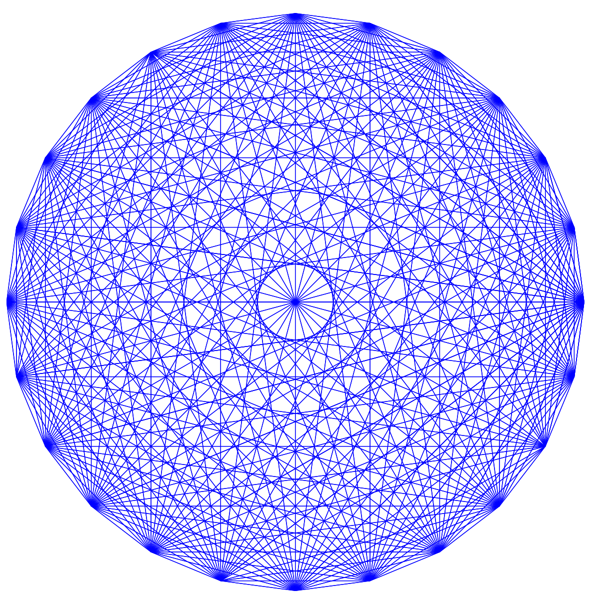
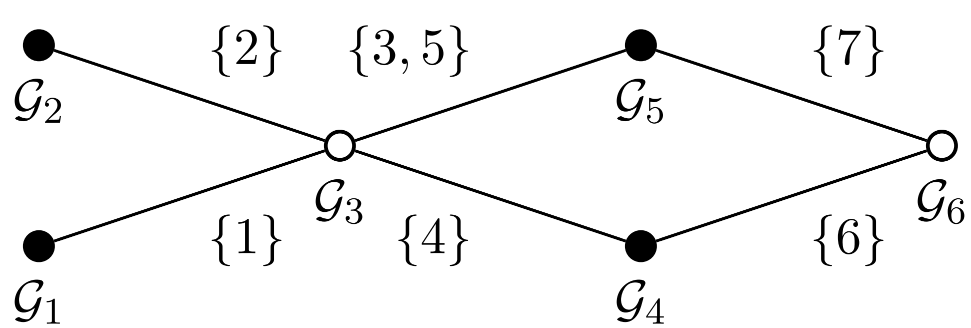
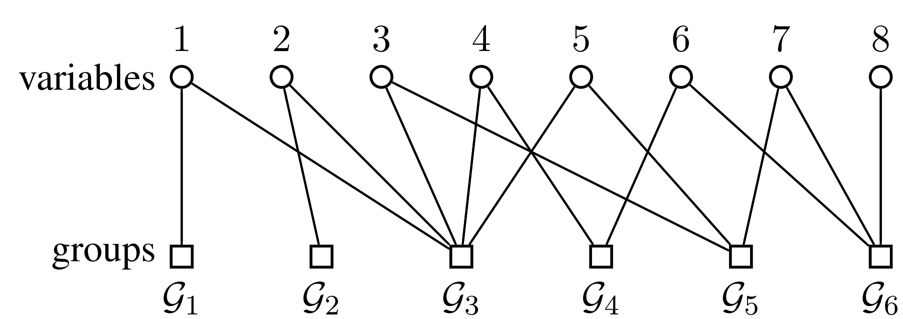
Tractability of Interpretability via Selection of Group-Sparse Models

A tale of NP-hard problems claimed to be solved by convex relaxations....



Luca Baldassarre
Laboratory for Information and Inference Systems
Ecole Polytechnique Federale de Lausanne
volkan.cevher@epfl.ch



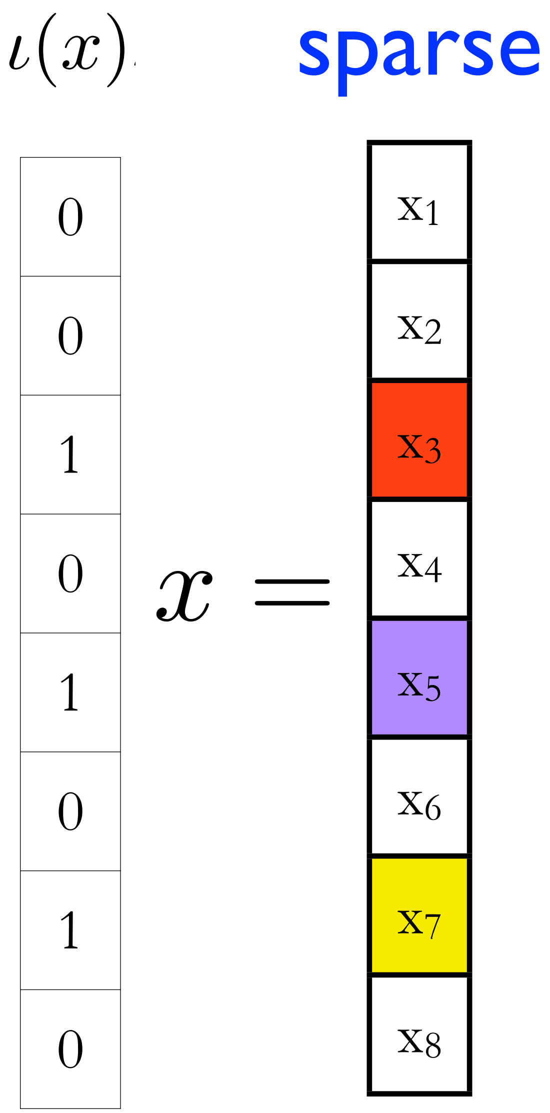


joint work with
Nirav Bhan
Volkan Cevher
Siddhartha Satpathi

lions@epfl

MARIE CURIE ACTIONS
FNSNF
FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION
erc
EPFL

A natural generalization of sparsity



support indicator vector:

$$\iota(x)_i = \begin{cases} 1 & \text{if } x_i \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

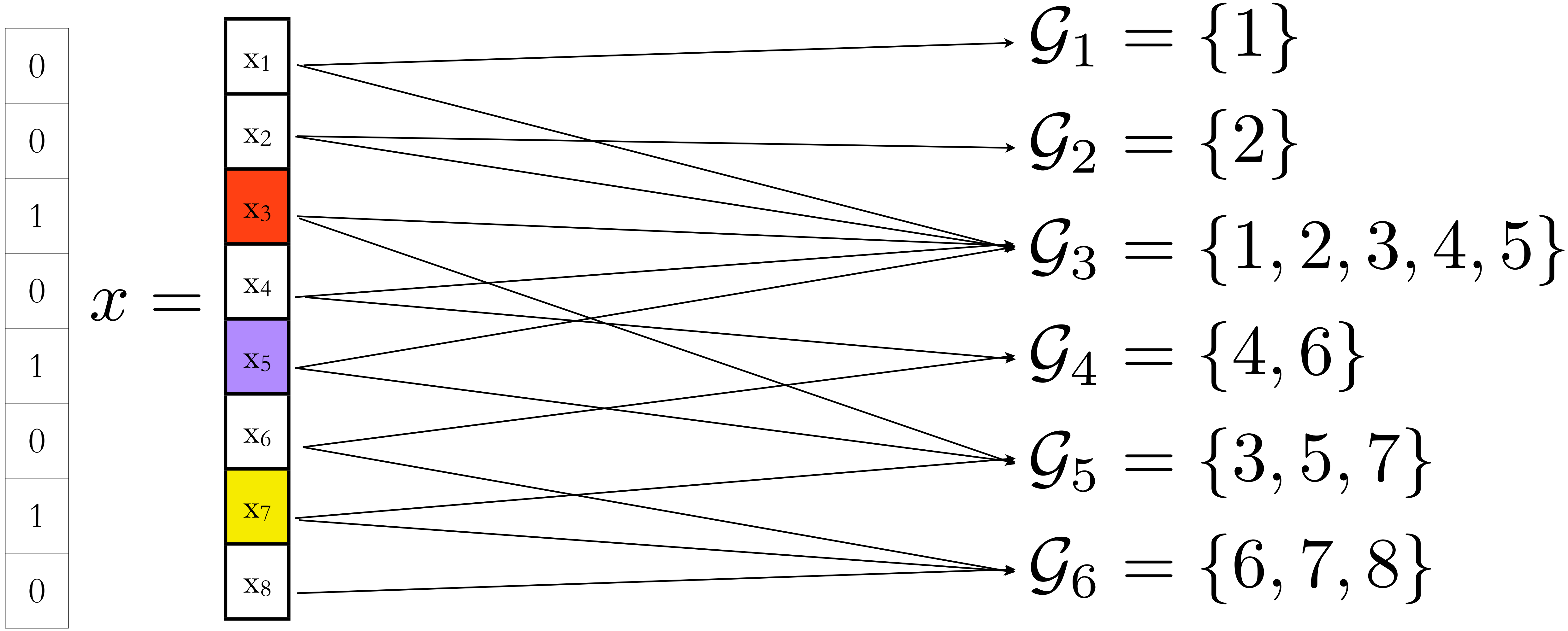
$$x \in \mathbb{R}^N$$

$$\sum_i \iota(x)_i := K \ll N$$

A natural generalization of sparsity

$\iota(x)$

sparse



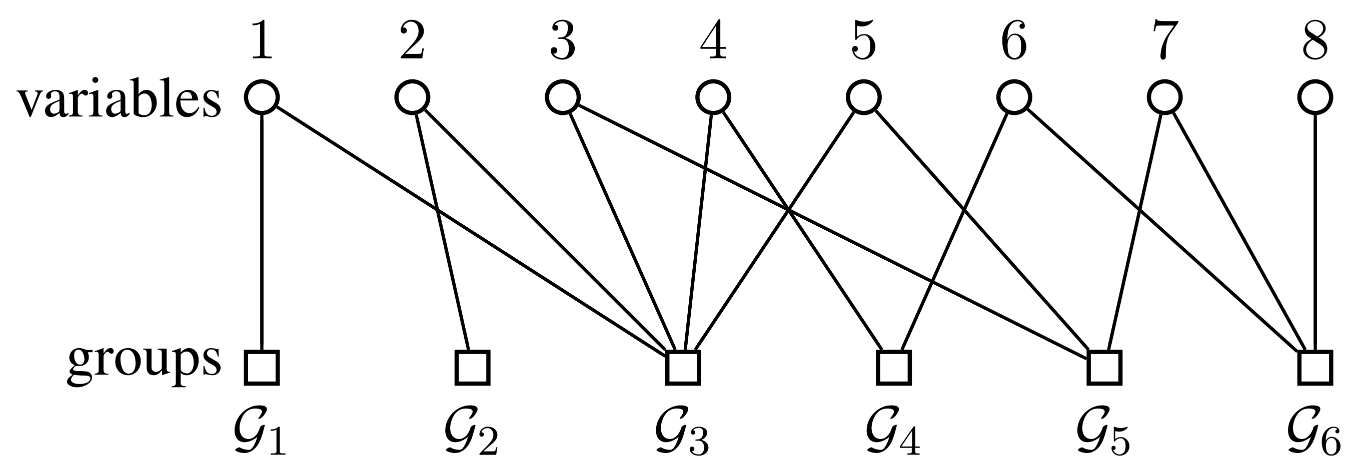
support indicator vector:

$$\iota(x)_i = \begin{cases} 1 & \text{if } x_i \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

