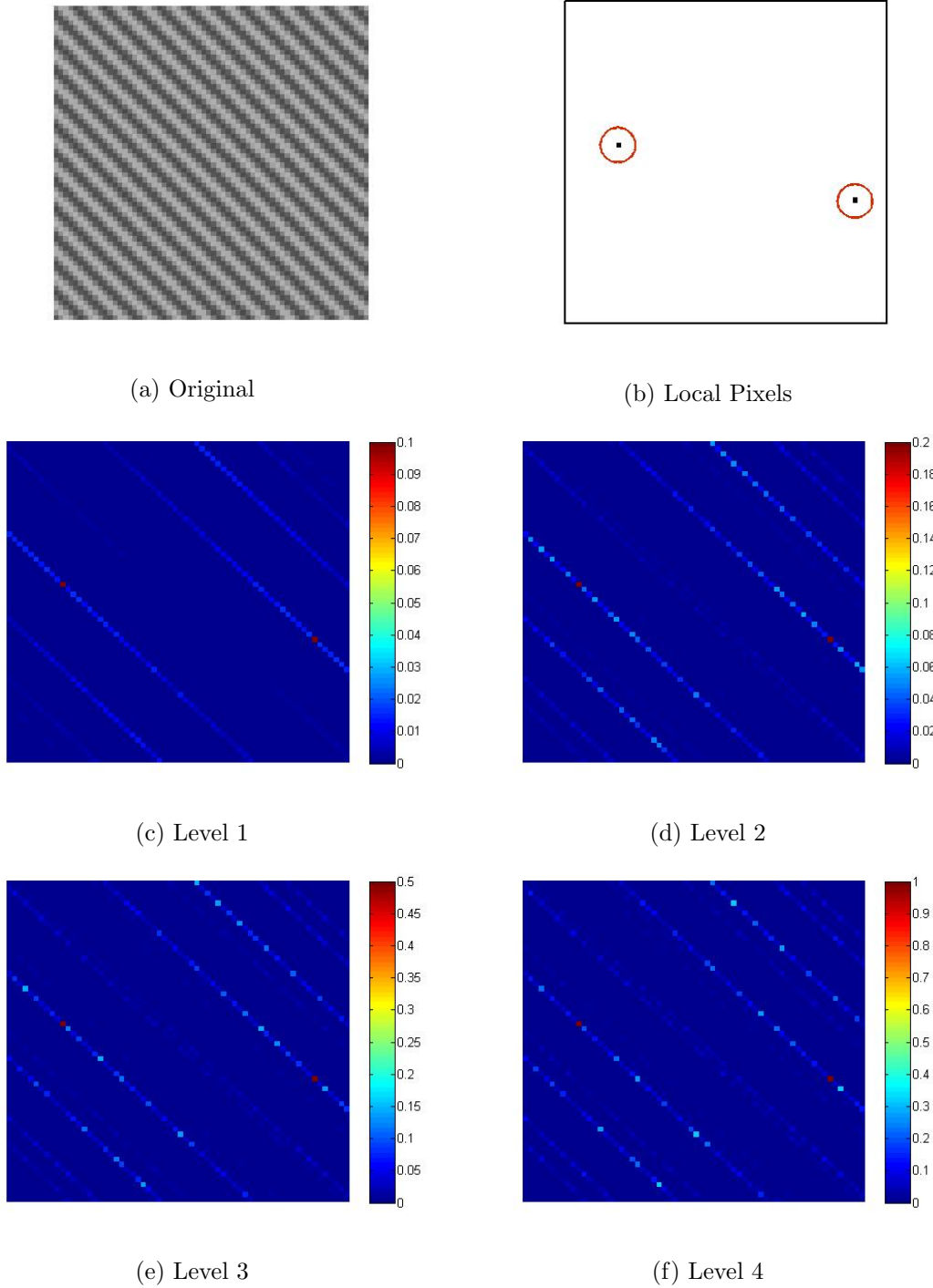
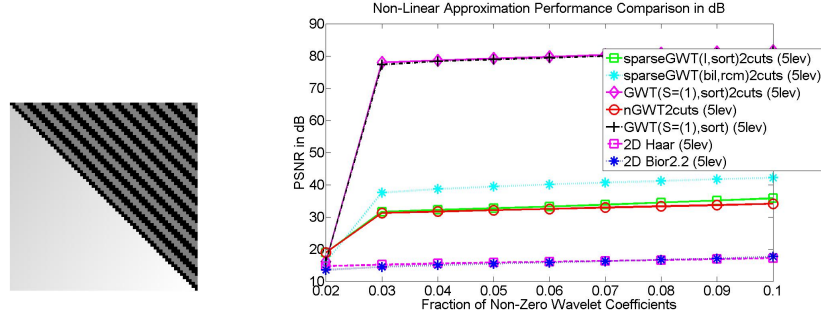
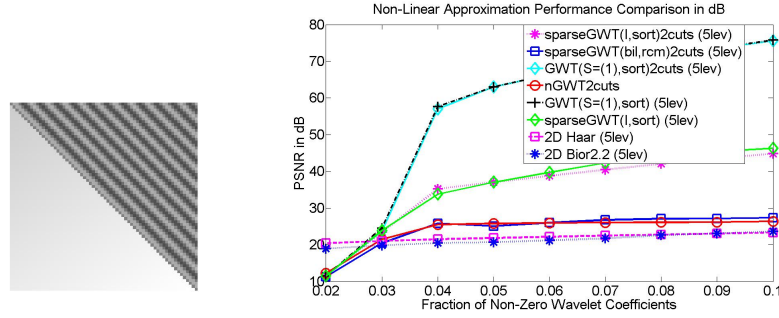


Figure 6.7: Localized Basis Functions of the $\text{sparseGWT}(I, \text{sort})$ depicted on selected (nodes) pixels (b) of the original image patch.


 Figure 6.8: Comparison of NLA performance at 5 levels for a 64×64 image patch.

 Figure 6.9: Comparison of NLA performance at 5 levels for a 64×64 image patch.

In summary, the $\text{sparseGWT}(\text{bil}, \text{RCM})$ provides a solid improvement over traditional methods as well as the simple execution of a complete weighted GWT, with spatially localized analysis basis functions, however, since it does not produce smooth sub-graph signals, it proves only near-optimal when multiple graph cuts are performed and/or the associated signals are (almost) constant. On the contrary, the $\text{sparseGWT}(\text{I}, \text{sort})$ exploits the relation between signal smoothness and bandwidth for a marginally better performance, with optimal results when it is reduced to the smoothest cycle, yet its basis functions are signal dependent and hence need to be individually tailored, while its parameters may require further adjustment to be effective for multiple graph cuts. Overall, it has transpired that both a basis with sparse support and a smooth labelling consistent with the graph structure facilitate superior performance, which, nevertheless, simultaneously shows that graphs are most effective when sparse.

While the preceding analysis has demonstrated the ability of graph wavelets, and its associated tools, to operate with respect to the inherent image geometry, resulting in highly sparse representations, one needs to emphasize that the discussed approaches are data-driven, i.e. their ultimate effectiveness depends on, and thus varies according to, the (image) data at hand. Furthermore, while obtained representations are sparse, and therefore of interest for applications such as image compression, one further needs to take into account the additional side information of the image graphs, such as the pixel-to-node mapping.

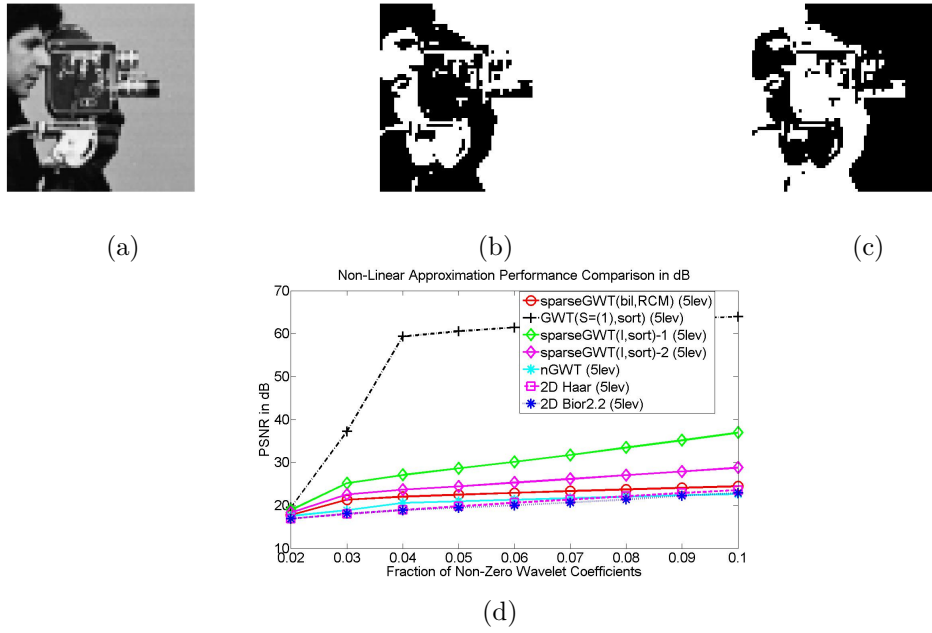


Figure 6.10: (a) Original 64×64 image patch, (b)-(c) Graph Cut Regions for 1 cut & (d) Comparison of NLA performance. ©2017 Elsevier Inc.

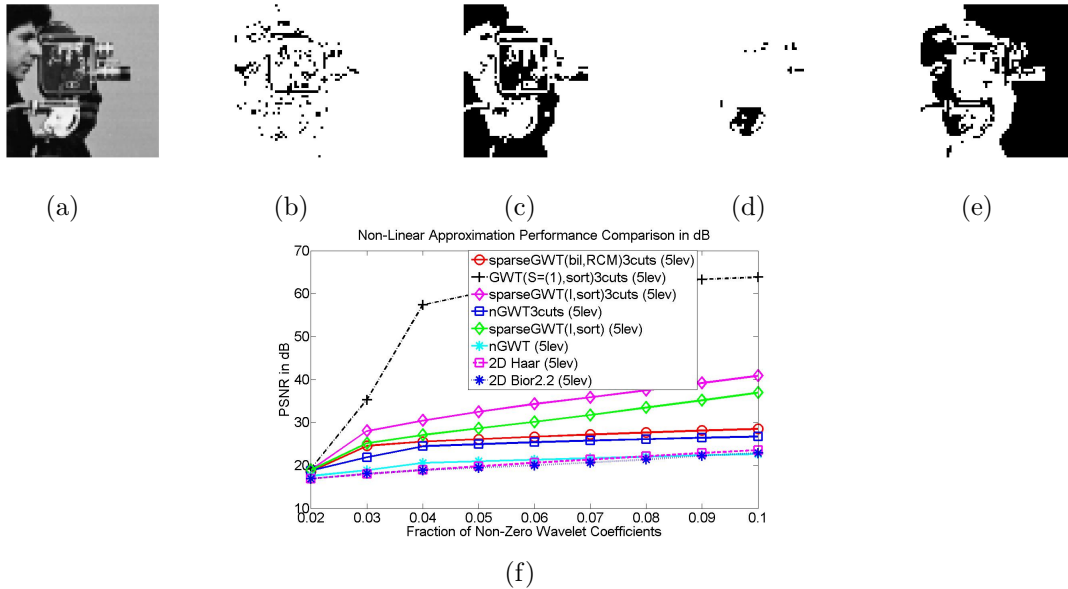


Figure 6.11: (a) Original 64×64 image patch, (b) – (e) Graph Cut Regions for 3 cuts & (f) Comparison of NLA performance.

6.3 Open Problem: Graph Labelling Under Noise

The underlying strengths of the proposed approach have been established as the partitioning and re-labelling steps. In the following, we briefly review other works that have operated under comparable notions, and discuss possible extensions.

The Easy Path Wavelet Transform (EPWT) for sparse image representation [137], while not a legitimate graph wavelet transform per se, exploits local data correlations in intensity images (and potentially higher-dimensional structures) by constructing a smooth path of data points on which a discrete 1-D wavelet transform can be subsequently applied for maximum sparsity; here individual path elements are chosen from within a spatially confined (grid-like) neighborhood, so that the absolute differences between their function values are minimized. In addition, the process can be repeated for multiple levels, however, each time a path is chosen from within an ‘upsampled’, smoothed version of the low-passed image. This approach is similar to the proposed image approximation scheme *sparse GWT(I, sort)* with the distinct difference that we operate directly in the graph domain, facilitating a graph cut and graph-based multiscale transform construction, as well as make use of the graph signal to determine the ideal path without spatial confinement.

When the signal at hand is uncorrupted, the optimal solution to the bandwidth-minimization problem is clearly given through a simple sorting of its sample values; nevertheless when samples are missing or contain noise, the traditional total-variation measure becomes unreliable as even small additions of noise can falsify the associated graph topology (or underlying geometry). The proposed framework still works fairly well under noise when the image patch features sharp contrasts, such as the example in Fig. 6.12, where denoising was performed by retaining a fixed percentage of highest magnitude wavelet coefficients (averaged over 10 trials of noise), and/or the noise level is relatively small, however, for general scenarios, one requires a more robust and, hence, less data-dependent labelling method.

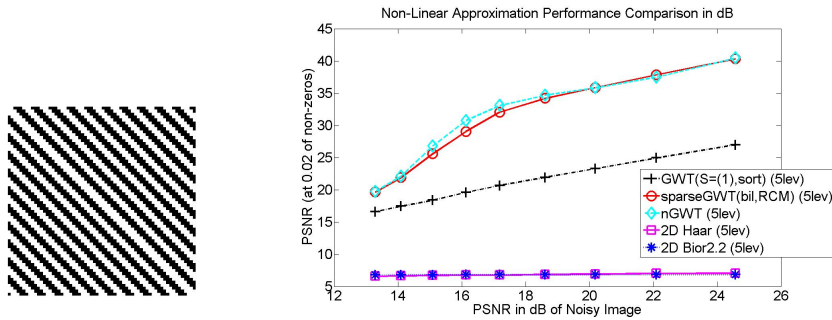


Figure 6.12: Denoising performance comparison for a 64×64 image patch.

