Tractability of Interpretability via Selection of Group-Sparse Models

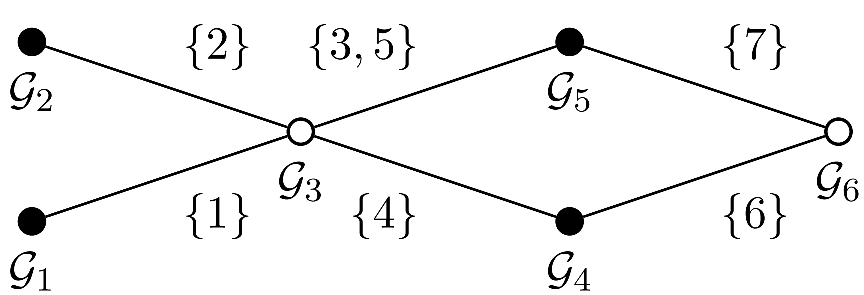
A tale of NP-hard problems claimed to be solved by convex relaxations....

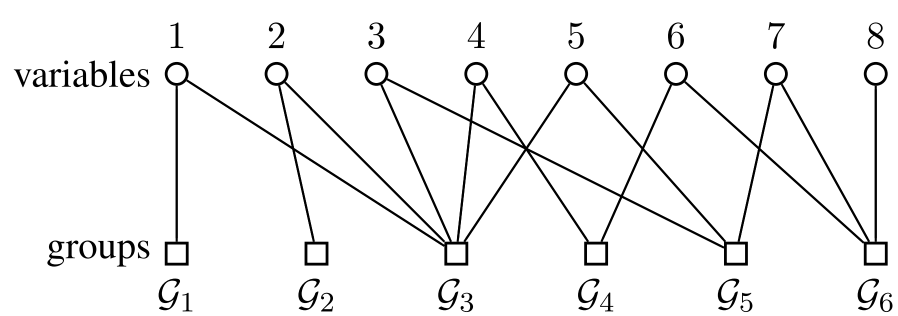
Nirav Bhan, Luca Baldassarre, Volkan Cevher

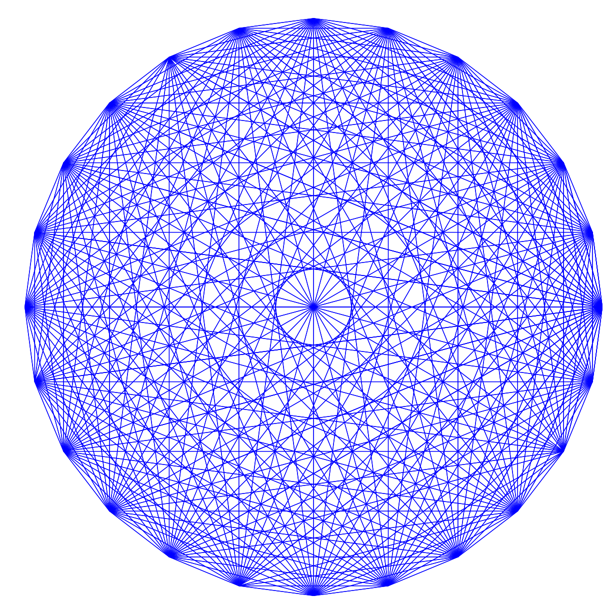
Laboratory for Information and Inference Systems