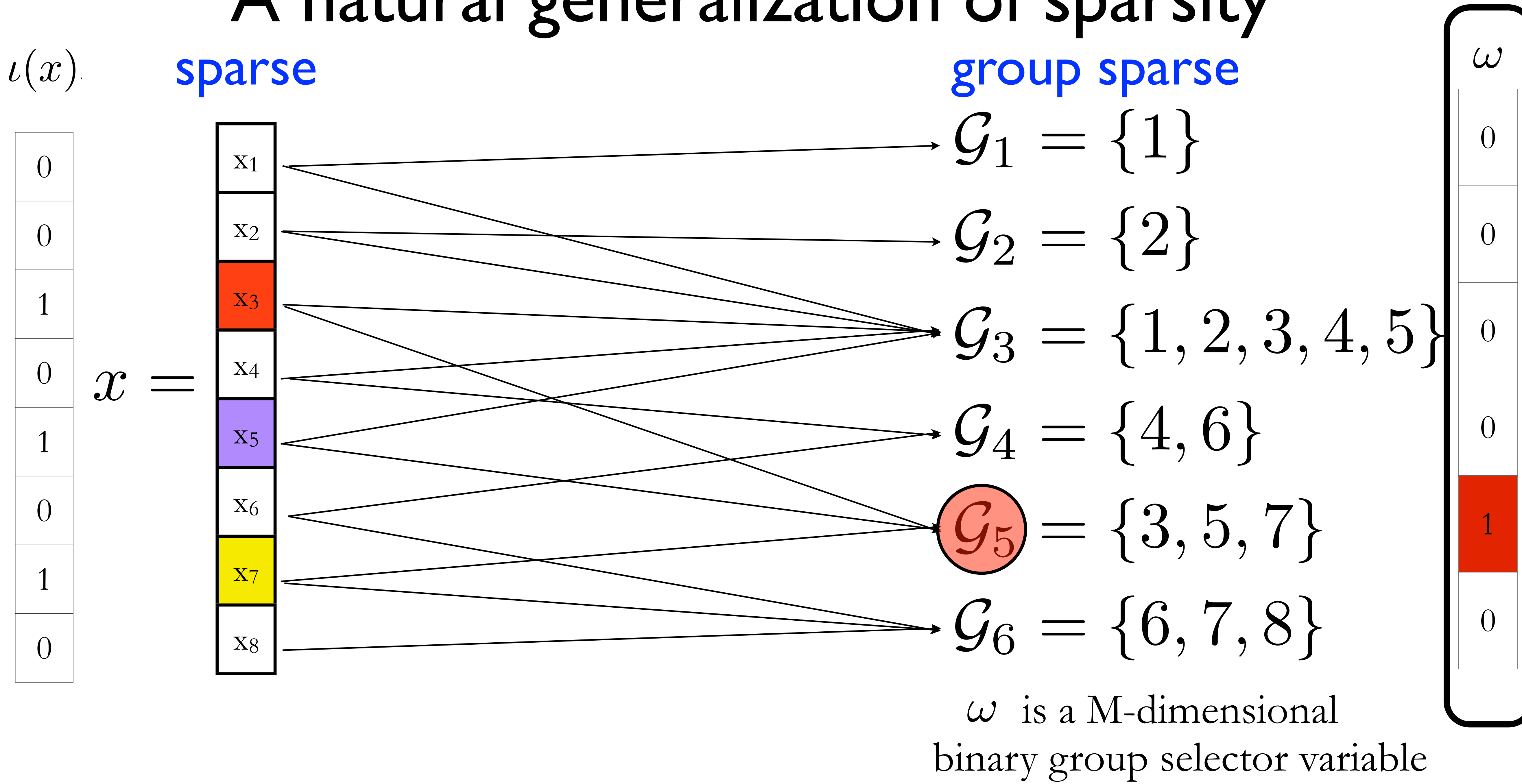
$$x \in \mathbb{R}^N$$

$$\sum_i \iota(x)_i := K \ll N$$

Group-structure: a collection of groups of variables



A natural generalization of sparsity



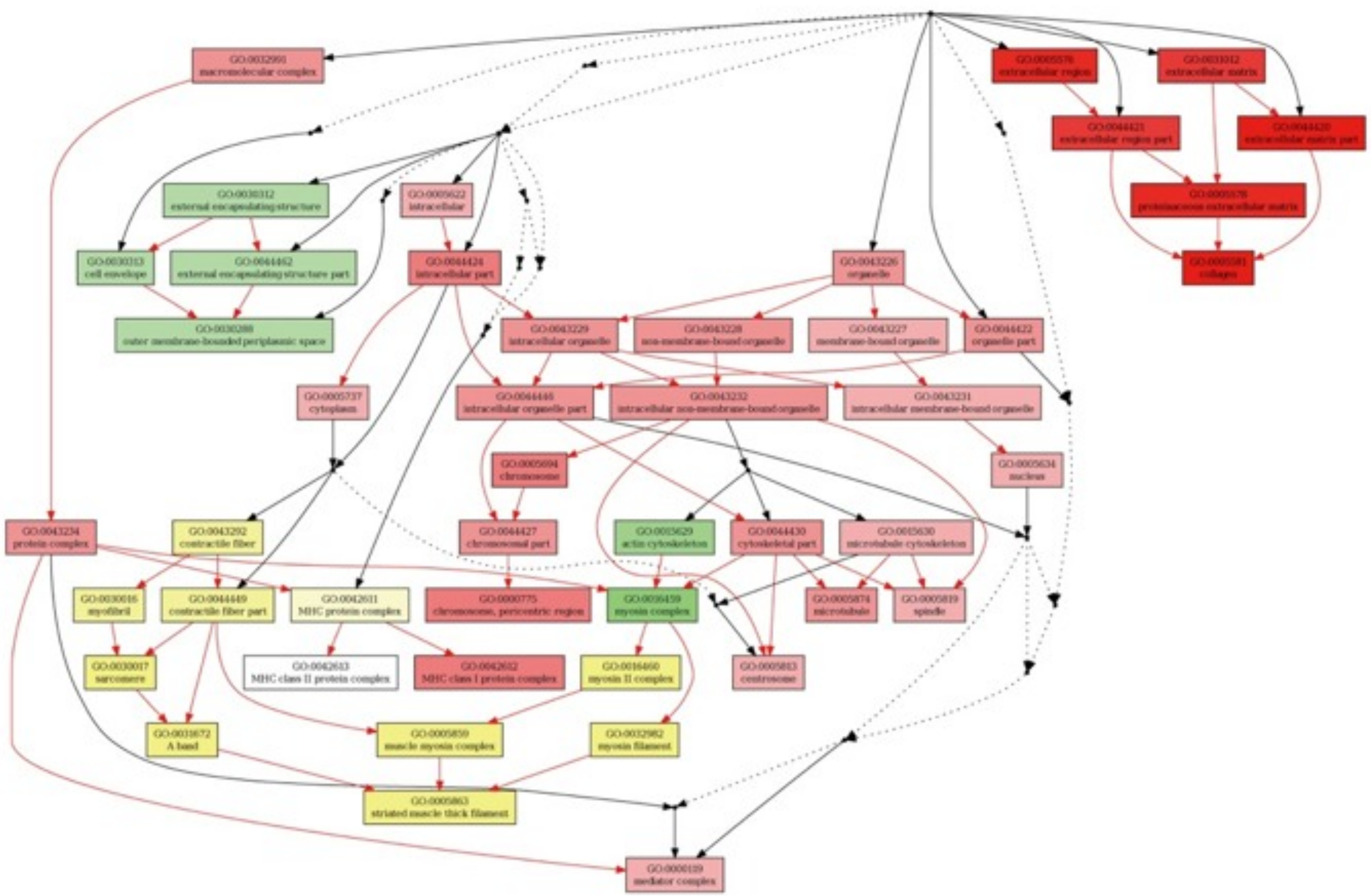
Goal:

Given a signal $x \in \mathbb{R}^N$, we want to find an approximation $\hat{x}$ whose support is contained in the union of **a few active groups** from the group-structure

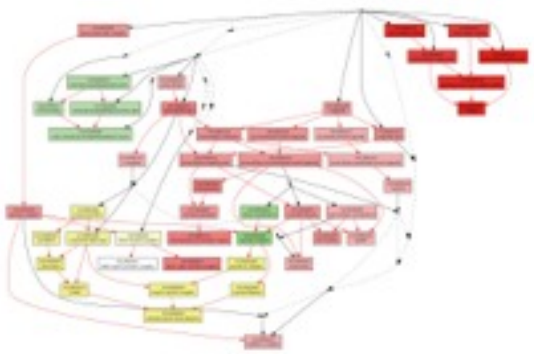
Group-based interpretations

Group models are ubiquitous.
Examples:

- Genetic Pathways in Microarray data analysis

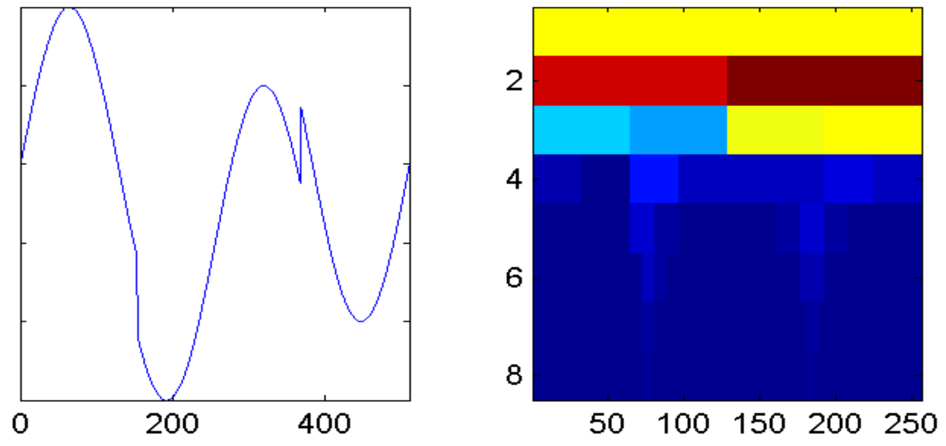


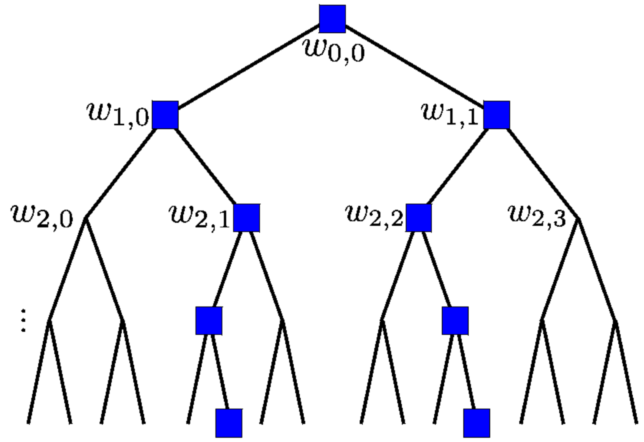
Group-based interpretations



Group models are ubiquitous.
Examples:

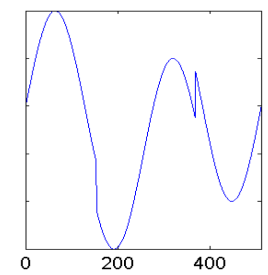
- Genetic Pathways in Microarray data analysis
- Wavelet models in image processing

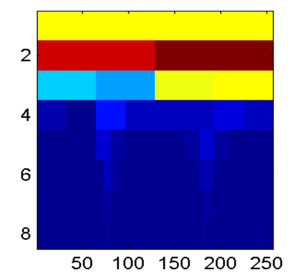


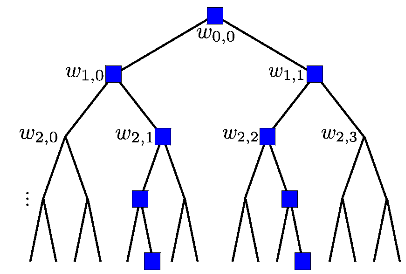


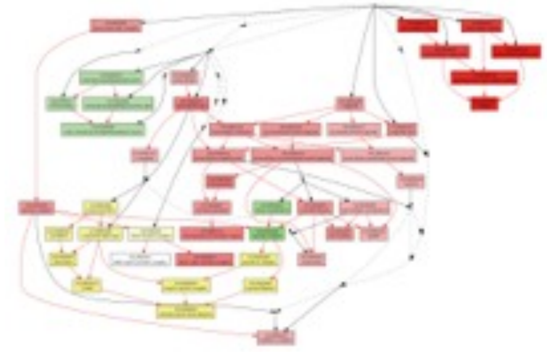
Group-based interpretations

Group models are ubiquitous.
Examples:

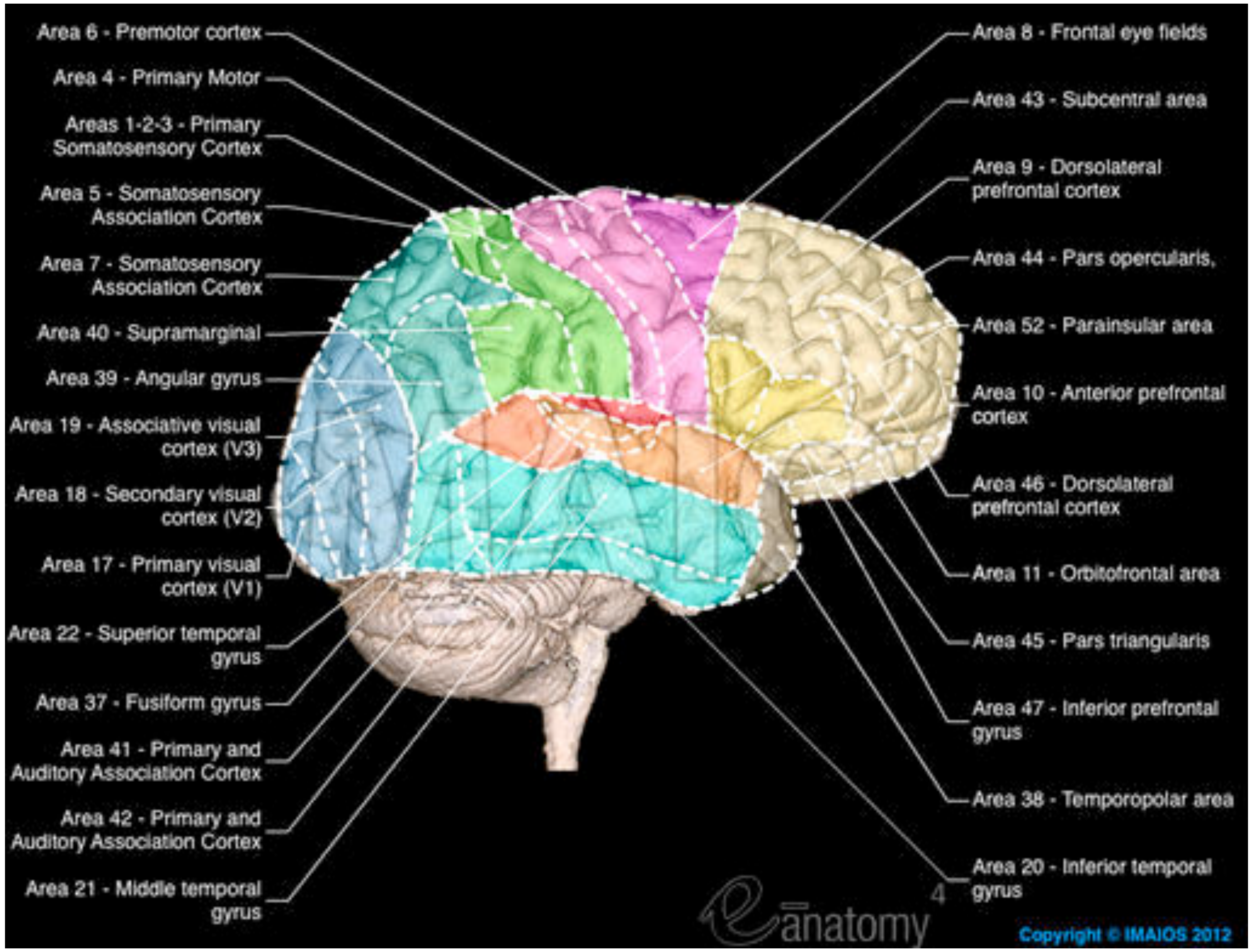






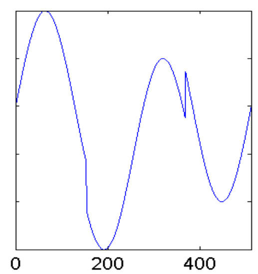


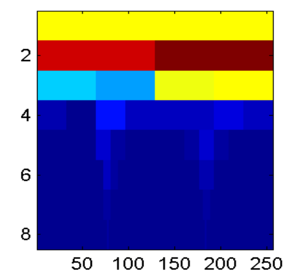
- Genetic Pathways in Microarray data analysis
- Wavelet models in image processing
- Brain regions in neuroimaging

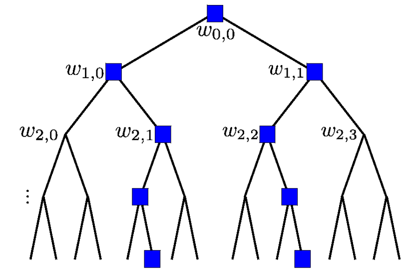


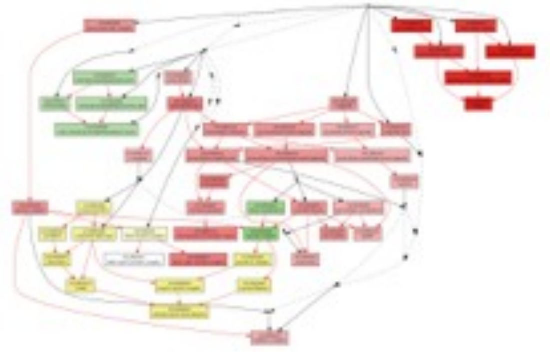
Group-based interpretations

Group models are ubiquitous.
Examples:

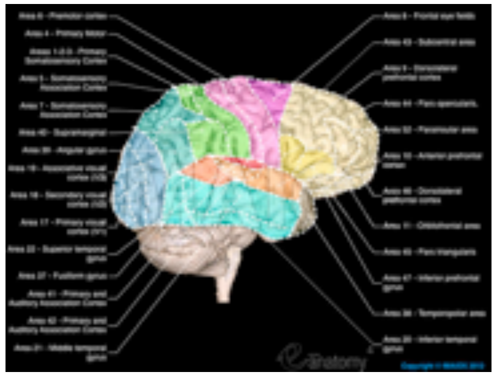








- Genetic Pathways in Microarray data analysis
- Wavelet models in image processing
- Brain regions in neuroimaging



Group models are motivated by **interpretability**.

- Which are the pathways that lead to correct diagnosis of cancer?
- Which coefficients capture most of the energy in the image?
- Which brain regions decode external stimula?

Outline

1. Definitions and graph-based representation of group structures.
2. **NP-hardness** of the group-sparse decompositions.
3. Special group-structures that allow tractable decompositions.
4. Relaxations:
   I. *Discrete relaxations* & totally unimodular constraints.
   II. Convex relaxations & their deficiencies.
5. Generalizations: group model + sparsity
6. Conclusions

Definitions I $x \in \mathbb{R}^N$

Ground set: $\mathcal{N} = \{1, \ldots, N\}$

Group structure: a collection of subsets of the ground set

$$\mathfrak{G} = \{\mathcal{G}_1, \ldots, \mathcal{G}_M\} \quad \mathcal{G}_j \subseteq \mathcal{N}, \ \forall j = 1, \ldots, M$$

Definitions I $x \in \mathbb{R}^N$

Ground set: $\mathcal{N} = \{1, \ldots, N\}$

Group structure: a collection of subsets of the ground set

$$\mathfrak{G} = \{\mathcal{G}_1, \ldots, \mathcal{G}_M\} \quad \mathcal{G}_j \subseteq \mathcal{N},\ \forall j = 1, \ldots, M$$

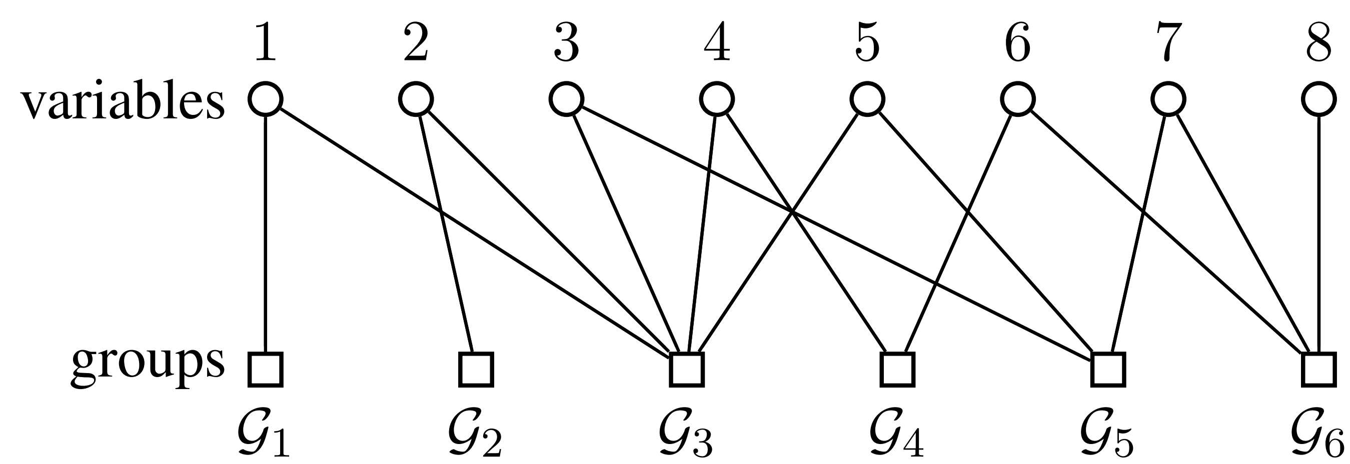
Binary matrix encoding the group structure $\mathbf{A}^{\mathfrak{G}}_{ij} = \begin{cases} 1 & \text{if } i \in \mathcal{G}_j \\ 0 & \text{otherwise} \end{cases}$

Example:

$\mathcal{G}_1 = \{1\}$ $\mathcal{G}_4 = \{4, 6\}$

$\mathcal{G}_2 = \{2\}$ $\mathcal{G}_5 = \{3, 5, 7\}$

$\mathcal{G}_3 = \{1, 2, 3, 4, 5\}$ $\mathcal{G}_6 = \{6, 7, 8\}$



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

Definitions I $\quad x \in \mathbb{R}^N$

Ground set: $\mathcal{N} = \{1, \ldots, N\}$

Group structure: a collection of subsets of the ground set

$$\mathfrak{G} = \{\mathcal{G}_1, \ldots, \mathcal{G}_M\} \qquad \mathcal{G}_j \subseteq \mathcal{N},\ \forall j = 1, \ldots, M$$

Binary matrix encoding the group structure $\quad \mathbf{A}^{\mathfrak{G}}_{ij} = \begin{cases} 1 & \text{if } i \in \mathcal{G}_j \\ 0 & \text{otherwise} \end{cases}$