Several works by Elad et al. that mainly targeted the problem of image processing, and

denoising in particular, have considered the issue of finding the best ‘noisy’ signal path through a vectorized version of the image in order to apply subsequent operations on an appropriate smooth permutation, which should be close to that of the clean version, and thus, robust to noise. A significant difference to our graph-based framework is the inclusion of high-dimensional feature vectors, such as patches, as opposed to pixels in the reordering scheme, while spatial localization of pixels (and their surrounding patches) takes on a more substantial role, as it circumvents the bias of an exclusively data-dependent labelling scheme by incorporating local geometric information.

In particular, in [138], [139], a tree-based wavelet transform for functions defined on high-dimensional data or graphs is proposed which employs reordering based on the distance measure between given data points within a hierarchical tree structure, and can be considered a generalization of the work in [137]. It is applicable to image denoising by using the proximity between feature vectors (images patches) as a predictor of similarity between the clean versions of its middle pixels. Further, in [140], similar graph-like notions are applied in image denoising, as the vector of corrupted/noisy image pixels is reordered according to the smoothest (or shortest) path through higher-dimensional data points (image patches) \mathbf{x}_j^p with total variation measure $x_{TV}^p = \sum_{j=2}^N w(\mathbf{x}_j^p, \mathbf{x}_{j-1}^p)$ for distance measure $w(\cdot, \cdot)$, in the spirit of the travelling salesman problem. The latter can be approximately minimized by conducting spatially confined searches within image regions and choosing the nearest neighbor (patch) according to a probability proportional to its degree of similarity. Subsequent smoothing and filtering operations are then applied to the reordered 1D-signal before converting it back to denoised image form.

Following similar notions, the formerly introduced EPWT has been further extended for image denoising by creating more robust and adaptive path selection methods which i. a. consider the proximity between associated feature vectors, within confined neighborhoods, in addition to similarity between function values in both deterministic and probabilistic schemes [141].

In order to demonstrate directly the significance of the underlying labelling (or ideal path selection) for a graph signal, consider a noisy version of the cameraman-image patch and compare an optimally labelled, ‘segmentation-based’ graph wavelet transform on the simple cycle to the BM3D [142], as a state of the art image denoising algorithm. The underlying graph labelling of the former is obtained from a simple sorting of the clean sub-graph signals, following one graph cut, and GWT values at and below 3σ in magnitude are thresholded, for standard deviation σ of the noise (see Fig. 6.13). It can be seen that the smoothest simple cycle with the oracle labelling largely outperforms the BM3D at the indicated noise range, which exemplifies the promise behind recovering a (near-)optimal labelling and provides an intriguing avenue for extensions of the presented framework.

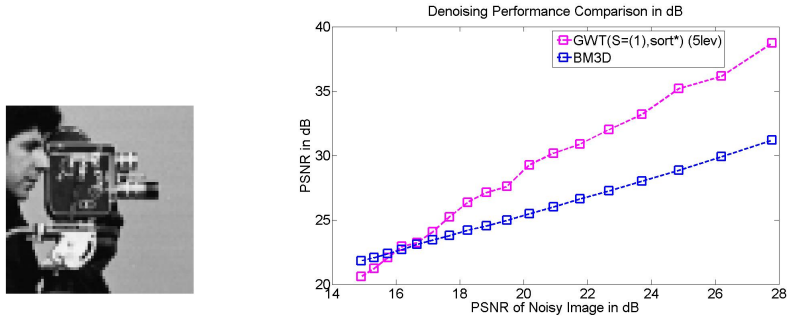


Figure 6.13: Denoising performance comparison between BM3D and simple cycle GWT with ideal labelling on the 64×64 cameraman image patch.

While the use of graphs in image processing has facilitated data-specific operations by capturing and exploiting more (geometric) information, it is nevertheless also more susceptible to instability as a result of noise in the data. Several works, both directly and indirectly affiliated with GSP, have attempted to remedy that by i. a. associating entire feature vectors to each node to incorporate information beyond the pixel intensity, however, methods to date are largely heuristic, with the problem of finding the optimum labelling under noise remaining an open issue.

Chapter 7

Conclusion

7.1 Summary

This thesis has provided a sparsity-driven tour of graph signal processing with particular focus on circulant graphs and their implications for both generalized graph scenarios and traditional signal processing concepts. In an effort to elucidate how sparse representations on graphs can be both induced and leveraged as well as comprehend the links between the emerging GSP theory and traditional signal processing theory, we have conducted a theoretical analysis of sparsity on graphs by delving into the (inter-related) topics of wavelet and sampling theory on graphs.

Hereby, we have discovered and leveraged the underlying polynomial and annihilation properties of circulant graph Laplacian matrices (and their generalizations) for the development of graph spline wavelet theory, in the style of its traditional counterpart, while exploiting the circulant nature of traditional signal processing operations to solidify the connection between the Euclidean and graph domain. Having explored both a range of sparsifying wavelet transforms as well as classes of (piecewise) smooth signals on circulant graphs and beyond, along with their properties and limitations, we have further shifted the focus to sampling of (wavelet-)sparse signals within the GFRI framework, inspired by Finite Rate of Innovation theory, which facilitates their perfect recovery from dimensionality-reduced spectral representations, while simultaneously identifying an associated coarsened graph. Eventually, we have devised a graph-based image approximation algorithm, which employs graph partitioning with subsequent sparsity-driven graph wavelet analysis as a superior method, compared with traditional tensor-product bases, due to its (conceptually simple) potential to operate with respect to the inherent geometry of images with patterns.

In order to motivate both the problem context as well as its mathematical frame, in Chapter 2, we reviewed notable contributions and insights from both traditional sparse signal processing and graph signal processing, along with relevant definitions and concepts

from graph and matrix theory, and with particular regard to wavelet and sampling theory. We further presented the theory of circulant matrices and graphs, and, in particular, its inherently polynomial characteristics.

In Chapter 3, we presented several contributions in form of special operators and wavelet transforms on circulant graphs with inherent and/or imposed desirable properties, while introducing the notion of a spline on a graph and its (structural) relation to the traditional spline. Specifically, we leveraged the vanishing moment property of circulant graph Laplacian matrices to derive the graph spline and e-spline wavelet transforms, along with a range of complementary constructions based on spectral factorization, which exhibit reproduction and/or annihilation properties for classes of (exponential) polynomial signals, and further discussed their significance and special cases.

Chapter 4 commenced with a contextualization of previous derivations for circulants, followed by the introduction of a variety of sparsity-inducing graph wavelet transforms for (piecewise) smooth graph signals and their properties on arbitrary undirected graphs, including the generalized bandlimiting, space-variant and time-varying graph wavelet transforms, as well as multi-dimensional extensions on the basis of graph products. In the course of this analysis, further properties such as the transform condition number and degree of annihilation were compared.

Chapter 5 introduced a novel sampling and graph coarsening framework for (wavelet-)sparse signals on circulant graphs and beyond, with an excursion into the problem of sampling under noise, featuring a perturbation analysis. Inspired by Finite Rate of Innovation (FRI) sampling framework, we devised its counterpart and extension in the graph domain, which facilitates the perfect reconstruction of K -sparse graph signals on circulant graphs from its dimensionality-reduced GFT vector of minimum size $2K$, while simultaneously inferring an associated coarse graph on the basis of the previously derived graph (e-)spline wavelet theory. Extensions to arbitrary graphs can be enforced via suitable (graph) approximation schemes, while an alternative approach of sampling under connectivity noise which employs perturbation models was also explored.

At last, Chapter 6 presented an application to image processing in form of a sparsity-driven framework for wavelet analysis, paired with novel insights on the impact of graph labelling, and concluding with the open problem of denoising on graphs. By modelling a given image with distinct discontinuities and patterns as a bilaterally weighted graph, the proposed algorithm conducts an initial partition into homogeneous regions before applying sparsifying graph wavelet transforms on circulant approximations of the resulting subgraphs. Maximum sparsity can be achieved by transform-variations which are localized with respect to the image-intensity content and further exploit a smooth node-relabelling, which is shown to be related to the optimal band-form of the graph. The effectiveness of the algorithm and superiority over traditional tensor-product wavelets was demonstrated

on the basis of real and artificial image examples.

Overall, the collective of derivations and results has elucidated how the connectivity of circulant and suitable general graphs can be leveraged to induce a sparse graph signal representation of certain classes of signals as well as how sparsity on graphs can be further appropriated for dimensionality reduction (or sampling), while providing a range of examples and ultimately an application in image processing.

At the same time, the analysis has uncovered certain challenges and ensuing new problems, which need to be tackled in further investigations.

7.2 Open Problems and Future Work

Chapters 3 and 4 have focused on the identification of transforms and bases with predominant regard to inducing a sparse graph signal representation when no noise or corruption is present in the signal. Nevertheless, preliminary experiments and a closer investigation of the transform condition number revealed that certain developed constructions may become susceptible to noise. Therefore, it would be of interest to further investigate how these may be amended to accommodate noise as well as explore alternatives. Due to the complex connectivity of graphs and present variability, it nevertheless remains an open issue to identify the ideal noise-robust basis (or a set of necessary properties or constraints), beyond an individualized and/or application-dependent study.

In order to leverage the desirable properties of circulant graphs, or more generally, conveniently structured graphs, it is usually necessary to resort to suitable graph approximation schemes. As noted in Chapters 5 and 6, in the current analysis, this has required the use of the nearest circulant approximation or (circulant) Kronecker product approximation schemes. In particular, while for matrices the minimization of an error norm may be a conventional approximation step and measure, in the case of a graph this entails the addition of edges (with possibly negative weights) which may change the inherent graph properties, and which raises the question of how suitable such a scheme eventually is from a theoretical viewpoint. While a prevailing number of GSP works is resorting to matrix-theory driven approximation or optimization approaches and results, relatively few are concerned with exploring the degree of graph-invasiveness or -alteration of such processes as well as how this may be measured and/or circumvented.

The presence of noise on networks and/or network data is a notable open issue in GSP, and Chapter 6 conclusively addressed the challenges of transferring the developed image approximation framework to the noisy setting. Current limitations are imposed by the fact that one of its driving components, the ideal node labelling of the graph for maximum sparsity is selected based on a clean version of the graph signal (image); the addition of

noise however renders the scheme unstable. We have discovered that if given access to the ideal ordering for the clean signal, imposed on the noisy case, our current scheme can outperform state-of-the-art algorithms such as the BM3D by a high margin. It is therefore of interest to further explore the issue of an approximate smooth labelling under noise.

Appendix A

Proofs of Chapter 3

A.1 Proof of Theorem 3.1

Theorem 3.1. *Given the undirected, and connected circulant graph $G = (V, E)$ of dimension N , with adjacency matrix \mathbf{A} and degree d per node, define the higher-order graph-spline wavelet transform (HGSWT), composed of the low-and high-pass filters*

$$\mathbf{H}_{LP} = \frac{1}{2^k} \left(\mathbf{I}_N + \frac{\mathbf{A}}{d} \right)^k$$

$$\mathbf{H}_{HP} = \frac{1}{2^k} \left(\mathbf{I}_N - \frac{\mathbf{A}}{d} \right)^k$$

whose associated high-pass representer polynomial $H_{HP}(z)$ has $2k$ vanishing moments. This filterbank is invertible for any downsampling pattern, as long as at least one node retains the low-pass component, while the complementary set of nodes retains the high-pass components.

Proof of Theorem 3.1. (appears in [9] and [7]) It is self-evident that since the high-pass filter of Eq. (3.2) is a power k of the graph Laplacian matrix, whose associated polynomial representation has 2 vanishing moments, the annihilation property is generalized to higher order of $2k$ vanishing moments; thus we proceed to demonstrate invertibility of the filterbank. The core of the proof follows a similar line of argumentation as the one provided in [55] for $k = 1$ with generalizations pertaining to the parameter k . For completeness we present the entire proof here.

Applying the binomial theorem, we observe that

$$\frac{1}{2^k} \left(\mathbf{I}_N \pm \frac{\mathbf{A}}{d} \right)^k = \frac{1}{2^k} \sum_{j=0}^k (\pm 1)^j \binom{k}{j} \left(\frac{\mathbf{A}}{d} \right)^j$$

and so we need to show that the nullspace of

$$\frac{1}{2^k} \left(\sum_{j \in \mathbb{Z}} \binom{k}{2j} \left(\frac{\mathbf{A}}{d} \right)^{2j} + \mathbf{K} \sum_{j \in \mathbb{Z}} \binom{k}{2j+1} \left(\frac{\mathbf{A}}{d} \right)^{2j+1} \right)$$

is empty, where \mathbf{K} is the diagonal matrix with entries $K(i, i) = 1$ if node i retains the low- and $K(i, i) = -1$ if it retains the high-pass component. Assume the contrary and define the vector $\mathbf{z} = \mathbf{V}\mathbf{r}$ to lie in the nullspace, given eigendecomposition $\frac{\mathbf{A}}{d} = \mathbf{V}\mathbf{\Gamma}\mathbf{V}^H$, which yields the following simplifications:

$$\frac{1}{2^k} \left(\sum_{j \in \mathbb{Z}} \binom{k}{2j} \left(\frac{\mathbf{A}}{d} \right)^{2j} + \mathbf{K} \sum_{j \in \mathbb{Z}} \binom{k}{2j+1} \left(\frac{\mathbf{A}}{d} \right)^{2j+1} \right) \mathbf{V}\mathbf{r} = \mathbf{0}_N \quad (\text{A.1})$$

$$\Leftrightarrow \|\mathbf{V} \sum_{j \in \mathbb{Z}} \binom{k}{2j} \mathbf{\Gamma}^{2j} \mathbf{r}\|_2^2 = \|\mathbf{KV} \sum_{j \in \mathbb{Z}} \binom{k}{2j+1} \mathbf{\Gamma}^{2j+1} \mathbf{r}\|_2^2 \quad (\text{A.2})$$

where Eq. (A.2) is the result of a rearrangement of terms in Eq. (A.1) and subsequent application of the l_2 -vector norm on both sides of the equality. After further simplification, we obtain

$$\sum_{i=0}^{N-1} r(i)^2 \left(\left(\sum_{j \in \mathbb{Z}} \binom{k}{2j} \gamma_i^{2j} \right)^2 - \left(\sum_{j \in \mathbb{Z}} \binom{k}{2j+1} \gamma_i^{2j+1} \right)^2 \right) \stackrel{(a)}{=} \sum_{i=0}^{N-1} r(i)^2 (A_i^2 - B_i^2) = 0, \quad (\text{A.3})$$

where in (a), we let A_i and B_i represent the sum of even and odd terms in the binomial series respectively. For the nullspace to be empty, we need to show that $\mathbf{r} = \mathbf{0}_N$, which follows if $(A_i^2 - B_i^2) \neq 0$ and is strictly positive or negative $\forall i$. By utilizing the fact that for a general binomial series $(x + a)^n$, with terms A_i and B_i , the following holds: $(x^2 - a^2)^n = A_i^2 - B_i^2$, we obtain

$$\sum_{i=0}^{N-1} r(i)^2 (A_i^2 - B_i^2) = \sum_{i=0}^{N-1} r(i)^2 (1 - \gamma_i^2)^k = 0.$$

