Generalized Haar-Walsh Dictionaries: Extensions and Applications

Graph Signal Processing Workshop 2025
Mila - Quebec AI Institute, Montréal, Canada

Naoki Saito

UC Davis TETRAPODS Institute of Data Science
Department of Mathematics, University of California, Davis

May 16, 2025

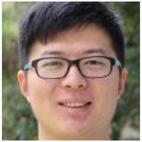
> *"We ought to give greater attention and greater support to unfashionable research. At any particular moment in the history of science, the most important and fruitful ideas are often lying dormant merely because they are unfashionable." — Freeman Dyson: "Unfashionable Pursuits," Math. Intell., vol.5, pp.47–54, 1983*

Outline

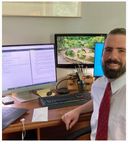
Acknowledgment

- NSF Grants: DMS-1819222, DMS-1912747, CCF-1934568, DMS-2012266
- ONR Grants: N00014-16-1-2255, N00014-20-1-2381
- My collaborators (former PhD students and a postdoc)

Jeff Irion
(Zoox)



Yiqun Shao
(Meta)



S. Schonsheck
(CA Ener. Safety)

Eugene Shvarts
(Typesafe)

Outline

Motivations

The Generalized Haar-Walsh Transform (GHWT)

The *extended* GHWT (eGHWT)

Applications

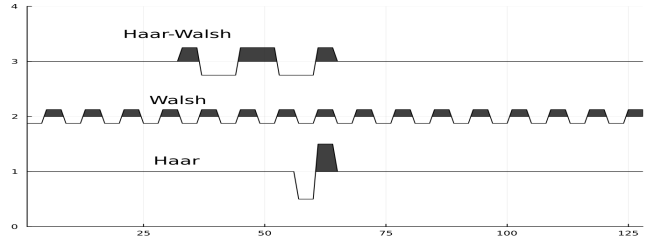
Generalization to Simplicial Complexes

Summary

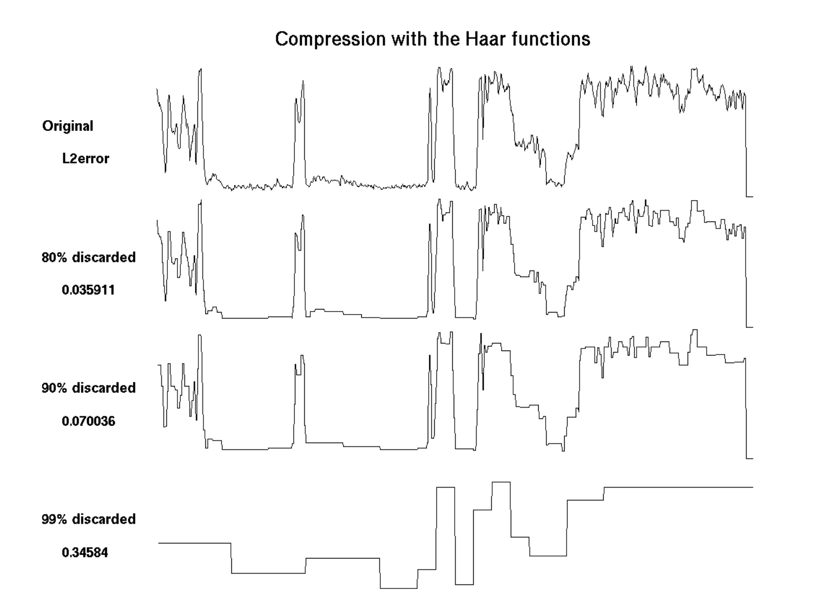
References

Motivations: Wavelets and Wavelet Packets

- For usual signals and images, *wavelet transforms* and their generalization *wavelet packet transforms* have a proven track record of success, e.g., JPEG *2000* Image Compression Standard.
- The *Haar* wavelet transform is the simplest among them; it decomposes a given signal into *translations* and *dilations* of a difference of *blocky* functions.
- The *Walsh* transform decomposes a given signal into more oscillatory global square waves.
- The *Haar-Walsh* wavelet packet transform decomposes a given signal into all sorts of local, global, and/or oscillatory blocky functions (hence, it is a *redundant* transform).

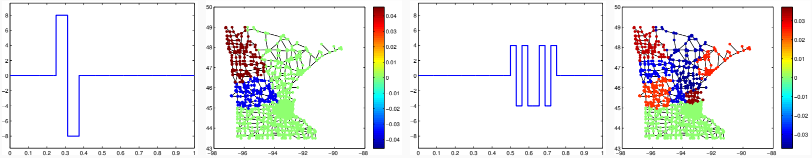


Motivations: Wavelets and Wavelet Packets ...



Motivations: Lifting the Haar-Walsh Wavelet Packets to Graphs

- Want to lift the Haar-Walsh wavelet packet transform to the graph setting
- The Haar-Walsh wavelet transform is the most amenable to graphs and networks among all the wavelets and wavelet packets family due to its operational simplicity (straightforward *sum and difference computation*)



(a) Haar function (b) Haar vector (c) Walsh fcn. ($j = 2$) (d) Walsh vec. ($j = 1$)

A Bit of History

(a) Alfred Haar
(1910)

A Bit of History

(a) Alfred Haar (1910)

(b) Joseph L. Walsh (1923)

A Bit of History

(a) Alfred Haar (1910)

(b) Joseph L. Walsh (1923)

(c) Raphy Coifman (1989)

(d) Yves Meyer (1989)

A Bit of History

(a) Alfred Haar (1910)

(b) Joseph L. Walsh (1923)

(c) Raphy Coifman (1989)

(d) Yves Meyer (1989)

(e) Christoph Thiele (1996)

(f) Lars Villemoes (1996)

A Bit of History

(a) Alfred Haar (1910)

(b) Joseph L. Walsh (1923)

(c) Raphy Coifman (1989)

(d) Yves Meyer (1989)

(e) Christoph Thiele (1996)

(f) Lars Villemoes (1996)

(g) Jeff Irion (2014)

(h) Naoki Saito (2014)

A Bit of History

(a) Alfred Haar (1910)

(b) Joseph L. Walsh (1923)

(c) Raphy Coifman (1989)

(d) Yves Meyer (1989)

(e) Christoph Thiele (1996)

(f) Lars Villemoes (1996)

(g) Jeff Irion (2014)

(h) Naoki Saito (2014)

(i) Naoki Saito (2019-22)

(j) Yiqun Shao (2019-22)

Outline

Generalized Haar-Walsh Transform (GHWT)

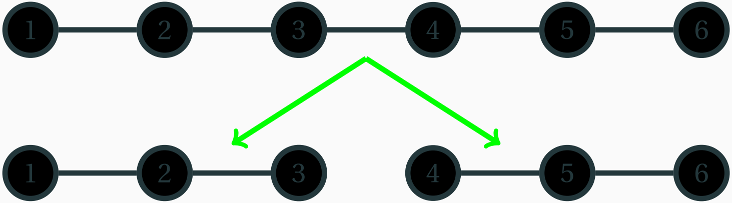
The *GHWT* is a true generalization of the classical Haar-Walsh Wavelet Packet Transform, and generates a *dictionary* (i.e., a redundant set) of basis vectors that are *piecewise-constant* on their support.

The algorithm using the Fiedler vectors can be summarized as follows although any other graph partitioning algorithm can be used:

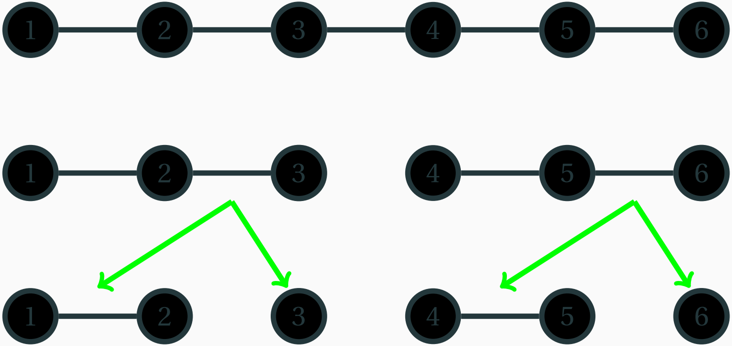
1. Generate a full recursive bipartitioning of the graph using *Fiedler vectors* $\phi^j_{k,1}$ of $L_{\mathrm{rw}}(G^j_k) := I - D^{-1}(G^j_k)W(G^j_k)$, where $k = 0, \ldots, K^j - 1$ indicates a region, $j = 0, \ldots, j_{\max}$ indicates a level (or scale), $V = V^0_0 = V^1_0 \cup V^1_1 = \cdots$
2. Generate an orthonormal basis for level $j_{\max}$ (the finest level) $\Rightarrow$ ***scaling vectors*** on the single-node regions
3. Using the basis for level $j_{\max}$, generate an orthonormal basis for level $j_{\max} - 1 \Rightarrow$ ***scaling*** and ***Haar*** vectors
4. For $j = j_{\max} - 1 : -1 : 1$ Using the basis for level j, generate an orthonormal basis for level $j - 1 \Rightarrow$ ***scaling***, ***Haar***, and ***Walsh*** vectors