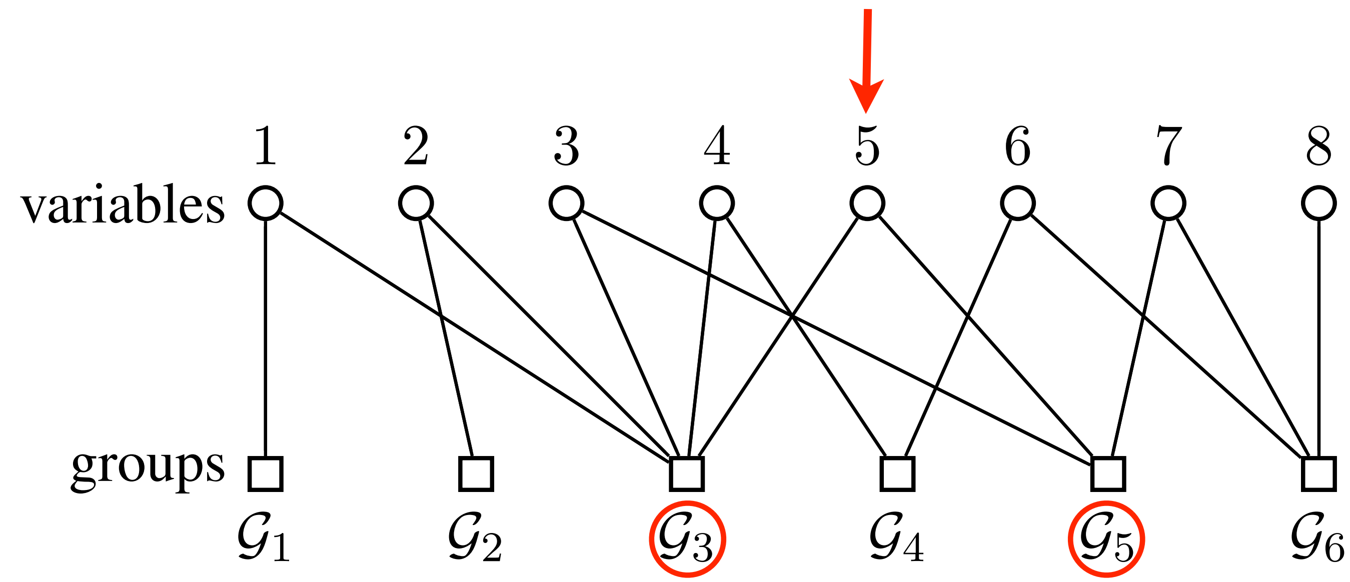
Example:

$\mathcal{G}_1 = \{1\}$ $\qquad \mathcal{G}_4 = \{4, 6\}$

$\mathcal{G}_2 = \{2\}$ $\qquad \mathcal{G}_5 = \{3, 5, 7\}$

$\mathcal{G}_3 = \{1, 2, 3, 4, 5\}$ $\qquad \mathcal{G}_6 = \{6, 7, 8\}$

$$\begin{array}{cc} & \begin{array}{cccccc} 1 & 2 & \textcircled{3} & 4 & \textcircled{5} & 6 \end{array} \\ \mathbf{A} = & \left[\begin{array}{cccccc} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & \color{red}{1} & 0 & \color{red}{1} & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \leftarrow \\ 6 \\ 7 \\ 8 \end{array} \end{array}$$



Definitions I $x \in \mathbb{R}^N$

Ground set: $\mathcal{N} = \{1, \ldots, N\}$

Group structure: a collection of subsets of the ground set

$$\mathfrak{G} = \{\mathcal{G}_1, \ldots, \mathcal{G}_M\} \quad \mathcal{G}_j \subseteq \mathcal{N}, \ \forall j = 1, \ldots, M$$

Binary matrix encoding the group structure $\mathbf{A}^{\mathfrak{G}}_{ij} = \begin{cases} 1 & \text{if } i \in \mathcal{G}_j \\ 0 & \text{otherwise} \end{cases}$

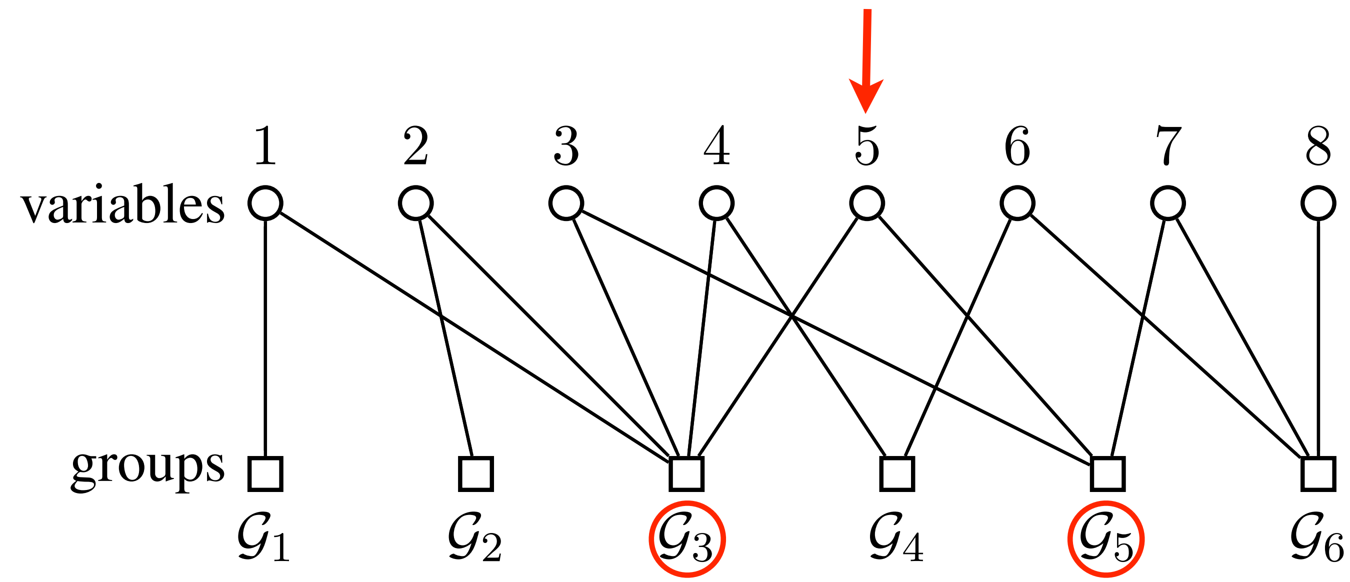
Example:

$\mathcal{G}_1 = \{1\}$ $\quad$ $\mathcal{G}_4 = \{4, 6\}$

$\mathcal{G}_2 = \{2\}$ $\quad$ $\mathcal{G}_5 = \{3, 5, 7\}$

$\mathcal{G}_3 = \{1, 2, 3, 4, 5\}$ $\quad$ $\mathcal{G}_6 = \{6, 7, 8\}$

$$\mathbf{A} = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \\ \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{matrix}$$



Usage: group cover

$$A^{\mathfrak{G}} \omega \geq \iota(x)$$

support indicator vector:

$$\iota(x)_i = \begin{cases} 1 & \text{if } x_i \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

ω is a M-dimensional binary group selector variable

Definitions II

Group cover: $S(x) \subseteq \mathfrak{G}$ s.t. $\text{supp}(x) \subseteq \bigcup_{\mathcal{G} \in S} \mathcal{G}$

G-group cover: $S^G(x) \subseteq \mathfrak{G}$ s.t. $|S| \leq G, \ \text{supp}(x) \subseteq \bigcup_{\mathcal{G} \in S} \mathcal{G}$ Might not exist!

Minimal group-cover: $\mathcal{M}(x)$ smallest group cover Might not be unique!

Definitions II

Group cover: $S(x) \subseteq \mathfrak{G}$ s.t. $\operatorname{supp}(x) \subseteq \bigcup_{\mathcal{G} \in S} \mathcal{G}$

G-group cover: $S^G(x) \subseteq \mathfrak{G}$ s.t. $|S| \leq G, \ \operatorname{supp}(x) \subseteq \bigcup_{\mathcal{G} \in S} \mathcal{G}$ Might not exist!

Minimal group-cover: $\mathcal{M}(x)$ smallest group cover Might not be unique!

$$\mathbf{A}^{\mathfrak{G}}_{ij} = \begin{cases} 1 & \text{if } i \in \mathcal{G}_j \\ 0 & \text{otherwise} \end{cases} \qquad \iota(x)_i = \begin{cases} 1 & \text{if } x_i \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

ω is a M-dimensional binary variable

Group cover: $$S(x) = \{\mathcal{G}_j \in \mathfrak{G} : \omega_j = 1, A^{\mathfrak{G}}\omega \geq \iota(x)\}$$

G-group cover: $$S^G(x) = \left\{\mathcal{G}_j \in \mathfrak{G} : \omega_j = 1, A^{\mathfrak{G}}\omega \geq \iota(x), \sum_j \omega_j \leq G\right\}$$

Minimal group-cover: $$\mathcal{M}(x) = \{\mathcal{G}_j \in \mathfrak{G} : \hat{\omega}_j = 1\}$$

$$\hat{\omega} \in \underset{\omega}{\operatorname{argmin}} \left\{\sum \omega_j : A^{\mathfrak{G}}\omega \geq \iota(x)\right\}$$

Signal approximation with the group-ℓ_0 "norm"

$$\|x\|_{\mathfrak{G},0} := \min_{\omega \in \mathbb{B}^M} \left\{ \sum_{j=1}^{M} \omega_j : A^{\mathfrak{G}} \omega \geq \iota(x) \right\}$$

x is **G-group sparse** if $\|x\|_{\mathfrak{G},0} \leq G$

G-group sparse approximation of a signal $\hat{x} \in \underset{z \in \mathbb{R}^N}{\operatorname{argmin}} \left\{ \|x - z\|_2^2 : \|z\|_{\mathfrak{G},0} \leq G \right\}$

Signal approximation with the group-ℓ_0 "norm"

$$\|x\|_{\mathfrak{G},0} := \min_{\omega \in \mathbb{B}^M} \left\{ \sum_{j=1}^{M} \omega_j : A^{\mathfrak{G}} \omega \geq \iota(x) \right\}$$

x is **G-group sparse** if $\|x\|_{\mathfrak{G},0} \leq G$

G-group sparse approximation of a signal $\hat{x} \in \operatorname*{argmin}_{z \in \mathbb{R}^N} \left\{ \|x - z\|_2^2 : \|z\|_{\mathfrak{G},0} \leq G \right\}$

It is sufficient to find the **G-group cover** for the approximation

$$\hat{x}_I = x_I, \ \hat{x}_{I^c} = 0$$

$$S^G(\hat{x}) \in \operatorname*{argmax}_{\substack{S \subseteq \mathfrak{G} \\ |S| \leq G}} \left\{ \sum_{i \in I} x_i^2 : I = \bigcup_{\mathcal{G} \in S} \mathcal{G} \right\}$$

Signal approximation with the group-ℓ_0 "norm"

$$\|x\|_{\mathfrak{G},0} := \min_{\omega \in \mathbb{B}^M} \left\{ \sum_{j=1}^{M} \omega_j : A^{\mathfrak{G}} \omega \geq \iota(x) \right\}$$

x is **G-group sparse** if $\|x\|_{\mathfrak{G},0} \leq G$

G-group sparse approximation of a signal $\hat{x} \in \underset{z \in \mathbb{R}^N}{\operatorname{argmin}} \left\{ \|x - z\|_2^2 : \|z\|_{\mathfrak{G},0} \leq G \right\}$

It is sufficient to find the **G-group cover** for the approximation

$\hat{x}_I = x_I, \ \hat{x}_{I^c} = 0$

$$S^G(\hat{x}) \in \underset{\substack{S \subseteq \mathfrak{G} \\ |S| \leq G}}{\operatorname{argmax}} \left\{ \sum_{i \in I} x_i^2 : I = \bigcup_{\mathcal{G} \in S} \mathcal{G} \right\}$$

Weighted Maximum Coverage problem $S^G(\hat{x}) = \{\mathcal{G}_j \in \mathfrak{G} : \omega_j^G = 1\}$

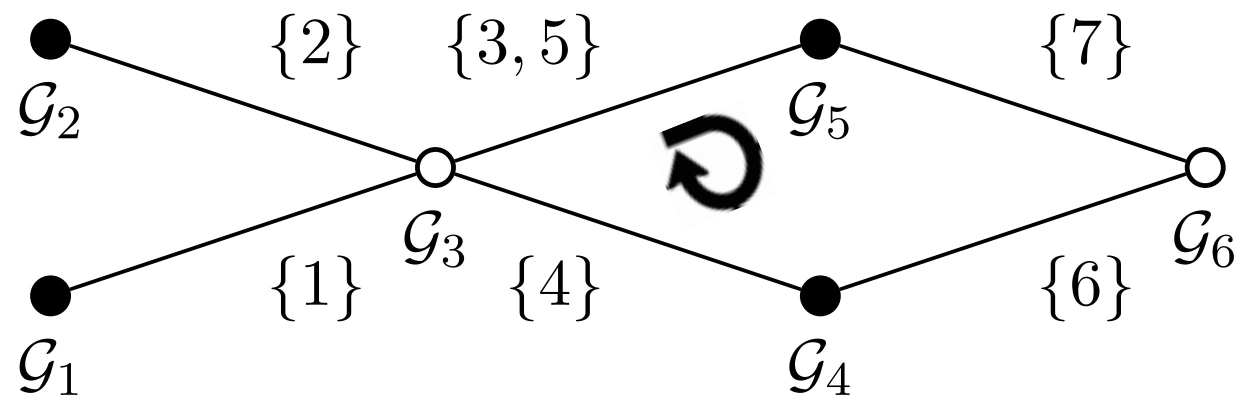
$$(\omega^G, y^G) \in \underset{\omega \in \mathbb{B}^M, \ y \in \mathbb{B}^N}{\operatorname{argmax}} \left\{ \sum_{i=1}^{N} y_i x_i^2 : A^{\mathfrak{G}} \omega \geq y, \sum_{j=1}^{M} \omega_j \leq G \right\}$$

NP-HARD!