The eigenvalues are given by $|\gamma_i| \leq 1$, via the Gershgorin circle theorem, where $\gamma_i > -1$ unless the graph is bipartite [3]; thus we have that $|r(i)| > 0$ only if $|\gamma_i| = 1$ and $r(i) = 0$ otherwise. To examine these special cases, let the corresponding eigenvectors for $\gamma_1 = -1$ and $\gamma_2 = 1$ be given by $\tilde{\mathbf{V}}$ and $\frac{r(0)}{\sqrt{N}} \mathbf{1}_N$ respectively, such that $\mathbf{z} = \frac{r(0)}{\sqrt{N}} \mathbf{1}_N + \tilde{\mathbf{V}}\tilde{\mathbf{r}}$, and substitute into Eq. (A.1). Here, the multiplicity of $\gamma_2 = 1$ is one, since the graph is connected [3]. We consider the case of a non-bipartite graph first:

$$\frac{r(0)}{\sqrt{N}} \left(\sum_{j \in \mathbb{Z}} \binom{k}{2j} \mathbf{1}_N + \mathbf{K} \sum_{j \in \mathbb{Z}} \binom{k}{2j+1} \mathbf{1}_N \right) = \mathbf{0}_N$$

Noting that $\sum_{j \in \mathbb{Z}} \binom{k}{2j} = \sum_{j \in \mathbb{Z}} \binom{k}{2j+1}$, we need at least one entry $K(i, i) = 1$, such that $r(0) = 0$.

In the bipartite case, due to spectral folding, if γ is an eigenvalue of \mathbf{A} with eigenvector $\begin{bmatrix} \mathbf{v}_B \\ \mathbf{v}_{B^c} \end{bmatrix}$, so is $-\gamma$ with eigenvector $\begin{bmatrix} \mathbf{v}_B \\ -\mathbf{v}_{B^c} \end{bmatrix}$, where B is the set of the node indices in one bipartite set [3], and partitions \mathbf{v}_B and \mathbf{v}_{B^c} interlace. Then $\gamma_1 = 1$ and $\gamma_2 = -1$ each have multiplicity one with respective eigenvectors $\mathbf{1}_N$ and $\begin{bmatrix} \mathbf{1}_B \\ -\mathbf{1}_{B^c} \end{bmatrix}$, where $|B| = |B^c| = N/2$, giving

$$\begin{aligned} & \frac{r(0)}{\sqrt{N}} \underbrace{\left(\sum_{j \in \mathbb{Z}} \binom{k}{2j} \mathbf{I}_N + \mathbf{K} \sum_{j \in \mathbb{Z}} \binom{k}{2j+1} \mathbf{I}_N \right)}_{T_1} \mathbf{1}_N \\ & + \frac{r(1)}{\sqrt{N}} \underbrace{\left(\sum_{j \in \mathbb{Z}} \binom{k}{2j} \mathbf{I}_N - \mathbf{K} \sum_{j \in \mathbb{Z}} \binom{k}{2j+1} \mathbf{I}_N \right)}_{T_2} \begin{bmatrix} \mathbf{1}_B \\ -\mathbf{1}_{B^c} \end{bmatrix} = \mathbf{0}_N. \end{aligned} \quad (\text{A.4})$$

Here we have used the property $\frac{\mathbf{A}}{d} \mathbf{v} = \gamma \mathbf{v}$, in $(\frac{\mathbf{A}}{d})^j \begin{bmatrix} \mathbf{1}_B \\ -\mathbf{1}_{B^c} \end{bmatrix} = (-1)^j \begin{bmatrix} \mathbf{1}_B \\ -\mathbf{1}_{B^c} \end{bmatrix}$, leading to an alternating pattern on the RHS when j is odd.

In particular, for any choice of downsampling pattern \mathbf{K} , the terms T_1 and T_2 in the first and second summands in Eq. (A.4), will respectively have zero entries along the main diagonal, which lie in complementary index sets. Therefore, as long as at least one node retains the low-pass component $K(i, i) = 1$, it follows that $r(0) = 0$ and $r(1) = 0$, which again implies $\mathbf{z} = \mathbf{0}$, completing the proof. \square

A.2 Proof of Theorem 3.2

Theorem 3.2. *The higher-order graph e-spline wavelet transform (HGESWT) on a connected, undirected circulant graph G , is composed of the low-and high-pass filters*

$$\begin{aligned} \mathbf{H}_{LP_{\vec{\alpha}}} &= \prod_{n=1}^T \frac{1}{2^k} \left(\beta_n \mathbf{I}_N + \frac{\mathbf{A}}{d} \right)^k \\ \mathbf{H}_{HP_{\vec{\alpha}}} &= \prod_{n=1}^T \frac{1}{2^k} \left(\beta_n \mathbf{I}_N - \frac{\mathbf{A}}{d} \right)^k \end{aligned}$$

where \mathbf{A} is the adjacency matrix, d the degree per node and parameter β_n is given by $\beta_n = \frac{\tilde{d}_{\alpha_n}}{d}$ with $\tilde{d}_{\alpha_n} = \sum_{j=1}^M 2d_j \cos(\alpha_n j)$ and $\vec{\alpha} = (\alpha_1, \dots, \alpha_T)$. Then the high-pass filter annihilates complex exponential polynomials (of $\deg(p(t)) \leq k-1$) with exponent $\pm i\alpha_n$ for $n = 1, \dots, T$. The transform is invertible for any downsampling pattern as long as the

eigenvalues γ_i of $\frac{\mathbf{A}}{d}$ satisfy $|\beta_n| \neq |\gamma_i|$, $i = 0, \dots, N-1$ for $n = 1, \dots, T$, under either of the sufficient conditions

(i) $k \in 2\mathbb{N}$, or

(ii) $k \in \mathbb{N}$ and β_n, T such that $\forall \gamma_i, f(\gamma_i) = \prod_{n=1}^T (\beta_n^2 - \gamma_i^2)^k > 0$ or $f(\gamma_i) < 0$.

If parameters β_n , are such that $\beta_n = \gamma_i$, for up to T distinct values, the filterbank continues to be invertible under the above as long as $\beta_n \neq 0$ and at least $\sum_{i=1}^T m_i$ low-pass components are retained at nodes in set V_α such that $\{\mathbf{v}_{+i,k}(V_\alpha)\}_{i=1,k=1}^{i=T,k=m_i}$ (and, if eigenvalue $-\gamma_i$ exists, complement $\{\mathbf{v}_{-i,k}(V_\alpha^c)\}_{i=1,k=1}^{i=T,k=m_i}$) form linearly independent sets, where m_i is the multiplicity of γ_i and $\{\mathbf{v}_{\pm i,k}\}_{k=1}^{m_i}$ are the eigenvectors respectively associated with $\pm \gamma_i$.

Proof of Theorem 3.2. (appears in [7]) We can rewrite the simplified filters

$$\mathbf{H}_{LP_\alpha} = \prod_{n=1}^{Tk} \left(\beta_n \mathbf{I}_N + \frac{\mathbf{A}}{d} \right) \quad (\text{A.5})$$

$$\mathbf{H}_{HP_\alpha} = \prod_{n=1}^{Tk} \left(\beta_n \mathbf{I}_N - \frac{\mathbf{A}}{d} \right) \quad (\text{A.6})$$

noting that the new indices incorporate multiplicities, as follows

$$\begin{aligned} \mathbf{H}_{LP_\alpha} &= \sum_{i=0}^{Tk} s_i \left(\frac{\mathbf{A}}{d} \right)^i \\ \mathbf{H}_{HP_\alpha} &= (-1)^{Tk} \sum_{i=0}^{Tk} (-1)^{i+Tk} s_i \left(\frac{\mathbf{A}}{d} \right)^i = \sum_{i=0}^{Tk} (-1)^i s_i \left(\frac{\mathbf{A}}{d} \right)^i \end{aligned}$$

where the coefficients s_i are the elementary symmetric polynomials $e_n(\beta_1, \dots, \beta_{Tk})$ in β_n :

$$s_0 = e_{Tk}(\beta_1, \dots, \beta_{Tk}) = \beta_1 \beta_2 \dots \beta_{Tk}$$

$$s_k = \dots$$

$$s_{Tk-1} = e_1(\beta_1, \dots, \beta_{Tk}) = \beta_1 + \beta_2 + \dots + \beta_{Tk}$$

$$s_{Tk} = e_0(\beta_1, \dots, \beta_{Tk}) = 1.$$

We need to prove that the filterbank

$$\sum_{j \in \mathbb{Z}} s_{2j} \left(\frac{\mathbf{A}}{d} \right)^{2j} + \mathbf{K} \sum_{j \in \mathbb{Z}} s_{2j+1} \left(\frac{\mathbf{A}}{d} \right)^{2j+1}$$

with diagonal downsampling matrix \mathbf{K} is invertible by showing that its nullspace is empty. Similarly, as in A.1, we assume the contrary and let $\mathbf{z} = \mathbf{V}\mathbf{r}$ lie in its nullspace, where

$\mathbf{V}\mathbf{\Gamma}^j\mathbf{V}^H = \left(\frac{\mathbf{A}}{d}\right)^j$, such that

$$\left(\sum_{j \in \mathbb{Z}} s_{2j} \left(\frac{\mathbf{A}}{d}\right)^{2j} + \mathbf{K} \sum_{j \in \mathbb{Z}} s_{2j+1} \left(\frac{\mathbf{A}}{d}\right)^{2j+1}\right) \mathbf{V}\mathbf{r} = \mathbf{0}_N \quad (\text{A.7})$$

$$\Leftrightarrow \|\mathbf{V} \sum_{j \in \mathbb{Z}} s_{2j} \mathbf{\Gamma}^{2j} \mathbf{r}\|_2^2 = \|\mathbf{K} \mathbf{V} \sum_{j \in \mathbb{Z}} s_{2j+1} \mathbf{\Gamma}^{2j+1} \mathbf{r}\|_2^2 \quad (\text{A.8})$$

where Eq. (A.8) results from rearranging Eq. (A.7) and taking norms of both sides, and gives rise to

$$\begin{aligned} \sum_{i=0}^{N-1} r(i)^2 \left(\left(\sum_{j \in \mathbb{Z}} s_{2j} \gamma_i^{2j} \right)^2 - \left(\sum_{j \in \mathbb{Z}} s_{2j+1} \gamma_i^{2j+1} \right)^2 \right) &= 0, \text{ and hence} \\ \sum_{i=0}^{N-1} r(i)^2 \prod_{n=1}^{Tk} (\beta_n - \gamma_i) \prod_{n=1}^{Tk} (\beta_n + \gamma_i) &= \sum_{i=0}^{N-1} r(i)^2 \prod_{n=1}^T (\beta_n^2 - \gamma_i^2)^k = 0. \end{aligned} \quad (\text{A.9})$$

Given parameters β_n such that $|\beta_n| \leq 1$, $n = 1, \dots, T$, and with eigenvalues satisfying $|\gamma_i| \leq 1$, $i = 0, \dots, N-1$ by the Perron-Frobenius Theorem [93], we assume $|\beta_n| \neq |\gamma_i|$. Thus, all summands in Eq. (A.9) need to be of the same sign to guarantee $\mathbf{r} = \mathbf{0}_N$. As the function $f(\gamma_i) = \prod_{n=1}^T (\beta_n^2 - \gamma_i^2)^k$, for spectrum $\gamma = \{\gamma_i\}_{i=0}^{N-1}$, does not have exclusively positive or negative range for odd k , we require $k \in 2\mathbb{N}$. Furthermore, all terms remain of the same sign at any k as long as parameters β_n and T are suitably chosen. This is a sufficient condition for guaranteeing invertibility at any downsampling pattern.

If, for some n , we have $|\beta_n| = |\gamma_i|$ with $i \in [0, N-1]$, giving $|r(i)| \geq 0$, we can show that for certain downsampling patterns, the transform continues to be invertible. In particular, this is the case when parameter α in $\beta_n = \frac{\tilde{d}_\alpha}{d}$ is such that $\alpha = \frac{2\pi k}{N}$ for some $k \in [0, N-1]$, i.e. \mathbf{H}_{HP_α} annihilates the k -th (eigen-)vector in the DFT matrix. For eigenvalues $\lambda_k = \sum_{j=1}^M 2d_j \cos\left(\frac{2\pi jk}{N}\right)$ of the non-normalized symmetric and circulant adjacency matrix \mathbf{A} with first row $[0 \ d_1 \ d_2 \ \dots \ d_2 \ d_1]$, we thus have $\tilde{d}_\alpha = \lambda_k$ for some $k \in [0, N-1]$. We proceed to show that the filterbank at hand is invertible for such α as long as downsampling is conducted with respect to $s = 1 \in S$, and more generally, when at least m_i suitably chosen low-pass components are retained (for multiplicity m_i of γ_i), and by extension, $\sum_{i=1}^P m_i$ components for P distinct β_n that satisfy $|\beta_n| = |\gamma_i|$.