GHWT on P_6

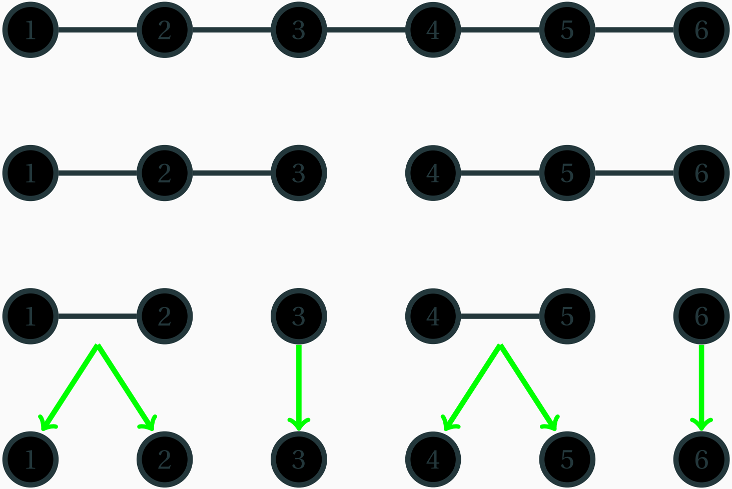
GHWT on P_6



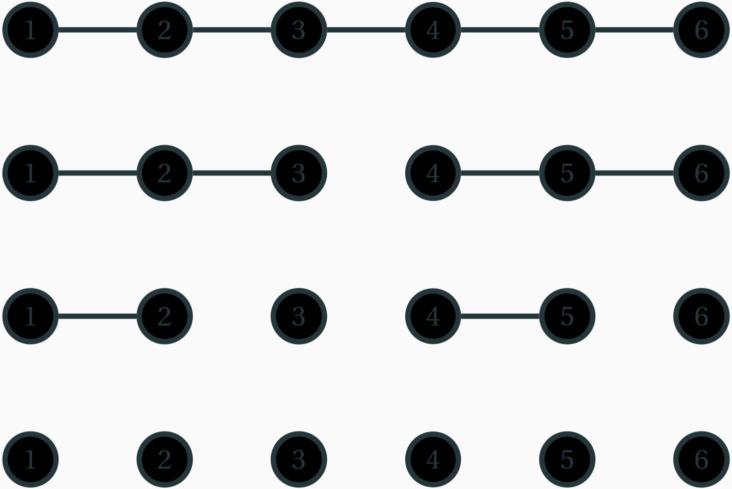
GHWT on P_6



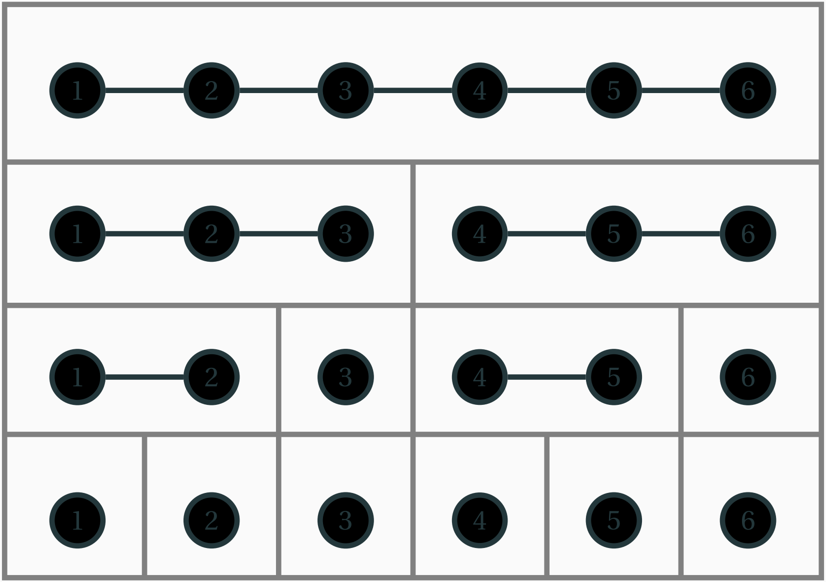
GHWT on P_6



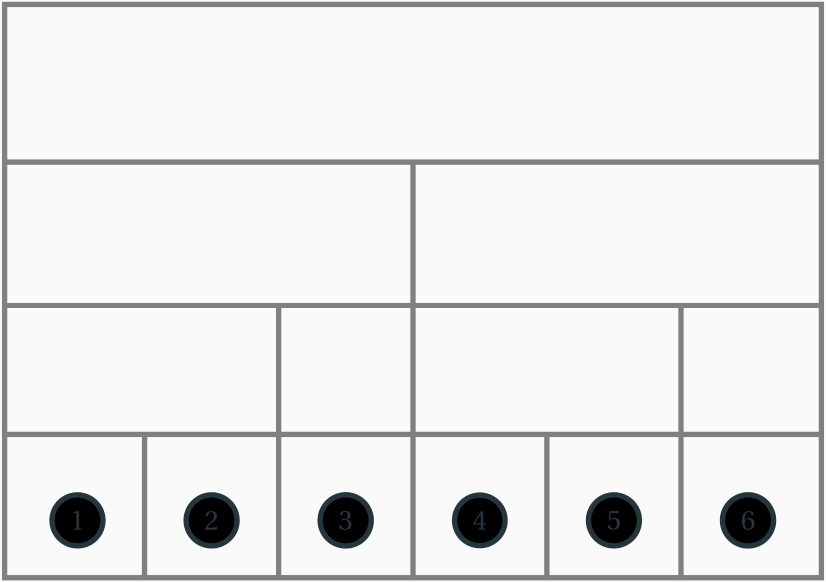
GHWT on P_6



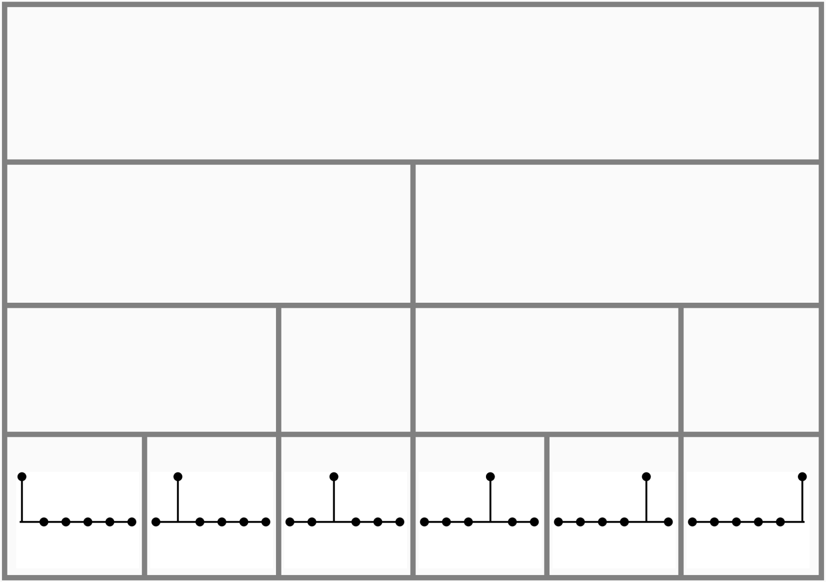
GHWT on P_6



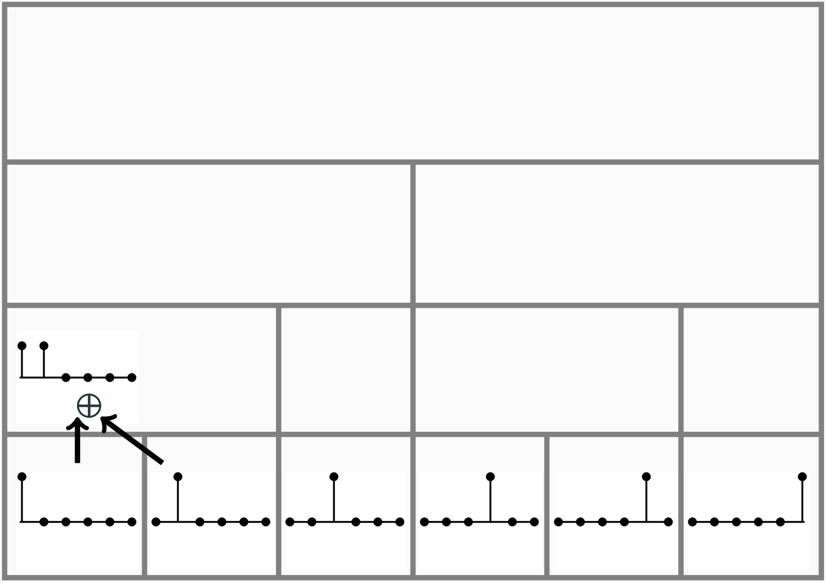
GHWT on P_6



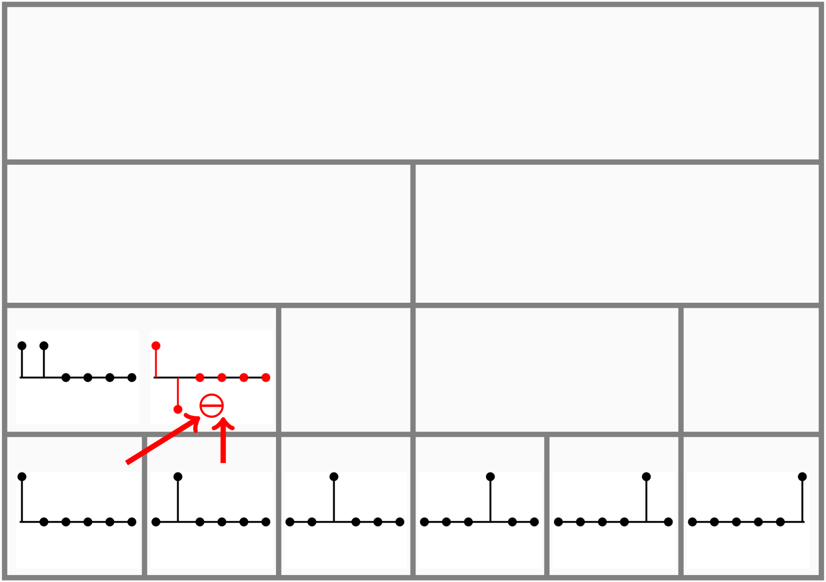
GHWT on P_6



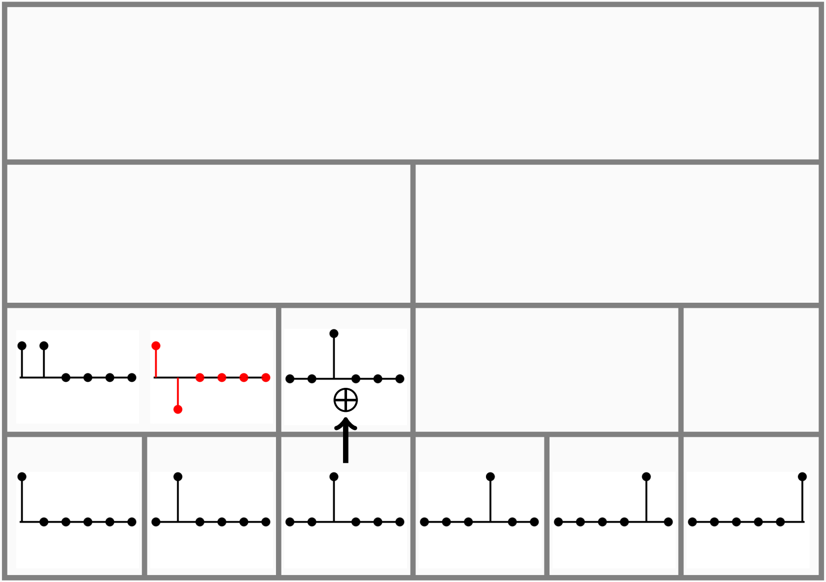
GHWT on P_6



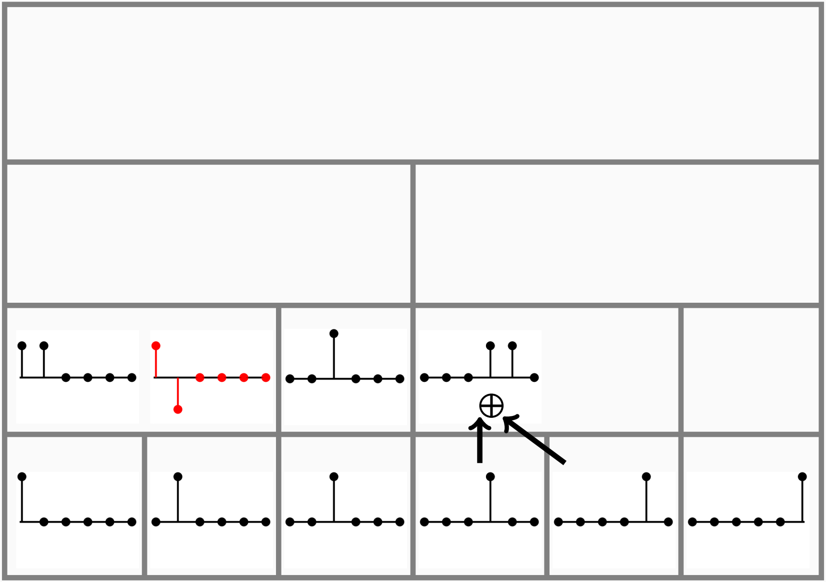
GHWT on P_6



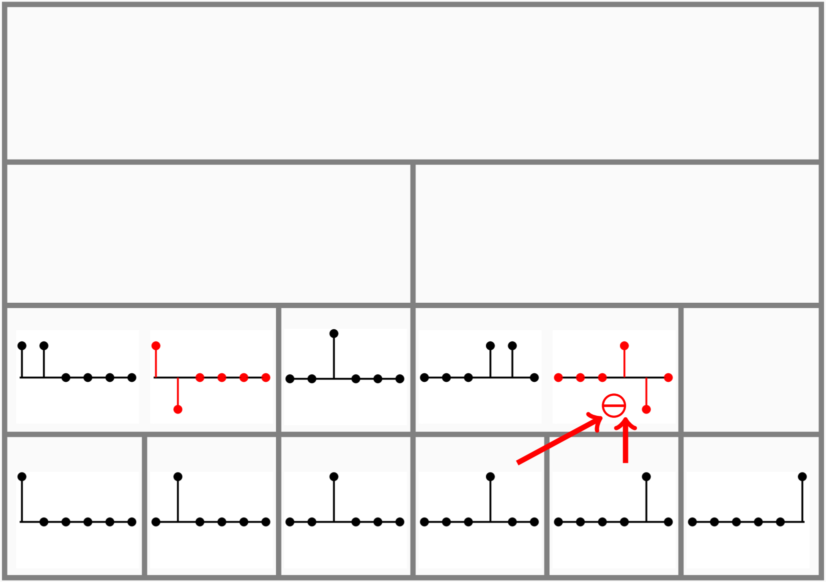
GHWT on P_6



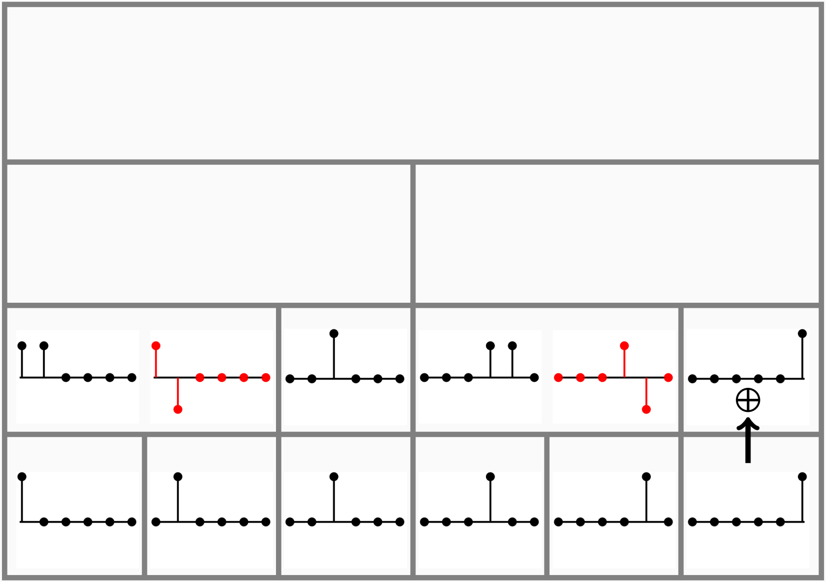
GHWT on P_6



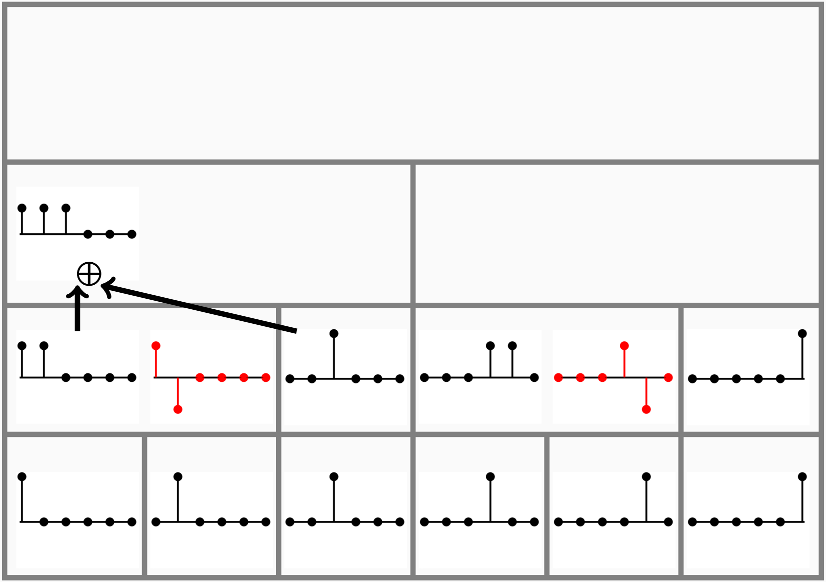