Ecole Polytechnique Federale de Lausanne

volkan.cevher@epfl.ch

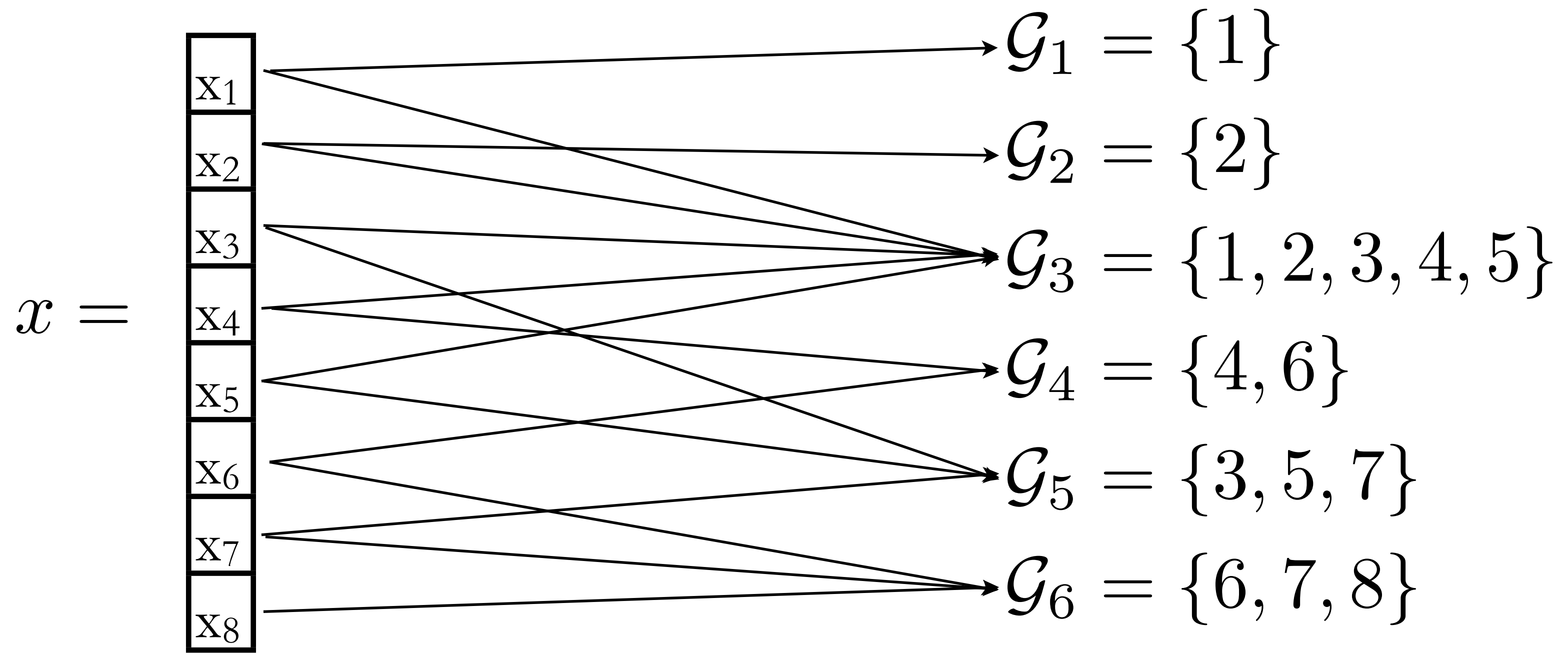






lions@epfl

A generalization of sparsity



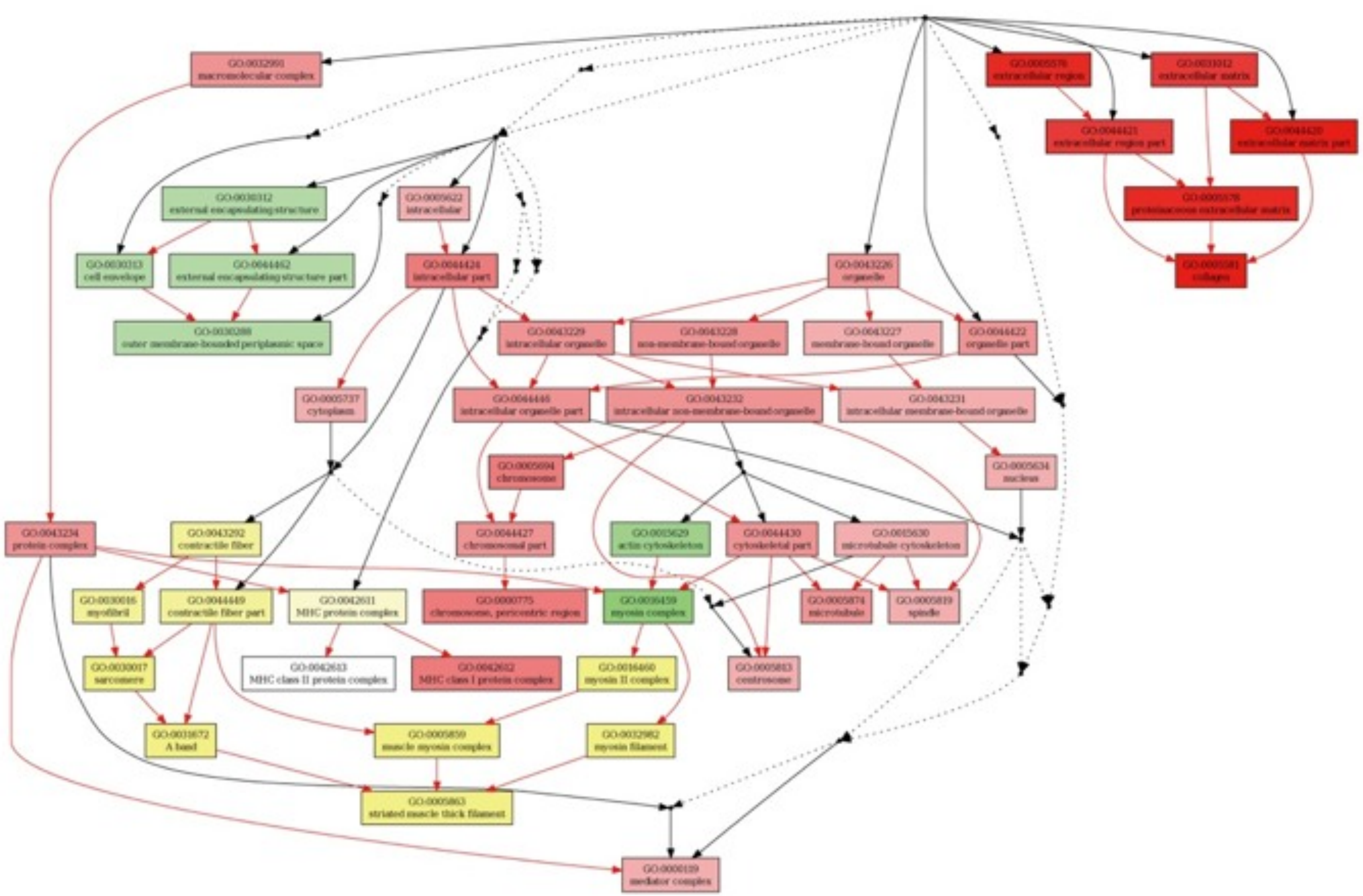
Group-structure: a collection of groups of variables

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

Group-based interpretations

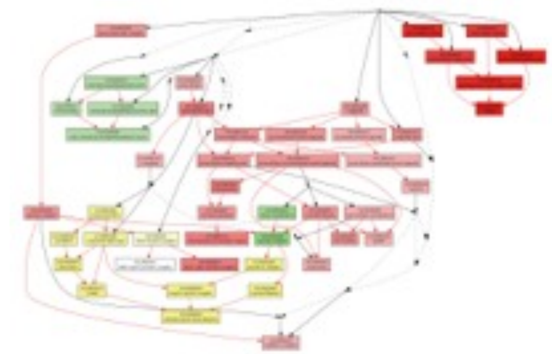
Group models are ubiquitous.
Examples:

- Genetic Pathways in Microarray data analysis

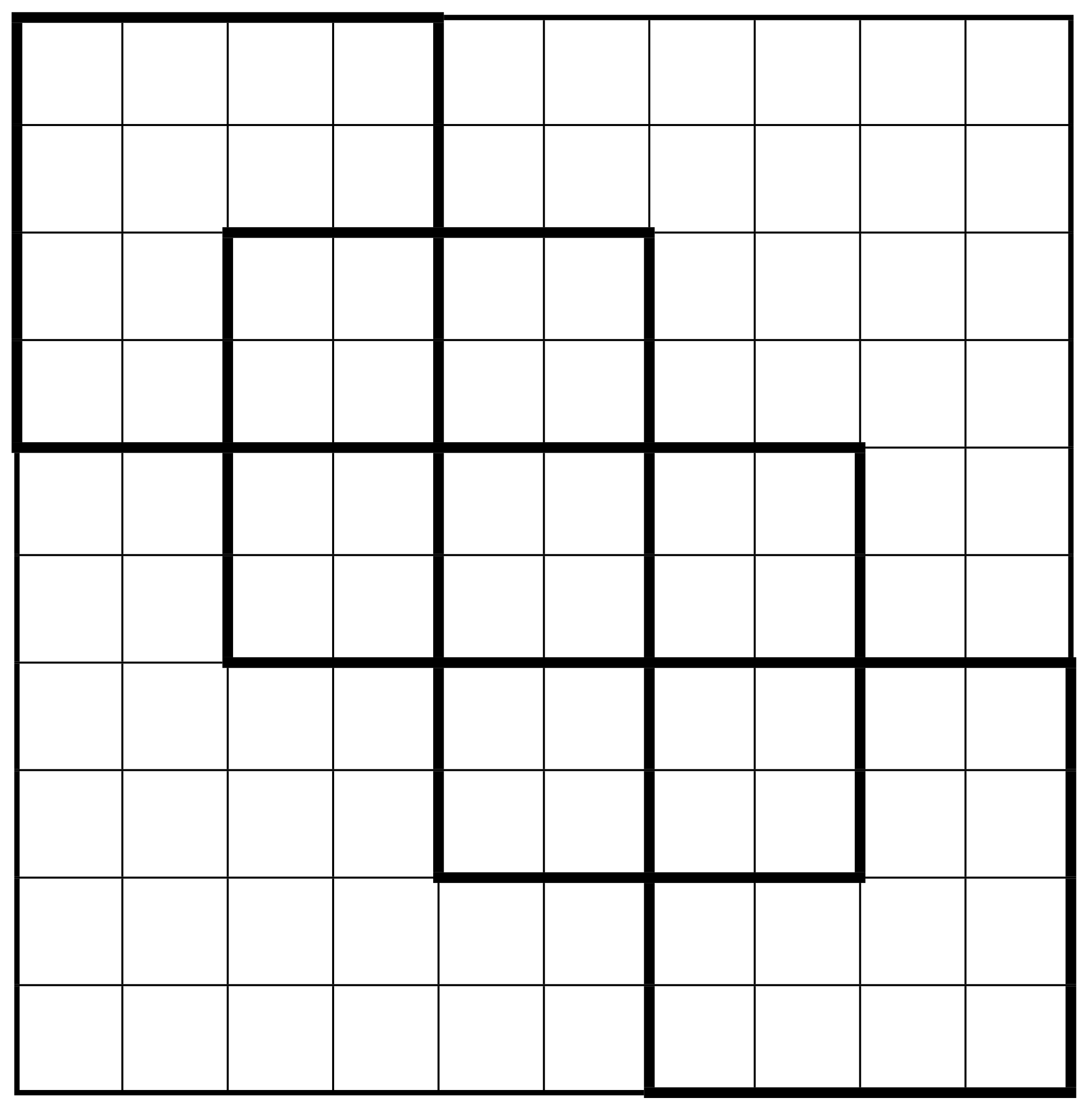


Group-based interpretations

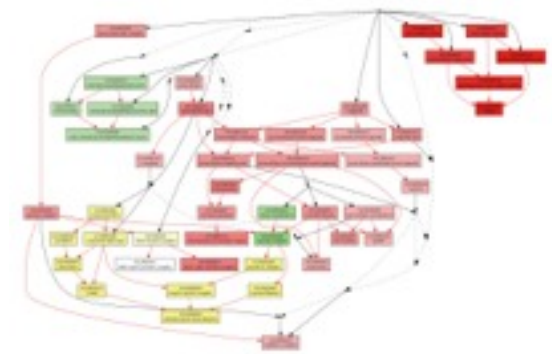
Group models are ubiquitous.
Examples:

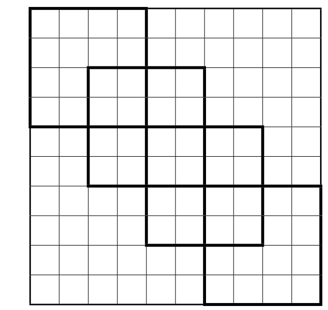


- Genetic Pathways in Microarray data analysis
- Image patches in image processing



