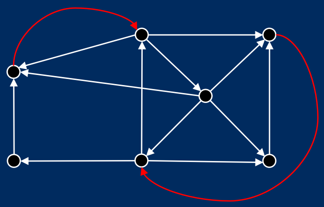
Fourier Analysis with Direction

M. Püschel



Department of Computer Science

ETH zürich

An Approach to Generalizing Signal Processing

M. Püschel and J. M. F. Moura
Algebraic Signal Processing Theory
http://arxiv.org/abs/cs.IT/0612077, 2006

M. Püschel and J. M. F. Moura
Algebraic Signal Processing Theory: Foundation and 1-D Time
IEEE TSP, 2008

M. Püschel and J. M. F. Moura
Algebraic Signal Processing Theory: 1-D Space
IEEE TSP, 2008

J. Kovacevic and M. Püschel
Algebraic Signal Processing Theory: Sampling for Infinite and Finite 1-D Space
IEEE TSP, 2010

M. Püschel and M. Rötteler
Algebraic Signal Processing Theory: 2-D Spatial Hexagonal Lattice
IEEE TIP, 2007

A. Sandryhaila, J. Kovacevic and M. Püschel
Algebraic Signal Processing Theory: 1-D Nearest-Neighbor Models
IEEE TSP, 2012

M. Püschel and M. Rötteler
Fourier Transform for the Directed Quincunx Lattice
Proc. ICASSP, 2005

M. Püschel and M. Rötteler
Fourier Transform for the Spatial Quincunx Lattice
Proc. ICASSP, 2005

Concise Derivation of Fast Transforms

M. Püschel and J. M. F. Moura
Algebraic Signal Processing Theory: Cooley-Tukey Type Algorithms for DCTs
IEEE TSP, 2008

Y. Voronenko and M. Püschel
Algebraic Signal Processing Theory: Cooley-Tukey Type Algorithms for Real DFTs
IEEE TSP, 2009

M. Püschel and M. Rötteler
Algebraic Signal Processing Theory: Cooley-Tukey Type Algorithms on the 2-D Spatial Hexagonal Lattice
Applicable Algebra, 2008

A. Sandryhaila, J. Kovacevic and M. Püschel
Algebraic Signal Processing Theory: Cooley-Tukey Type Algorithms for Polynomial Transforms Based on Induction
SIAM JMAA, 2011

Key Insights

Axiomatic foundation

Signal model: $(\mathcal{A}, \mathcal{M}, \Phi)$

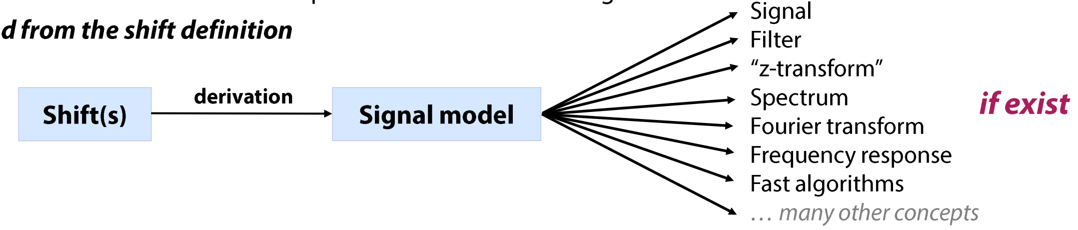
algebra of filters

signal ***module***

associated "z-transform"

Signal model defined: Other concepts follow from abstract algebra

Derived from the shift definition



Shift(s)

derivation

Signal model

- Signal
- Filter
- "z-transform"
- Spectrum
- Fourier transform
- Frequency response
- Fast algorithms
- *… many other concepts*

if exist

The shift is everything!

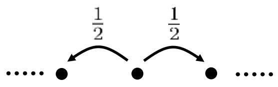
$(\mathcal{A}, \mathcal{M}, \Phi)$ shift invariant $\Longleftrightarrow$ $\mathcal{A}$ is commutative = a polynomial algebra

Detailed insight into boundary conditions: why needed and which are good

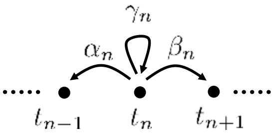
---

Considered Shifts Back Then (boundary conditions omitted)

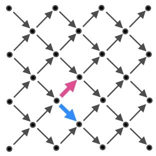
discrete time



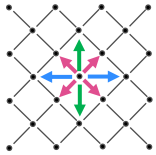
discrete space
(undirected path graph)



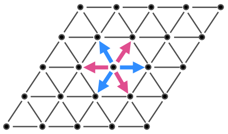
weighted path graph



directed quincunx



undirected quincunx



undirected triangle grid

M. Püschel and J. M. F. Moura
Algebraic Signal Processing Theory
http://arxiv.org/abs/cs.IT/0612077, 2006

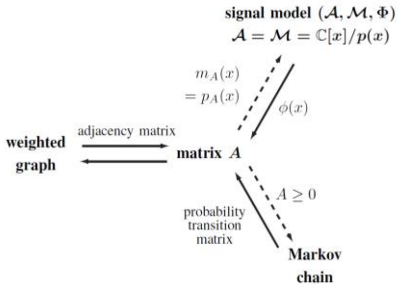


Fig. 22. The connection between square matrices A, weighted graphs, regular signal models, and finite Markov chains.

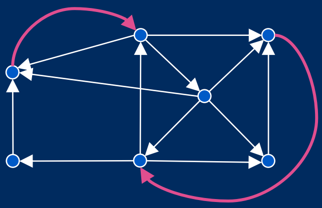
Every weighted graph defines a signal model

Signal model with one shift = (adjacency) GSP

“Detailed insight into boundary conditions:
why needed and which are good”

Part 1:
Fourier Analysis on Directed Graphs

with
Bastian Seifert



Digraph Signal Processing with Generalized Boundary Conditions, IEEE TSP 2021

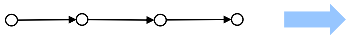
Graph Signal Processing…

Concept	Undirected Graphs
Shift/Variation operator	✓ (Adjacency or Laplacian) Symmetric
Convolution	✓
Fourier Basis/Transform (eigendecomposition)	✓
Orthogonality	✓

Graph Signal Processing…

Concept	**Undirected Graphs**	**Directed Graphs (Digraphs)**
Shift/Variation operator	✓ (Adjacency or Laplacian) Symmetric	✓ (Adjacency or Laplacian) Not symmetric
Convolution	✓	✓
Fourier Basis/Transform (eigendecomposition)	✓	✗ **Does not exist in general**
Orthogonality	✓	✗ (In general no)

Digraph Example:



$$A = \begin{bmatrix} 0&1&0&0\\ 0&0&1&0\\ 0&0&0&1\\ 0&0&0&0 \end{bmatrix}, \quad L = D - A = \begin{bmatrix} 1&-1&0&0\\ 0&1&-1&0\\ 0&0&1&-1\\ 0&0&0&1 \end{bmatrix}$$

No Eigendecomposition/Fourier basis
Only one eigenvalue (Jordan Block)

Digraph Signal Processing is an Open Problem

Graph Signal Processing: Overview, Challenges, and Applications, Ortega et al., Proc. IEEE, vol. 106(5), pp. 808–828, 2018

Jordan normal form + tricks:

Spectral Projector-Based Graph Fourier Transforms
Joya A. Deri, *Member, IEEE,* and José M. F. Moura, *Fellow, IEEE*

Agile Inexact Methods for Spectral Projector-Based Graph Fourier Transforms
Joya A. Deri, *Member, IEEE,* and José M. F. Moura, *Fellow, IEEE*

Graph Fourier Transform based on Directed Laplacian
Rahul Singh, Abhishek Chakraborty, *Graduate Student Member, IEEE,* and B. S. Manoj, *Senior Member, IEEE*

Fourier transform via optimization:

On the Graph Fourier Transform for Directed Graphs
Stefania Sardellitti, *Member, IEEE,* Sergio Barbarossa, *Fellow, IEEE,* and Paolo Di Lorenzo, *Member, IEEE*

A Directed Graph Fourier Transform With Spread Frequency Components
Rasoul Shafipour, *Student Member, IEEE,* Ali Khodabakhsh, *Student Member, IEEE,* Gonzalo Mateos, *Senior Member, IEEE,* and Evdokia Nikolova

DIGRAPH FOURIER TRANSFORM VIA SPECTRAL DISPERSION MINIMIZATION
Rasoul Shafipour[†], Ali Khodabakhsh[‡], Gonzalo Mateos[†], and Evdokia Nikolova[‡]

Approximate diagonalization:

Graph Fourier Transform: A Stable Approximation
João Domingos and José M. F. Moura

Complex-valued Shift:

Graph Signal Processing for Directed Graphs based on the Hermitian Laplacian
Satoshi Furutani[1] (✉), Toshiki Shibahara[1], Mitsuaki Akiyama[1], Kunio Hato[1], and Masaki Aida[2]

ORTHOGONAL TRANSFORMS FOR SIGNALS ON DIRECTED GRAPHS
Julia Barrufet[1], Antonio Ortega[2]

Subset of diagonalizable filters:

DIAGONALIZABLE SHIFT AND FILTERS FOR DIRECTED GRAPHS BASED ON THE JORDAN-CHEVALLEY DECOMPOSITION
*Panagiotis Misiakos**
Chris Wendler, Markus Püschel

Our Key Idea: Boundary Conditions

Classical Signal Processing

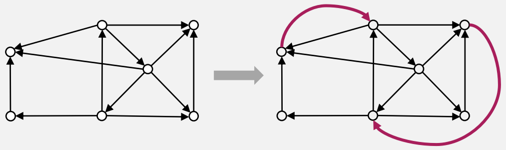
Digraph for finite discrete time

$$A + B = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathrm{DFT}_4 (A + B)\, \mathrm{DFT}_4^{-1} = \begin{bmatrix} 1 & \cdot & \cdot & \cdot \\ \cdot & \mathrm{i} & \cdot & \cdot \\ \cdot & \cdot & -1 & \cdot \\ \cdot & \cdot & \cdot & -\mathrm{i} \end{bmatrix}$$

This is done even though
the signal is usually not periodic!

Can we do this for other digraphs?