Tractability of Interpretability I

Weighted Maximum Coverage

$$\max_{\omega \in \mathbb{B}^M,\ y \in \mathbb{B}^N} \left\{ \sum_{i=1}^{N} y_i x_i^2 \,:\, A^{\mathfrak{G}} \omega \geq y, \sum_{j=1}^{M} \omega_j \leq G \right\}$$



NP-HARD!

Tractability of Interpretability I

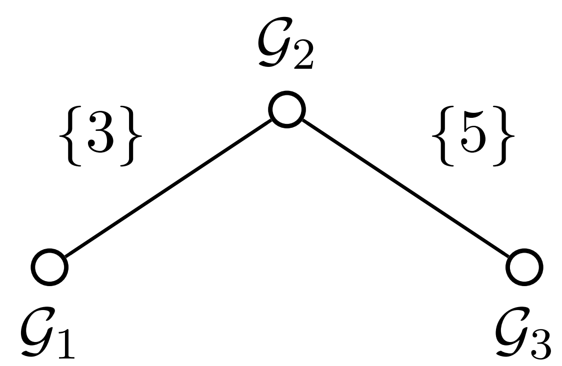
Weighted Maximum Coverage

$$\max_{\omega \in \mathbb{B}^M,\ y \in \mathbb{B}^N} \left\{ \sum_{i=1}^{N} y_i x_i^2 : A^{\mathfrak{G}} \omega \geq y, \sum_{j=1}^{M} \omega_j \leq G \right\}$$

Solve with ***D**ynamic **P**rogramming* for loopless pairwise overlapping groups

- Only pairwise overlaps
- No loops

$$\mathcal{G}_1 = \{1, 2, 3\}$$
$$\mathcal{G}_2 = \{3, 4, 5\}$$
$$\mathcal{G}_3 = \{5, 6, 7\}$$



The DP is polynomial-time $O(GM^2)$

Tractability of Interpretability I

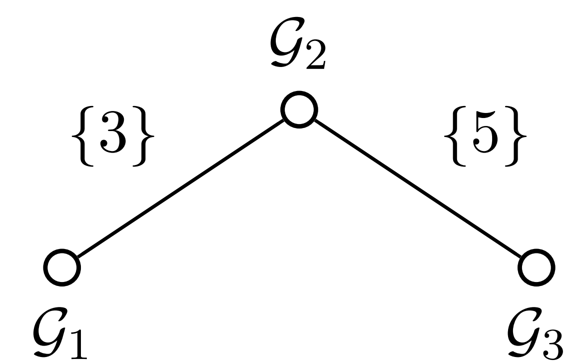
Weighted Maximum Coverage

$$\max_{\omega \in \mathbb{B}^M,\ y \in \mathbb{B}^N} \left\{ \sum_{i=1}^{N} y_i x_i^2 : A^{\mathfrak{G}} \omega \geq y, \sum_{j=1}^{M} \omega_j \leq G \right\}$$

Solve with ***D**ynamic **P**rogramming* for loopless pairwise overlapping groups

- Only pairwise overlaps
- No loops

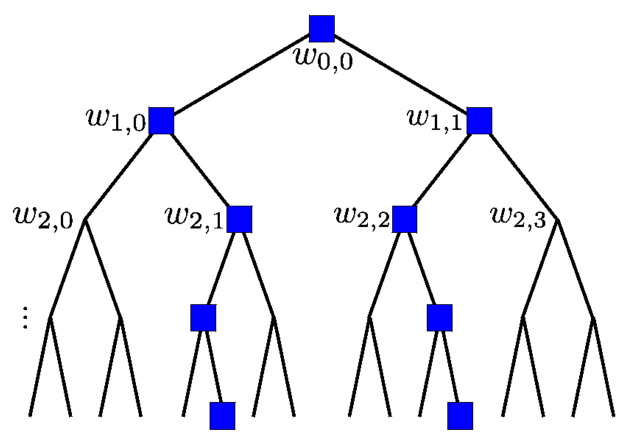
$$\mathcal{G}_1 = \{1, 2, 3\}$$
$$\mathcal{G}_2 = \{3, 4, 5\}$$
$$\mathcal{G}_3 = \{5, 6, 7\}$$

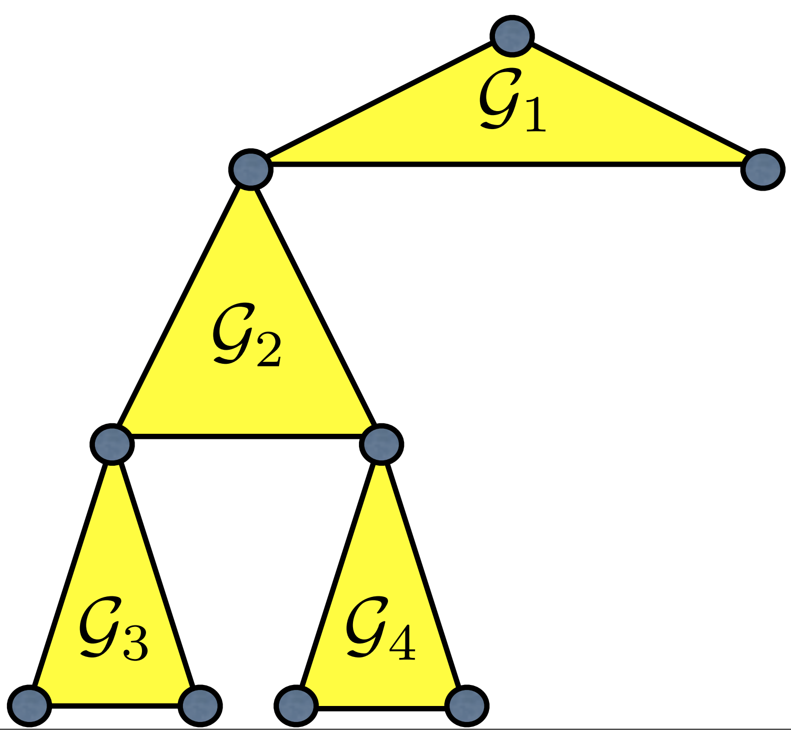


The DP is polynomial-time $O(GM^2)$

Example: a new wavelet **family** model

(where no child is left behind)





Tractability of Interpretability II

Weighted Maximum Coverage

$$\max_{\omega \in \mathbb{B}^M,\ y \in \mathbb{B}^N} \left\{ \sum_{i=1}^{N} y_i x_i^2 : A^{\mathfrak{G}} \omega \geq y, \sum_{j=1}^{M} \omega_j \leq G \right\}$$

Discrete relaxation: linear program

$$\max_{\omega \in \mathbb{B}^M,\ y \in \mathbb{B}^N} \left\{ \sum_{i=1}^{N} y_i x_i^2 - \lambda \sum_{j=1}^{M} \omega_j : A^{\mathfrak{G}} \omega \geq y \right\}$$

Tractability of Interpretability II

Weighted Maximum Coverage

$$\max_{\omega \in \mathbb{B}^M,\ y \in \mathbb{B}^N} \left\{ \sum_{i=1}^{N} y_i x_i^2 : A^{\mathfrak{G}} \omega \geq y, \sum_{j=1}^{M} \omega_j \leq G \right\}$$

Discrete relaxation: linear program

$$\max_{\omega \in \mathbb{B}^M,\ y \in \mathbb{B}^N} \left\{ \sum_{i=1}^{N} y_i x_i^2 - \lambda \sum_{j=1}^{M} \omega_j : A^{\mathfrak{G}} \omega \geq y \right\}$$

if $A^{\mathfrak{G}}$ is Totally Unimodular → Polynomial-time solvers

*TU: every square submatrix has determinant ±1 or 0

Tractability of Interpretability II

Weighted Maximum Coverage

$$\max_{\omega \in \mathbb{B}^M,\ y \in \mathbb{B}^N} \left\{ \sum_{i=1}^{N} y_i x_i^2 : A^{\mathfrak{G}} \omega \geq y, \sum_{j=1}^{M} \omega_j \leq G \right\}$$

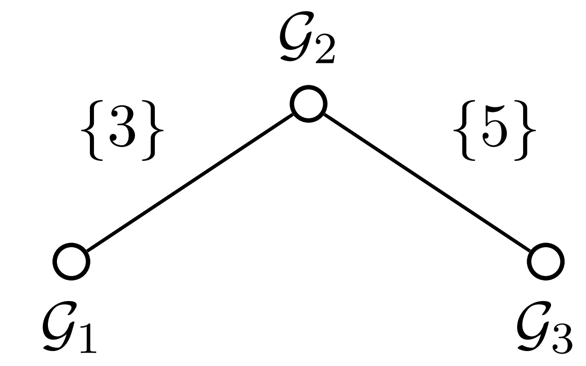
Discrete relaxation: linear program

$$\max_{\omega \in \mathbb{B}^M,\ y \in \mathbb{B}^N} \left\{ \sum_{i=1}^{N} y_i x_i^2 - \lambda \sum_{j=1}^{M} \omega_j : A^{\mathfrak{G}} \omega \geq y \right\}$$

if $A^{\mathfrak{G}}$ is Totally Unimodular → Polynomial-time solvers

***TU: every square submatrix has determinant ±1 or 0**

Theorem: Any loopless group model is TU!