GHWT on P_6



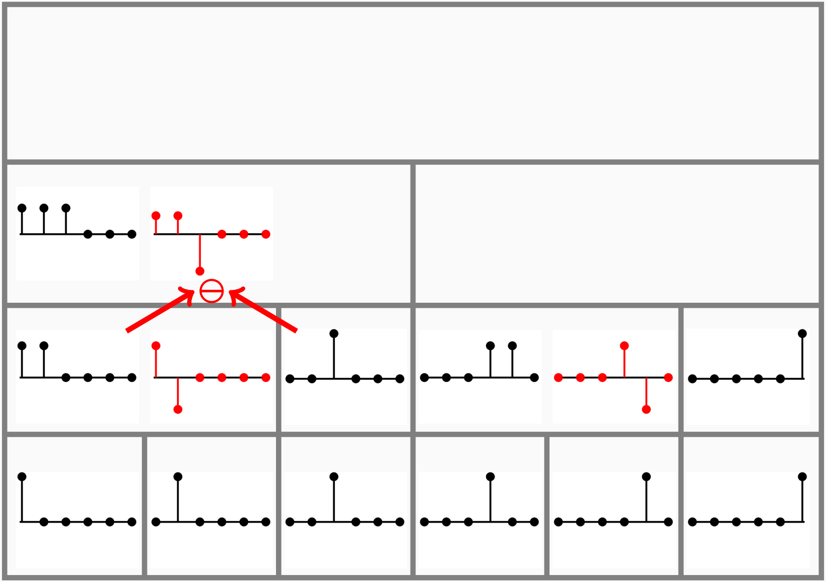
GHWT on P_6



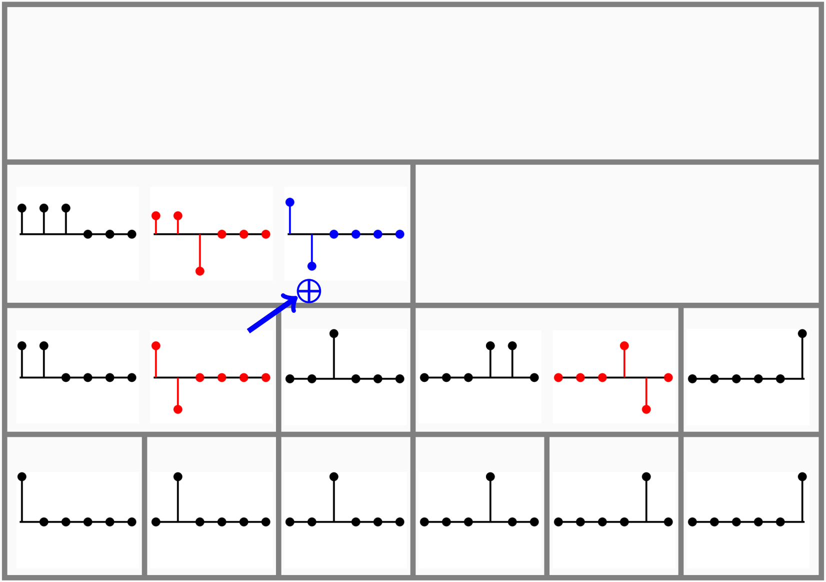
GHWT on P_6



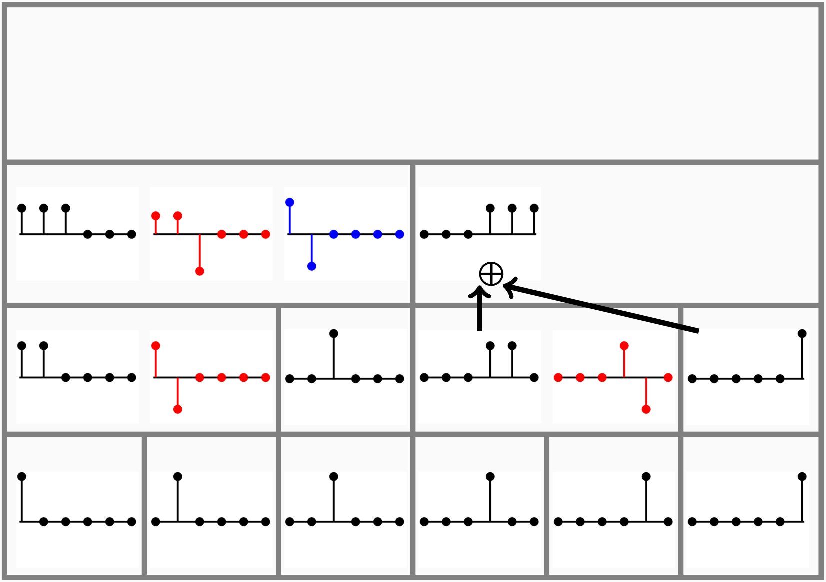
GHWT on P_6



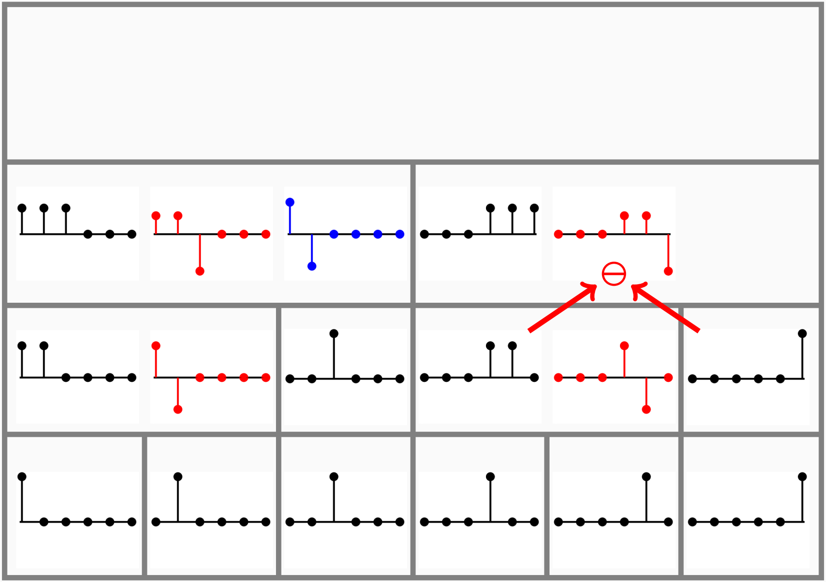
GHWT on P_6



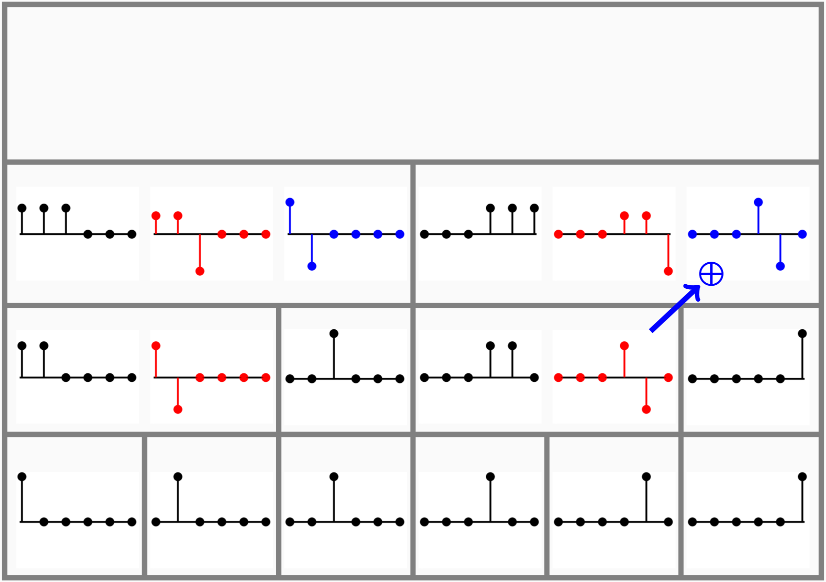
GHWT on P_6



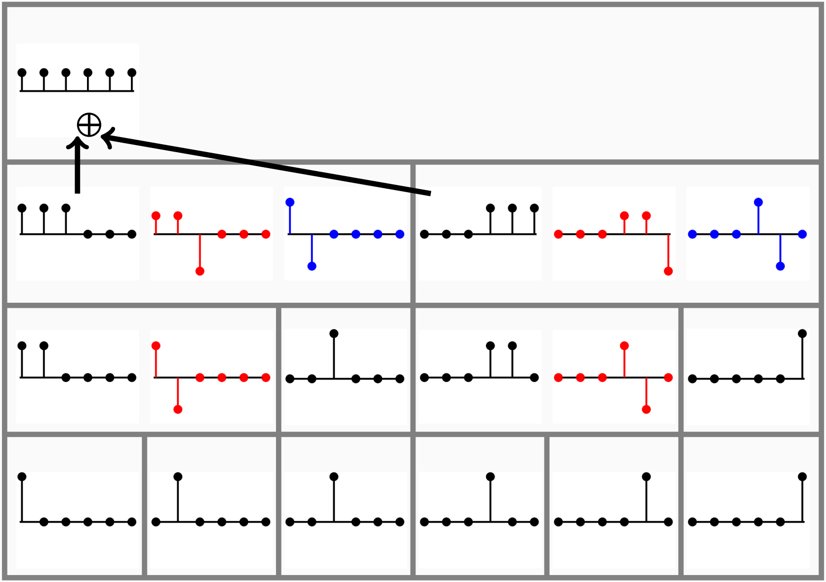
GHWT on P_6



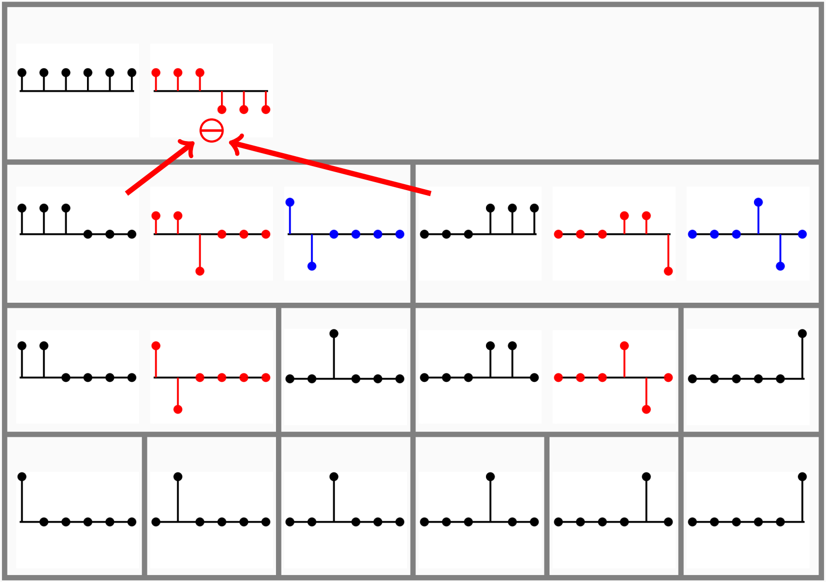
GHWT on P_6



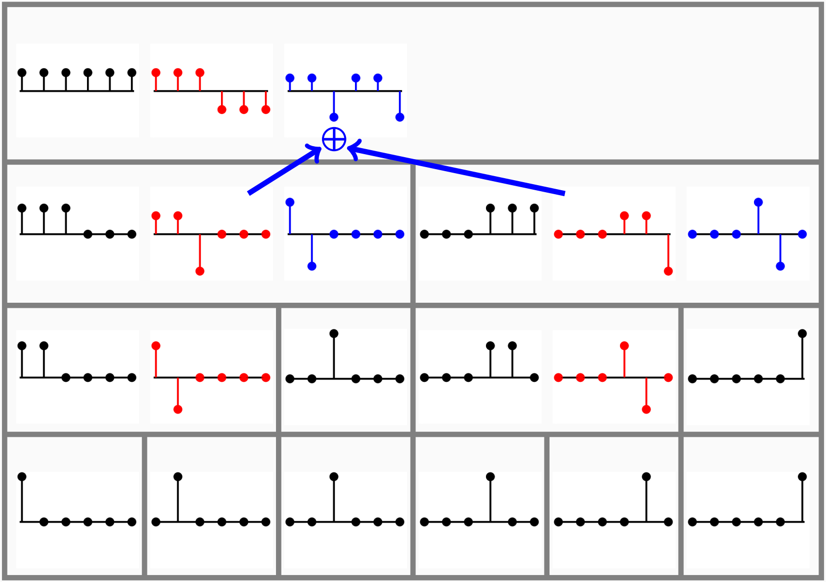
GHWT on P_6



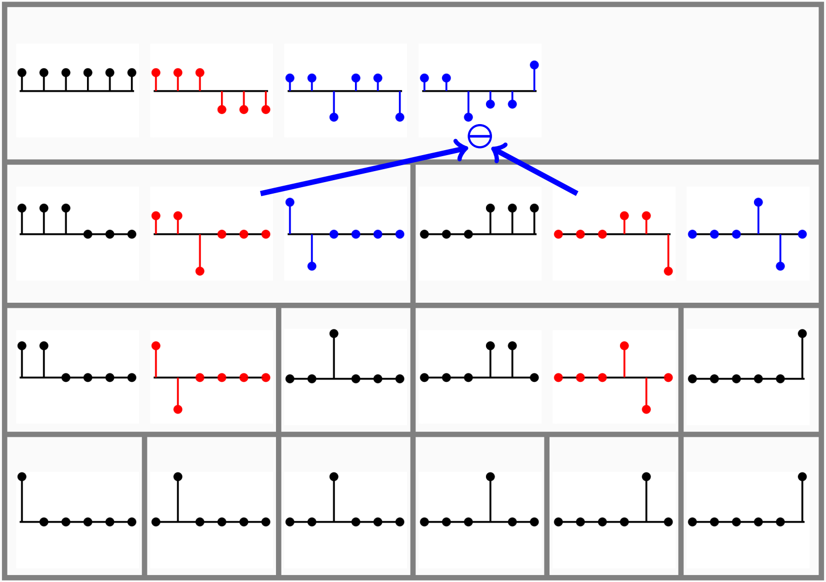
GHWT on P_6



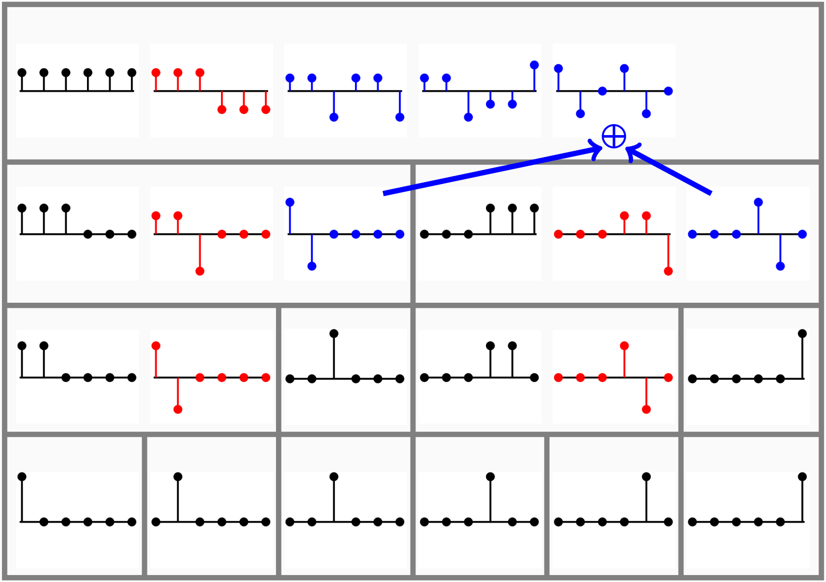
GHWT on P_6



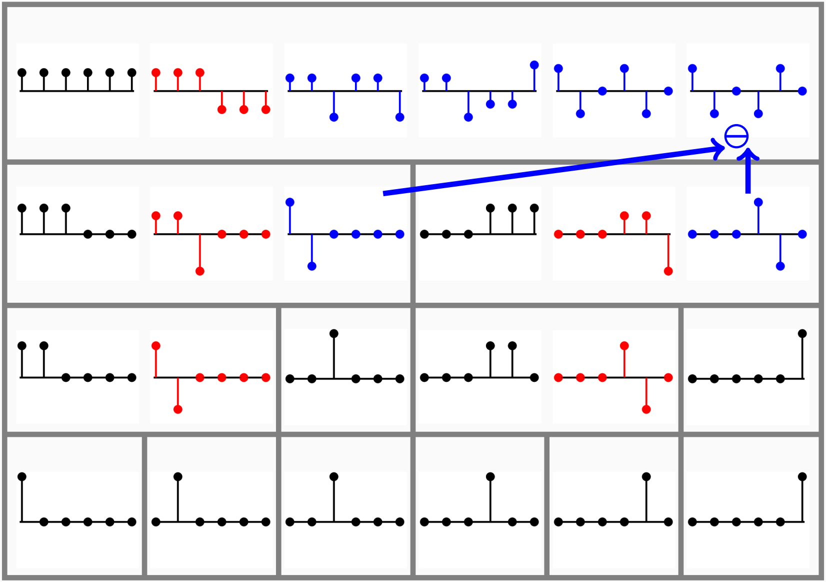
GHWT on P_6



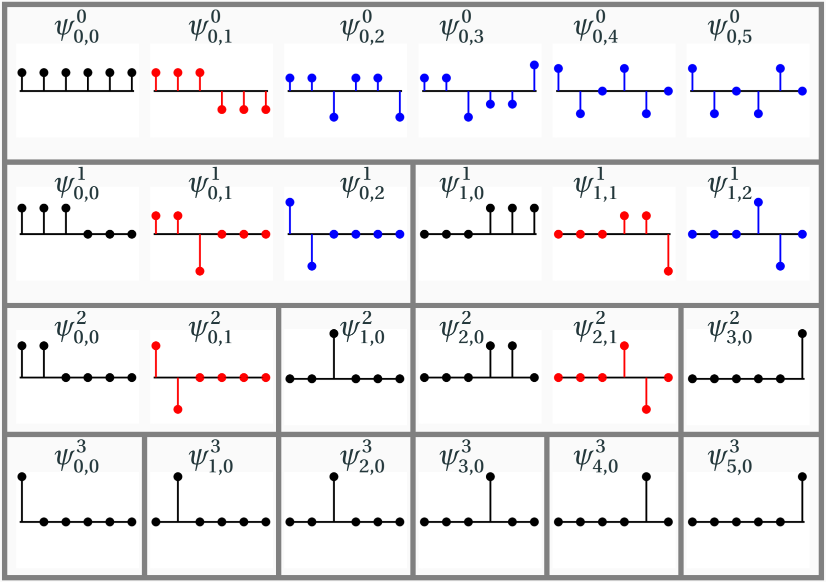