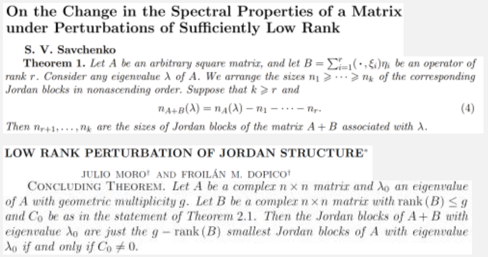


$$\begin{bmatrix} \boxed{\begin{matrix} 0 & 1 \\ 0 & 0 \end{matrix}} & & & & \\ & \boxed{\begin{matrix} 0 & 1 \\ 0 & 0 \end{matrix}} & & & \\ & & 1 & & \\ & & & \omega_3^2 & \\ & & & & \omega_3 \end{bmatrix} \rightarrow \begin{bmatrix} \lambda_1 & & & & & & \\ & \lambda_2 & & & & & \\ & & \lambda_3 & & & & \\ & & & \lambda_4 & & & \\ & & & & \lambda_5 & & \\ & & & & & \lambda_6 & \\ & & & & & & \lambda_7 \end{bmatrix}$$

Goal: Add small number of edges to obtain Fourier basis of eigenvectors

Edges to Destroy Jordan Blocks

Tool: Perturbation Theory



On the Change in the Spectral Properties of a Matrix under Perturbations of Sufficiently Low Rank

S. V. Savchenko

Theorem 1. *Let A be an arbitrary square matrix, and let $B = \sum_{i=1}^{r} (\cdot, \xi_i)\eta_i$ be an operator of rank r. Consider any eigenvalue λ of A. We arrange the sizes $n_1 \geqslant \cdots \geqslant n_k$ of the corresponding Jordan blocks in nonascending order. Suppose that $k \geqslant r$ and*

$$n_{A+B}(\lambda) = n_A(\lambda) - n_1 - \cdots - n_r. \quad (4)$$

Then $n_{r+1}, \ldots, n_k$ are the sizes of Jordan blocks of the matrix $A + B$ associated with λ.

LOW RANK PERTURBATION OF JORDAN STRUCTURE*

JULIO MORO† AND FROILÁN M. DOPICO†

CONCLUDING THEOREM. *Let A be a complex $n \times n$ matrix and λ_0 an eigenvalue of A with geometric multiplicity g. Let B be a complex $n \times n$ matrix with rank $(B) \leq g$ and C_0 be as in the statement of Theorem 2.1. Then the Jordan blocks of $A + B$ with eigenvalue λ_0 are just the $g -$ rank (B) smallest Jordan blocks of A with eigenvalue λ_0 if and only if $C_0 \neq 0$.*

Our work: Specialize to Adjacency/Laplacian matrices

Corollary: Adding an edge is enough to **destroy** the largest **Jordan block** to a choosen eigenvalue.

Q: How to find these edges?

Theorem: Let $u_1, \ldots, u_r, v_1, \ldots, v_r$ be left/right eigenvectors of Jordan blocks to the eigenvalue λ and B the matrix containig only the new edge, then if

$$\sum_{k=1}^{r} u_k^T B v_k \neq 0$$

the largest Jordan block of λ gets destroyed in $A+B$.

$$\begin{tikzcd} A \arrow[r] \arrow[d] & A + B \arrow[d] \\ \begin{bmatrix} \boxed{\begin{matrix} \lambda & 1 & & \\ & \lambda & 1 & \\ & & \ddots & 1 \\ & & & \lambda \end{matrix}} & \\ & \ddots \end{bmatrix} \arrow[r] & \begin{bmatrix} \lambda_1 & & & & \\ & \lambda_2 & & & \\ & & \ddots & & \\ & & & \lambda_n & \\ & & & & \ddots \end{bmatrix} \end{tikzcd}$$

Jordan-Normal-Form

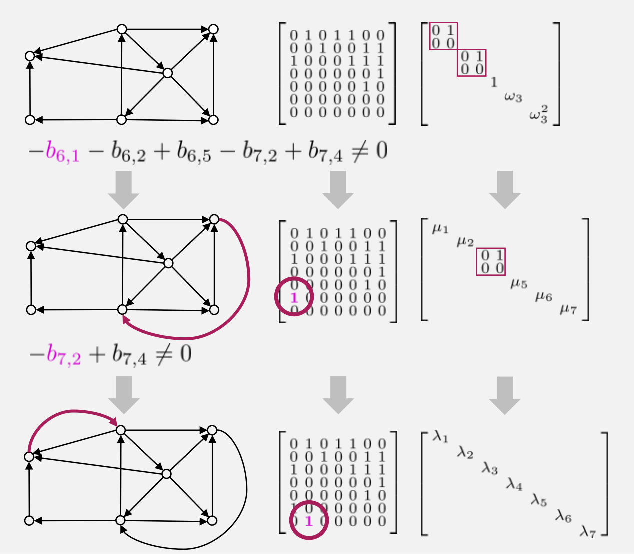
The Algorithm

```
function DestroyAllJordanBlocks($A$)
    while $A$ not diagonalizable do
        $u_1, \ldots, u_r \leftarrow$ left EVs to largest Jordan blocks
        $v_1, \ldots, v_r \leftarrow$ right EVs to largest Jordan blocks
        if $\exists(i,j)$ s.t. $\sum_k u_{k,j} v_{k,i} \neq 0$ and $A_{i,j} = 0$ then
            $A_{i,j} \leftarrow 1$
        else
            select $(i,j)$ random s.t. $A_{i,j} = 0$
            $A_{i,j} \leftarrow 1$
        end if
    end while
    return $A$
end function
```

Numerical Implementation:
Needs some tricks, see paper

Faster, Inexact Implementation:
Eigenvalue 0 appears often, use that

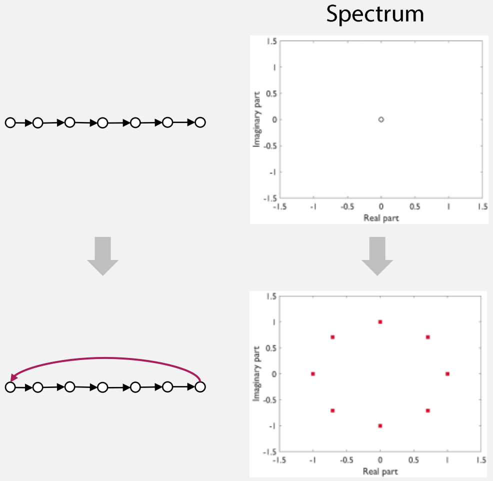
Variant: Destroy zero eigenvalues so Fourier transform is numerically stable



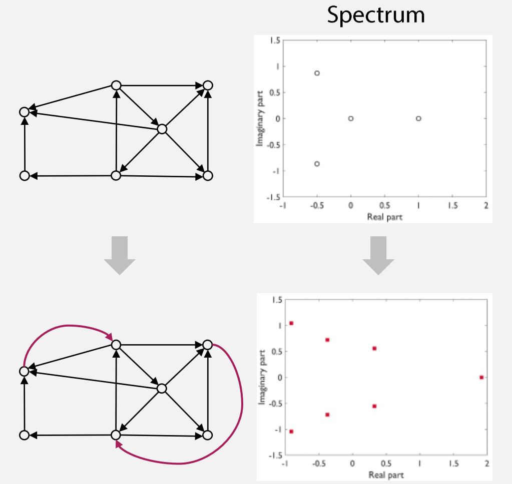
Also works with Laplacian D – A …

Behaviour of Eigenvalues? (Spectrum)

Finite Discrete Time



with cyclic boundary = periodic extension



Generally Applicable & Fast

Random digraphs with different properties, **500 nodes & ~5000 edges**

	min		median		max	
	edges	time	edges	time	edges	time
Watts-Strogatz	0	0.2s	1	0.5s	3	1.3s
Barabási-Albert	36	4.4s	44	10s	55	31s
Klemm-Eguílez	10	2.2s	27	6s	47	9s

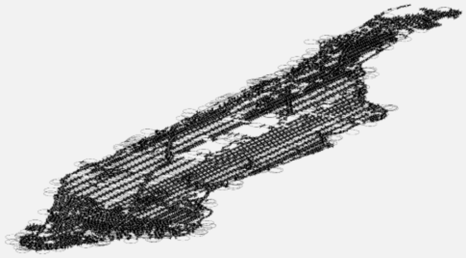
Medium number of edges added: 27
Median runtime: 6 seconds