Tractability of Interpretability II

Weighted Maximum Coverage

$$\max_{\omega \in \mathbb{B}^M,\ y \in \mathbb{B}^N} \left\{ \sum_{i=1}^{N} y_i x_i^2 : A^{\mathfrak{G}} \omega \geq y, \sum_{j=1}^{M} \omega_j \leq G \right\}$$

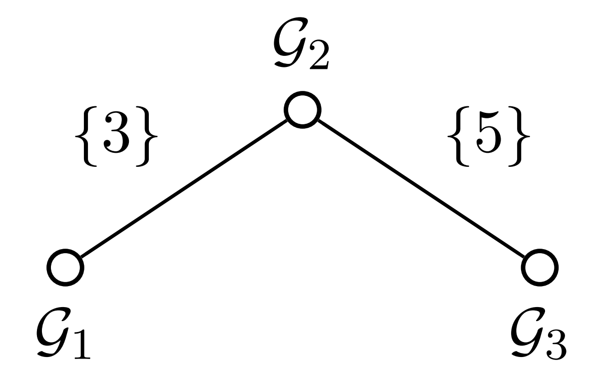
Discrete relaxation: linear program

$$\max_{\omega \in \mathbb{B}^M,\ y \in \mathbb{B}^N} \left\{ \sum_{i=1}^{N} y_i x_i^2 - \lambda \sum_{j=1}^{M} \omega_j : A^{\mathfrak{G}} \omega \geq y \right\}$$

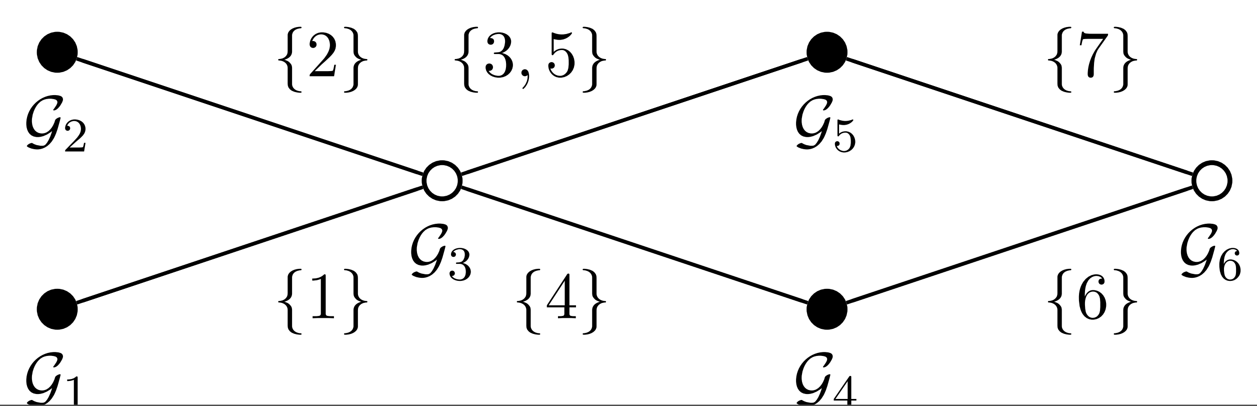
if $A^{\mathfrak{G}}$ is Totally Unimodular → Polynomial-time solvers

***TU: every square submatrix has determinant ±1 or 0**

Theorem: Any loopless group model is TU!



Theorem: Any bipartite group model is TU!



Example: Approximation via wavelet trees

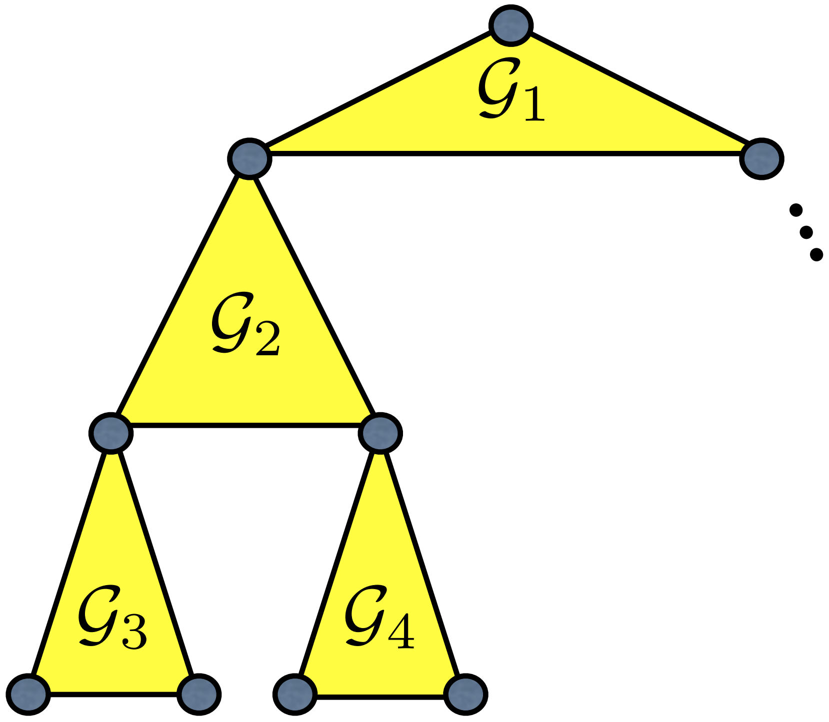
Pareto frontier of **WMC:**
set of optimal values as parameter G is varied

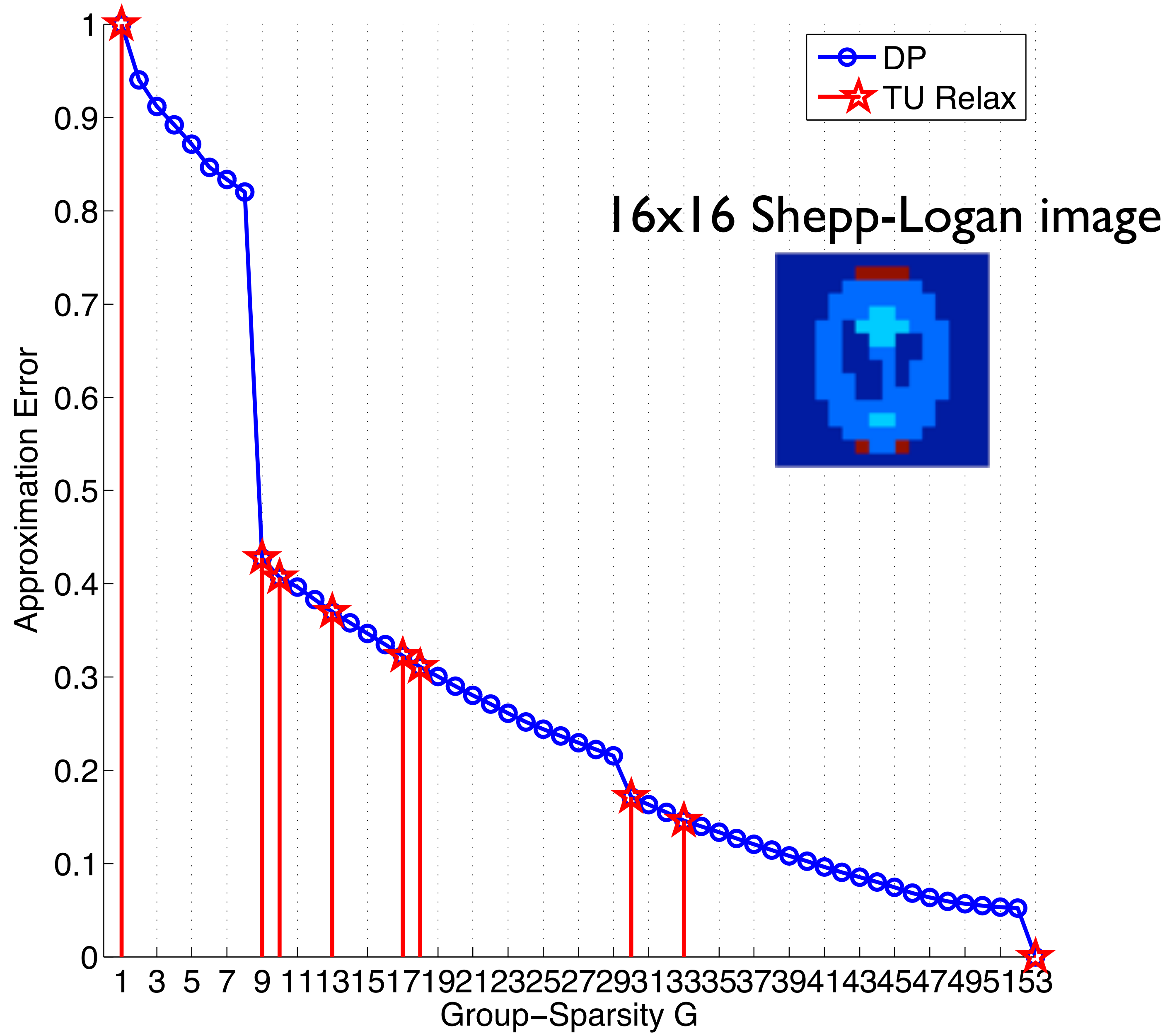
$$\mathcal{P} = \left\{ G,\ \sum_i x_i^2 y_i \right\}_{G=1}^{M}$$

Theorem: The discrete relaxation finds only the solutions in the intersection between $\mathcal{P}$ and the boundary of the convex hull of $\mathcal{P}$

Methods:

1. Dynamic Programming (DP)
2. Discrete relaxation with TU constraints (TU)





Existing convex relaxations: the fine print

Latent Group Lasso Norm for overlapping groups
promotes sparsity at the group level via decompositions [Obozinski et al., 2011]

$$\|x\|_{\mathfrak{G},\{1,2\}} := \inf_{\substack{\mathbf{v}^1,\ldots,\mathbf{v}^M \in \mathbb{R}^N \\ \forall j, \operatorname{supp}(\mathbf{v}^j) = \mathcal{G}_j}} \left\{ \sum_{j=1}^{M} d_j \|\mathbf{v}^j\|_2 : \sum_{j=1}^{M} \mathbf{v}^j = x \right\} \qquad (*)$$

The group cover $\breve{\mathcal{S}}(x)$ is defined by the non-zero terms in any optimal decomposition:

$$\breve{\mathcal{S}}(x) := \{\mathcal{G}_j \in \mathfrak{G} : \exists\ \mathbf{v} \in \mathcal{V}(x) \text{ s.t. } \mathbf{v}_j \neq 0\}$$

$\mathcal{V}(x)$ is the set of solutions of (*)

$\breve{\mathcal{S}}(x)$ **might not** contain the minimal group cover for x

Existing convex relaxations: the fine print

Latent Group Lasso Norm for overlapping groups

promotes sparsity at the group level via decompositions [Obozinski et al., 2011]

$$\|x\|_{\mathfrak{G},\{1,2\}} := \inf_{\substack{\mathbf{v}^1,\ldots,\mathbf{v}^M \in \mathbb{R}^N \\ \forall j,\, \mathrm{supp}(\mathbf{v}^j) = \mathcal{G}_j}} \left\{ \sum_{j=1}^{M} d_j \|\mathbf{v}^j\|_2 : \sum_{j=1}^{M} \mathbf{v}^j = x \right\} \qquad (*)$$

The group cover $\breve{\mathcal{S}}(x)$ is defined by the non-zero terms in any optimal decomposition:

$$\breve{\mathcal{S}}(x) := \{\mathcal{G}_j \in \mathfrak{G} : \exists\ \mathbf{v} \in \mathcal{V}(x) \text{ s.t. } \mathbf{v}_j \neq 0\}$$

$\mathcal{V}(x)$ is the set of solutions of (*)

$\breve{\mathcal{S}}(x)$ **might not** contain the minimal group cover for x

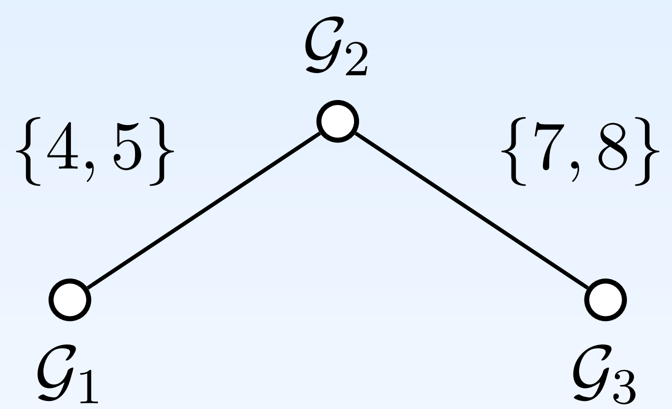
Group-sparse approximation: $\hat{x} = \underset{z \in \mathbb{R}^N}{\mathrm{argmin}} \left\{ \|x - z\|_2^2 : \|z\|_{\mathfrak{G},\{1,2\}} \leq \lambda \right\}$

Group-support recovery guarantees are given with respect to $\breve{\mathcal{S}}(x)$ and not to the underlying discrete problem (WMC).