GHWT on P_6



GHWT on P_6



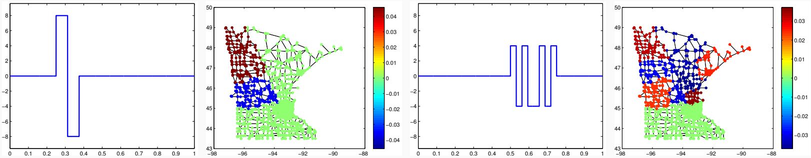
GHWT on P_6


$\psi^0_{0,0}$ $\psi^0_{0,1}$ $\psi^0_{0,2}$ $\psi^0_{0,3}$ $\psi^0_{0,4}$ $\psi^0_{0,5}$
$\psi^1_{0,0}$ $\psi^1_{0,1}$ $\psi^1_{0,2}$ $\psi^1_{1,0}$ $\psi^1_{1,1}$ $\psi^1_{1,2}$
$\psi^2_{0,0}$ $\psi^2_{0,1}$ $\psi^2_{1,0}$ $\psi^2_{2,0}$ $\psi^2_{2,1}$ $\psi^2_{3,0}$
$\psi^3_{0,0}$ $\psi^3_{1,0}$ $\psi^3_{2,0}$ $\psi^3_{3,0}$ $\psi^3_{4,0}$ $\psi^3_{5,0}$

Basis Vector & Coefficient Notation

GHWT basis vectors and coefficients are written as $\boldsymbol{\psi}^{j}_{k,l}$ and $c^{j}_{k,l}$, respectively, where j and k correspond to level and region and l is the **tag**.