Scalable

Manhattan Taxi Graph

Li & Moura, ECAI, 2020

5464 nodes & 11568 edges



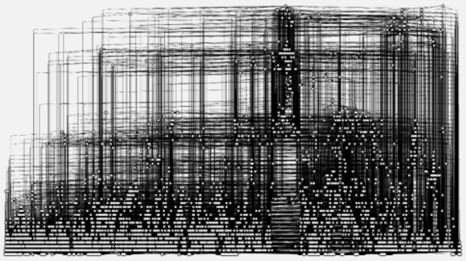
Runtime: 19 hours, 243 edges added

Runtime (inexact algo): **5 min**, 772 edges added

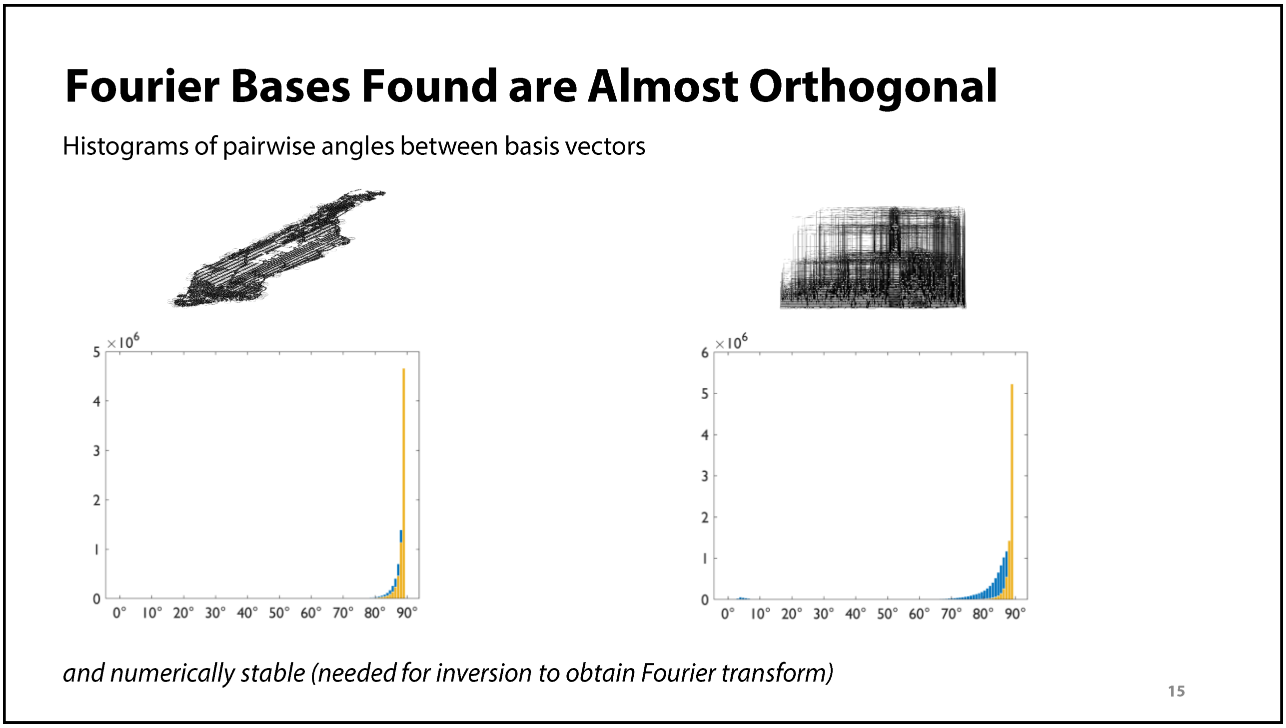
Citation Graph

https://snap.stanford.edu/data/cit-HepPh.html

4989 nodes & 17840 edges
(almost acyclic)



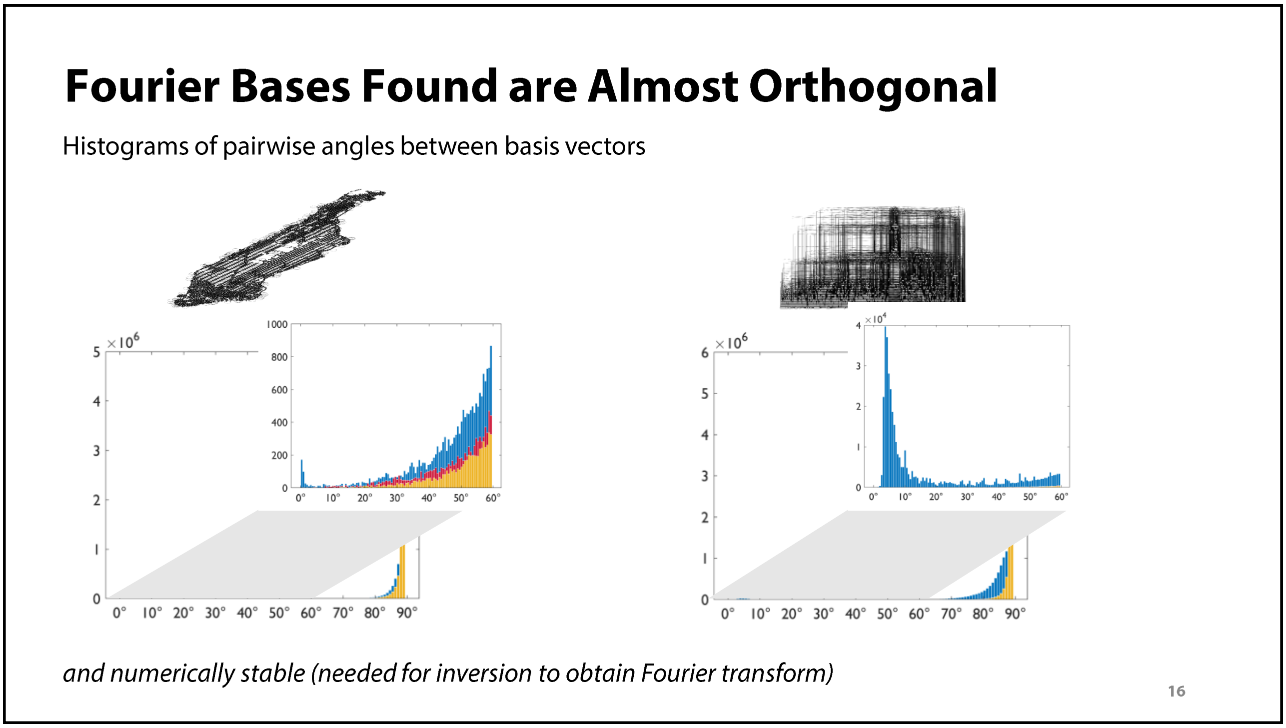
Runtime (inexact algo): **31.5 min**, 1911 edges added



Fourier Bases Found are Almost Orthogonal

Histograms of pairwise angles between basis vectors

and numerically stable (needed for inversion to obtain Fourier transform)



Fourier Bases Found are Almost Orthogonal

Histograms of pairwise angles between basis vectors

and numerically stable (needed for inversion to obtain Fourier transform)

One Application: Denoising with Neural Networks

S. Chen, Y. C. Eldar, and L. Zhao, **Graph Unrolling Networks: Interpretable Neural Networks for Graph Signal Denoising**, IEEE TSP 2021

ported to digraphs (needs a Fourier basis)

V. Mihal, B. Seifert, M. Püschel, **Porting Signal Processing from Undirected to Directed Graphs: Case Study Signal Denoising with Unrolling Networks**, Proc. EUSIPCO 2022

	ForgetDir	AdHoc	SelfLoops	UndShift	AssocShift
$p = 0.5$					
SprFreq	0.43 ± 0.01	0.46 ± 0.01	0.35 ± 0.01	$\mathbf{0.19} \pm 0.00$	-
HermL	0.28 ± 0.01	0.31 ± 0.01	0.19 ± 0.00	$\mathbf{0.07} \pm 0.01$	0.19 ± 0.04
StabApp	0.13 ± 0.00	$\mathbf{0.08} \pm 0.00$	0.09 ± 0.00	0.10 ± 0.00	-
GenBC1	0.10 ± 0.00	$\mathbf{0.01} \pm 0.00$	$\mathbf{0.01} \pm 0.00$	0.03 ± 0.00	$\mathbf{0.01} \pm 0.00$
GenBC2	0.10 ± 0.00	$\mathbf{0.01} \pm 0.00$	$\mathbf{0.01} \pm 0.00$	0.03 ± 0.00	$\mathbf{0.01} \pm 0.00$
$p = 0.9$					
SprFreq	0.19 ± 0.01	0.19 ± 0.01	0.12 ± 0.00	$\mathbf{0.06} \pm 0.00$	-
HermL	0.21 ± 0.01	0.28 ± 0.01	0.16 ± 0.01	$\mathbf{0.07} \pm 0.00$	0.19 ± 0.01
StabApp	0.24 ± 0.00	$\mathbf{0.10} \pm 0.00$	0.11 ± 0.00	0.11 ± 0.00	-
GenBC1	0.15 ± 0.01	$\mathbf{0.03} \pm 0.00$	$\mathbf{0.03} \pm 0.01$	0.06 ± 0.01	0.05 ± 0.00
GenBC2	0.15 ± 0.00	0.06 ± 0.00	0.06 ± 0.00	0.05 ± 0.00	$\mathbf{0.01} \pm 0.00$

TABLE I: NMSE results of denoising low-frequency signals.

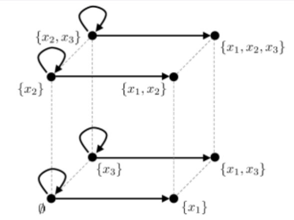
works well

"shift defines signal model"

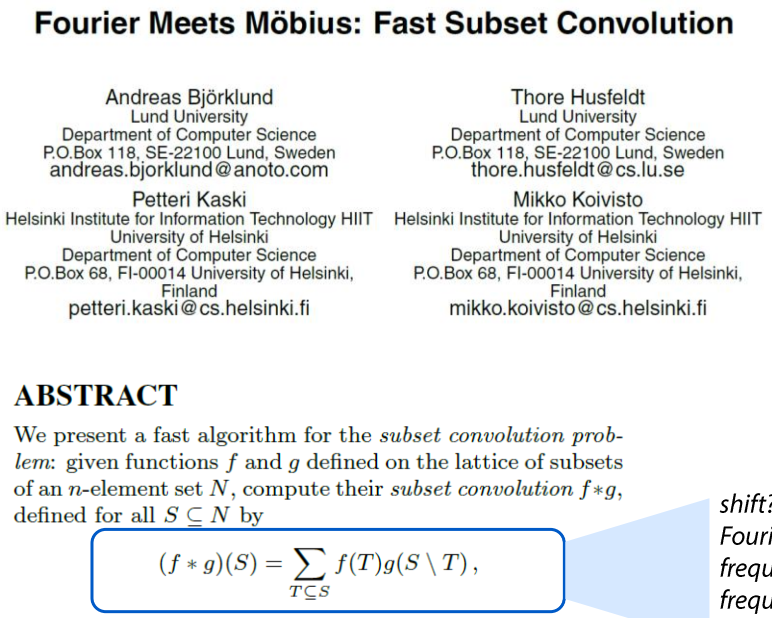
"SP with one shift = GSP"

Part 2.1:
Fourier Analysis on Powersets

with
Chris Wendler



Discrete Signal Processing with Set Functions, IEEE SP 2021
Learning Set Functions that are Sparse in Non-Orthogonal Fourier Bases. Proc. AAAI 2021



Fourier Meets Möbius: Fast Subset Convolution

Andreas Björklund
Lund University
Department of Computer Science
P.O.Box 118, SE-22100 Lund, Sweden
andreas.bjorklund@anoto.com

Thore Husfeldt
Lund University
Department of Computer Science
P.O.Box 118, SE-22100 Lund, Sweden
thore.husfeldt@cs.lu.se

Petteri Kaski
Helsinki Institute for Information Technology HIIT
University of Helsinki
Department of Computer Science
P.O.Box 68, FI-00014 University of Helsinki,
Finland
petteri.kaski@cs.helsinki.fi

Mikko Koivisto
Helsinki Institute for Information Technology HIIT
University of Helsinki
Department of Computer Science
P.O.Box 68, FI-00014 University of Helsinki,
Finland
mikko.koivisto@cs.helsinki.fi

Proc. STOC 2007

ABSTRACT

We present a fast algorithm for the *subset convolution problem*: given functions f and g defined on the lattice of subsets of an n-element set N, compute their *subset convolution* $f * g$, defined for all $S \subseteq N$ by

$$(f * g)(S) = \sum_{T \subseteq S} f(T)g(S \setminus T)\,,$$

shift?
Fourier transform?
frequency ordering?
frequency response?
applications?