Assuming wlog $\beta_n = \gamma_i$, for P distinct eigenvalues ($1 \leq P < N$), each of multiplicity m_i with corresponding eigenvector(s) $\{\mathbf{v}_{i,l}\}_{l=0}^{m_i-1}$, we consider, for the case of a non-bipartite

graph, the following nullspace representation

$$\mathbf{z} = \sum_{n=1}^P \sum_{l=0}^{m_n-1} r_{n,l} \mathbf{v}_{n,l}, \quad \forall n, m_n \geq 1,$$

where index n signifies distinct eigenvalues (or β_n) and $r_{n,l}$ are scalar coefficients. With $\left(\frac{\mathbf{A}}{d}\right)^k \mathbf{v}_i = \gamma_i^k \mathbf{v}_i$, we obtain via substitution into Eq. (A.7)

$$\sum_{j \in \mathbb{Z}} s_{2j} \sum_{n=1}^P \sum_{l=0}^{m_n-1} r_{n,l} \beta_n^{2j} \mathbf{v}_{n,l} + \mathbf{K} \sum_{j \in \mathbb{Z}} s_{2j+1} \sum_{n=1}^P \sum_{l=0}^{m_n-1} r_{n,l} \beta_n^{2j+1} \mathbf{v}_{n,l} = \mathbf{0}_N.$$

If $K_{i,i} = 1$, $i = 0, \dots, N-1$, then

$$\sum_{n=1}^P \sum_{l=0}^{m_n-1} \prod_{q=1}^T (\beta_q + \beta_n)^k r_{n,l} \mathbf{v}_{n,l} = \mathbf{0}_N$$

and if $\forall i, K_{i,i} = -1$, $i = 0, \dots, N-1$, then

$$\sum_{n=1}^P \sum_{l=0}^{m_n-1} \prod_{q=1}^T (\beta_q - \beta_n)^k r_{n,l} \mathbf{v}_{n,l} = \mathbf{0}_N$$

where in the case of the latter we observe that $\prod_{q=1}^T (\beta_q - \beta_n)^k$ is always zero since $\beta_q = \beta_n$ for $q = n$. Therefore, we need the number of low-pass components to be greater than or equal to the sum of multiplicities m_n for all P distinct eigenvalues and, in addition, their (node) locations D need to be suitably chosen to facilitate linearly independent partitions $\mathbf{v}_{n,l}(D)$ such that $r_{n,l} = 0$. Note that for $\beta_n = -\gamma_i$ and $-\gamma_i \notin \gamma$, the opposite is the case, i.e. we need at least $\sum_{n=1}^P m_n$ suitably chosen high-pass components to facilitate linear independence of partitions $\{\mathbf{v}_{n,l}(D^c)\}_{n,l}$. If both $\pm\gamma_i$ exist, similar reasoning as for the following bipartite case applies.

For bipartite graphs, let m_i denote the multiplicity of eigenvalue γ_i and $-\gamma_i$ respectively (due to symmetry of the spectrum [3]), and $\{\mathbf{v}_{i,l}\}_{l=0}^{m_i-1}$ and $\left\{ \begin{bmatrix} \mathbf{v}_{i,l}^B \\ -\mathbf{v}_{i,l}^{B^c} \end{bmatrix} \right\}_{l=0}^{m_i-1}$ the corresponding eigenvectors, resulting in a nullspace representation of the form:

$$\mathbf{z} = \sum_{n=1}^P \sum_{l=0}^{m_n-1} \left(r_{n,l} \mathbf{v}_{n,l} + \tilde{r}_{n,l} \begin{bmatrix} \mathbf{v}_{n,l}^B \\ -\mathbf{v}_{n,l}^{B^c} \end{bmatrix} \right) \quad m_n \geq 1,$$

whose substitution into Eq. (A.7) yields

$$\sum_{j \in \mathbb{Z}} s_{2j} \sum_{n=1}^P \sum_{l=0}^{m_n-1} r_{n,l} \beta_n^{2j} \mathbf{v}_{n,l} + \mathbf{K} \sum_{j \in \mathbb{Z}} s_{2j+1} \sum_{n=1}^P \sum_{l=0}^{m_n-1} r_{n,l} \beta_n^{2j+1} \mathbf{v}_{n,l}$$

$$+ \sum_{j \in \mathbb{Z}} s_{2j} \sum_{n=1}^P \sum_{l=0}^{m_n-1} \tilde{r}_{n,l} \beta_n^{2j} \begin{bmatrix} \mathbf{v}_{n,l}^B \\ -\mathbf{v}_{n,l}^{B^c} \end{bmatrix} - \mathbf{K} \sum_{j \in \mathbb{Z}} s_{2j+1} \sum_{n=1}^P \sum_{l=0}^{m_n-1} \tilde{r}_{n,l} \beta_n^{2j+1} \begin{bmatrix} \mathbf{v}_{n,l}^B \\ -\mathbf{v}_{n,l}^{B^c} \end{bmatrix} = \mathbf{0}_N.$$

Thus, for $K_{i,i} = 1$, $i = 0, \dots, N-1$, we obtain

$$\sum_{n=1}^P \sum_{l=0}^{m_n-1} \prod_{q=1}^T (\beta_q + \beta_n)^k r_{n,l} \mathbf{v}_{n,l} + \sum_{n=1}^P \sum_{l=0}^{m_n-1} \prod_{q=1}^T (\beta_q - \beta_n)^k \tilde{r}_{n,l} \begin{bmatrix} \mathbf{v}_{n,l}^B \\ -\mathbf{v}_{n,l}^{B^c} \end{bmatrix} = \mathbf{0}_N$$

and for $K_{i,i} = -1$, $i = 0, \dots, N-1$

$$\sum_{n=1}^P \sum_{l=0}^{m_n-1} \prod_{q=1}^T (\beta_q - \beta_n)^k r_{n,l} \mathbf{v}_{n,l} + \sum_{n=1}^P \sum_{l=0}^{m_n-1} \prod_{q=1}^T (\beta_q + \beta_n)^k \tilde{r}_{n,l} \begin{bmatrix} \mathbf{v}_{n,l}^B \\ -\mathbf{v}_{n,l}^{B^c} \end{bmatrix} = \mathbf{0}_N.$$

Hence, we require at least m_n low-and at most $N-m_n$ high-pass components per β_n at suitably chosen locations D and D^c such that the corresponding partitions of $\{\mathbf{v}_{n,l}(D)\}_{l=0}^{m_n-1}$ and $\left\{ \begin{bmatrix} \mathbf{v}_{n,l}^B \\ -\mathbf{v}_{n,l}^{B^c} \end{bmatrix} (D^c) \right\}_{l=0}^{m_n-1}$ are linearly independent, leading to $r_{n,l} = 0$ and $\tilde{r}_{n,l} = 0$. In general, we need the number of retained low-pass components to be $\sum_{n=1}^P m_n \leq N/2$ and D such that the above partitions form linearly independent sets for $n = 1, \dots, P$.

Consider the relevant case, when we downsample by 2 w.r.t. $s = 1$ such that $D = (0 : 2 : N-1)$ (corresponding to $D = B$ in the bipartite case) and $|D| = |D^c| = N/2$ (see also Cor. 3.5). When eigenbasis \mathbf{V} is represented as the $N \times N$ -DFT matrix, the relation $\mathbf{V}_{D,0:N-1} = [\tilde{\mathbf{V}} \tilde{\mathbf{V}}]$ holds, where $\tilde{\mathbf{V}}$ denotes the $N/2 \times N/2$ DFT-matrix. Thus, the transform is invertible for $\beta_n = \gamma_i$, if corresponding eigenvalue(s) γ_i with multiplicity m_i are suitably located in the DFT-ordered spectrum $\gamma_{DFT} = \{\gamma_i\}_{i=0}^{N-1}$ such that the associated eigenvectors $\{\mathbf{v}_{i,l}\}_{l=0}^{m_i-1}$ remain linearly independent after downsampling, i.e. their pairwise column positions (j, j') in $\mathbf{V}_{D,0:N-1}$ are not of the form $(j, N/2 + j)$. Equivalently, the above can be extended for the bipartite case when $-\gamma_i$ exists, since we also have $\mathbf{V}_{D^c,0:N-1} = [\tilde{\mathbf{V}} \tilde{\mathbf{V}}]$ up to a normalization constant per column.

A special case occurs, when $\beta = \tilde{d} = 0$; we need to show that $\mathbf{K} \frac{\mathbf{A}}{d} \mathbf{z} = \mathbf{0}_N$ as long as $\sum_{i=0}^{N-1} r(i)^2 \gamma_i^2 = 0$. The latter yields $r(i) = 0$, except when $\gamma_i = 0$. Since, however, the eigenvector(s) for $\gamma_i = 0$ lie in the nullspace of $\frac{\mathbf{A}}{d}$, we have $r(0) \neq 0$, thus the filterbank is not invertible for any downsampling pattern, including the case when downsampling is conducted with respect to $s = 1$. When $\frac{\mathbf{A}}{d}$ on the other hand is invertible with $\gamma_i \neq 0$, so is the filterbank. \square

A.3 Proof of Corollary 3.3

Corollary 3.3. *Let $G = (V, E)$ be an undirected, circulant graph with adjacency matrix \mathbf{A} and degree $d = \sum_{j=1}^M 2d_j$ per node with symmetric weights $d_j = A_{i, (j+i)_N}$. Then the low-pass filter \mathbf{H}_{LP_α} in Eq. (3.3) is invertible unless (i) G is bipartite while β_n satisfies $|\beta_n| = |\gamma_i|$ or (ii) $\beta_n = -\gamma_i$, $i \in [0 \ N - 1]$.*

Proof of Corollary 3.3. (appears in [7]) Consider the simplest case with $k, T = 1$ and one parameter β : we need to show that the nullspace \mathbf{z} of low-pass filter \mathbf{H}_{LP_α} in

$$\frac{1}{2} \left(\beta \mathbf{I}_N + \frac{\mathbf{A}}{d} \right) \mathbf{z} = \mathbf{0}_N \quad (\text{A.10})$$

is empty by contradiction. In a similar fashion as in previous proofs, we let $\mathbf{z} = \mathbf{V}\mathbf{r}$, with $\frac{\mathbf{A}}{d} = \mathbf{V}\mathbf{\Gamma}\mathbf{V}^H$ and obtain

$$\begin{aligned} & \|\beta \mathbf{V}\mathbf{r} + \mathbf{V}\mathbf{\Gamma}\mathbf{r}\|_2^2 = 0 \\ \Leftrightarrow & \sum_{i=0}^{N-1} r(i)^2 (\beta^2 + 2\beta\gamma_i + \gamma_i^2) = \sum_{i=0}^{N-1} r(i)^2 (\beta + \gamma_i)^2 = 0. \end{aligned}$$

Hence, it follows from inspection that the low-pass filter is invertible unless $\beta = -\gamma_i$. If $-\gamma_i \in \gamma$, similar reasoning as for the bipartite case applies. In the case of a bipartite graph, where $|\beta| = |\gamma_i|$ and $\gamma_i, -\gamma_i$ of respective multiplicity m_i exist, with eigenvectors

$$\mathbf{z} = \sum_{l=0}^{m_i-1} \left(r(l)\mathbf{v}_l + \tilde{r}(l) \begin{bmatrix} \mathbf{v}_l^B \\ -\mathbf{v}_l^{B^c} \end{bmatrix} \right),$$

we observe after substitution into Eq. (A.10) that

$$\sum_{l=0}^{m_i-1} r(l)(\beta \mathbf{v}_l + \gamma_i \mathbf{v}_l) + \sum_{l=0}^{m_i-1} \tilde{r}(l) \left(\beta \begin{bmatrix} \mathbf{v}_l^B \\ -\mathbf{v}_l^{B^c} \end{bmatrix} - \gamma_i \begin{bmatrix} \mathbf{v}_l^B \\ -\mathbf{v}_l^{B^c} \end{bmatrix} \right) = \mathbf{0}_N.$$

Due to spectral folding, one eigenvector-set always cancels out for $\beta = \pm\gamma_i$, so that we cannot guarantee zero coefficients, and hence invertibility. By extension, $\mathbf{H}_{LP_\alpha}^k$ is invertible, while $\mathbf{H}_{LP_\alpha}^k$ requires invertibility of each individual factor $\mathbf{H}_{LP_{\alpha_n}}$ for parameters β_n , under the above.