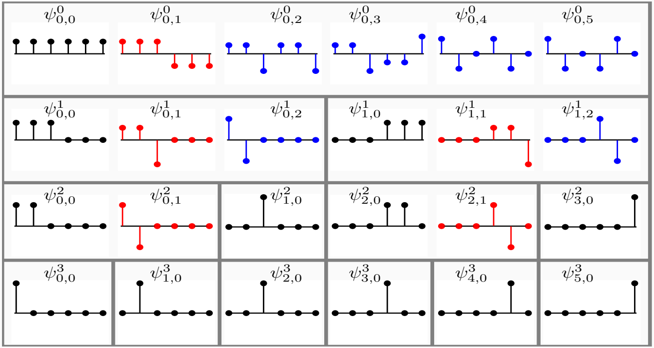
- $l = 0 \Rightarrow$ **scaling coefficient/basis vector**
- $l = 1 \Rightarrow$ **Haar coefficient/basis vector**
- $l \geq 2 \Rightarrow$ **Walsh coefficient/basis vector**



(a) Haar function (b) Haar vector (c) Walsh fcn. ($j = 2$) (d) Walsh vec. ($j = 1$)

Remarks

- For an unweighted path graph of dyadic length, this yields *exactly* a dictionary of the conventional Haar-Walsh wavelet packets.
- Recursive Partitioning (RP) via Fiedler vectors costs $O(N^2)$ in general.
- Given a recursive partitioning with $O(\log N)$ levels, the computational cost of expanding an input data into the GHWT is $O(N \log N)$.
- We can select an orthonormal basis for the entire graph by taking the union of orthonormal bases on disjoint regions.

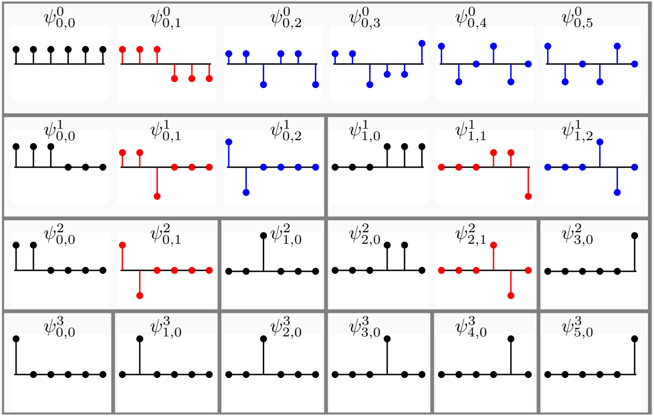


Remarks ...

- We can also reorder and regroup the vectors on each level of the GHWT dictionary according to their type (**scaling**, **Haar**, or **Walsh**).

Remarks ...

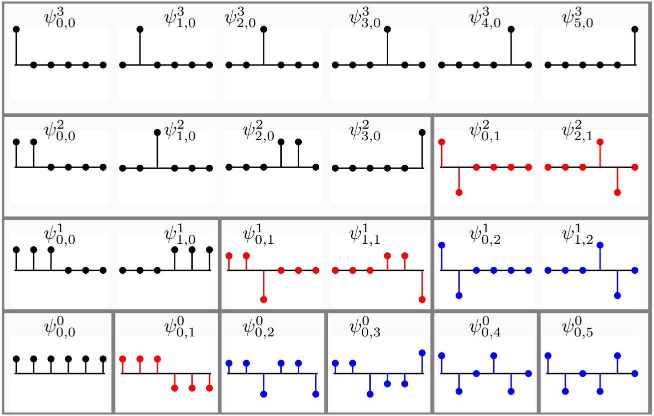
- We can also reorder and regroup the vectors on each level of the GHWT dictionary according to their type (**scaling**, **Haar**, or **Walsh**).



Default dictionary: *coarse-to-fine*

Remarks ...

- We can also reorder and regroup the vectors on each level of the GHWT dictionary according to their type (**scaling**, **Haar**, or **Walsh**).



Reordered & regrouped dictionary: *fine-to-coarse*

Outline

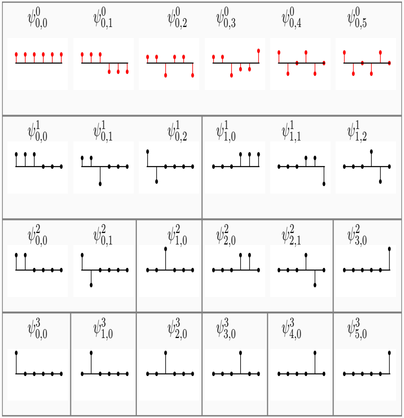
Best-Basis Algorithms for GHWT

- Coifman and Wickerhauser (1992) developed the *best-basis* algorithm as a means of selecting the basis from a dictionary of wavelet packets that is “best” for approximation/compression.
- We generalize this algorithm for selecting the basis from the GHWT dictionary in the *bottom-up* manner that is “best” for approximation/compression.
- We require an appropriate cost functional $\mathscr{J}$, e.g.,

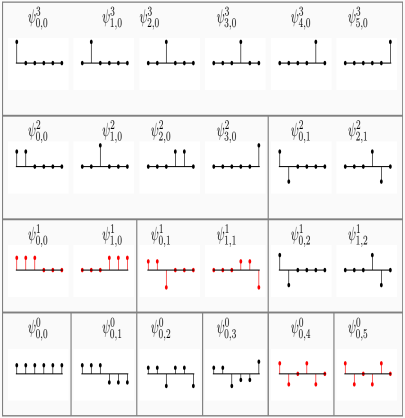
$$\mathscr{J}\left(\boldsymbol{c}_k^j\right) = \left\|\boldsymbol{c}_k^j\right\|_p := \left(\sum_{l=0}^{N_k^j-1} \left|c_{k,l}^j\right|^p\right)^{1/p} \qquad 0 < p \le 1,$$

  to seek the *sparsest* representation of the input graph signal.
- For other tasks, e.g., classification and regression, see our own work on *Local Discriminant Basis*, *Local Regression Basis*, *Least Statistically-Dependent Basis*, …, all of which use different cost functionals and can also be used in the graph setting.

Consider a simple graph signal $\boldsymbol{f} = [2, -2, 1, 3, -1, -2]^{\mathsf{T}} \in \mathbb{R}^6$ on $G = P_6$. Note $\|\boldsymbol{f}\|_1 = 11$. Running the best-basis algorithm with the ℓ^1-norm minimization generates the following best bases:



(a) The GHWT c2f best basis: $\|\widehat{\boldsymbol{f}}\|_1 \approx \mathbf{8.28}$



(b) The GHWT f2c best basis: $\|\widehat{\boldsymbol{f}}\|_1 \approx \mathbf{7.84}$

Outline

Motivation of developing eGHWT

- We want to complete the lifting of the Haar-Walsh wavelet packets to the graph setting:

Regular Lattice	$\subset$ Graphs	# Choosable Bases	Costs
H-W Wavelet Packets[1]	$\subset$ GHWT[2]	$> (1.5)^N$	$O(N \log N)$
$\cap$	$\cap$	$\wedge$	$\|\|$
Adaptive H-W Tilings[3]	$\subset$ eGHWT[4]	$> 0.618 \cdot (1.84)^N$	$O(N \log N)$

- The difference between these two could be huge: for $N = 1024$, eGHWT searches 10^{270} possible bases whereas GHWT does 10^{181} bases.

[1]Coifman-Meyer (1989)
[2]Irion-Saito (2014)
[3]Thiele-Villemoes (1996)
[4]Saito-Shao (2019, 2022)

Outline

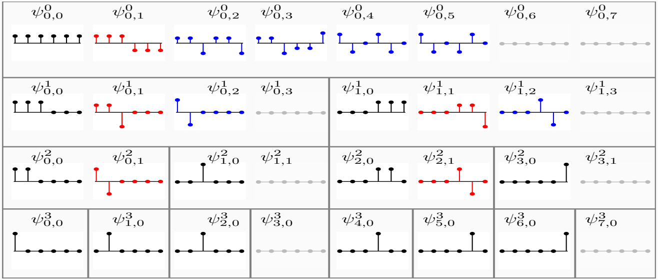
Time-Frequency Adapted Haar-Walsh Tilings

- Thiele and Villemoes (1996) proposed an $O(N \log N)$ algorithm to search the best basis among much larger collection of orthonormal bases than the conventional best-basis algorithm due to Coifman and Wickerhauser (1992) can search.
- The essence of this algorithm is that at each step of the recursive evaluation of the costs of subspaces, it compares the cost of the parent subspace with not only its two children subspaces partitioned in the "frequency" domain (like the wavelet packets), but also its two children subspaces partitioned in the "time" (or "space") domain (like the local cosines).
- Lindberg and Villemoes (2000) extended this algorithm for 2D signals and got quite good compression/approximation of various digital images.

Outline

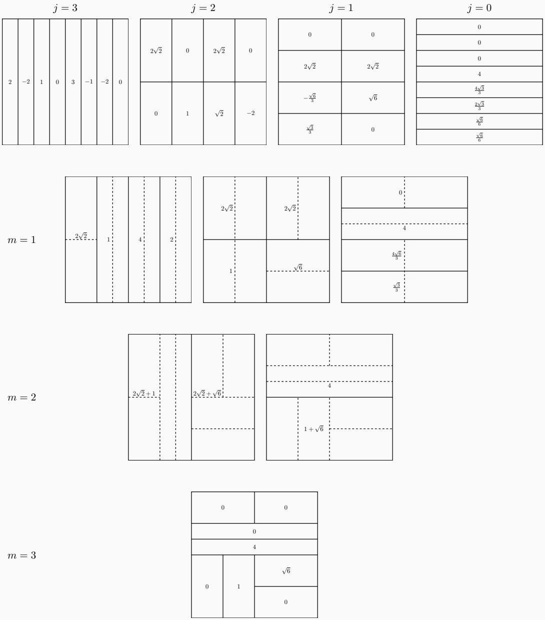
The eGHWT best-basis algorithm

1. Given the c2f GHWT dictionary, add *fictitious leaves* to the bottom of the tree so that all the leaves are in pair
2. Proceed upward as the GHWT to make it a *completely balanced tree*
3. Adjust the tag $\boldsymbol{l}$ and region index $\boldsymbol{k}$ of each $\boldsymbol{\psi}^j_{k,l}$ accordingly. $\mathbf{0}$ will be assigned as the expansion coefficients of an input graph signal relative to the newly added basis vectors due to those fictitious leaves:

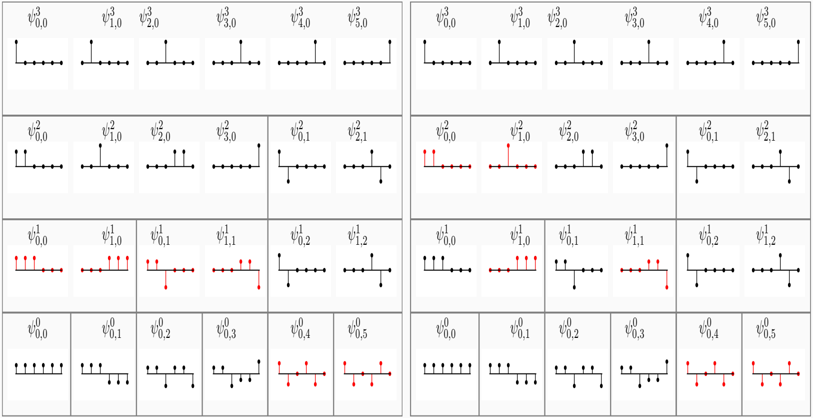


4. *Apply the Thiele-Villemoes algorithm* on this modified binary tree
5. *Restrict the support* of the best basis vectors selected from this balanced tree to the original nodes $\Longrightarrow$ *the best basis we want!*