Signals on Powersets = Set Functions

Finite set: $N = \{x_1, \dots, x_n\}$

Its power set: $2^N = \{A \mid A \subseteq N\}$

Set function: $\mathbf{s} = (s_A)_{A \subseteq N}$

Shift by $x_i \in N$ $\quad \mathbf{s} \mapsto (s_{A \setminus \{x_i\}})$

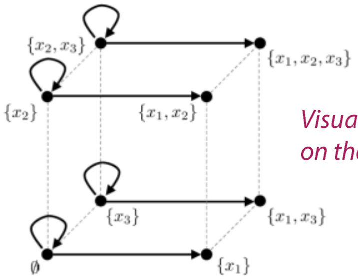
"Delay by x_i"

not invertible

n shifts!

A. Krause and D. Golovin, **Tractability: Practical Approaches to Hard Problems. ch. Submodular function maximization**, pp. 71–104, 2014

Applications: Recommender systems, image processing, facility location, online auctions etc.



Visualization of shift by x_1 on the associated "z-transform"

Signal domain = directed hypercube

Derived Signal Model and SP Concepts

Shift(s) by $x_i \in N$ $\quad \mathbf{s} \mapsto (s_{A\setminus\{x_i\}})$ *n shifts commute: shift-invariance*

Convolution (filter): $\quad (\mathbf{h} * \mathbf{s})_A = \sum_{X\subseteq N} h_X s_{X\setminus A}$

Fourier basis (as matrix): $\left[\begin{smallmatrix}1 & 0\\ 1 & -1\end{smallmatrix}\right] \otimes \ldots \otimes \left[\begin{smallmatrix}1 & 0\\ 1 & -1\end{smallmatrix}\right]$

Fourier transform: $\quad \widehat{s}_B = \sum_{A\subseteq B} (-1)^{|A|} s_A$

as matrix: $\quad \mathrm{DSFT}_{2^n} = \left[\begin{smallmatrix}1 & 0\\ 1 & -1\end{smallmatrix}\right] \otimes \ldots \otimes \left[\begin{smallmatrix}1 & 0\\ 1 & -1\end{smallmatrix}\right]$ *diagonalizes all shifts and filters*
not orthogonal, lower triangular

$n2^{n-1}$ ops

Frequency response: $\quad \overline{\mathbf{h}}_B = \sum_{A\subseteq N, A\cap B=\emptyset} (-1)^{|A|} h_A$

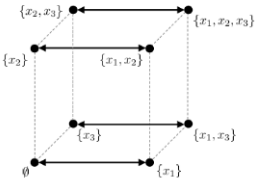
Convolution theorem: $\quad \widehat{\mathbf{h} * \mathbf{s}} = \overline{\mathbf{h}} \odot \widehat{\mathbf{s}}$

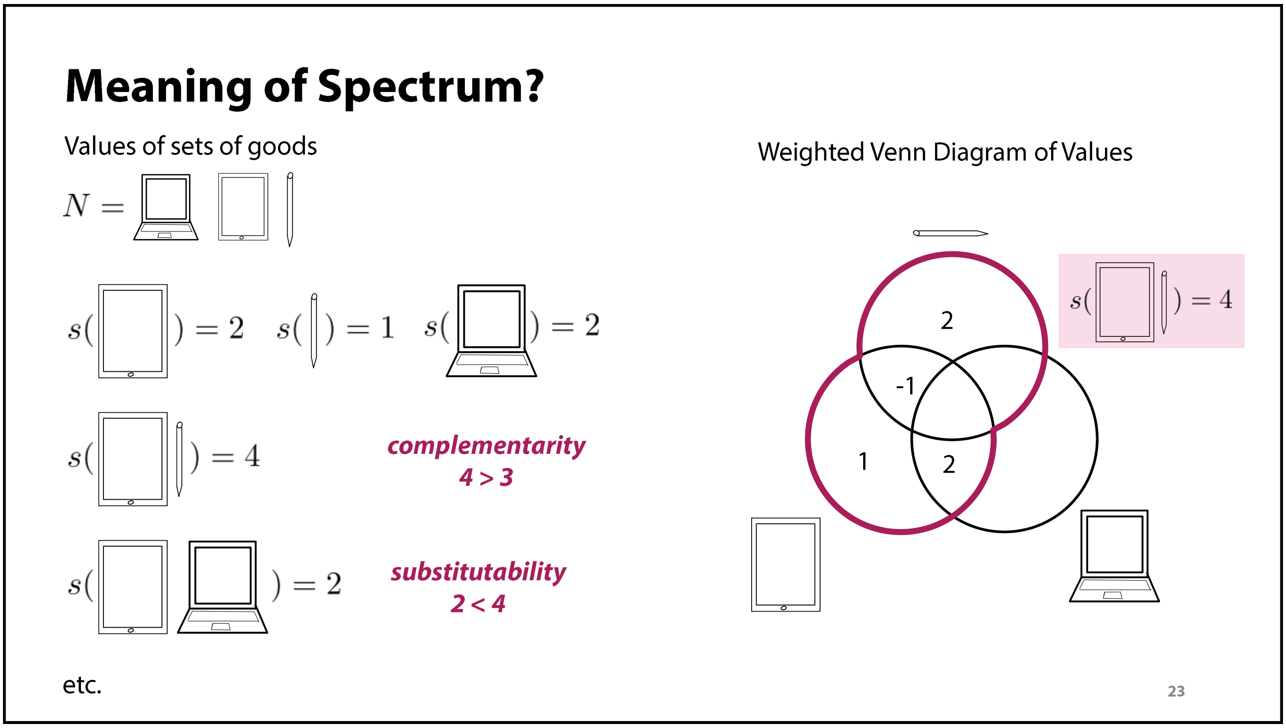
Four Variants of Signal Models

model	$q\,A$	on signal	$Q\,A$	on signal	$(\mathbf{h} * \mathbf{s})_A$	matrix for q
1	$A \cup \{q\}$	$\begin{cases} s_A + s_{A\setminus\{q\}}, & q \in A \\ 0, & \text{else}\end{cases}$	$A \cup Q$	$\begin{cases} \sum_{A\setminus Q\subseteq B\subseteq A} s_B, & Q \subseteq A \\ 0, & \text{else}\end{cases}$	$\sum_{Q\cup B=A} h_Q s_B$	$\begin{bmatrix}0 & 0\\ 1 & 1\end{bmatrix}$
2	$A \setminus \{q\}$	$\begin{cases} s_A + s_{A\cup\{q\}}, & q \notin A \\ 0, & \text{else}\end{cases}$	$A \setminus Q$	$\begin{cases} \sum_{B\subseteq Q} s_{A\cup B}, & Q \subseteq N \setminus A \\ 0, & \text{else}\end{cases}$	$\sum_{Q\subseteq N\setminus A} \sum_{B\subseteq Q} h_Q s_{A\cup B}$	$\begin{bmatrix}1 & 1\\ 0 & 0\end{bmatrix}$
3	$\begin{cases} A + A \cup \{q\}, & q \notin A \\ 0, & \text{else}\end{cases}$	$s_{A\setminus\{q\}}$	$\begin{cases} \sum_{B\subseteq Q} A \cup B, & Q \subseteq N \setminus A \\ 0, & \text{else}\end{cases}$	$s_{A\setminus Q}$	$\sum_{Q\subseteq N} h_Q s_{A\setminus Q}$	$\begin{bmatrix}1 & 0\\ 1 & 0\end{bmatrix}$
4	$\begin{cases} A + A \setminus \{q\}, & q \in A \\ 0, & \text{else}\end{cases}$	$s_{A\cup\{q\}}$	$\begin{cases} \sum_{B\subseteq Q} A \setminus B, & Q \subseteq A \\ 0, & \text{else}\end{cases}$	$s_{A\cup Q}$	$\sum_{Q\subseteq N} h_Q s_{A\cup Q}$	$\begin{bmatrix}0 & 1\\ 0 & 1\end{bmatrix}$
5	$A \setminus \{q\} \cup \{q\} \setminus A = \begin{cases} A \cup \{q\}, & q \notin A \\ A \setminus \{q\}, & \text{else}\end{cases}$	$s_{A\setminus\{q\}\cup\{q\}\setminus A} = \begin{cases} s_{A\cup\{q\}}, & q \notin A \\ s_{A\setminus\{q\}}, & \text{else}\end{cases}$	$A \setminus Q \cup Q \setminus A$	$s_{A\setminus Q\cup Q\setminus A}$	$\sum_{Q\subseteq N} h_Q s_{A\setminus Q\cup Q\setminus A}$	$\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$

Classical model for set functions
Fourier transform = Walsh-Hadamard transform

shift(s): $\quad \mathbf{s} \mapsto (s_{A\setminus\{x_i\}} \cup s_{\{x_i\}\setminus A})_{A\subseteq N}$





Meaning of Spectrum?