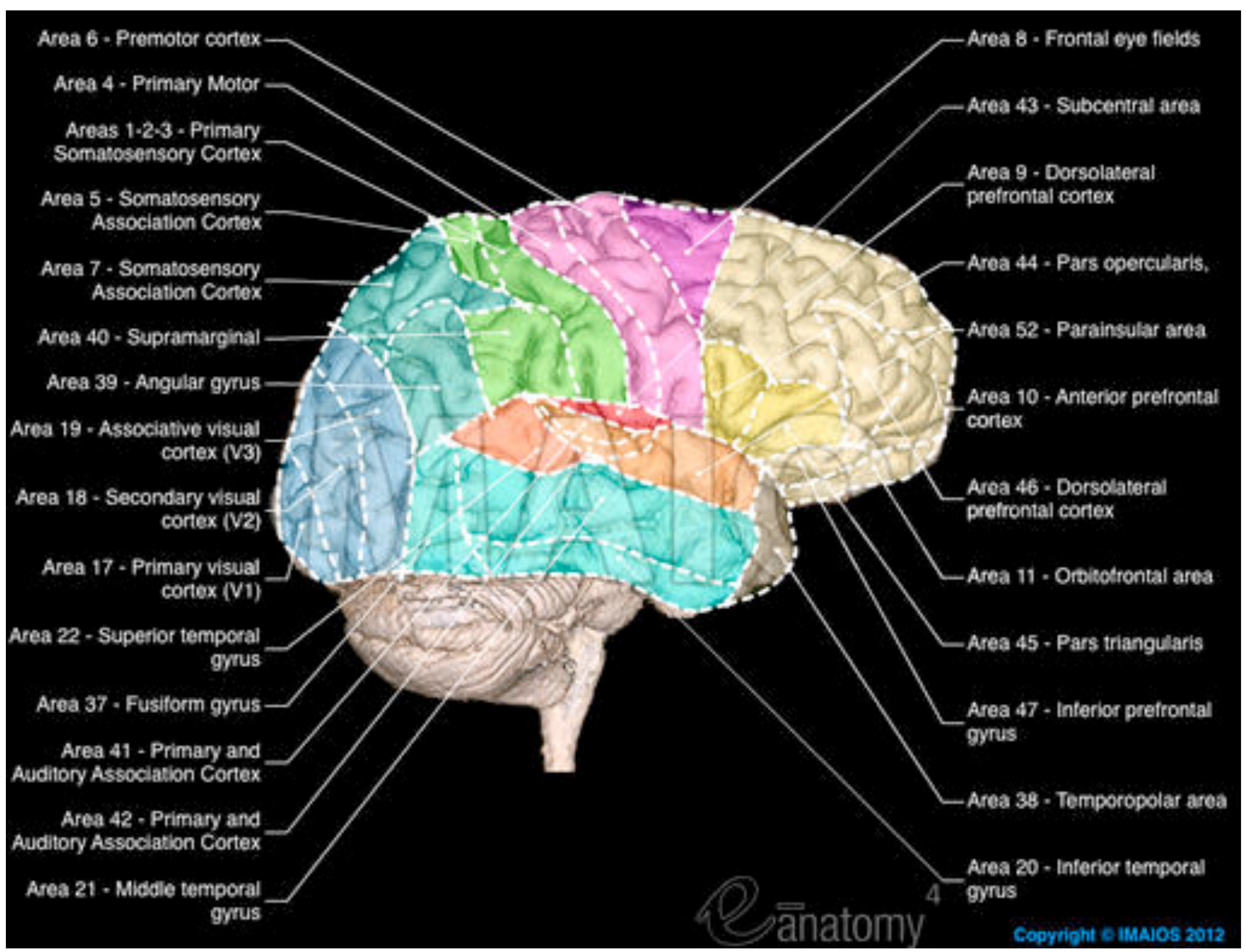
Group-based interpretations



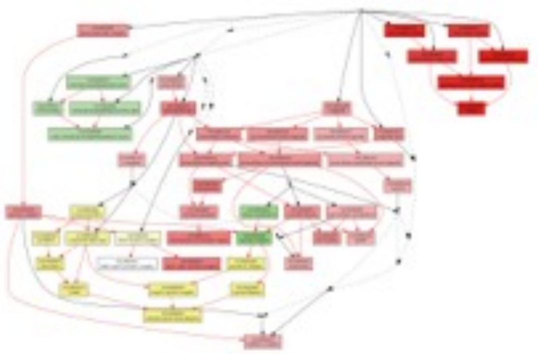


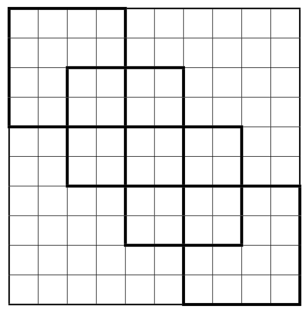
Group models are ubiquitous.
Examples:

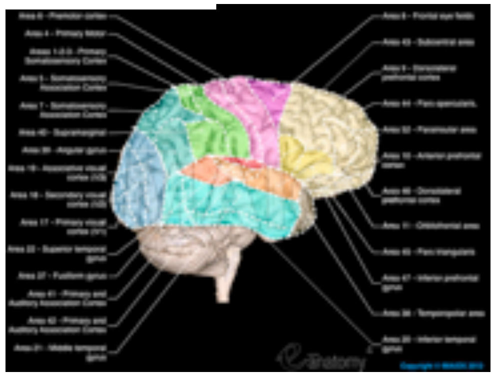
- Genetic Pathways in Microarray data analysis
- Image patches in image processing
- Brain regions in neuroimaging



Group-based interpretations







Group models are ubiquitous.
Examples:

- Genetic Pathways in Microarray data analysis
- Image patches 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 patches contain foregound objects?
- 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