Graphical Illustration with $f = [2, -2, 1, 3, -1, -2]^\top$ on P_6



eGHWT vs GHWT f2c best bases for $f = [2, -2, 1, 3, -1, -2]^{\mathsf{T}}$ on P_6

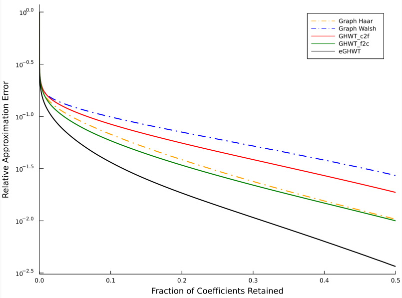


(a) The GHWT f2c best basis: $\|\hat{\boldsymbol{f}}\|_1 \approx \mathbf{7.84}$

(b) The eGHWT best basis: $\|\hat{\boldsymbol{f}}\|_1 \approx \mathbf{7.45}$

The Lindberg-Villemoes Algorithm (2D) via eGHWT

(a) Barbara



(b) Relative ℓ^2 approximation error

(a) The original Barbara image of size 512×512 pixels; (b) Relative ℓ^2 approximation errors by various graph bases

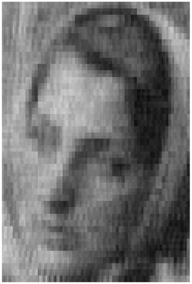
(a) Haar, PSNR = 24.50dB

(b) GHWT c2f, PSNR = 23.51dB

(c) GHWT f2c, PSNR = 25.27dB

(d) eGHWT, PSNR = 27.78dB

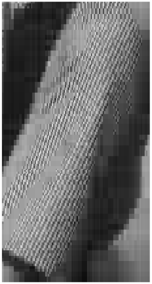
(a) Haar



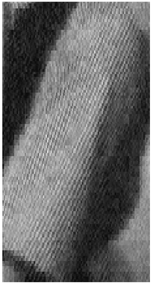
(b) GHWT c2f

(c) GHWT f2c

(d) eGHWT

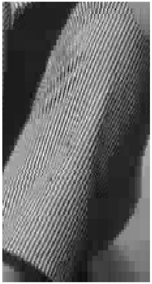


(a) Haar



(b) GHWT c2f

(c) GHWT f2c



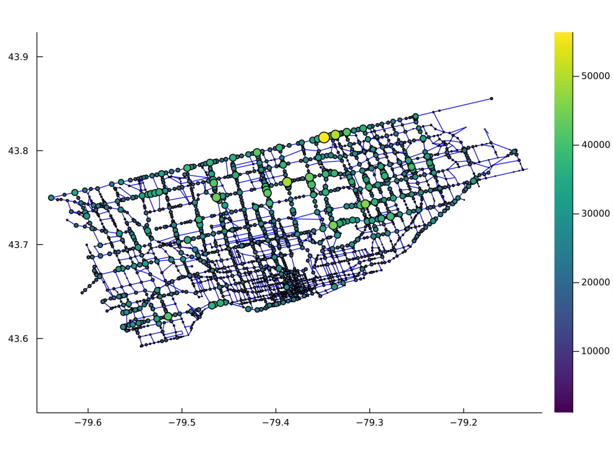
(d) eGHWT

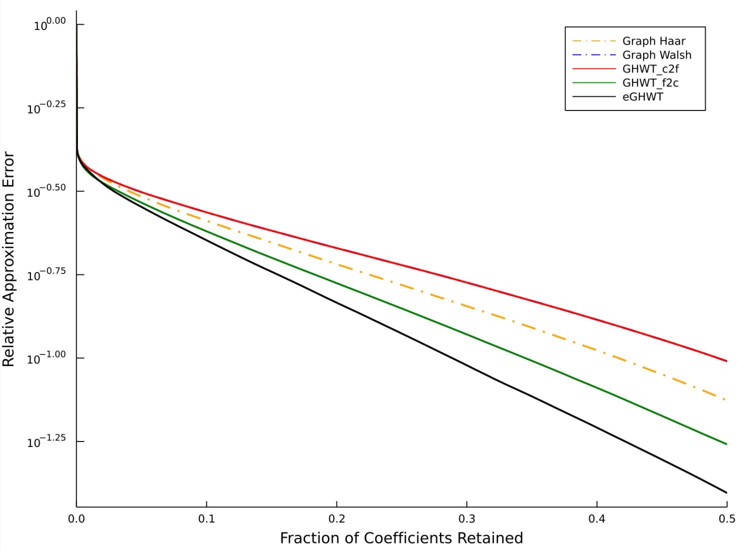
Outline

Outline

Graph Signal Compression

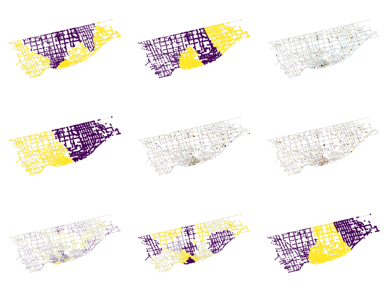
- Vehicular traffic volume data over 8 peak-hours at intersections in the street network of Toronto ($N = 2275$ nodes and $M = 3381$ edges) are used for comparing the performance of various graph bases
- Edge weights = 1/the Euclidean distances between the nodes



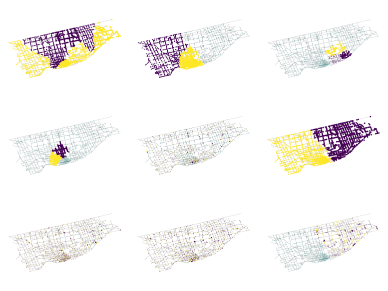


Relative ℓ^2 approximation error of the Toronto traffic data; GHWT-c2f $\equiv$ Graph Walsh for this example

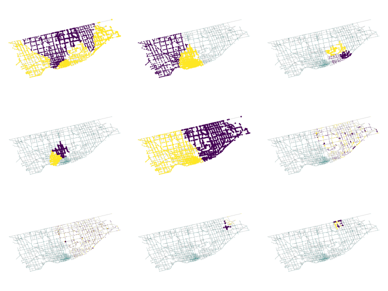
(a) Haar



(b) GHWT c2f = Walsh



(c) GHWT f2c



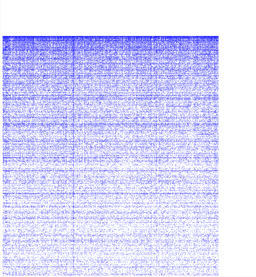
(d) eGHWT

Comparing nine most significant basis vectors

Outline

Matrix Data Analysis

- We often encounter data in the form of a *matrix*, e.g., a term-document matrix; a questionnaire; multiple sensor measurements; ...
- For example, look at the following Science News database where
  - Rows → preselected words
  - Columns → articles from 8 fields: Anthropology; Astronomy; Behavioral Sciences; Earth Sciences; Life Sciences; Math & CS; Medicine; Physics
  - a_{ij} → the relative frequency of word i appears in article j ⇒ all column sums are 1



Science News database (1153×1042)

Matrix Data Analysis ...

1. View matrix data as a *bipartite graph* and apply the *spectral co-clustering* to recursively partition the rows and the columns *simultaneously*
2. Analyze column vectors of the input matrix using the GHWT dictionary based on the row partitions and extract the best basis for handling columns as a whole, which we call the *row* best basis
3. Analyze row vectors of the input matrix using the GHWT dictionary based on the column partitions and extract the best basis for handling rows as a whole, which we call the *column* best basis
4. Expand the input matrix w.r.t. the *tensor product* of the row and column best bases
5. Analyze the expansion coefficients for a variety of tasks, e.g., compression, classification, regression, etc.

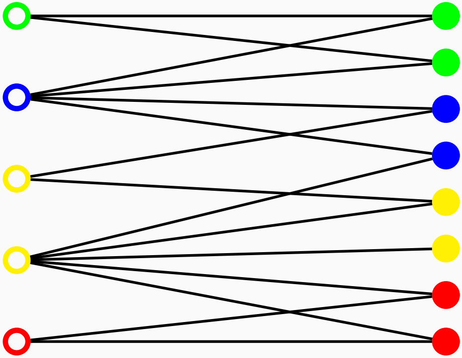
Spectral Co-Clustering (Dhillon, 2001)[1]

- Given a matrix $A \in \mathbb{R}_{\geq 0}^{N_r \times N_c}$ (e.g., a term-document matrix), the rows and columns are viewed as the two sets of nodes in a *bipartite* graph.
- a_{ij} denotes the edge weight between the ith row and the jth column.

[1] I. S. Dhillon: “Co-clustering documents and words using Bipartite Spectral Graph Partitioning,” *Proc. 7th ACM SIGKDD*, pp. 269–274, 2001.

Spectral Co-Clustering (Dhillon, 2001)[1]

- Given a matrix $A \in \mathbb{R}_{\geq 0}^{N_r \times N_c}$ (e.g., a term-document matrix), the rows and columns are viewed as the two sets of nodes in a *bipartite* graph.
- a_{ij} denotes the edge weight between the ith row and the jth column.



[1] I. S. Dhillon: “Co-clustering documents and words using Bipartite Spectral Graph Partitioning,” *Proc. 7th ACM SIGKDD*, pp. 269–274, 2001.

Spectral Co-Clustering (Dhillon, 2001) ...

- Then, matrices associated with this bipartite graph can be written as:

$$W = \begin{bmatrix} O & A \\ A^\top & O \end{bmatrix}$$ weighted adjacency matrix

$$D = \begin{bmatrix} D_r & O \\ O & D_c \end{bmatrix} \quad \begin{matrix} D_r := \mathrm{diag}(A\mathbf{1}) \\ D_c := \mathrm{diag}(A^\top \mathbf{1}) \end{matrix}$$ degree matrix

$$L := D - W = \begin{bmatrix} D_r & -A \\ -A^\top & D_c \end{bmatrix}$$ (unnormalized) graph Laplacian

$$L_{\mathrm{rw}} := D^{-1} L = I - D^{-1} W$$ random-walk normalized graph Laplacian

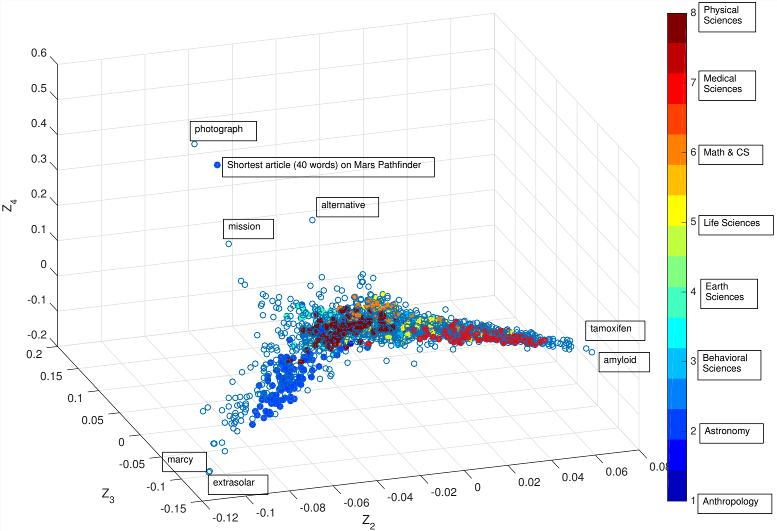
Spectral Co-Clustering (Dhillon, 2001) ...

- The *Fiedler vector* of L_{rw} bipartitions the bipartite graph:

$$\boldsymbol{\phi}_1 = \begin{bmatrix} D_r^{-1/2}\boldsymbol{u}_1 \\ D_c^{-1/2}\boldsymbol{v}_1 \end{bmatrix},$$

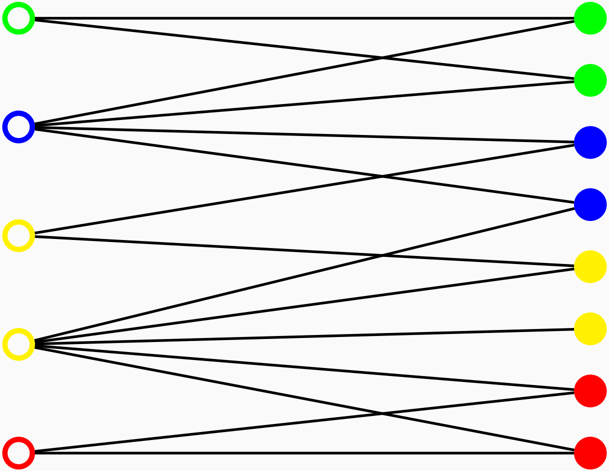
  where $\boldsymbol{u}_1$ and $\boldsymbol{v}_1$ are the second left and right singular vectors of $\tilde{A} = D_r^{-1/2} A D_c^{-1/2}$.
- The rows and the columns are partitioned *simultaneously*.
- This also allows the analysis of rows and columns *on an equal footing*, i.e., we can see not only which columns are similar but also which rows are closely related to a specific group of columns, etc.