Values of sets of goods

$N =$

$s(\;) = 2$ $s(\;) = 1$ $s(\;) = 2$

$s(\;) = 4$

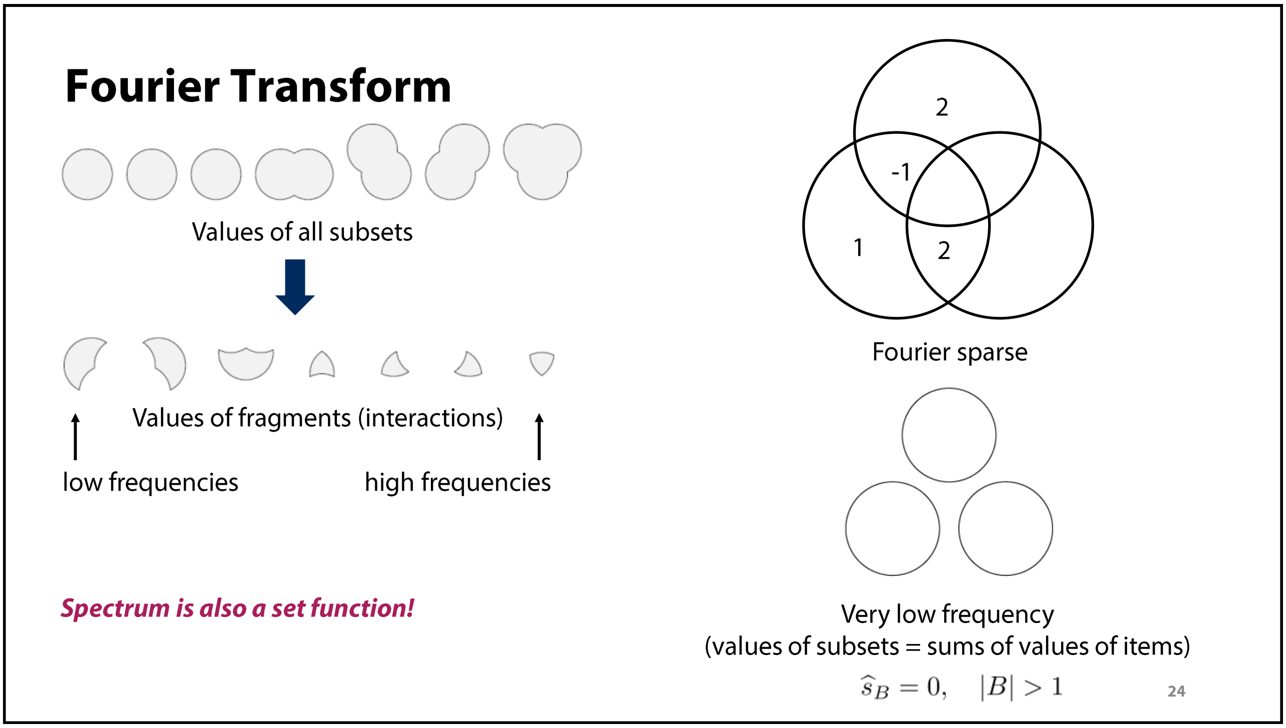
complementarity
4 > 3

$s(\;) = 2$

substitutability
2 < 4

etc.

Weighted Venn Diagram of Values



Fourier Transform

Values of all subsets

Values of fragments (interactions)

low frequencies

high frequencies

Spectrum is also a set function!

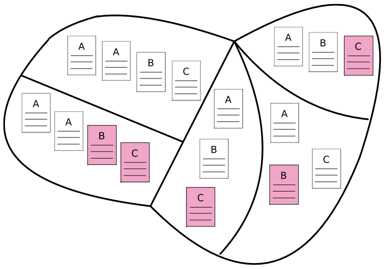
Fourier sparse

Very low frequency
(values of subsets = sums of values of items)

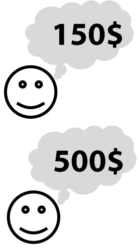
$\widehat{s}_B = 0, \quad |B| > 1$

Application: Preference Elicitation in Spectrum Auctions

N = licenses of bands electromagnetic spectrum

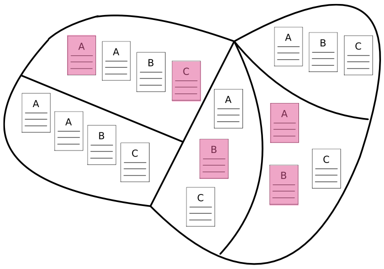


Set functions = bidders

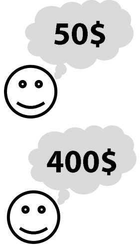


Application: Preference Elicitation in Spectrum Auctions

N = licenses of bands electromagnetic spectrum

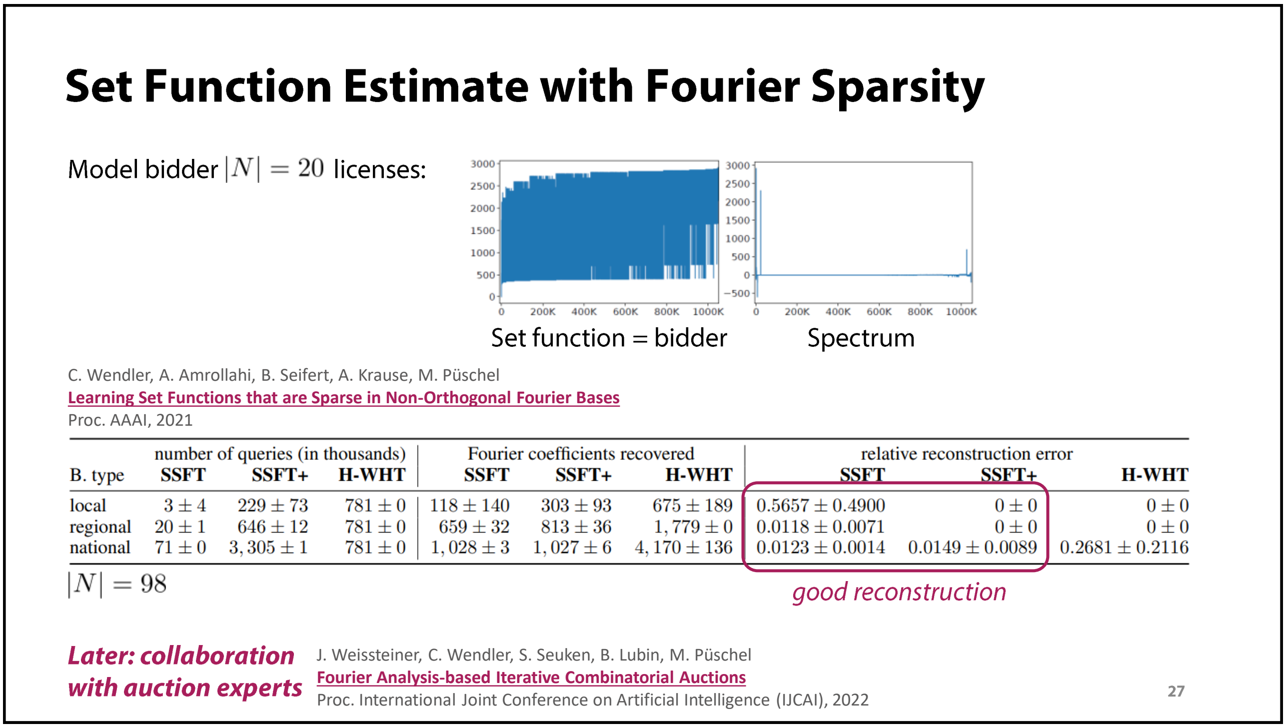


Set functions = bidders



Preference elicitation: estimate bidders from few (500 say) samples (values of subsets)

Idea: assume Fourier sparsity



Set Function Estimate with Fourier Sparsity

Model bidder $|N| = 20$ licenses:

Set function = bidder

Spectrum

C. Wendler, A. Amrollahi, B. Seifert, A. Krause, M. Püschel
Learning Set Functions that are Sparse in Non-Orthogonal Fourier Bases
Proc. AAAI, 2021

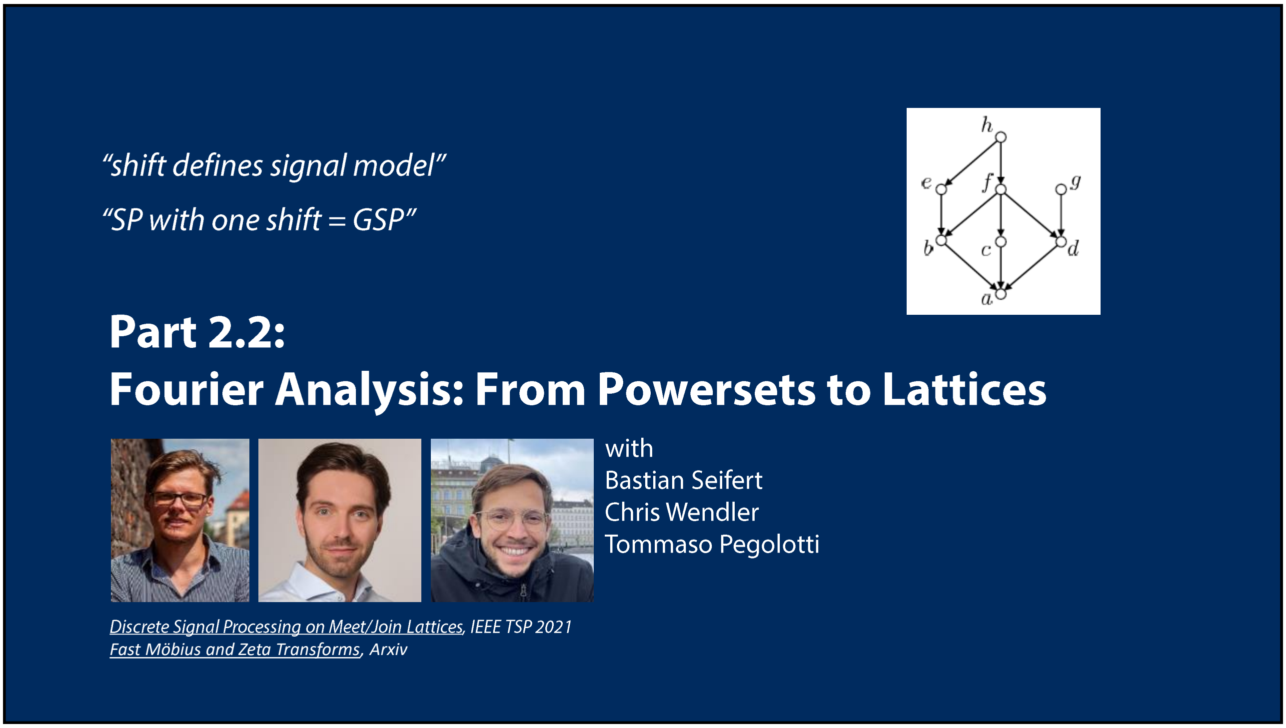
	number of queries (in thousands)			Fourier coefficients recovered			relative reconstruction error		
B. type	**SSFT**	**SSFT+**	**H-WHT**	**SSFT**	**SSFT+**	**H-WHT**	**SSFT**	**SSFT+**	**H-WHT**
local	3 ± 4	229 ± 73	781 ± 0	118 ± 140	303 ± 93	675 ± 189	0.5657 ± 0.4900	0 ± 0	0 ± 0
regional	20 ± 1	646 ± 12	781 ± 0	659 ± 32	813 ± 36	$1,779 \pm 0$	0.0118 ± 0.0071	0 ± 0	0 ± 0
national	71 ± 0	$3,305 \pm 1$	781 ± 0	$1,028 \pm 3$	$1,027 \pm 6$	$4,170 \pm 136$	0.0123 ± 0.0014	0.0149 ± 0.0089	0.2681 ± 0.2116