To be (consistent), or not to be

$$\mathcal{N} = \{1, \ldots, 11\}$$

$$\begin{aligned} \mathcal{G}_1 &= \{1, \ldots, 5\} \\ \mathcal{G}_2 &= \{4, \ldots, 8\} \\ \mathcal{G}_3 &= \{7, \ldots, 11\} \end{aligned}$$



Loopless pairwise overlapping groups

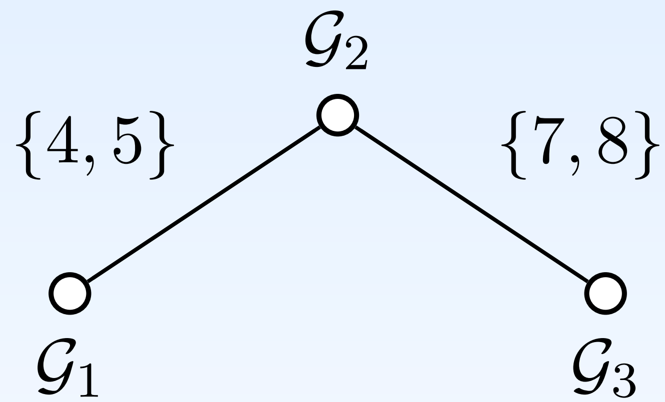
$$x = [\overbrace{0\ 0\ 1\ 1\ 1}^{\mathcal{G}_1}\ 0\ \overbrace{1\ 1\ 1\ 0\ 0}^{\mathcal{G}_3}]^\top, \quad \mathcal{G}_2 = \text{entries } 4, \ldots, 8$$

Minimal Group Cover: $\mathcal{M}(x) = \{\mathcal{G}_1, \mathcal{G}_3\}$

To be (consistent), or not to be

$\mathcal{N} = \{1, \ldots, 11\}$

$$\begin{aligned} \mathcal{G}_1 &= \{1, \ldots, 5\} \\ \mathcal{G}_2 &= \{4, \ldots, 8\} \\ \mathcal{G}_3 &= \{7, \ldots, 11\} \end{aligned}$$



Loopless pairwise overlapping groups

$$x = [\overbrace{0\ 0\ 1\ \underbrace{1\ 1}_{}}^{\mathcal{G}_1}\ \underbrace{0\ \overbrace{1\ 1}^{}}_{\mathcal{G}_2}\ \overbrace{1\ 0\ 0}^{\mathcal{G}_3}]^\top$$

Minimal Group Cover: $\mathcal{M}(x) = \{\mathcal{G}_1, \mathcal{G}_3\}$

DP solution: $S^G(\hat{x}) = \{\mathcal{G}_1, \mathcal{G}_3\}$ for $G = 2$

Discrete relaxation solution: $S^\lambda(\hat{x}) = \{\mathcal{G}_1, \mathcal{G}_3\}$ for $0 < \lambda \leq 2$

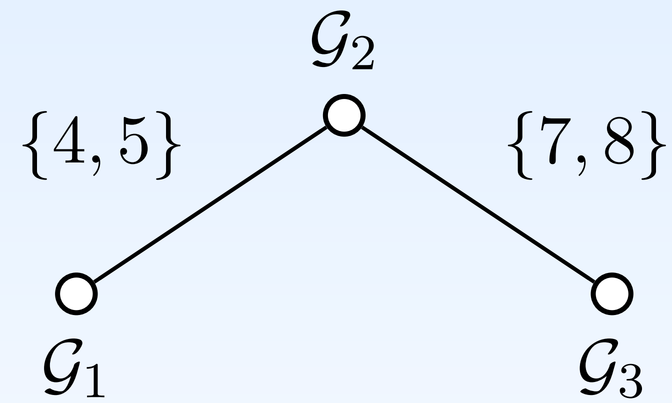
Convex relaxation solution: $\breve{\mathcal{S}}(x) = \mathfrak{G}$ with $d_j = \text{constant}$

$\breve{\mathcal{S}}(x) = \{\mathcal{G}_1, \mathcal{G}_3\}$ with $d_1 = d_3 = 1$ and $d_2 > \frac{2}{\sqrt{3}}$

To be (consistent), or not to be

$\mathcal{N} = \{1, \ldots, 11\}$

$$\begin{aligned} \mathcal{G}_1 &= \{1, \ldots, 5\} \\ \mathcal{G}_2 &= \{4, \ldots, 8\} \\ \mathcal{G}_3 &= \{7, \ldots, 11\} \end{aligned}$$



Loopless pairwise overlapping groups

$$x = [\overbrace{0\ 0\ 1\ 1\ 1}^{\mathcal{G}_1}\ 0\ \overbrace{1\ 1\ 1\ 0\ 0}^{\mathcal{G}_3}]^\top, \quad \mathcal{G}_2 = \text{entries } 4, \ldots, 8$$

Minimal Group Cover: $\mathcal{M}(x) = \{\mathcal{G}_1, \mathcal{G}_3\}$

DP solution: $S^G(\hat{x}) = \{\mathcal{G}_1, \mathcal{G}_3\}$ for $G = 2$

Discrete relaxation solution: $S^\lambda(\hat{x}) = \{\mathcal{G}_1, \mathcal{G}_3\}$ for $0 < \lambda \leq 2$

Convex relaxation solution: $\breve{\mathcal{S}}(x) = \mathfrak{G}$ with $d_j = \text{constant}$

$\breve{\mathcal{S}}(x) = \{\mathcal{G}_1, \mathcal{G}_3\}$ with $d_1 = d_3 = 1$ and $d_2 > \frac{2}{\sqrt{3}}$

Need to know beforehand which groups are irrelevant!

Two important generalizations

1. Introduce sparsity budget K.
   - Allow individual elements within a group to be selected.
   - Discrete problem (still NP-HARD in general)

Generalized Weighted Maximum Coverage

$$\begin{aligned} \underset{\omega \in \mathbb{B}^M,\ y \in \mathbb{B}^N}{\text{maximize}} \quad & \textstyle\sum_{i=1}^{N} y_i x_i^2 \\ \text{subject to} \quad & A^{\mathfrak{G}} \omega \geq y \\ & \textstyle\sum_{j=1}^{M} \omega_j \leq G \\ & \textstyle\sum_{i=1}^{N} y_i \leq K \end{aligned}$$

- Dynamic Program for loopless: $O(K^2 M^2 G)$ - *can be improved for regular models*

Two important generalizations

1. Introduce sparsity budget K.
   - Allow individual elements within a group to be selected.
   - Discrete problem (still NP-HARD in general)

Generalized Weighted Maximum Coverage

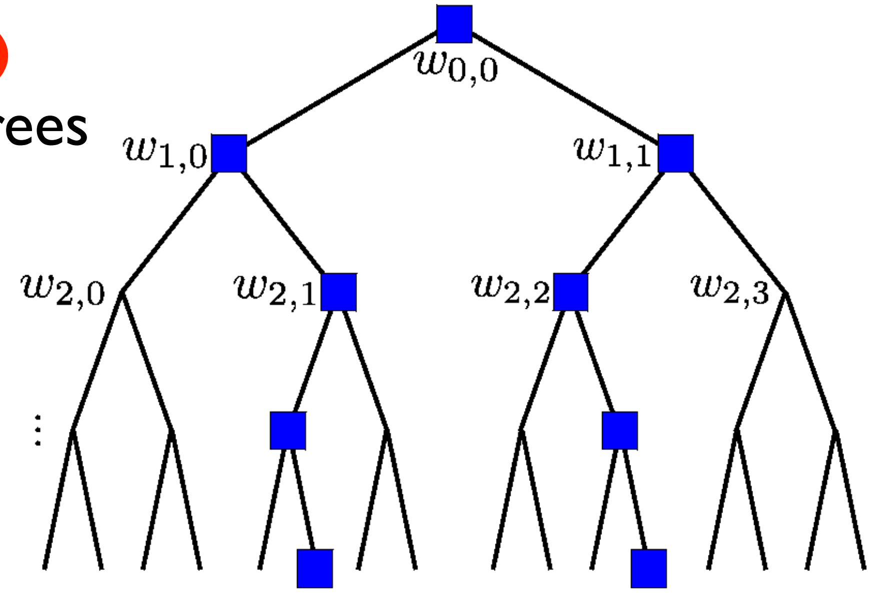
$$\begin{aligned} \underset{\omega \in \mathbb{B}^M,\ y \in \mathbb{B}^N}{\text{maximize}} \quad & \textstyle\sum_{i=1}^N y_i x_i^2 \\ \text{subject to} \quad & A^{\mathfrak{G}} \omega \geq y \\ & \textstyle\sum_{j=1}^M \omega_j \leq G \\ & \textstyle\sum_{i=1}^N y_i \leq K \end{aligned}$$

   - Dynamic Program for loopless: $O(K^2M^2G)$ - *can be improved for regular models*

2. Introduce hierarchical constraints to be encoded in the Generalized WMC.
   - Polynomial-time Dynamic Program.
   - Hierarchical constraints are TU.

O(KDN)
for regular trees

$$\max_{\omega \in \mathbb{B}^M,\ y \in \mathbb{B}^N} \left\{ \sum_{i=1}^N y_i x_i^2 - \lambda_G \sum_{j=1}^M \omega_j - \lambda_K \sum_{i=1}^N y_i : A^{\mathfrak{G}} \omega \geq y \right\}$$

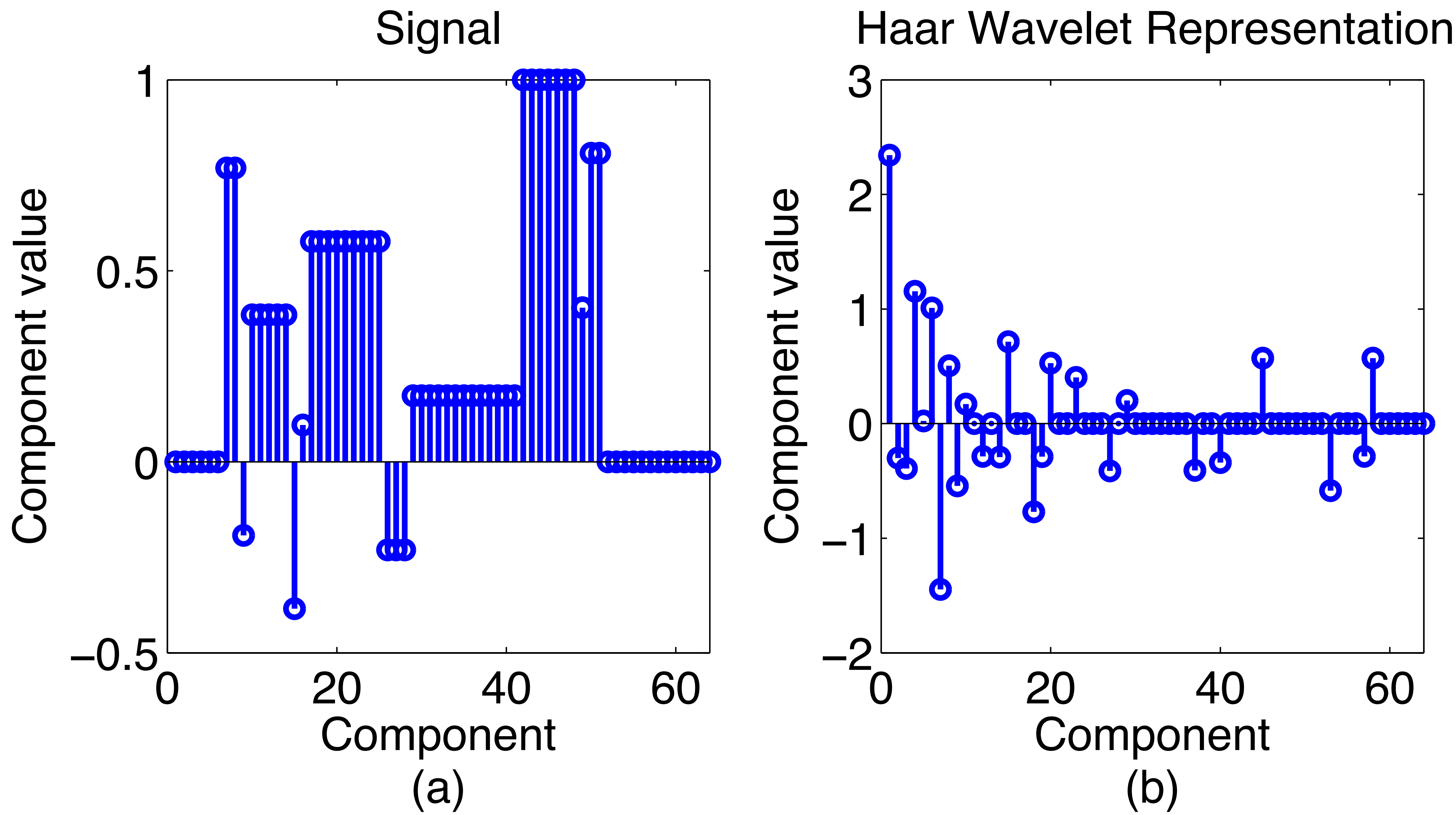


Example: Approximation via wavelet trees

Pareto frontier of **WMC:**
set of optimal values as parameter K is varied

$$\mathcal{P} = \left\{ K,\ \sum_i x_i^2 y_i \right\}_{K=1}^{N}$$

- Block signal of size N = 64
- Sparse representation in Haar wavelet coefficients that satisfy hierarchical constraints