□

Appendix B

Proofs of Chapter 4

B.1 Proof of Corollary 4.3

Corollary 4.3. *The time-varying graph wavelet transform, consisting respectively of the filters*

(i) *in Eqs. (4.8) and (4.9), with parameters $\beta_{t,n} = 1$ and $S = 1 \forall t$, is invertible for any downsampling pattern, as long as at least one node retains the low-pass component for connected graphs with non-negative weights, and otherwise, for a suitable set of retained nodes V_α such that the partitioned eigenvectors $\{\mathbf{v}_{t,j}(V_\alpha)\}_{t,j}$ corresponding to $\gamma_{t,i} = 1$ (and for bipartite graphs, also $\{\tilde{\mathbf{v}}_{t,j}(V_\alpha^{\mathbb{L}})\}_{t,j}$ for $\gamma_{t,i} = -1$) of multiplicity j are linearly independent. If the graphs are bipartite, non-negativity is sufficient.*

(ii) *in Eqs. (4.8) and (4.9), is invertible for any downsampling pattern as long as $|\beta_{t,n}| \neq |\gamma_{t,i}|$ and subject to restrictions (a) the eigenvalues of $\mathbf{H}_{LP}\mathbf{H}_{HP}$ are non-zero and of the same sign or (b) $k \in 2\mathbb{N}$; otherwise, if for some $\beta_{t,n} = \gamma_{t,i}$, and the graph is non-bipartite, invertibility follows, under the previous conditions, from linear independence of the corresponding partitions $\{\mathbf{v}_{t,j}(V_\alpha)\}_{t,j}$ for suitable V_α .*

(iii) *in Eqs. (4.6) and (4.7), is invertible under the same conditions as (ii), where in (b) it is additionally required that $h_t, \tilde{h}_t \leq 0 \forall t$; otherwise, if $\beta_{t,n} = \gamma_{t,i}$ holds on each graph G_t such that all \mathbf{H}_{HP_t} share the same nullspace \mathbf{v} , invertibility follows under the preceding provided the corresponding partitions $\mathbf{v}(V_\alpha)$ (and $\tilde{\mathbf{v}}(V_\alpha^{\mathbb{L}})$ in the bipartite case, for $\tilde{\mathbf{v}}$ inducing $-\gamma_{t,i}$ on each graph G_t) are linearly independent for suitable V_α .*

Proof of Cor. 4.3. (i)

Consider a sequence of symmetric normalized adjacency matrices $\{\tilde{\mathbf{A}}_t\}_{t=0}^T$ which possess

the same eigenbasis, i.e. $\tilde{\mathbf{A}}_t = \mathbf{V}\mathbf{\Gamma}_t\mathbf{V}^H$ and construct the following filters

$$\mathbf{H}_{LP} = C \prod_{t=0}^T (\mathbf{I}_N + \tilde{\mathbf{A}}_t)^k$$

$$\mathbf{H}_{HP} = \tilde{C} \prod_{t=0}^T (\mathbf{I}_N - \tilde{\mathbf{A}}_t)^k,$$

whose coefficients $C = \prod_{t=0}^T h_t$ and $\tilde{C} = \prod_{t=0}^T \tilde{h}_t$, we set to $C = \tilde{C} = 1$ for simplicity and wlog. We proceed to prove that the transform $\mathbf{W} = \frac{1}{2}(\mathbf{I}_N + \mathbf{K})\mathbf{H}_{LP} + \frac{1}{2}(\mathbf{I}_N - \mathbf{K})\mathbf{H}_{HP}$ is invertible by showing that its nullspace $\mathbf{z} = \mathbf{V}\mathbf{r}$ is empty, evaluating the system:

$$\begin{aligned} & (\mathbf{H}_{LP} + \mathbf{H}_{HP})\mathbf{V}\mathbf{r} + \mathbf{K}(\mathbf{H}_{LP} - \mathbf{H}_{HP})\mathbf{V}\mathbf{r} = \mathbf{0}_N \\ \Leftrightarrow & \mathbf{V} \left(\prod_{t=0}^T (\mathbf{I}_N + \mathbf{\Gamma}_t)^k + \prod_{t=0}^T (\mathbf{I}_N - \mathbf{\Gamma}_t)^k \right) \mathbf{r} + \mathbf{KV} \left(\prod_{t=0}^T (\mathbf{I}_N + \mathbf{\Gamma}_t)^k - \prod_{t=0}^T (\mathbf{I}_N - \mathbf{\Gamma}_t)^k \right) \mathbf{r} = \mathbf{0}_N \\ \Leftrightarrow & \|\mathbf{V} \left(\prod_{t=0}^T (\mathbf{I}_N + \mathbf{\Gamma}_t)^k + \prod_{t=0}^T (\mathbf{I}_N - \mathbf{\Gamma}_t)^k \right) \mathbf{r}\|_2^2 = \|\mathbf{KV} \left(\prod_{t=0}^T (\mathbf{I}_N + \mathbf{\Gamma}_t)^k - \prod_{t=0}^T (\mathbf{I}_N - \mathbf{\Gamma}_t)^k \right) \mathbf{r}\|_2^2 \\ \Leftrightarrow & \mathbf{r}^H \left(\prod_{t=0}^T (\mathbf{I}_N + \mathbf{\Gamma}_t)^k + \prod_{t=0}^T (\mathbf{I}_N - \mathbf{\Gamma}_t)^k \right)^2 \mathbf{r} = \mathbf{r}^H \left(\prod_{t=0}^T (\mathbf{I}_N + \mathbf{\Gamma}_t)^k - \prod_{t=0}^T (\mathbf{I}_N - \mathbf{\Gamma}_t)^k \right)^2 \mathbf{r} \\ \Leftrightarrow & \mathbf{r}^H 4 \left(\prod_{t=0}^T (\mathbf{I}_N + \mathbf{\Gamma}_t)^k (\mathbf{I}_N - \mathbf{\Gamma}_t)^k \right) \mathbf{r} = \mathbf{0}_N \\ \Leftrightarrow & \sum_{i=0}^{N-1} |r(i)|^2 \prod_{t=0}^T (1 + \gamma_{t,i})^k (1 - \gamma_{t,i})^k = 0. \end{aligned}$$

For any symmetric normalized (or otherwise suitably normalized) adjacency matrix of an undirected graph, we have $|\gamma_{t,i}| \leq 1$, noting that i refers to the position in the spectrum whereas t denotes the t -th graph, hence all we need to show is that $r(i) = 0$ when $\gamma_{t,i} = \pm 1$ for some t, i , to prove that $\mathbf{r} = \mathbf{0}_N$.

The multiplicity of ± 1 is one if the graph is connected (acc. to the Perron Frobenius Thm.), however, as a result of the differing graph topology the cases $|\gamma_{t,i}| = 1, \forall t$ may not always be associated with the same eigenvector of the shared basis \mathbf{V} . When all $\tilde{\mathbf{A}}_t$ are non-negative and connected (irreducible), one can further deduce from Perron-Frobenius theory that the eigenvector associated with the spectral radius 1 is a unique positive vector (Thms. 8.3.4, 8.4.4 [97]), (or, if a positive eigenvector exists, it is associated with 1), hence, $\gamma_{t,i} = 1$ must be shared by the same eigenvector across graphs. In particular, if the graph is regular and connected, the corresponding common eigenvector is the all-constant vector with $\mathbf{z} = r(i)\mathbf{1}_N$ (and its bipartite complement, where applicable), requiring at least one node to be retained (the proof of Thm. 3.1 applies). Otherwise, if e.g. negative weights

arise, the latter no longer holds, which would require for the linear independence of a system of (at most) $T + 1$ downsampled eigenvectors, that at least $T + 1$ suitably chosen nodes are retained (or more, with multiplicities).

For the remaining graph cases, assume the non-bipartite case first. Let $\mathbf{z} = \sum_{t,j} r_{t,j} \mathbf{v}_{t,j}$ where $\mathbf{v}_{t,j}$ is the eigenvector for $\gamma_{t,i} = 1$ of the t -th graph, including multiplicities j , and $r_{t,j} \in \mathbf{r}$ denote the re-labelled coefficients for simplicity. In the case, when eigenvectors are shared by selected eigenvalues across graphs, the latter may be reduced. Hence

$$\begin{aligned} \mathbf{W}\mathbf{z} &= \sum_{t,j} r_{t,j} \left(\prod_{t=0}^T (1 + \gamma_{t,j})^k + \prod_{t=0}^T (1 - \gamma_{t,j})^k \right) \mathbf{v}_{t,j} + \sum_{t,j} r_{t,j} \left(\prod_{t=0}^T (1 + \gamma_{t,j})^k - \prod_{t=0}^T (1 - \gamma_{t,j})^k \right) \mathbf{K}\mathbf{v}_{t,j} \\ &= \mathbf{0}_N, \end{aligned}$$

thus if $K_{i,i} = 1, \forall i$

$$\sum_{t,j} 2r_{t,j} \prod_{t=0}^T (1 + \gamma_{t,j})^k \mathbf{v}_{t,j} = \mathbf{0}_N$$

and if $K_{i,i} = -1, \forall i$

$$\sum_{j,t} 2r_{t,j} \prod_{t=0}^T (1 - \gamma_{t,j})^k \mathbf{v}_{t,j} = \mathbf{0}_N$$

the latter of which always vanishes since at least one factor is always 0, with $\gamma_{t,i} = 1$. Thus one needs to ensure that the downsampled $\frac{1}{2}(\mathbf{I}_N + \mathbf{K})\mathbf{v}_{t,j}$ are linearly independent for a suitable pattern \mathbf{K} which governs all graphs; in the simplest case of $\mathbf{v} = \mathbf{1}_N$ and multiplicity one, it is easy to deduce that $\mathbf{r} = \mathbf{0}_N$ must hold.

When the graph is also bipartite, we have $\mathbf{z} = \sum_{t,j} (r_{t,j} \mathbf{v}_{t,j} + \tilde{r}_{t,j} \tilde{\mathbf{v}}_{t,j})$, for $\tilde{\mathbf{v}}_{t,j}$ corresponding to $-\gamma_{t,i} = -1$, and therefore,

$$\begin{aligned} \mathbf{W}\mathbf{z} &= \sum_{t,j} r_{t,j} \left(\prod_{t=0}^T (1 + \gamma_{t,j})^k + \prod_{t=0}^T (1 - \gamma_{t,j})^k \right) \mathbf{v}_{t,j} + \sum_{t,j} r_{t,j} \left(\prod_{t=0}^T (1 + \gamma_{t,j})^k - \prod_{t=0}^T (1 - \gamma_{t,j})^k \right) \mathbf{K}\mathbf{v}_{t,j} \\ &\quad + \sum_{t,j} \tilde{r}_{t,j} \left(\prod_{t=0}^T (1 - \gamma_{t,j})^k + \prod_{t=0}^T (1 + \gamma_{t,j})^k \right) \tilde{\mathbf{v}}_{t,j} + \sum_{t,j} \tilde{r}_{t,j} \left(\prod_{t=0}^T (1 - \gamma_{t,j})^k - \prod_{t=0}^T (1 + \gamma_{t,j})^k \right) \mathbf{K}\tilde{\mathbf{v}}_{t,j} \\ &= \mathbf{0}_N, \end{aligned}$$

thus if $K_{i,i} = 1, \forall i$

$$\sum_{t,j} 2r_{t,j} \prod_{t=0}^T (1 + \gamma_{t,j})^k \mathbf{v}_{t,j} + \sum_{t,j} 2\tilde{r}_{t,j} \prod_{t=0}^T (1 - \gamma_{t,j})^k \tilde{\mathbf{v}}_{t,j} = \sum_{t,j} 2r_{t,j} \prod_{t=0}^T (1 + \gamma_{t,j})^k \mathbf{v}_{t,j} = \mathbf{0}_N$$

and if $K_{i,i} = -1, \forall i$

$$\sum_{t,j} 2r_{t,j} \prod_{t=0}^T (1 - \gamma_{t,j})^k \mathbf{v}_{t,j} + \sum_{t,j} 2\tilde{r}_{t,j} \prod_{t=0}^T (1 + \gamma_{t,j})^k \tilde{\mathbf{v}}_{t,j} = \sum_{t,j} 2\tilde{r}_{t,j} \prod_{t=0}^T (1 + \gamma_{t,j})^k \tilde{\mathbf{v}}_{t,j} = \mathbf{0}_N$$

and we need to ensure that the corresponding complementary partitions of $\{\mathbf{v}_{t,j}\}_{t,j}$ and $\{\tilde{\mathbf{v}}_{t,j}\}_{t,j}$ after downsampling are respectively linearly independent. From the Perron-Frobenius Thm. [47], it is known that the eigenvector associated with the maximum (simple) eigenvalue of the adjacency matrix of a connected undirected (irreducible) graph is positive; hence, in the case of bipartite graphs, the eigenvector associated with $\gamma_{t,i} = -1$ cannot be positive, as per spectral folding. This reveals that even when $\gamma_{t,i} = 1$ were associated with a different eigenvector on each bipartite graph, it cannot correspond to an eigenvector which would induce $\gamma_{t,i} = -1$ on any other graph and thus possibly invalidate the above derivation. Further, if the graph has nonnegative weights, and is not necessarily connected, the eigenvector associated with the spectral radius ($\gamma_{max} = 1$) is nonnegative (Thm. 8.3.1, [97]), similarly re-affirming the former. Hence, it is only for nonnegative bipartite graphs that this destructive case can be excluded.

(ii) Consider the product transform with filters

$$\mathbf{H}_{LP} = C \prod_{t=0}^T \prod_{n=1}^S (\beta_{t,n} \mathbf{I}_N + \tilde{\mathbf{A}}_t)^k$$

$$\mathbf{H}_{HP} = \tilde{C} \prod_{t=0}^T \prod_{n=1}^S (\beta_{t,n} \mathbf{I}_N - \tilde{\mathbf{A}}_t)^k$$

and coefficients as above. Following the above and previous proofs, the transform \mathbf{W} may be reduced to the system

$$\mathbf{r}^H \left(\prod_{t=0}^T \prod_{n=1}^S (\beta_{t,n} \mathbf{I}_N + \mathbf{\Gamma}_t)^k (\beta_{t,n} \mathbf{I}_N - \mathbf{\Gamma}_t)^k \right) \mathbf{r} = \mathbf{0}_N$$

$$\sum_{i=0}^{N-1} |r(i)|^2 f(\gamma_{0,i}, \dots, \gamma_{T,i}) = \sum_{i=0}^{N-1} |r(i)|^2 \prod_{t=0}^T \prod_{n=1}^S (\beta_{t,n} + \gamma_{t,i})^k (\beta_{t,n} - \gamma_{t,i})^k = 0$$

where $\gamma_{t,i}$ is the i -th eigenvalue of the t -th graph, and thus we deduce that, provided $k \in 2\mathbb{N}$ (or alternative conditions that ensure $f(\gamma_{0,i}, \dots, \gamma_{T,i}) > 0$ or $f(\gamma_{0,i}, \dots, \gamma_{T,i}) < 0$, $\forall \gamma_{t,i}$), invertibility is upheld for $\beta_{t,n} \neq \pm \gamma_{t,i}$ at any downsampling pattern, since $\mathbf{r} = \mathbf{0}_N$.