$|N| = 98$

good reconstruction

Later: collaboration with auction experts

J. Weissteiner, C. Wendler, S. Seuken, B. Lubin, M. Püschel
Fourier Analysis-based Iterative Combinatorial Auctions
Proc. International Joint Conference on Artificial Intelligence (IJCAI), 2022



"shift defines signal model"

"SP with one shift = GSP"

Part 2.2:
Fourier Analysis: From Powersets to Lattices

with
Bastian Seifert
Chris Wendler
Tommaso Pegolotti

Discrete Signal Processing on Meet/Join Lattices, IEEE TSP 2021
Fast Möbius and Zeta Transforms, Arxiv

Generalization to (Meet) Lattices

Power set convolution:

$$\begin{aligned}(\mathbf{h} * \mathbf{s})_A &= \sum_{X \subseteq N} h_X s_{X \setminus A} \\ &= \sum_{X \subseteq N} h_X s_{X \cap N \setminus A} \\ &\approx \sum_{X \subseteq N} h_X s_{X \cap A}\end{aligned}$$

Power set is an example of a meet lattice
meet operation = $\cap$

Meet lattice convolution:

$$(\mathbf{h} * \mathbf{s})_a = \sum_{x \in \mathcal{L}} h_x s_{x \wedge a}$$

639 pages
No convolution or Fourier transform

Prior work:
A. Björklund, T. Husfeldt, P. Kaski, M. Koivisto, J. Nederlof, and P. Parviainen, **Fast zeta transforms for lattices with few irreducibles**, ACM Trans. on Algorithms, 2015.

(Meet) Lattices

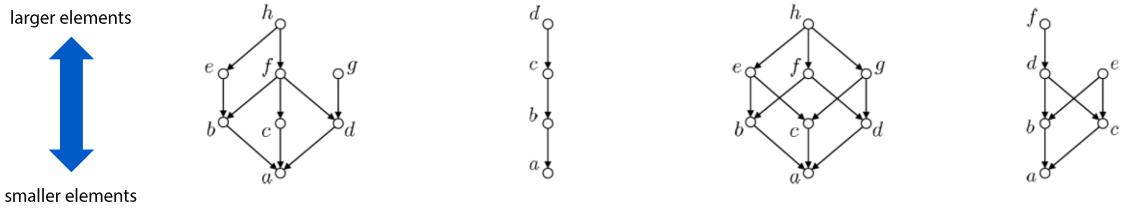
A lattice $\mathcal{L}$ is a (here finite) poset:
set with a partial order $\leq$

For $a, b, c \in \mathcal{L}$:

1. $a \leq a$,
2. $a \leq b$ and $b \leq a$ implies $a = b$, and
3. $a \leq b$ and $b \leq c$ implies $a \leq c$

For each $a, b \in \mathcal{L}$ there is a unique greatest lower bound $a \wedge b$

1. $a \wedge a = a$,
2. $a \wedge b = b \wedge a$, and
3. $(a \wedge b) \wedge c = a \wedge (b \wedge c)$



Some lattice — Total order lattice — Subset lattice (hypercube) — Not a lattice

Lattices $\subset$ DAGs (directed acyclic graphs)

Derived Signal Model and SP Concepts

Lattice signal: $\mathbf{s} = (s_x)_{x \in \mathcal{L}}$

Shift(s) by $a \in \mathcal{L}$ $\mathbf{s} \mapsto (s_{x \wedge a})_{x \in \mathcal{L}}$

"x gets delayed by a"
shifts commute: shift-invariance

Convolution (filter): $(\mathbf{h} * \mathbf{s})_a = \sum_{x \in \mathcal{L}} h_x s_{x \wedge a}$

Inverse Fourier transform: $s_x = \sum_{y \le x} \widehat{s}_y$

partial order

Fourier basis consists of 0/1 vectors
IFT sums spectral values of all predecessors

Fourier transform: $\widehat{s}_y = \sum_{x \le y} \mu(x, y) s_x$

$\mu(x,x) = 1,$ *Moebius function*

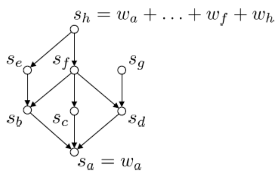
$\mu(x,y) = -\sum_{x \le z < y} \mu(x,z), \quad x \neq y$

diagonalizes all shifts and filters
not orthogonal, lower triangular

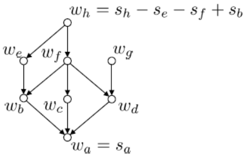
Spectrum: is partially ordered isomorphic to $\mathcal{L}$

G.-C. Rota, **On the foundations of combinatorial theory. I. theory of Möbius functions**, Z. Wahrscheinlichkeitstheorie und Verwandte Gebiete, 1964.

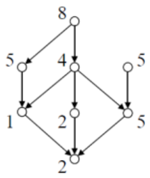
Example



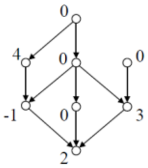
Signal
$s_h = w_a + w_b + w_c + w_d + w_e + w_f + w_h$
$s_g = w_a + w_d + w_g$
$s_f = w_a + w_b + w_c + w_d + w_f$
$s_e = w_a + w_b + w_e$
$s_d = w_a + w_d$
$s_c = w_a + w_c$
$s_b = w_a + w_b$
$s_a = w_a$



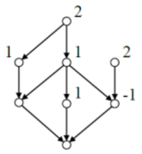
Spectrum
$w_h = s_h - s_f - s_e + s_b$
$w_g = s_g - s_d$
$w_f = s_f - s_d - s_c - s_b + 2s_a$
$w_e = s_e - s_b$
$w_d = s_d - s_a$
$w_c = s_c - s_a$
$w_b = s_b - s_a$
$w_a = s_a$



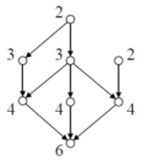
signal



spectrum



basic low pass filter
1 + sum of all shifts



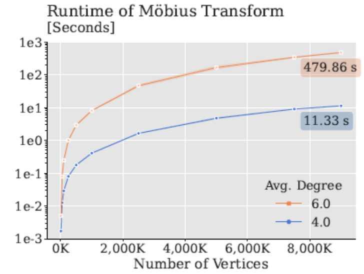
its frequency response

Applications: see paper

Fast Lattice Fourier Transform

Can be computed in O(nk), k = longest antichain in lattice

Enables computation for millions of nodes if DAG is sparse = low average degree



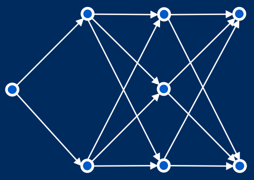
+ close to linear speedup when parallelized

T. Pegolotti, B. Seifert, M. Püschel, **Fast Möbius and Zeta Transforms,** Arxiv

---

“shift defines signal model”

“SP with one shift = GSP”



Part 2.3:
Fourier Analysis: From Lattices to DAGs

with
Bastian Seifert
Chris Wendler
Panos Misiakos
Vedran Mihal

Causal Fourier Analysis on Directed Acyclic Graphs and Posets, Arxiv
Learning DAGs from Data with Few Root Causes, Arxiv

Lattices once more:

Convolution $(\mathbf{h} * \mathbf{s})_a \;=\; \sum_{x \in \mathcal{L}} h_x s_{x \wedge a}$

Inverse Fourier transform: $s_x = \sum_{y \leq x} \widehat{s}_y$

Fourier transform: $\widehat{s}_y = \sum_{x \leq y} \mu(x, y) s_x$

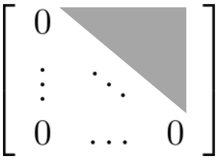
Chris Wendler

"Uses only the partial order. I think we don't need the lattice property. Partial order/DAG is enough"

But how to do the shift without $\wedge$?
Also, we need weighted edges for broader applicability ….

DAGs are, in a sense, "worst case" in GSP:

But highly relevant in causal reasoning, Bayesian networks etc.

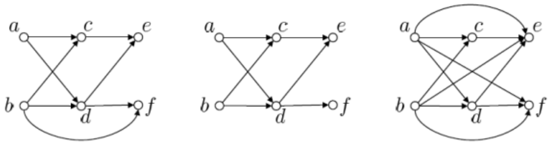


adjacency matrix
only 0 eigenvalue

DAGs $\approx$ Posets

Every DAG defines a unique partial order: $a \leq b \Longleftrightarrow a \text{ is a predecessor of } b \text{ in the DAG}$

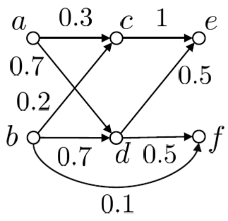
A partial order can be represented by several DAGs:



some DAG *transitive reduction* *transitive closure*

all define the same partial order

Signal Model for DAGs



Weighted DAG: $D = (V, E, A)$ with induced $\leq$

Observed signal on DAG: $\mathbf{s} = (s_x)_{x \in V}$

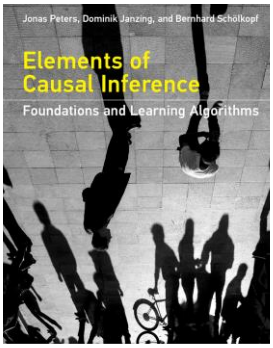
The signal satisfies: $s_x = \sum_{y \leq x} w_{y,x} c_y$

Unknown value produced by each node that we call ***cause***

Weights
How obtained from A?