An Example: Science News Dataset ...

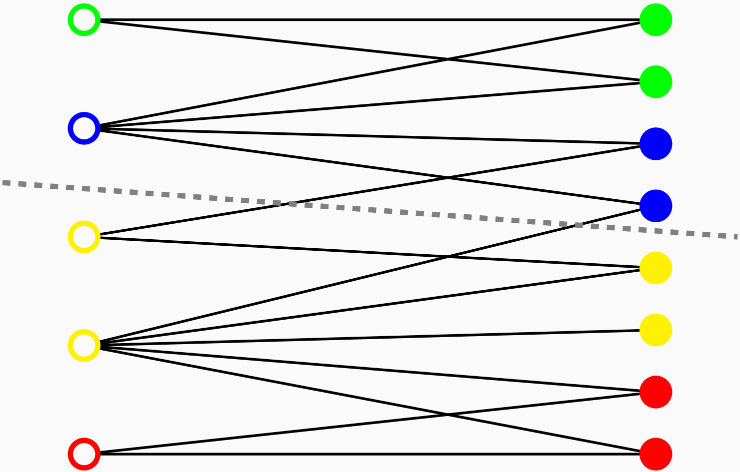


Words and articles embedded in $\{\phi_1, \phi_2, \phi_3\}$ at the top level.

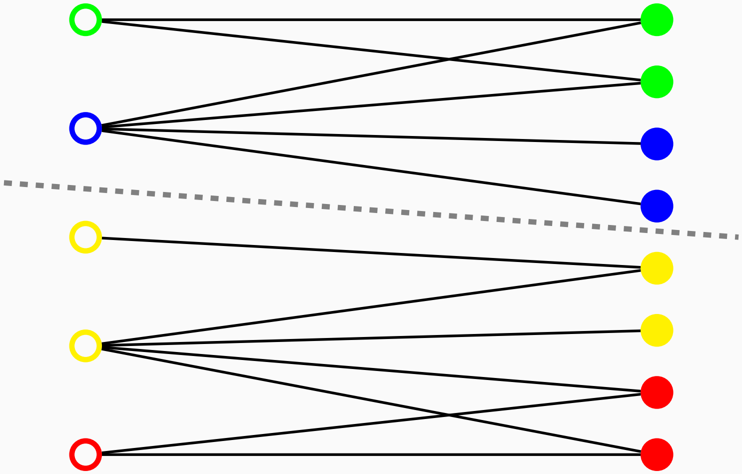
Matrix Partitioning à la Dhillon (2001)



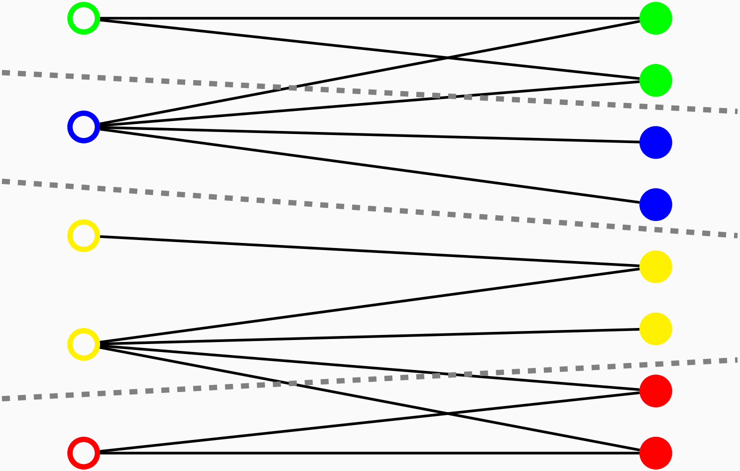
Matrix Partitioning à la Dhillon (2001)



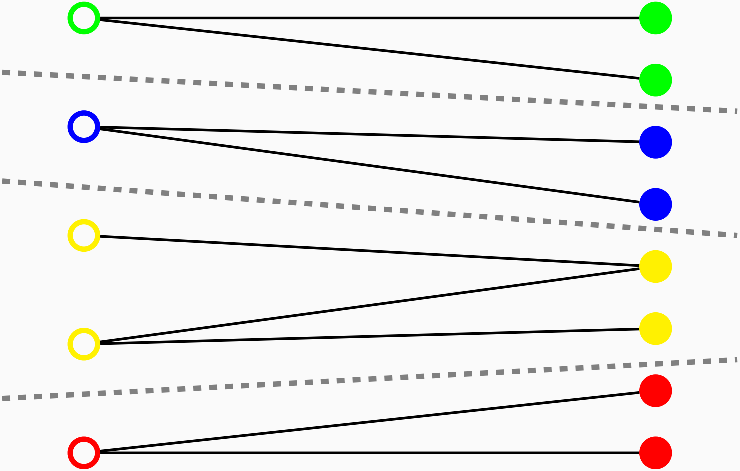
Matrix Partitioning à la Dhillon (2001)



Matrix Partitioning à la Dhillon (2001)



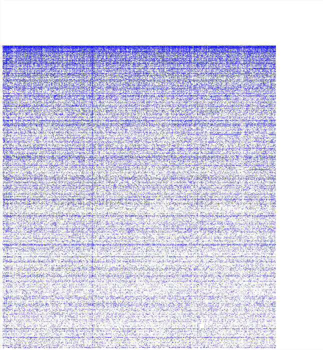
Matrix Partitioning à la Dhillon (2001)



Term-Document Matrix Analysis

Dataset: the Science News database (1153×1042)

- Rows → preselected words
- Columns → articles from 8 fields: Anthropology; Astronomy; Behavioral Sciences; Earth Sciences; Life Sciences; Math & CS; Medicine; Physics
- a_{ij} → the relative frequency of word i appears in article j ⇒ all column sums are 1
- GHWT/eGHWT best basis vectors for rows analyze meaningful groupings of words while those for columns do the same for documents

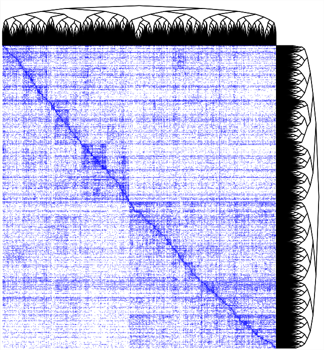


Science News database (original order)

Term-Document Matrix Analysis

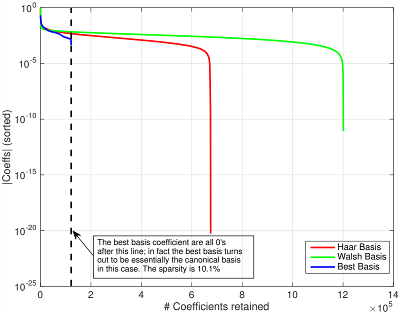
Dataset: the Science News database (1153×1042)

- Rows → preselected words
- Columns → articles from 8 fields: Anthropology; Astronomy; Behavioral Sciences; Earth Sciences; Life Sciences; Math & CS; Medicine; Physics
- a_{ij} → the relative frequency of word i appears in article j ⇒ all column sums are 1
- GHWT/eGHWT best basis vectors for rows analyze meaningful groupings of words while those for columns do the same for documents



Science News database (reordered rows and columns)

Example: Science News Dataset



Decay of the expansion coefficients w.r.t. Haar basis, Walsh basis, and GHWT best basis. The vertical line denotes the percentage of nonzero entries in the matrix (**10.1%**).

- Cost functional: 1-norm
- Total number of orthonormal bases searched: $> 10^{370}$
- **62.3%** of the Haar coefficients and **100%** of the Walsh coefficients must be kept to achieve perfect reconstruction, compared to **10.1%** for the GHWT best basis

$\Rightarrow$ The Haar and Walsh bases could not efficiently capture the underlying structure of this Science News dataset under the current matrix partitioning strategy!

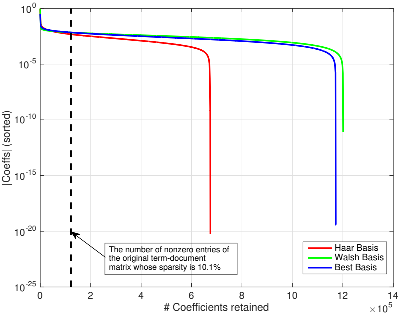
Example: Science News Dataset

- Since the sparsity was used as the cost functional, the best basis is in fact almost the canonical basis; the fine scale information was too much emphasized, which may be sensitive to 'noise'.
- We are interested in the *medium scale* information in this database, e.g., clustering structures both in words (rows) and articles (cols).
- Hence, we *weight* the coefficients in the GHWT dictionary:

$$\begin{aligned} c^j_{k,l} &\leftarrow c^j_{k,l} \cdot \left(\operatorname{supp}(G^0_0) / \operatorname{supp}(G^j_k)\right)^\alpha \\ &= c^j_{k,l} \cdot (N/N^j_k)^\alpha \end{aligned}$$

  where $\alpha \geq 0$ is chosen empirically to make the magnitude of the finer coefficients bigger, which discourages the best-basis algorithm to select fine scale subgraphs.
- See also Coifman-Leeb (2016); Ankenman-Leeb (2018) for such weighting scheme and its relation to the *Earth Mover's Distance*.

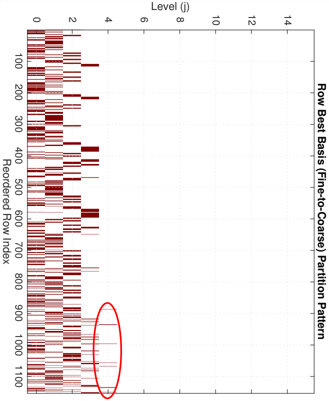
Example: Science News Dataset



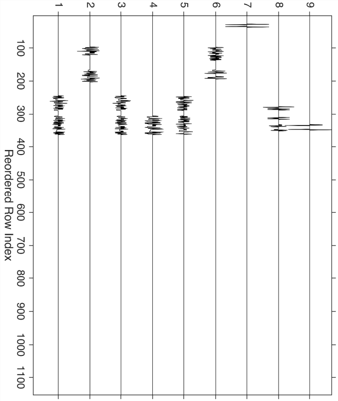
Decay of the expansion coefficients w.r.t. Haar basis, Walsh basis, and GHWT best basis. The vertical line denotes the percentage of nonzero entries in the matrix (**10.1%**).

- Cost functional: 1-norm
- $\alpha^{\text{row}} = 1.0, \alpha^{\text{col}} = 0.15$
- This best basis is less sparse than before, and is between the Haar and the Walsh bases, i.e., well captures information on intermediate scales.

Example: Science News Dataset

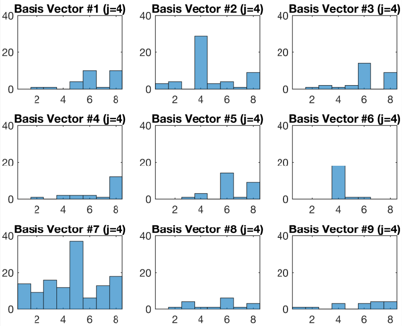


The *row* best basis (f2c) partition pattern.



The *row* best basis vectors at $j = 4$.