For $\beta_{t,n} = \gamma_{t,i}$, and assuming $S = 1$ for simplicity, let $\mathbf{z} = \sum_{t,j} (r_{t,j} \mathbf{v}_{t,j} + \tilde{r}_{t,j} \tilde{\mathbf{v}}_{t,j})$ denote the nullspace with eigenvectors $(\mathbf{v}_{t,j}, \tilde{\mathbf{v}}_{t,j})$ respectively corresponding to $(\gamma_{t,j}, -\gamma_{t,j})$, and index j indicating multiplicities for a single chosen parameter $\beta_{t,n}$ (per graph). Thus if

$$K_{i,i} = 1, \forall i$$

$$\begin{aligned} & \sum_{t,j} 2r_{t,j} \prod_{t=0}^T \prod_{n=1}^S (\beta_{t,n} + \gamma_{t,j})^k \mathbf{v}_{t,j} + \sum_{t,j} 2\tilde{r}_{t,j} \prod_{t=0}^T \prod_{n=1}^S (\beta_{t,n} - \gamma_{t,j})^k \tilde{\mathbf{v}}_{t,j} \\ &= \sum_{t,j} 2r_{t,j} \prod_{t=0}^T \prod_{n=1}^S (\beta_{t,n} + \gamma_{t,j})^k \mathbf{v}_{t,j} = \mathbf{0}_N \end{aligned}$$

and if $K_{i,i} = -1, \forall i$

$$\begin{aligned} & \sum_{t,j} 2r_{t,j} \prod_{t=0}^T \prod_{n=1}^S (\beta_{t,n} - \gamma_{t,j})^k \mathbf{v}_{t,j} + \sum_{t,j} 2\tilde{r}_{t,j} \prod_{t=0}^T \prod_{n=1}^S (\beta_{t,n} + \gamma_{t,j})^k \tilde{\mathbf{v}}_{t,j} \\ &= \sum_{t,j} 2\tilde{r}_{t,j} \prod_{t=0}^T \prod_{n=1}^S (\beta_{t,n} + \gamma_{t,j})^k \tilde{\mathbf{v}}_{t,j} = \mathbf{0}_N. \end{aligned}$$

When the graph at hand is not bipartite and no opposing sign eigenvalues exist, the terms $\tilde{\mathbf{v}}_{t,j}$ vanish and prior analysis applies; hence a suitable downsampling pattern V_α is required so that the partitions of $\{\mathbf{v}_{t,j}(V_\alpha)\}_{t,j}$ are linearly independent. In the case of a bipartite graph however, there is ambiguity as to whether a certain eigenvector associated with $\beta_{t,n} = \gamma_{t,i}$ on one graph may correspond to $-\gamma_{t,i}$ on another, thus invalidating the proof.

(iii) The proof is similar to the preceding, except since the eigenvalues of each filter are given as a sum of the filters eigenvalues on each individual graph, the coefficient sets $\{h_t\}_t, \{\tilde{h}_t\}_t$ need to be respectively of the same sign in addition to having $k \in 2\mathbb{Z}^+$.

Under equivalent transformations as in previous proofs, one arrives at the following familiar condition on the filter eigenvalues

$$\sum_{i=0}^{N-1} |r(i)|^2 \left(\sum_{t=0}^T h_t \prod_{n=1}^M (\beta_{t,n} + \gamma_{t,i})^k \right) \left(\sum_{t=0}^T \tilde{h}_t \prod_{n=1}^M (\beta_{t,n} - \gamma_{t,i})^k \right) = 0,$$

for which we have $\mathbf{r} = \mathbf{0}_N$ provided their spectral product is non-zero and of the same sign; this can be i. a. achieved for positive coefficients and $k \in 2\mathbb{N}$. Otherwise, if at least one $\beta_{t,n}$ per graph is selected to be of the form $\beta_{t,n} = \gamma_{t,i}$ corresponding to the same eigenvector(s) \mathbf{v}_j , such that all $\{\mathbf{H}_{HP_t}\}_t$ share the same nullspace, one proceeds as before to discover that for non-bipartite graphs, if $K_{i,i} = 1, \forall i$

$$\sum_j 2r_j \left(\sum_{t=0}^T h_t \prod_{n=1}^M (\beta_{t,n} + \gamma_{t,j})^k \right) \mathbf{v}_j = \mathbf{0}_N$$

and if $K_{i,i} = -1, \forall i$:

$$\sum_j 2r_j \left(\sum_{t=0}^T h_t \prod_{n=1}^M (\beta_{t,n} - \gamma_{t,j})^k \right) \mathbf{v}_j = \mathbf{0}_N.$$

Here, the latter version vanishes, and invertibility follows for suitable downsampling pattern V_α , as before, i. e. the partitions $\{\mathbf{v}_j(V_\alpha)\}_j$ need to be linearly independent.

In the bipartite graph case, this is upheld, and additionally, the partitions $\{\tilde{\mathbf{v}}_j(V_\alpha^{\mathbb{C}})\}_j$, where $\tilde{\mathbf{v}}_j$ are the eigenvectors corresponding to $-\beta_{t,n} = -\gamma_{t,i}$, are required to be linearly independent for suitable $V_\alpha^{\mathbb{C}}$, however, nonnegativity is no longer a requirement since the individual graph filters are summed and at least one non-zero term always remains.

□

Appendix C

Proofs of Chapter 5

C.1 Proof of Corollary 5.1

Corollary 5.1. *Consider an undirected, circulant graph G of dimension N and bandwidth \tilde{B} , and let \mathbf{x} be the multiresolution decomposition of graph signal \mathbf{x}_W , which is a 1-piece polynomial of maximum degree $D \leq 2k - 1$, on G via the j -level GWT matrix \mathbf{W} .*

(i) *Let \mathbf{W} be the HGSWT of order $2k$, with corresponding low-and high-pass graph filter matrices each of bandwidth $B = k\tilde{B}$, and assume that B is sufficiently small such that $\sum_{n=0}^l \frac{B}{2^n} \leq \frac{N}{2^{l+1}}$ at each level $l \leq j - 1$. The resulting $\mathbf{x} = \mathbf{PW}\mathbf{x}_W$ is K -sparse, where $K = \frac{N}{2^j} + B(2(j - 1) + 2^{1-j})$, when $B = 2^{j-1}r$, $r \in \mathbb{Z}^+$.*

(ii) *Let \mathbf{W} be the HCGSWT of order $2k$, with corresponding low-and high-pass graph filter matrices of bandwidth T and $B = k\tilde{B}$ respectively, such that $B + \sum_{n=1}^l \frac{T}{2^n} \leq \frac{N}{2^{l+1}}$ at each level $l \leq j - 1$. The resulting \mathbf{x} is K -sparse, where $K = \frac{N}{2^j} + Bj + T(j + 2^{1-j} - 2)$, when $T = 2^{j-1}r$, $r \in \mathbb{Z}^+$.*

(iii) *Let \mathbf{W} be the HGSWT at $j = 0$, with the alternative ‘minimum’ downsampling pattern, which retains only one low-pass component. Then \mathbf{x} is K -sparse with $K = 2B$.*

Proof of Corollary 5.1. (appears in [6])

(i) The number of non-zero high-pass coefficients after applying one level of the HGSWT is B ; due to the additional ‘border effect’ of the low-pass filter at subsequent levels, we obtain the following series after j levels

$$S = B + \left(B + \frac{B}{2}\right) + \left(B + \frac{\frac{B}{2} + B}{2}\right) + \cdots = \sum_{n=0}^{j-1} (j - n) \frac{B}{2^n}.$$

Using the finite summation results

$$\sum_{n=0}^{j-1} \frac{j}{2^n} = j(2 - 2^{-j+1}), \quad \text{and} \quad \sum_{n=0}^{j-1} \frac{n}{2^n} = 2^{(1-j)}(-j - 1 + 2^j)$$

and considering the $\frac{N}{2^j}$ low-pass coefficients, we obtain $K = \frac{N}{2^j} + B(2(j-1) + 2^{1-j})$ as the total number of non-zeros. For large B , the number of high-pass coefficients at each level $l \leq j-1$ is bounded $\sum_{n=0}^l \frac{B}{2^n} \leq \frac{N}{2^{l+1}}$. If $B = 2^{j-1}r$, the formula for S is exact, otherwise, since S has to be an integer, we need to adjust the formula by adding/subtracting a term s_l at each level l , depending on whether downsampling requires rounding up or down. In particular, at each level, the high-pass filter is applied on the odd-numbered nodes $1, 3, \dots$ of the (previously) low-pass filtered and sampled graph signal $\tilde{\mathbf{y}}$; we thus note that if the length of the non-zero ‘border’ support (before downsampling) of the high-pass filtered $\tilde{\mathbf{y}}$ at the beginning of the resulting labelled sequence is an even number, while that at the end of the sequence is odd, we need to round up, and vice versa.

(ii) Following the reasoning of the previous proof, we need to consider the border effect caused by filtering with the low-pass filter of bandwidth T ; we therefore end up with the following series summation for the total number of non-zeros

$$K = \frac{N}{2^j} + B + \left(B + \frac{T}{2}\right) + \left(B + \frac{\frac{T}{2} + T}{2}\right) + \dots = \frac{N}{2^j} + jB + \sum_{n=1}^{j-1} \frac{T(j-n)}{2^n},$$

giving the formula $K = \frac{N}{2^j} + jB + T(j-2+2^{1-j})$ subject to a correction term $\pm s_l$ per level.

(iii) By Thm. 3.1, we need to retain at least one low-pass component for invertibility of the filterbank, therefore we choose to assign the low-pass component to only one node, while the remaining nodes retain the high-pass components. While this downsampling approach is not conducted with respect to the generating set of the circulant graph, and therefore less rigorous from a graph-theoretical perspective, it achieves a maximally sparse representation in the graph wavelet domain. The number of non-zeros is $2B$, where $2B-1$ is the number of non-zero high-pass coefficients. \square

C.2 Proof of Theorem 5.1

Theorem 5.1. (*Graph-FRI*) Define the permuted GFT basis \mathbf{U} of undirected circulant graph G such that \mathbf{U}^H is the DFT-matrix. We can sample and perfectly reconstruct a (wavelet-) K -sparse graph signal (with multiresolution) $\mathbf{x} \in \mathbb{C}^N$, on the vertices of circulant G using the dimensionality-reduced GFT representation $\mathbf{y} = \mathbf{U}_M^H \mathbf{x}$, $\mathbf{y} \in \mathbb{C}^M$, where \mathbf{U}_M^H are the first M rows of \mathbf{U}^H , as long as $M \geq 2K$.

Proof of Thm. 5.1. (Prony's method [120]):

Given the representation $\mathbf{y} = \mathbf{U}_M^H \mathbf{x}$, where $\|\mathbf{x}\|_0 = K$ and \mathbf{U}_M^H are the first M rows of the DFT-matrix, we can represent the n -th entry of \mathbf{y} as $y_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{K-1} x_{c_k} e^{-i2\pi c_k n/N}$ with weights x_{c_k} of \mathbf{x} at positions c_k , and apply Prony's method to recover \mathbf{x} , provided $M \geq 2K$. We redefine $y_n = \sum_{k=0}^{K-1} \alpha_k u_k^n$ with locations $u_k = e^{-i2\pi c_k/N}$ and amplitudes $\alpha_k = x_{c_k}/\sqrt{N}$, which are successively recovered. In the following, we summarise the reconstruction algorithm: given the samples y_n , we construct a Toeplitz matrix $\mathbf{T}_{K,l}$, and determine the vector \mathbf{h} , which lies in its nullspace, also known as 'the annihilating filter':

$$\mathbf{T}_{K,l} \mathbf{h} = \begin{pmatrix} y_{l+K} & y_{l+K-1} & \cdots & y_l \\ y_{l+K+1} & y_{l+K} & \cdots & y_{l+1} \\ \vdots & \ddots & \ddots & \vdots \\ y_{l+2K-2} & \ddots & \ddots & \vdots \\ y_{l+2K-1} & y_{l+2K-2} & \cdots & y_{l+K-1} \end{pmatrix} \begin{pmatrix} 1 \\ h_1 \\ h_2 \\ \vdots \\ h_K \end{pmatrix} = \mathbf{0}_K$$

which can be accomplished via the SVD-decomposition of $\mathbf{T}_{K,l}$. It can be shown that $\mathbf{T}_{K,l}$ is of rank K for distinct u_k (Prop. 1, [120]). In particular, this corresponds to the matrix-form expression of $\sum_{0 \leq k \leq K-1} \alpha_k u_k^n P(u_k) = 0$ for $l \leq n < l + K$, with polynomial

$$P(x) = x^K + \sum_{k=1}^K h_k x^{K-k} = \prod_{k=1}^K (x - u_{k-1})$$

whose roots $\{u_k\}_{k=0}^{K-1}$ can be subsequently determined from \mathbf{h} . At last, we can recover the corresponding amplitudes $\{\alpha_k\}_{k=0}^{K-1}$ by solving a system of K linear equations given by y_n . \square

Bibliography

- [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] R. Diestel, *Graph Theory, 4th Edition*, ser. Graduate texts in mathematics. Springer, 2012, vol. 173.
- [3] F. R. K. Chung, *Spectral Graph Theory*. American Mathematical Society, 1997.
- [4] M. Püschel and J. M. F. Moura, “Algebraic Signal Processing Theory: Foundation and 1-D Time,” *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3572–3585, Aug 2008.
- [5] G. H. Golub and C. F. V. Loan, *Matrix computations (3rd ed.)*. Baltimore, MD, USA: Johns Hopkins University Press, 1996. [Online]. Available: <http://portal.acm.org/citation.cfm?id=248979>
- [6] M. S. Kotzagiannidis and P. L. Dragotti, “Sampling and Reconstruction of Sparse Signals on Circulant Graphs - An Introduction to Graph-FRI,” *Appl. Comput. Harmon. Anal. (2017)*, in press, <http://dx.doi.org/10.1016/j.acha.2017.10.003>, available on arXiv: arXiv:1606.08085.
- [7] —, “Splines and wavelets on circulant graphs,” *Applied and Computational Harmonic Analysis (2017)*, in press, <https://doi.org/10.1016/j.acha.2017.10.002>, in press, available on arXiv: arXiv:1603.04917.
- [8] —, “The Graph FRI Framework-Spline Wavelet Theory and Sampling on Circulant Graphs,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, March 2016, pp. 6375–6379.
- [9] —, “Higher-Order Graph Wavelets and Sparsity on Circulant Graphs,” in *SPIE Optical Engineering+ Applications. Wavelets and Sparsity XVI*, vol. 9597. San Diego, USA: International Society for Optics and Photonics, 2015, pp. 95 971E–95 971E–9.