Weighted version of the previous inverse Fourier transform on lattices

This equation does not imply causality



Jonas Peters, Dominik Janzing, and Bernhard Schölkopf

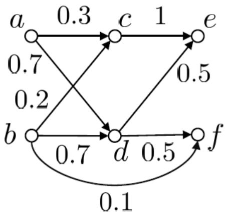
Elements of Causal Inference

Foundations and Learning Algorithms

$$s_x = \sum_{y \leq x} w_{y,x} c_y$$

$$\mathbf{s} = W\mathbf{c}$$

Weights
How obtained from A?



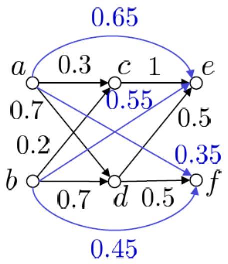
River network
Nodes: cities
Edges: rivers
Edge weights: fraction of pollution propagated

c_y unknown pollution inserted by city y
s_x measured pollution at city x

accumulated from all predecessor nodes

In this case: $W = I + A + A^2 + \dots A^{n-1} = (I - A)^{-1}$

$(+, \cdot)$-transitive closure of A



Fourier transform = $W^{-1} = I - A$:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ -0.3 & -0.2 & 1 & 0 & 0 & 0 \\ -0.7 & -0.7 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & -0.5 & 1 & 0 \\ 0 & -0.1 & 0 & -0.5 & 0 & 1 \end{bmatrix}$$

Transitive closures

S	$u \oplus v$	$u \odot v$	0_S	1_S	Meaning of edge weight $\overline{a}_{x,y}$ in closure
$\{0,1\}$	u or v	u and v	0	1	Reachability in graphs; x is cause of y
$[0,1]$	$u+v$	$u \cdot v$	0	1	Fraction of pollution from x reaching y
$[0,1]$	$\max(u,v)$	$u \cdot v$	0	1	Strongest influence/most reliable path from x to y
$\mathbb{R}^+ \cup \{\infty\}$	$\min(u,v)$	$u+v$	∞	0	Shortest path length from x to y
$\mathbb{R}^+ \cup \{\infty\}$	$\max(u,v)$	$\min(u,v)$	0	∞	Largest capacity path from x to y

← prior slide

S. K. Abdali and B. D. Saunders, **Transitive closure and related semiring properties via eliminants**, Theor. Comput. Sci., 1985

O(n^3) computation with generic Floyd-Warshall algorithm

```
function WeightedTransitiveClosure($W$)
    $H^{(0)} \leftarrow A$
    for $k = 1, \ldots, n$ do
        for $i = 1, \ldots, n$ do
            for $j = 1, \ldots, n$ do
                $h^{(k)}_{(x_i,x_j)} \leftarrow h^{(k-1)}_{(x_i,x_j)} \oplus (h^{(k-1)}_{(x_i,x_k)} \odot h^{(k-1)}_{(x_k,x_j)})$
            end for
        end for
    end for
    return $\overline{A} = H^{(n)}$
end function
```

What About the Shift?

Shift(s) by $q \in V$

$$s_x = \sum_{y \le x} w_{y,x} c_y$$

$$\mathbf{s} = W\mathbf{c}$$

signal

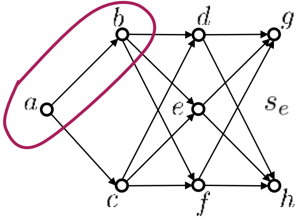
spectrum

chosen transitive closure of A
= inverse FT

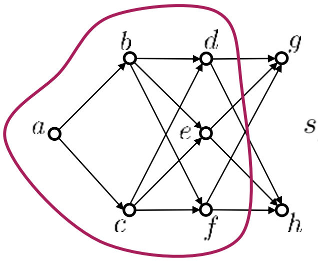
$$s_x = \sum_{y \le x} w_{y,x} c_y \mapsto \sum_{y \le x,\ y \le q} w_{y,x} c_y$$

In frequency domain:
Sum over common causes of x and q
=
Multiplies non-common causes by 0
Others by 1

Example:



s_e shifted by d: s_b



s_g shifted by h

Experiments

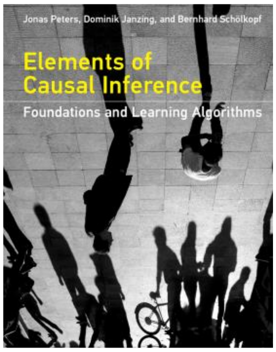
Fourier sparsity
=
Few causes

e.g., few cities pollute in one data set

Learn DAG signal from samples under the assumption of Fourier sparsity
Example infection data: see paper

Learn the DAG from data under the assumption of Fourier sparsity: next

Linear Structural Causal Models



Jonas Peters, Dominik Janzing, and Bernhard Schölkopf

Elements of Causal Inference

Foundations and Learning Algorithms

Linear SCM (also SEM) with DAG *A*: assumes data generation as

$$\mathbf{s} = A\mathbf{s} + \underset{\uparrow}{\mathbf{n}}$$

iid Gaussian noise

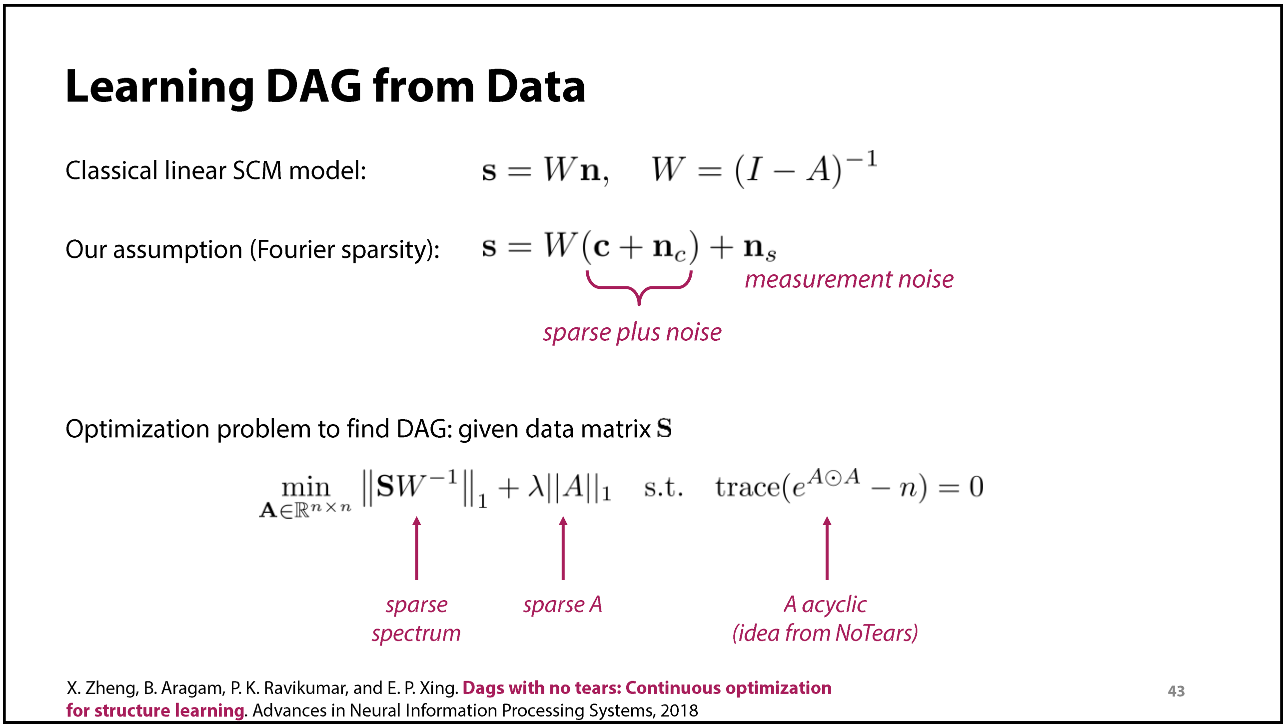
can represent any n-variate Gaussian distribution

In our language:

$$\Longleftrightarrow \mathbf{s} = \underset{\uparrow}{W}\underset{\uparrow}{\mathbf{n}}, \quad W = (I - A)^{-1}$$

dense, random spectrum

$(+,\cdot)$ transitive closure of A



Learning DAG from Data

Classical linear SCM model:

$$\mathbf{s} = W\mathbf{n}, \quad W = (I - A)^{-1}$$

Our assumption (Fourier sparsity):

$$\mathbf{s} = W\underbrace{(\mathbf{c} + \mathbf{n}_c)}_{\text{sparse plus noise}} + \mathbf{n}_s$$