Example: Approximation via wavelet trees

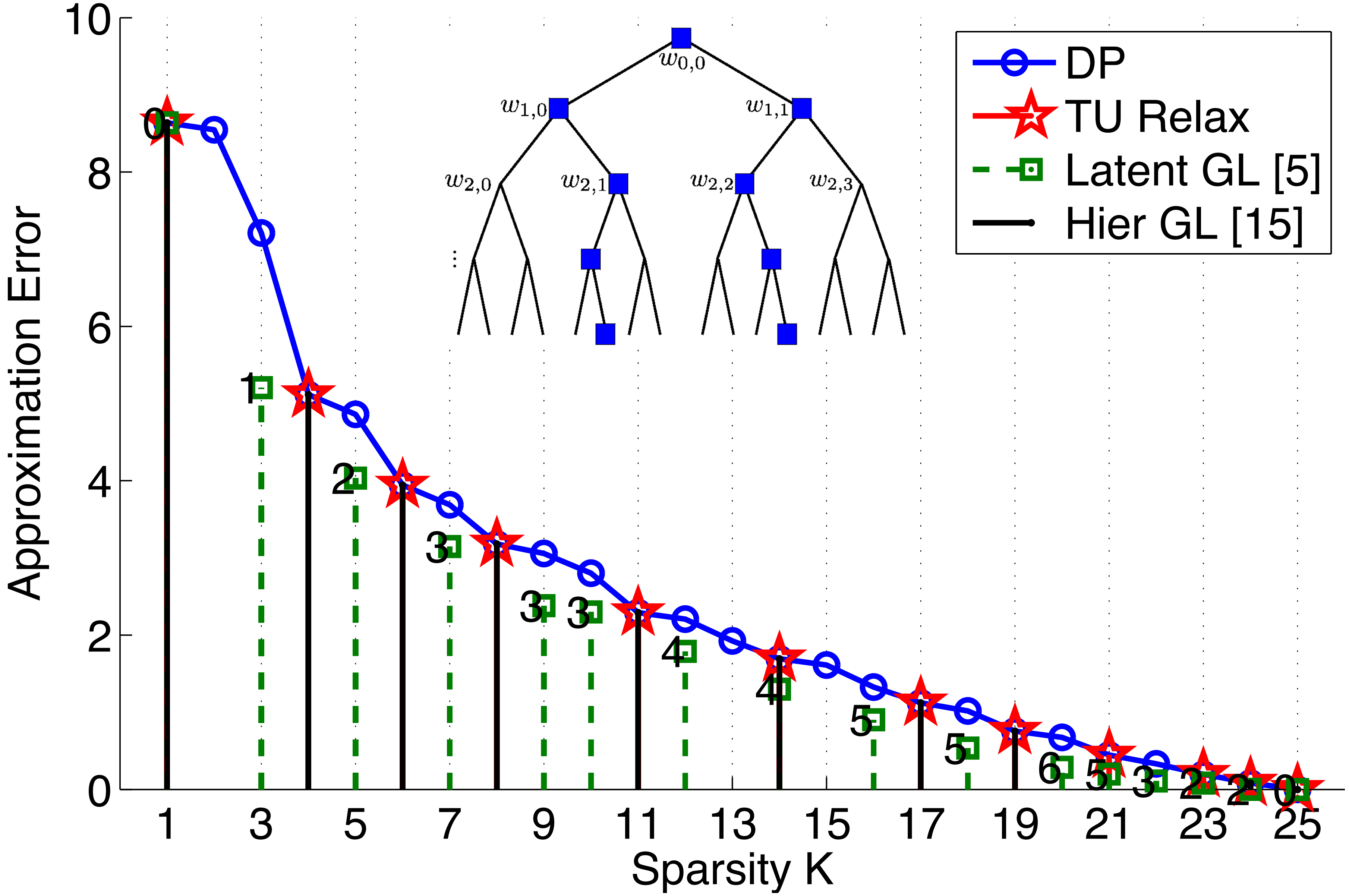
Pareto frontier of **WMC:**
set of optimal values as parameter K is varied

$$\mathcal{P} = \left\{ K,\ \sum_i x_i^2 y_i \right\}_{K=1}^{N}$$

- Block signal of size N = 64
- Sparse representation in Haar wavelet coefficients that satisfy hierarchical constraints

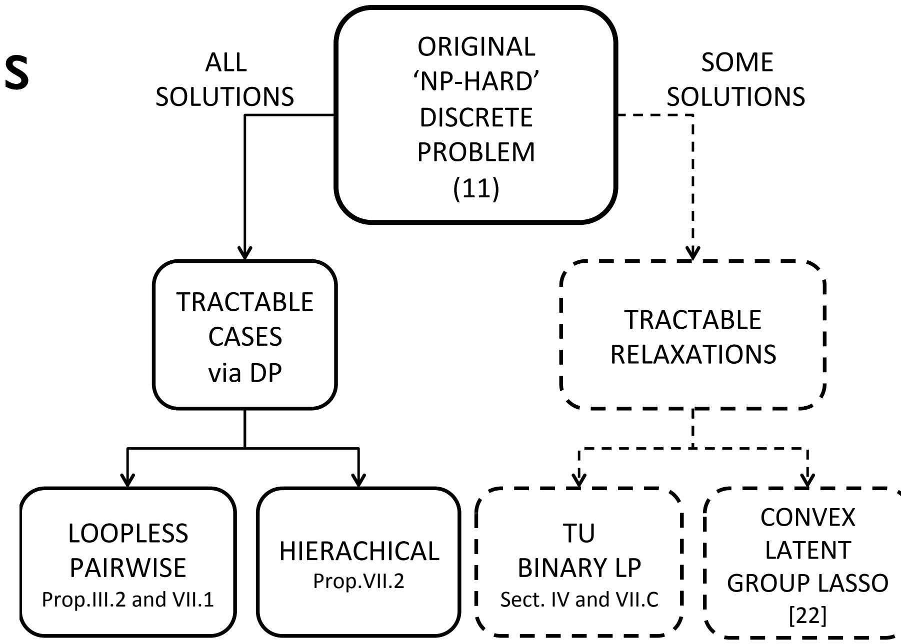
Methods:

1. Dynamic Programming (DP)
2. Discrete relaxation with TU constraints (TU)
3. Latent Group Lasso with groups given as all father-child pairs: not all constraints are satisfied [Rao et al., 2012]
4. Hierarchical group lasso [Jenatton et al., 2009]



Conclusions

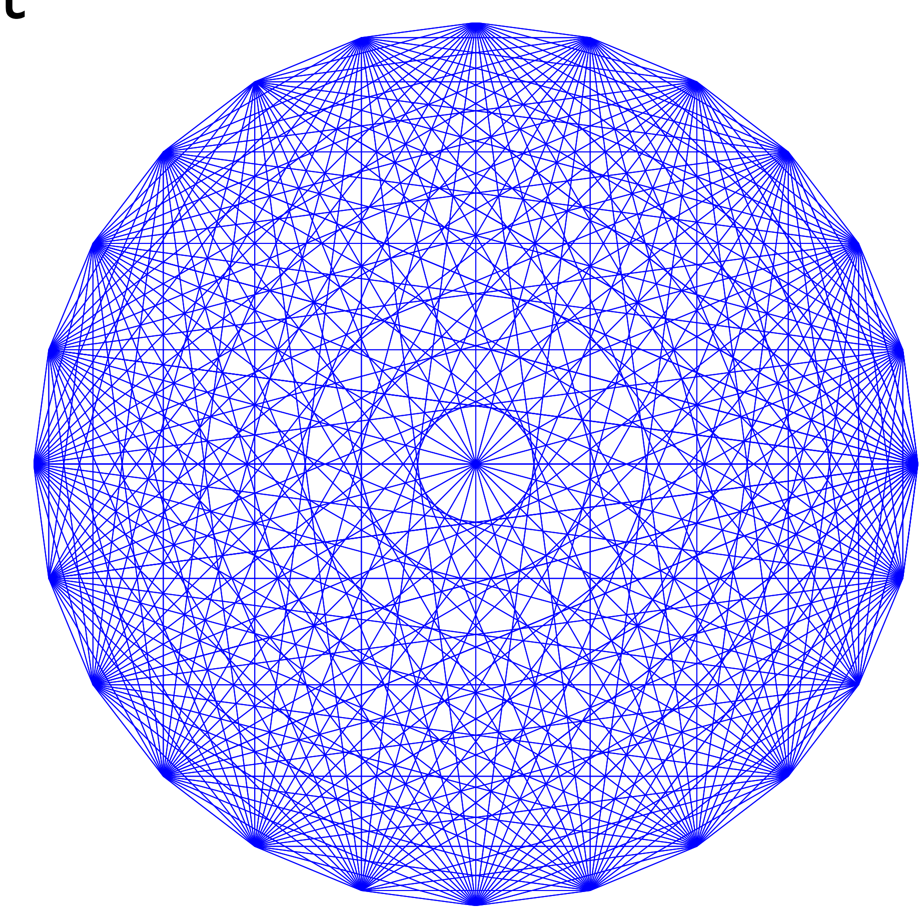
- Group sparse problems <> **NP-Hard** in general (GWMC)
  - Deceiving consistency results via convex relaxations



- ***Tractable group-based interpretations***
  - Loopless & hierarchical models with sparsity
  - Totally Unimodular group structures

- Pareto Frontier of solutions can be radically different
  - Dynamic Programming solutions (full frontier)
  - Linear Programming relaxations (convex hull)
  - Convex relaxations (sometimes! convex hull)

➡ Depressing problem: *breast cancer dataset* group-graph of the top 25 pathways from the Molecular Signature Database



References

- **Baldassarre et al., *Group-Sparse Model Selection: Hardness and Relaxations*, arXiv, 2013**
- Jenatton, R. et al., *Proximal methods for hierarchical sparse coding,* JMLR, 2011
- Obozinski, G. et al., *Group lasso with overlaps: The latent group lasso approach*, arXiv, 2011
- Rao, N. et al., *Signal recovery in unions of subspaces with applications to compressive imaging,* arXiv, 2012
- Subramanian, A. et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles*, PNAS, 2005.
- Cartis, C. and Thompson, A., *An exact tree projection algorithm for wavelets,* arXiv, 2013.