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

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

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

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

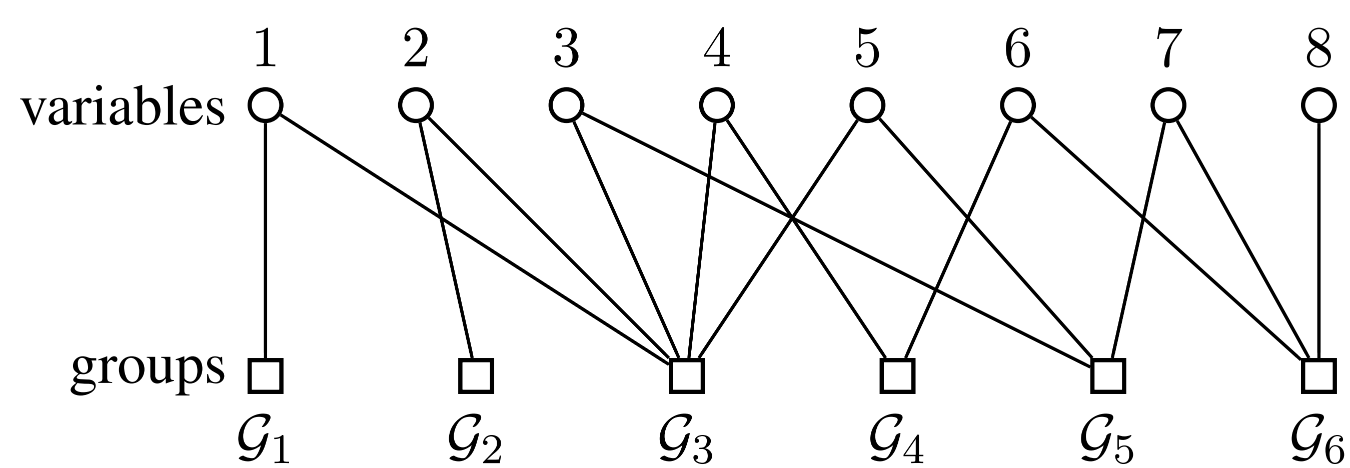
Group structure: a collection of subsets of the ground set

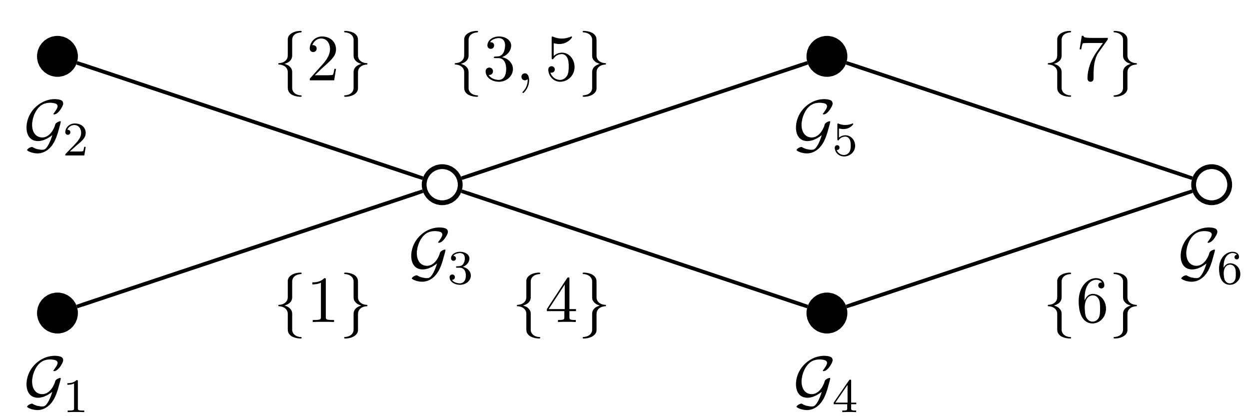
$$\mathfrak{G} = \{\mathcal{G}_1, \dots, \mathcal{G}_M\} \qquad \mathcal{G}_j \subseteq \mathcal{N},\ \forall j = 1, \dots, M$$

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:

$$\mathcal{G}_1 = \{1\}$$
$$\mathcal{G}_2 = \{2\}$$
$$\mathcal{G}_3 = \{1, 2, 3, 4, 5\}$$
$$\mathcal{G}_4 = \{4, 6\}$$
$$\mathcal{G}_5 = \{3, 5, 7\}$$
$$\mathcal{G}_6 = \{6, 7, 8\}$$





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. $\mathrm{supp}(x) \subseteq \bigcup_{\mathcal{G} \in S} \mathcal{G}$