-
- [10] —, “Sparse Graph Signal Reconstruction on Circulant Graphs with Perturbations,” in *10th IMA Conference on Mathematics in Signal Processing, Birmingham, UK*, 2014.
 - [11] —, “Sparse Graph Signal Reconstruction and Image Processing on Circulant Graphs,” in *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Atlanta, US, 2014, pp. 923–927.
 - [12] S. Mallat, *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*, 3rd ed. Academic Press, 2008.
 - [13] R. Rubinstein, A. M. Bruckstein, and M. Elad, “Dictionaries for Sparse Representation Modeling,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1045–1057, June 2010.
 - [14] M. Unser, “Ten Good Reasons For Using Spline Wavelets,” in *Proc. SPIE vol. 3169, Wavelet Applications in Signal and Image Processing V*, 1997, pp. 422–431.
 - [15] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, “Shiftable multiscale transforms,” *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 587–607, March 1992.
 - [16] G. Beylkin, “On the Representation of Operators in Bases of Compactly Supported Wavelets,” *SIAM J. Numer. Anal.*, vol. 29, no. 6, pp. 1716–1740, Dec. 1992. [Online]. Available: <http://dx.doi.org/10.1137/0729097>
 - [17] E. J. Candes and D. L. Donoho, “New tight frames of curvelets and optimal representations of objects with piecewise c^2 singularities,” *Communications on pure and applied mathematics*, vol. 57, no. 2, pp. 219–266, 2004.
 - [18] M. N. Do and M. Vetterli, “Contourlets: a new directional multiresolution image representation,” in *Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers, 2002.*, vol. 1, Nov 2002, pp. 497–501 vol.1.
 - [19] E. L. Pennec and S. Mallat, “Bandelet Image Approximation and Compression,” *SIAM Journal of Multiscale Modeling and Simulation*, vol. 4, p. 2005, 2005.
 - [20] M. Vetterli, “Wavelets, Approximation, and Compression,” *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 59–73, Sep 2001.
 - [21] M. Unser, “Sampling-50 years after Shannon,” *Proceedings of the IEEE*, vol. 88, no. 4, pp. 569–587, 2000.
 - [22] P. L. Dragotti, M. Vetterli, and T. Blu, “Sampling Moments and Reconstructing Signals of Finite Rate of Innovation: Shannon Meets Strang-Fix,” *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1741–1757, May 2007.

- [23] T. Blu, P. L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, “Sparse Sampling of Signal Innovations,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 31–40, March 2008.
- [24] M. Vetterli, P. Marziliano, and T. Blu, “Sampling Signals with Finite Rate of Innovation,” *IEEE Transactions on Signal Processing*, vol. 50, no. 6, pp. 1417–1428, Jun 2002.
- [25] C. Vonesch, T. Blu, and M. Unser, “Generalized Daubechies Wavelet Families,” *Signal Processing, IEEE Transactions on*, vol. 55, no. 9, pp. 4415–4429, Sept 2007.
- [26] J. A. Urigüen, T. Blu, and P. L. Dragotti, “FRI Sampling With Arbitrary Kernels,” *IEEE Transactions on Signal Processing*, vol. 61, no. 21, pp. 5310–5323, Nov 2013.
- [27] D. L. Donoho, “Compressed Sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.
- [28] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic Decomposition by Basis Pursuit,” *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, Jan. 2001. [Online]. Available: <http://dx.doi.org/10.1137/S003614450037906X>
- [29] E. J. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb 2006.
- [30] P. Stoica and R. L. Moses, *Introduction to spectral analysis*. Prentice hall Upper Saddle River, 1997, vol. 1.
- [31] E. J. Candes, Y. C. Eldar, D. Needell, and P. Randall, “Compressed sensing with coherent and redundant dictionaries,” *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59–73, 2011.
- [32] D. Thanou, D. I. Shuman, and P. Frossard, “Parametric dictionary learning for graph signals,” in *IEEE Global Conference on Signal and Information Processing, GlobalSIP 2013, Austin, TX, USA, December 3-5, 2013*, 2013, pp. 487–490. [Online]. Available: <http://dx.doi.org/10.1109/GlobalSIP.2013.6736921>
- [33] R. Coifman and M. Maggioni, “Diffusion wavelets,” *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 53–94, 2006.
- [34] V. N. Ekambaram, G. Fanti, B. Ayazifar, and K. Ramchandran, “Critically-sampled perfect-reconstruction spline-wavelet filterbanks for graph signals,” in *IEEE Global Conference on Signal and Information Processing, GlobalSIP 2013*, Dec 2013, pp. 475–478.

-
- [35] S. K. Narang and A. Ortega, “Perfect Reconstruction Two-Channel Wavelet Filter Banks for Graph Structured Data,” *Signal Processing, IEEE Transactions on*, vol. 60, no. 6, pp. 2786–2799, June 2012.
 - [36] —, “Compact Support Biorthogonal Wavelet Filterbanks for Arbitrary Undirected Graphs,” *Signal Processing, IEEE Transactions on*, vol. 61, no. 19, pp. 4673–4685, Oct 2013.
 - [37] D. K. Hammond, P. Vandergheynst, and R. Gribonval, “Wavelets on graphs via spectral graph theory,” *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1063520310000552>
 - [38] S. K. Narang, A. Gadde, and A. Ortega, “Signal processing techniques for interpolation in graph structured data,” in *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2013, Vancouver, BC, Canada, May 26-31, 2013*, 2013, pp. 5445–5449. [Online]. Available: <http://dx.doi.org/10.1109/ICASSP.2013.6638704>
 - [39] X. Wang, P. Liu, and Y. Gu, “Local-set-based graph signal reconstruction,” *IEEE Transactions on Signal Processing*, vol. 63, no. 9, pp. 2432–2444, 2015.
 - [40] S. Chen, A. Sandryhaila, J. M. Moura, and J. Kovačević, “Signal recovery on graphs: Variation minimization,” *IEEE Transactions on Signal Processing*, vol. 63, no. 17, pp. 4609–4624, 2015.
 - [41] S. K. Narang, Y. H. Chao, and A. Ortega, “Graph-wavelet filterbanks for edge-aware image processing,” in *Statistical Signal Processing Workshop (SSP), 2012*. IEEE, 2012, pp. 141–144.
 - [42] A. Gadde, A. Anis, and A. Ortega, “Active semi-supervised learning using sampling theory for graph signals,” in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2014, pp. 492–501.
 - [43] V. N. Ekambaram, G. Fanti, B. Ayazifar, and K. Ramchandran, “Wavelet-regularized graph semi-supervised learning,” in *Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE*. IEEE, 2013, pp. 423–426.
 - [44] R. Talmon, I. Cohen, S. Gannot, and R. R. Coifman, “Diffusion Maps for Signal Processing: A Deeper Look at Manifold-Learning Techniques Based on Kernels and Graphs,” *IEEE Signal Processing Magazine*, vol. 30, no. 4, pp. 75–86, July 2013.
 - [45] M. Belkin and P. Niyogi, “Towards a theoretical foundation for Laplacian-based manifold methods,” *Journal of Computer and System Sciences*, vol. 74, no. 8, pp. 1289 – 1308, 2008. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0022000007001274>

- [46] R. B. Bapat, *Graphs and Matrices*, 2nd ed., ser. Universitext, 2014.
- [47] P. F. Stadler, *Laplacian Eigenvectors of Graphs: Perron-Frobenius and Faber-Krahn Type Theorems*. Berlin; New York: Springer, 2007.
- [48] D. M. Cvetković, M. Doob, and H. Sachs, *Spectra of Graphs: Theory and Application*. New York: Academic Press, 1980.
- [49] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, “Random graphs with arbitrary degree distributions and their applications,” *Phys. Rev. E*, vol. 64, p. 026118, Jul 2001. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevE.64.026118>
- [50] A. Brouwer and W. Haemers, *Spectra of Graphs*, ser. Universitext. Springer New York, 2011. [Online]. Available: <https://books.google.gr/books?id=F98THwYgrXYC>
- [51] M. Fiedler, *Special matrices and their applications in numerical mathematics*. Springer Netherlands, 1986. [Online]. Available: <https://books.google.gr/books?id=lbQ95gPUrV8C>
- [52] N. Perraudin, J. Paratte, D. Shuman, V. Kalofolias, P. Vandergheynst, and D. K. Hammond, “GSPBOX: A toolbox for signal processing on graphs,” *ArXiv e-prints*, Aug. 2014.
- [53] A. Sandryhaila and J. M. F. Moura, “Discrete Signal Processing on Graphs: Frequency Analysis,” *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3042–3054, 2014.
- [54] A. Sandryhaila and J. Moura, “Discrete Signal Processing on Graphs,” *Signal Processing, IEEE Transactions on*, vol. 61, no. 7, pp. 1644–656, 2013.
- [55] V. Ekambaram, “Graph Structured Data Viewed Through a Fourier Lens,” Ph.D. dissertation, EECS Department, University of California, Berkeley, Dec 2013.
- [56] M. Crovella and E. Kolaczyk, “Graph wavelets for spatial traffic analysis,” in *IN-FOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, vol. 3. IEEE, 2003, pp. 1848–1857.
- [57] S. K. Narang and A. Ortega, “Lifting based wavelet transforms on graphs,” in *Proceedings: APSIPA ASC 2009: Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference*. Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference, International Organizing Committee, 2009, pp. 441–444.
- [58] M. Gavish, B. Nadler, and R. R. Coifman, “Multiscale wavelets on trees, graphs and high dimensional data: Theory and applications to semi supervised learning,”

- in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 367–374.
- [59] V. N. Ekambaram, G. C. Fanti, B. Ayazifar, and K. Ramchandran, “Circulant structures and graph signal processing,” in *IEEE International Conference on Image Processing, ICIP 2013*, Sept 2013, pp. 834–838.
 - [60] D. I. Shuman, M. J. Faraji, and P. Vandergheynst, “A Multiscale Pyramid Transform for Graph Signals,” *IEEE Transactions on Signal Processing*, vol. 64, no. 8, pp. 2119–2134, April 2016.
 - [61] B. Aspvall and J. R. Gilbert, “Graph Coloring Using Eigenvalue Decomposition,” *SIAM Journal on Algebraic Discrete Methods*, vol. 5, no. 4, pp. 526–538, Dec. 1984. [Online]. Available: <http://dx.doi.org/10.1137/0605051>
 - [62] U. Von Luxburg, “A tutorial on spectral clustering,” *Statistics and computing*, vol. 17, no. 4, pp. 395–416, 2007.
 - [63] B. Ricaud, D. I. Shuman, and P. Vandergheynst, “On the sparsity of wavelet coefficients for signals on graphs,” in *SPIE Optical Engineering+ Applications*. International Society for Optics and Photonics, 2013, pp. 88 581L–88 581L.
 - [64] B. Dong, “Sparse representation on graphs by tight wavelet frames and applications,” *Applied and Computational Harmonic Analysis*, vol. 42, no. 3, pp. 452 – 479, 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1063520315001372>
 - [65] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, “Sampling of Graph Signals With Successive Local Aggregations,” *IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1832–1843, April 2016.
 - [66] I. Z. Pesenson, “Sampling in Paley-Wiener spaces on combinatorial graphs,” *Trans. Amer. Math. Soc.*, vol. 360, no. 10, pp. 5603–5627, May 2008.
 - [67] A. Anis, A. Gadde, Akshay, and A. Ortega, “Towards a sampling theorem for signals on arbitrary graphs,” in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 3864–3868.
 - [68] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, “Discrete Signal Processing on Graphs: Sampling Theory,” *IEEE Transactions on Signal Processing*, vol. 63, no. 24, pp. 6510–6523, Dec 2015.
 - [69] P. Liu, X. Wang, and Y. Gu, “Coarsening graph signal with spectral invariance,” in *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2014, Florence, Italy, May 4-9, 2014*, 2014, pp. 1070–1074. [Online]. Available: <http://dx.doi.org/10.1109/ICASSP.2014.6853761>

- [70] F. Dorfler and F. Bullo, “Kron Reduction of Graphs With Applications to Electrical Networks,” *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 60, no. 1, pp. 150–163, Jan 2013.
- [71] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks,” *Nature*, vol. 393, no. 6684, pp. 440–442, 06 1998. [Online]. Available: <http://dx.doi.org/10.1038/30918>
- [72] S. Costa, J. Strapasson, M. Alves, and T. Carlos, “Circulant graphs and tessellations on flat tori,” *Linear Algebra and its Applications*, vol. 432, no. 1, pp. 369 – 382, 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0024379509004418>
- [73] G. J. Tee, “Eigenvectors of block circulant and alternating circulant matrices,” *New Zealand Journal of Mathematics*, vol. 36, no. 8, pp. 195–211, 2007.
- [74] D. Geller, I. Kra, S. Popescu, and S. Simanca, “On circulant matrices,” *Preprint, Stony Brook University*, 2004.
- [75] I. Kra and S. R. Simanca, “On circulant matrices,” *Notices of the AMS*, vol. 59, no. 3, pp. 368–377, 2012. [Online]. Available: <http://www.ams.org/notices/201203/rtx120300368p.pdf>
- [76] D. Kalman and J. E. White, “Polynomial Equations and Circulant Matrices,” *The American Mathematical Monthly*, vol. 108, no. 9, pp. 821–840, Nov. 2001. [Online]. Available: <http://www.jstor.org/stable/2695555>
- [77] D. S. G. Pollock, “Circulant matrices and time-series analysis,” *International Journal of Mathematical Education in Science and Technology*, vol. 33, no. 2, pp. 213–230, 2002. [Online]. Available: <http://dx.doi.org/10.1080/00207390110118953>
- [78] G. Strang and T. Q. Nguyen, *Wavelets and Filter banks*. Wellesley-Cambridge Press, 1997.
- [79] A. Agaskar and Y. M. Lu, “A Spectral Graph Uncertainty Principle,” *IEEE Transactions on Information Theory*, vol. 59, no. 7, pp. 4338–4356, July 2013.
- [80] N. Sharon and Y. Shkolnisky, “A class of Laplacian multiwavelets bases for high-dimensional data,” *Applied and Computational Harmonic Analysis*, vol. 38, no. 3, pp. 420 – 451, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1063520314000918>
- [81] M. Vetterli, J. Kovačević, and V. K. Goyal, *Foundations of Signal Processing*. <http://www.fourierandwavelets.org>: Cambridge University Press, 2014.