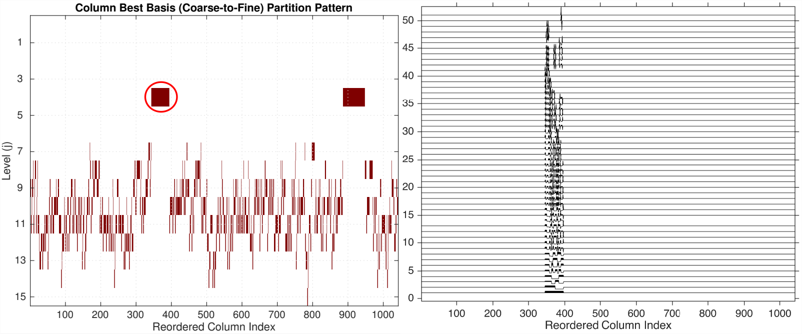
Example: Science News Dataset



The histograms of the article categories (1 to 8) of the expansion coefficients of column vectors w.r.t. those 9 row best basis vectors.

- For example, the positive components of the 6th basis vector correspond to the following words: `earthquake, down, california, dioxide, deep, warm, el, southern, crust, valley, once, geologist, bottom, tsunami, oxide, fault, antarctica, warning, tsunamis, prediction, greenhouse`
- On the other hand, the negative components of that vector correspond to: `temperature, ice, sea, layer, flow, around, survey, coast, warming, quake, past, nino, global, seismologist, cycle, cold, slow, recent, plate, thickness, meter, japan, forecast`
- Clearly, this basis vector is checking if a given article is in Category 4 (Earth Sciences).

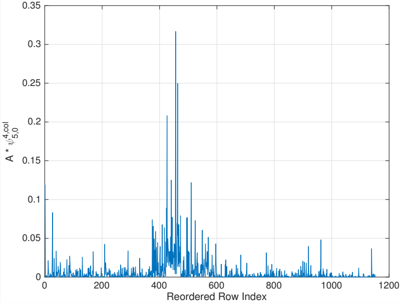
Example: Science News Dataset



The *column* best basis (c2f) partition pattern. The block indicated by a red circle corresponding to $(j,k) = (4,5)$.

The *column* best basis vectors with $(j,k) = (4,5)$ whose supports are 51 articles; 48 among 51 indicate 'Astronomy'.

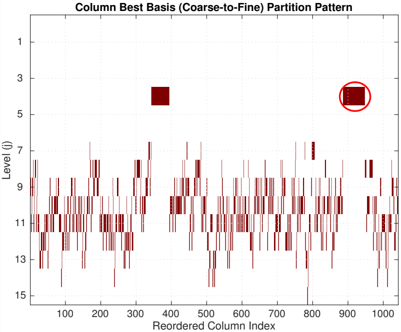
Example: Science News Dataset



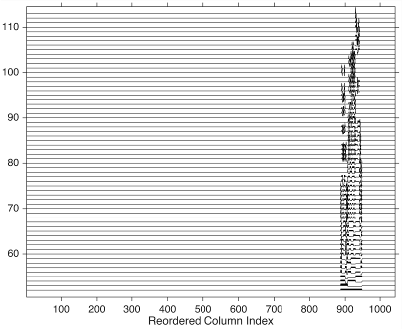
The expansion coefficients of row vectors w.r.t. the column best basis vector $\boldsymbol{\psi}^{4,\mathrm{col}}_{5,0}$ = the indicator vector of 51 articles.

- The 3 nonzero components in $\boldsymbol{\psi}^{4,\mathrm{col}}_{5,0}$ that are not in 'Astronomy' correspond to the following articles:
  - "`Old Glory, New Glory: The Star-Spangled Banner gets some tender loving care`" (Anthropology: on the preservation of the Star-Spangled Banner (flag) using the space-age technology);
  - "`Snouts: A star is born in a very odd way`" (Life Sciences: on star-nosed moles);
  - "`Gravity tugs at the center of a priority battle`" (Math & CS: on the priority war on the discovery of gravity between Newton, Halley, and Hooke).
- The expansion coefficients > 0.05 in the left figure correspond to the following words: `year, university, time, team, system, light, earth, star, planet, finding, astronomer, universe, galaxy, object, ray, telescope, orbit, mass, hole, dust, black, distance, disk, infrared`

Example: Science News Dataset

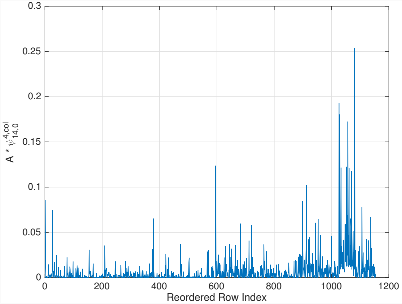


The *column* best basis (c2f) partition pattern. The block indicated by a red circle corresponding to $(j,k)=(4,14)$.



The *column* best basis vectors with $(j,k)=(4,14)$ whose supports are 62; 56 among 62 indicate 'Medical Sciences'.

Example: Science News Dataset



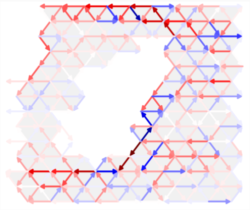
The expansion coefficients of row vectors w.r.t. the column basis vector $\boldsymbol{\psi}^{4,\mathrm{col}}_{14,0}$ = the indicator vector of 62 articles.

- Out of these 6 anomalies, 3 are in 'Life Sciences', i.e., not really surprising. The remaining 3 anomalies are:
  - `“In Silico Medicine: Computer simulations aid drug development and medical care”` (Math & CS);
  - `“Beyond Virtual Vaccinations: Developing a digital immune system in bits and bytes”` (Math & CS);
  - `“Paleopathological Puzzles: Researchers unearth ancient medical secrets”` (Anthropology).
- The expansion coefficients > 0.05 in the left figure correspond to the following words: `year, university, study, scientist, people, cell, group, disease, system, drug, protein, brain, human, blood, patient, test, immune, virus, strain, infection, vaccine, antibody, hiv, infected, aids, amyloid`

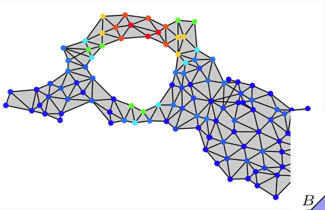
Outline

κ-GHWT

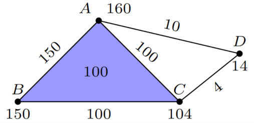
- Recently there has been great interest in analyzing and processing signals measured on *higher-order networks*.
- Data are sampled over C_κ, oriented *κ-simplices* of a graph, $\kappa \in \mathbb{N}$:
- For $\kappa = 0, 1, 2, 3, \ldots$, these signals take values over *nodes, edges, triangles, tetrahedra, ...*, respectively.
- The GHWT has been further generalized for *κ-simplices* via recursive bipartitions, which we call *κ-GHWT* [S.-Schonsheck-Shvarts (2024)]
- Recursive bipartitions are done by the Fiedler vectors of *Hodge Laplacians* (generalization of graph Laplacians)



Flows around Madagascar [Schaub et al. (2020)]

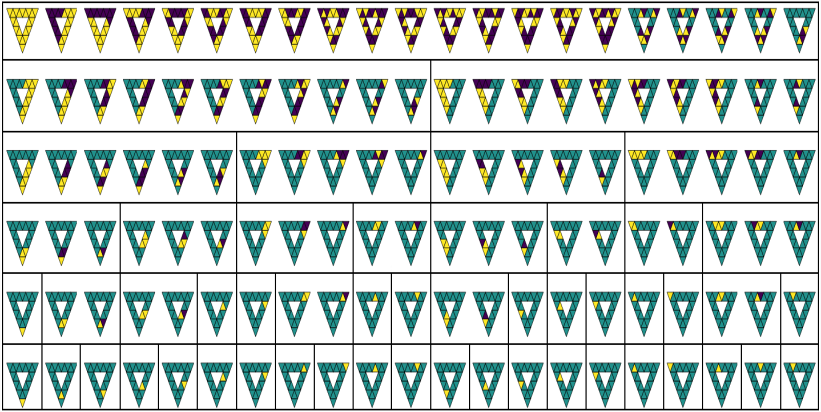


Gene expression correlations [Govek et al. (2019)]



Coauthorship graph [Ebli et al. (2022)]

The Coarse-to-Fine GHWT Dictionary on the Triangle Complex



Each row represents one level of the bipartition; Color represents the sign info

The *Fine-to-Coarse* GHWT Dictionary on the Triangle Complex

Color represents the sign info; the red boxes correspond to the *2-Haar Basis*

Applications of κ-GHWT

- Combined with the *scattering transform*, we proposed the *Multiscale Hodge Scattering Networks* [S.-Schonsheck-Shvarts (2024)]
- Application includes: Classification of Science News articles; graph classification; potential energy prediction in molecular dynamics simulation, ...
- Our results indicate that MHSNs provide comparable results with those by the state-of-the-art GNNs with up to a *two-order of magnitude reduction in number of learnable parameters*.
- We strongly believe that our success here comes from the *structure and organization of our multiscale basis dictionaries* that are conveniently arranged in terms of scales, locations, and frequencies.
- We also proposed how to extract *explainable features* from scattering transform coefficients [S.-Weber (2025)].
- No time to show these today, so please check our preprints: `arXiv:2311.10270 [cs.LG]` and `arXiv:2502.05722 [cs.LG]`!

Outline

Summary

- The eGHWT best-basis algorithm searches over an immense number of orthonormal bases, much more than the conventional GHWT best-basis algorithm does.
- When selected using an appropriate cost functional, the eGHWT best basis outperforms the graph Haar/Walsh bases, the conventional GHWT best basis.
- Graph signal compression and denoising demonstrate an advantage of a *data-adaptive basis dictionary* from which one can select the most suitable basis for one's task at hand!
- Combining the spectral co-clustering and GHWT leads to a powerful tool to analyze *matrix data*, e.g., term-document matrices, microarray data, etc.

Future Plan

- Explore different cost functionals than the sparsity $\Longrightarrow$ Local Discriminant Basis (LDB) and Local Regression Basis (LRB) of Saito and Coifman for classification and regression problems.
- Investigate the use of eGHWT for *operator compression* in numerical analysis (in collaboration with Raphy Coifman and Pei-Chun Su)
- What to do if your input data is of *tensor* form, i.e., $A = (a_{ijk}) \in \mathbb{R}^{I \times J \times K}$? $\Longrightarrow$ a *tripartite* graph (a.k.a. 3-uniform *hypergraph*)!

Outline

References

The following articles (and the other related ones) are available at
https://www.math.ucdavis.edu/~saito/publications/

- J. Irion & N. Saito: "The generalized Haar-Walsh transform," in *Proc. 2014 IEEE Workshop on Statistical Signal Processing*, pp. 472–475, 2014.
- J. Irion & N. Saito: "Applied and computational harmonic analysis on graphs and networks," in *Wavelets and Sparsity XVI, Proc. SPIE 9597*, #95971F, 2015.
- J. Irion & N. Saito: "Learning sparsity and structure of matrices with multiscale graph basis dictionaries," in *Proc. 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, 2016.
- J. Irion & N. Saito: "Efficient approximation and denoising of graph signals using the multiscale basis dictionaries,", *IEEE Trans. Signal Inform. Process. Netw.*, vol. 3, no. 3, pp. 607–616, 2017.
- N. Saito & Y. Shao: "eGHWT: The extended Generalized Haar-Walsh Transform," *J. Math. Imaging, Vis.*, vol. 64, no. 3, pp. 261–283, 2022.
- N. Saito, S. Schonsheck, & E. Shvarts: "Multiscale transforms for signals on simplicial complexes," *Sampling Theory, Signal Processing, and Data Analysis*, vol. 22, no. 1, Article #2, 2024.
- N. Saito, S. Schonsheck, & E. Shvarts: "Multiscale Hodge scattering networks for data analysis," *arXiv:2311.10270 [cs.LG]*, 2024.

Please check our Julia codes on GitHub!!

https://github.com/UCD4IDS/MultiscaleGraphSignalTransforms.jl

https://github.com/UCD4IDS/MultiscaleSimplexSignalTransforms.jl

Split $\Longrightarrow$ "Organize" $\Longrightarrow$ Merge

Thank you very much for your attention!