measurement noise

Optimization problem to find DAG: given data matrix $\mathbf{S}$

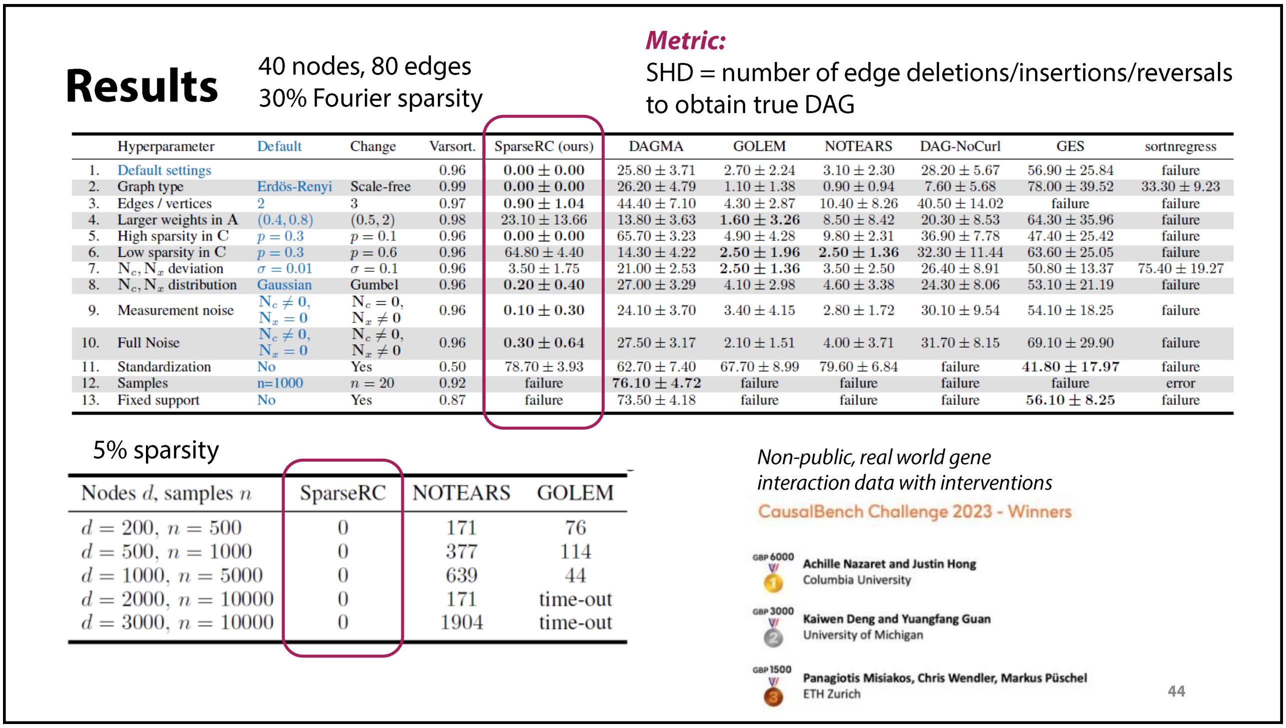
$$\min_{\mathbf{A}\in\mathbb{R}^{n\times n}} \left\|\mathbf{S}W^{-1}\right\|_1 + \lambda\|A\|_1 \quad \text{s.t.} \quad \operatorname{trace}(e^{A\odot A} - n) = 0$$

sparse spectrum

sparse A

A acyclic (idea from NoTears)

X. Zheng, B. Aragam, P. K. Ravikumar, and E. P. Xing. **Dags with no tears: Continuous optimization for structure learning**. Advances in Neural Information Processing Systems, 2018



Results

40 nodes, 80 edges
30% Fourier sparsity

Metric:
SHD = number of edge deletions/insertions/reversals to obtain true DAG

	Hyperparameter	Default	Change	Varsort.	SparseRC (ours)	DAGMA	GOLEM	NOTEARS	DAG-NoCurl	GES	sortnregress
1.	Default settings			0.96	**0.00 ± 0.00**	25.80 ± 3.71	2.70 ± 2.24	3.10 ± 2.30	28.20 ± 5.67	56.90 ± 25.84	failure
2.	Graph type	Erdös-Renyi	Scale-free	0.99	**0.00 ± 0.00**	26.20 ± 4.79	1.10 ± 1.38	0.90 ± 0.94	7.60 ± 5.68	78.00 ± 39.52	33.30 ± 9.23
3.	Edges / vertices	2	3	0.97	**0.90 ± 1.04**	44.40 ± 7.10	4.30 ± 2.87	10.40 ± 8.26	40.50 ± 14.02	failure	failure
4.	Larger weights in A	$(0.4, 0.8)$	$(0.5, 2)$	0.98	23.10 ± 13.66	13.80 ± 3.63	**1.60 ± 3.26**	8.50 ± 8.42	20.30 ± 8.53	64.30 ± 35.96	failure
5.	High sparsity in C	$p = 0.3$	$p = 0.1$	0.96	**0.00 ± 0.00**	65.70 ± 3.23	4.90 ± 4.28	9.80 ± 2.31	36.90 ± 7.78	47.40 ± 25.42	failure
6.	Low sparsity in C	$p = 0.3$	$p = 0.6$	0.96	64.80 ± 4.40	14.30 ± 4.22	**2.50 ± 1.96**	**2.50 ± 1.36**	32.30 ± 11.44	63.60 ± 25.05	failure
7.	$\mathrm{N}_c, \mathrm{N}_x$ deviation	$\sigma = 0.01$	$\sigma = 0.1$	0.96	3.50 ± 1.75	21.00 ± 2.53	**2.50 ± 1.36**	3.50 ± 2.50	26.40 ± 8.91	50.80 ± 13.37	75.40 ± 19.27
8.	$\mathrm{N}_c, \mathrm{N}_x$ distribution	Gaussian	Gumbel	0.96	**0.20 ± 0.40**	27.00 ± 3.29	4.10 ± 2.98	4.60 ± 3.38	24.30 ± 8.06	53.10 ± 21.19	failure
9.	Measurement noise	$\mathrm{N}_c \neq 0,$ $\mathrm{N}_x = 0$	$\mathrm{N}_c = 0,$ $\mathrm{N}_x \neq 0$	0.96	**0.10 ± 0.30**	24.10 ± 3.70	3.40 ± 4.15	2.80 ± 1.72	30.10 ± 9.54	54.10 ± 18.25	failure
10.	Full Noise	$\mathrm{N}_c \neq 0,$ $\mathrm{N}_x = 0$	$\mathrm{N}_c \neq 0,$ $\mathrm{N}_x \neq 0$	0.96	**0.30 ± 0.64**	27.50 ± 3.17	2.10 ± 1.51	4.00 ± 3.71	31.70 ± 8.15	69.10 ± 29.90	failure
11.	Standardization	No	Yes	0.50	78.70 ± 3.93	62.70 ± 7.40	67.70 ± 8.99	79.60 ± 6.84	failure	**41.80 ± 17.97**	failure
12.	Samples	n=1000	$n = 20$	0.92	failure	**76.10 ± 4.72**	failure	failure	failure	failure	error
13.	Fixed support	No	Yes	0.87	failure	73.50 ± 4.18	failure	failure	failure	**56.10 ± 8.25**	failure

5% sparsity

Nodes d, samples n	SparseRC	NOTEARS	GOLEM
$d = 200,\ n = 500$	0	171	76
$d = 500,\ n = 1000$	0	377	114
$d = 1000,\ n = 5000$	0	639	44
$d = 2000,\ n = 10000$	0	171	time-out
$d = 3000,\ n = 10000$	0	1904	time-out

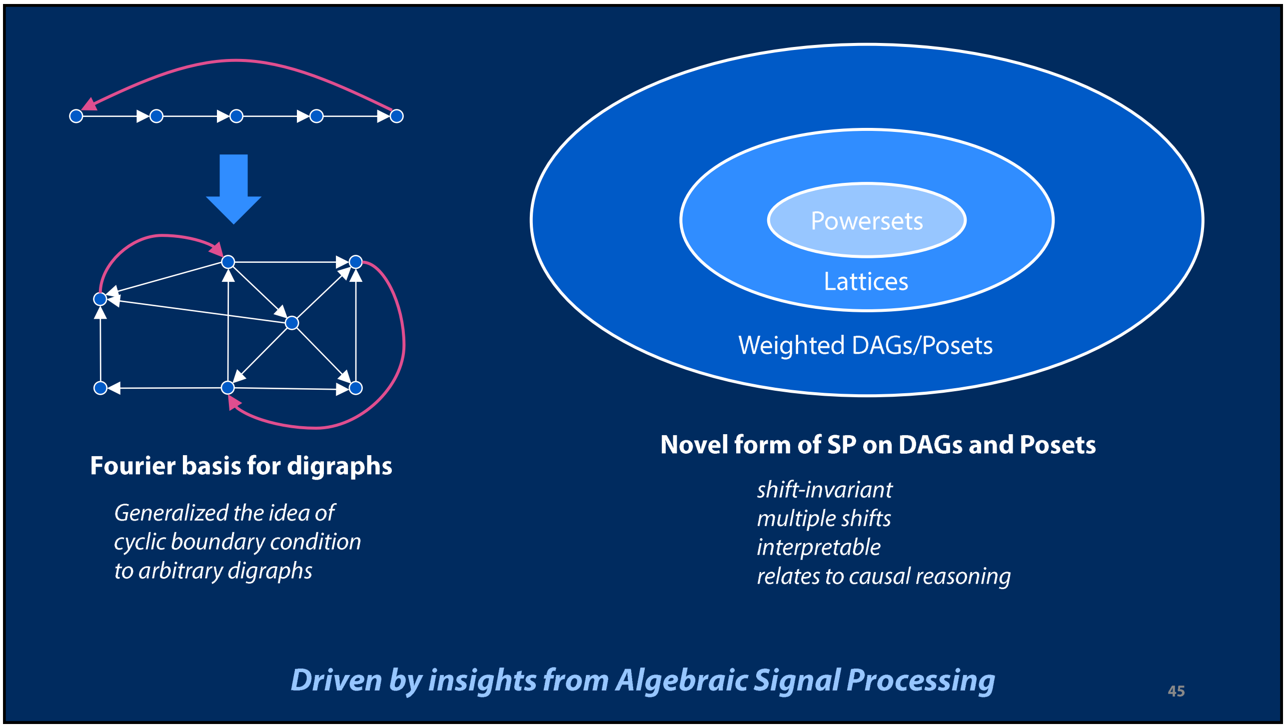
Non-public, real world gene interaction data with interventions

CausalBench Challenge 2023 - Winners

GBP 6000 **Achille Nazaret and Justin Hong**
Columbia University

GBP 3000 **Kaiwen Deng and Yuangfang Guan**
University of Michigan

GBP 1500 **Panagiotis Misiakos, Chris Wendler, Markus Püschel**
ETH Zurich



Powersets

Lattices

Weighted DAGs/Posets

Fourier basis for digraphs

Generalized the idea of
cyclic boundary condition
to arbitrary digraphs

Novel form of SP on DAGs and Posets

shift-invariant
multiple shifts
interpretable
relates to causal reasoning

Driven by insights from Algebraic Signal Processing