-
- [82] G. Strang and G. Fix, “A Fourier analysis of the finite element variational method,” in *Constructive Aspects of Functional Analysis*. Rome, Italy: Edizioni Cremonese, 1973, pp. 795–840.
 - [83] D. Cvetković, P. Rowlinson, and S. K. Simić, “Signless Laplacians of finite graphs,” *Linear Algebra and its Applications*, vol. 423, no. 1, pp. 155 – 171, 2007. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0024379507000316>
 - [84] D. Cvetković, P. Rowlinson, and S. Simić, “Eigenvalue Bounds for the Signless Laplacian,” *Publications de l’Institut Mathématique*, vol. 81(95), no. 101, pp. 11–27, 2007.
 - [85] M. Unser and T. Blu, “Cardinal exponential splines: Part I—Theory and filtering algorithms,” *IEEE Trans. Signal Process*, vol. 53, pp. 1425–1438, 2005.
 - [86] —, “Self-Similarity: Part I—Splines and Operators,” *IEEE Transactions on Signal Processing*, vol. 55, no. 4, pp. 1352–1363, April 2007.
 - [87] G. Micula, “A variational approach to spline functions theory.” *Rendiconti del Seminario Matematico*, vol. 61, no. 3, pp. 209–227, 2003. [Online]. Available: <http://eudml.org/doc/127983>
 - [88] M. Unser, “Splines: A perfect fit for signal and image processing,” *Signal Processing Magazine, IEEE*, vol. 16, no. 6, pp. 22–38, 1999.
 - [89] —, “Cardinal exponential splines: Part II - Think analog, act digital,” *IEEE Transactions on Signal Processing*, vol. 53, no. 4, pp. 1439–1449, April 2005.
 - [90] I. Pesenson, “Variational Splines and Paley–Wiener Spaces on Combinatorial Graphs,” *Constructive Approximation*, vol. 29, no. 1, pp. 1–21, 2008.
 - [91] F. Chung and S.-T. Yau, “Discrete Green’s Functions,” *Journal of Combinatorial Theory, Series A*, vol. 91, no. 1–2, pp. 191 – 214, 2000. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0097316500930942>
 - [92] R. Ellis, “Discrete green’s functions for products of regular graphs,” *arXiv preprint math/0309080*, 2003.
 - [93] A. Berman and R. J. Plemmons, *Nonnegative matrices in the mathematical sciences*, ser. Computer science and applied mathematics. New York: Academic Press, 1979, includes index. [Online]. Available: <http://opac.inria.fr/record=b1085532>
 - [94] F. Chung, “Laplacians and the Cheeger Inequality for Directed Graphs,” *Annals of Combinatorics*, vol. 9, pp. 1–19, 2005.
 - [95] M. Vetterli, J. Kovačević, and V. K. Goyal, *Fourier and Wavelet Signal Processing*, 2014. [Online]. Available: <http://www.fourierandwavelets.org/>

- [96] A. Gavili and X.-P. Zhang, “On the Shift Operator, Graph Frequency and Optimal Filtering in Graph Signal Processing,” <https://arxiv.org/abs/1511.03512>, 2016.
- [97] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. New York, NY, USA: Cambridge University Press, 2012.
- [98] S. Safavi and U. A. Khan, “Revisiting Finite-Time Distributed Algorithms via Successive Nulling of Eigenvalues,” *IEEE Signal Processing Letters*, vol. 22, no. 1, pp. 54–57, Jan 2015.
- [99] S. Segarra, A. G. Marques, and A. Ribeiro, “Linear network operators using node-variant graph filters,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2016, Shanghai, China, March 20-25, 2016*, 2016, pp. 4850–4854. [Online]. Available: <http://dx.doi.org/10.1109/ICASSP.2016.7472599>
- [100] D. Eynard, A. Kovnatsky, M. M. Bronstein, K. Glashoff, and A. M. Bronstein, “Multimodal Manifold Analysis by Simultaneous Diagonalization of Laplacians,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 12, pp. 2505–2517, Dec 2015.
- [101] R. Brualdi and D. Cvetkovic, *A Combinatorial Approach to Matrix Theory and Its Applications*, ser. Discrete Mathematics and Its Applications. CRC Press, 2008. [Online]. Available: <https://books.google.gr/books?id=pwx6t8QfZU8C>
- [102] P. Muthukrishnan, D. Radev, and Q. Mei, “Edge Weight Regularization over Multiple Graphs for Similarity Learning,” in *2010 IEEE International Conference on Data Mining*, Dec 2010, pp. 374–383.
- [103] M. M. Bronstein, K. Glashoff, and T. A. Loring, “Making Laplacians Commute,” *arXiv preprint arXiv:1307.6549*, 2013.
- [104] I. Sciriha, “A characterization of singular graphs,” *Electronic Journal of Linear Algebra*, vol. 16, no. 1, p. 38, 2007.
- [105] R. Hammack, W. Imrich, and S. Klavzar, *Handbook of Product Graphs, Second Edition*, 2nd ed. Boca Raton, FL, USA: CRC Press, Inc., 2011.
- [106] J. Kamm and J. G. Nagy, “Kronecker product and SVD approximations in image restoration,” *Linear Algebra and its Applications*, vol. 284, no. 1–3, pp. 177 – 192, 1998, international Linear Algebra Society (ILAS) Symposium on Fast Algorithms for Control, Signals and Image Processing. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0024379598100241>
- [107] J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, and Z. Ghahramani, “Kronecker Graphs: An Approach to Modeling Networks,” *J.*

- Mach. Learn. Res.*, vol. 11, pp. 985–1042, Mar. 2010. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1756006.1756039>
- [108] N. P. Pitsianis, “The Kronecker Product in Approximation and Fast Transform Generation,” Ph.D. dissertation, Ithaca, NY, USA, 1997, uMI Order No. GAX97-16143.
 - [109] A. Sandryhaila and J. M. F. Moura, “Big Data Analysis with Signal Processing on Graphs: Representation and processing of massive data sets with irregular structure,” *IEEE Signal Process. Mag.*, vol. 31, no. 5, pp. 80–90, 2014.
 - [110] C. V. Loan and N. Pitsianis, “Approximation with Kronecker Products,” in *Linear Algebra for Large Scale and Real Time Applications*. Kluwer Publications, 1993, pp. 293–314.
 - [111] R. Hoshino, “Independence Polynomials of Circulant Graphs,” Ph.D. dissertation, Dalhousie University, 2008.
 - [112] J. C. George and R. S. Sanders., “When is a Tensor Product of Circulant Graphs Circulant?” *In eprint arXiv:math/9907119 (07/1999)*.
 - [113] H. S. Sayama, “Estimation of Laplacian spectra of direct and strong product graphs,” *Discrete Applied Mathematics*, vol. 205, pp. 160–170, 2016.
 - [114] S. Barik, R. Bapat, and S. Pati, “On the Laplacian spectra of product graphs,” *Appl. Anal. Discrete Math.*, vol. 58, pp. 9–39, 2015.
 - [115] W. Imrich and S. Klavžar, *Product graphs, structure and recognition*. Wiley-Interscience, 2000, vol. 56.
 - [116] M. Vetterli, P. Marziliano, and T. Blu, “Sampling Signals with Finite Rate of Innovation,” *IEEE Transactions on Signal Processing*, vol. 50, no. 6, pp. 1417–1428, 2002.
 - [117] J. A. Uriguen, T. Blu, and P. L. Dragotti, “FRI Sampling With Arbitrary Kernels,” *IEEE Transactions on Signal Processing*, vol. 61, no. 21, pp. 5310–5323, Nov 2013.
 - [118] T. Blu and M. Unser, “Approximation Error for Quasi-Interpolators and (Multi-)Wavelet Expansions,” *Applied and Computational Harmonic Analysis*, vol. 6, no. 2, pp. 219 – 251, 1999. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1063520398902494>
 - [119] G. Strang, “The Discrete Cosine Transform,” *SIAM Review*, vol. 41, pp. 135–147, 1999.

- [120] P. L. Dragotti and Y. M. Lu, “On Sparse representation in Fourier and Local Bases,” *IEEE Transactions on Information Theory*, vol. 60, pp. 7888–7899, 2014. [Online]. Available: <http://dx.doi.org/10.1109/TIT.2014.2361858>
- [121] T. F. Chan, “An optimal circulant preconditioner for Toeplitz systems,” *SIAM journal on scientific and statistical computing*, vol. 9, no. 4, pp. 766–771, 1988.
- [122] E. Cuthill and J. McKee, “Reducing the Bandwidth of Sparse Symmetric Matrices,” in *Proceedings of the 1969 24th National Conference*, ser. ACM ’69. ACM, 1969, pp. 157–172. [Online]. Available: <http://doi.acm.org/10.1145/800195.805928>
- [123] J. A. Cadzow, “Signal enhancement: A composite property mapping algorithm,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, pp. 49–62, 1988.
- [124] T. Zhang, G. Golub, and K. Law, “Eigenvalue perturbation and generalized Krylov subspace method,” *Applied Numerical Mathematics*, vol. 27, no. 2, pp. 185 – 202, 1998. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0168927498000026>
- [125] J. R. Bunch, C. P. Nielsen, and D. C. Sorensen, “Rank-one modification of the symmetric eigenproblem,” *Numerische Mathematik*, vol. 31, no. 1, pp. 31–48, 1978. [Online]. Available: <http://dx.doi.org/10.1007/BF01396012>
- [126] J. K. Thomas, L. L. Scharf, and D. W. Tufts, “The probability of a subspace swap in the SVD,” *IEEE Transactions on Signal Processing*, vol. 43, no. 3, pp. 730–736, Mar 1995.
- [127] A. Gadde, S. K. Narang, and A. Ortega, “Bilateral Filter: Graph spectral interpretation and extensions,” in *2013 IEEE International Conference on Image Processing*, Sept 2013, pp. 1222–1226.
- [128] O. L  zoray and L. Grady, “Image processing and analysis with graphs: Theory and practice,” 2012.
- [129] K. Guo and D. Labate, “Optimally Sparse Multidimensional Representation Using Shearlets,” *SIAM J. Math. Analysis*, vol. 39, no. 1, pp. 298–318, 2007. [Online]. Available: <http://dx.doi.org/10.1137/060649781>
- [130] O. Lezoray, A. Elmoataz, and S. Bougleux, “Graph regularization for color image processing,” *Computer Vision and Image Understanding*, vol. 107, no. 1–2, pp. 38 – 55, 2007, special issue on color image processing. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1077314206002098>
- [131] J. Shi and J. Malik, “Normalized Cuts and Image Segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 888–905, 1997.

-
- [132] A. George and J. W. Liu, *Computer Solution of Large Sparse Positive Definite Systems*. Prentice Hall Professional Technical Reference, 1981.
 - [133] F. R. K. Chung, *Labelings of graphs*. Academic Press, San Diego, CA, 1988, vol. 3.
 - [134] Y. Shimada, T. Ikeguchi, and T. Shigehara, “From Networks to Time Series,” *Phys. Rev. Lett.*, vol. 109, p. 158701, Oct. 2012. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevLett.109.158701>
 - [135] A. S. L. O. Campanharo, M. I. Sirer, R. D. Malmgren, F. M. Ramos, and L. A. N. Amaral, “Duality between Time Series and Networks,” *PLoS ONE*, vol. 6, no. 8, p. e23378, Aug. 2011.
 - [136] R. Hamon, P. Borgnat, P. Flandrin, and C. Robardet, “Relabelling vertices according to the network structure by minimizing the cyclic bandwidth sum,” *Journal of Complex Networks*, vol. 4, no. 4, p. 534, 2016. [Online]. Available: <http://dx.doi.org/10.1093/comnet/cnw006>
 - [137] G. Plonka, “The Easy Path Wavelet Transform: A New Adaptive Wavelet Transform for Sparse Representation of Two-Dimensional Data.” *Multiscale Modeling Simulation*, vol. 7, no. 3, pp. 1474–1496, 2008. [Online]. Available: <http://dblp.uni-trier.de/db/journals/mmas/mmas7.htmlPlonka08>
 - [138] I. Ram, M. Elad, and I. Cohen, “Generalized Tree-Based Wavelet Transform,” *IEEE Trans. Signal Processing*, vol. 59, no. 9, pp. 4199–4209, 2011. [Online]. Available: <http://dblp.uni-trier.de/db/journals/tsp/tsp59.htmlRamEC11>
 - [139] —, “Redundant Wavelets on Graphs and High Dimensional Data Clouds,” *CoRR*, vol. abs/1111.4619, 2011. [Online]. Available: <http://dblp.uni-trier.de/db/journals/corr/corr1111.htmlabs-1111-4619>
 - [140] —, “Image Processing Using Smooth Ordering of its Patches,” *IEEE Trans. Image Processing*, vol. 22, no. 7, pp. 2764–2774, 2013. [Online]. Available: <http://dblp.uni-trier.de/db/journals/tip/tip22.htmlRamEC13>
 - [141] D. Heinen and G. Plonka, “Wavelet Shrinkage on Paths for Denoising of Scattered Data,” *Results in Mathematics*, vol. 62, no. 3, pp. 337–354, 2012. [Online]. Available: <http://dx.doi.org/10.1007/s00025-012-0285-3>
 - [142] K. Dabov, A. Foi, V. Katkovnik, and K. O. Egiazarian, “Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering.” *IEEE Trans. Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007. [Online]. Available: <http://dblp.uni-trier.de/db/journals/tip/tip16.htmlDabovFKE07>