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

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

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

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

ω is a M-dimensional binary variable

Definitions II

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

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

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

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

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

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

Group ell0 ”norm” and approximations

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

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

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

Group ell0 ”norm” and approximations

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

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

G-group sparse **<u>approximation</u>** 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

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

Group ell0 "norm" and approximations

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

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

G-group sparse **<u>approximation</u>** 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

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

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

Group ell0 ”norm” and approximations

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

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

G-group sparse <u>approximation</u> 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 $S^G(\hat{x}) = \{ \mathcal{G}_j \in \mathfrak{G} : \omega_j^G = 1 \}$

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

Group ell0 ”norm” and approximations

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

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

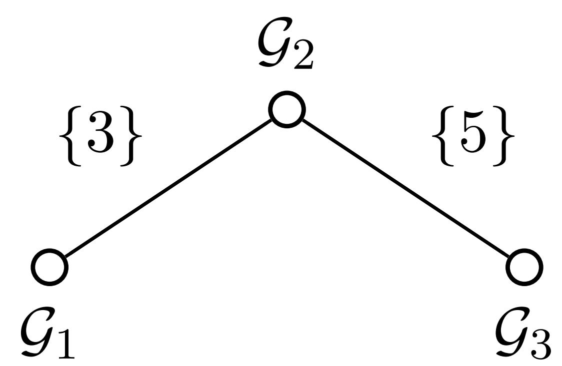
Weighted Maximum Coverage

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

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

- Only pairwise overlaps
- No loops

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



Tractability of Interpretability I

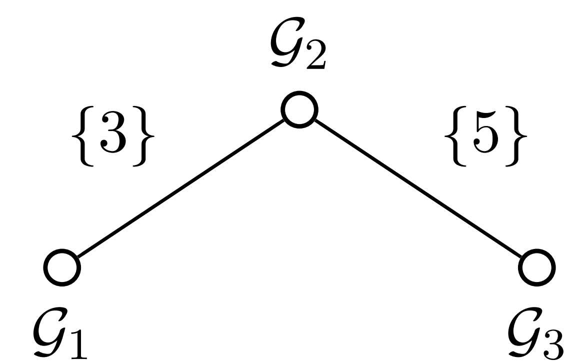
Weighted Maximum Coverage

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

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

- 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

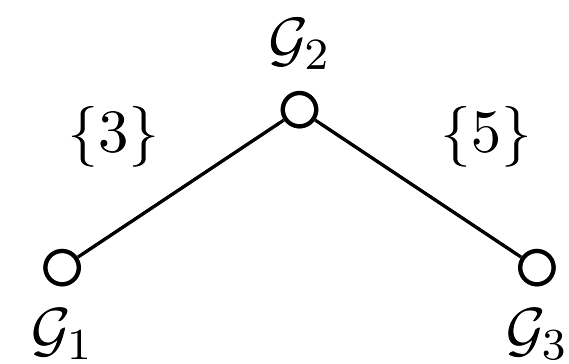
Weighted Maximum Coverage

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

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

- 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 $O(GM^2)$

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 I

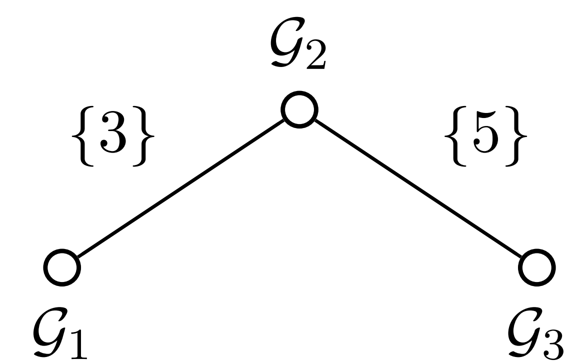
Weighted Maximum Coverage

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

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

- 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(GM²)

Discrete relaxation: linear program

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

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

Tractability of Interpretability II

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 $\mathcal{P}$

Tractability of Interpretability II

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

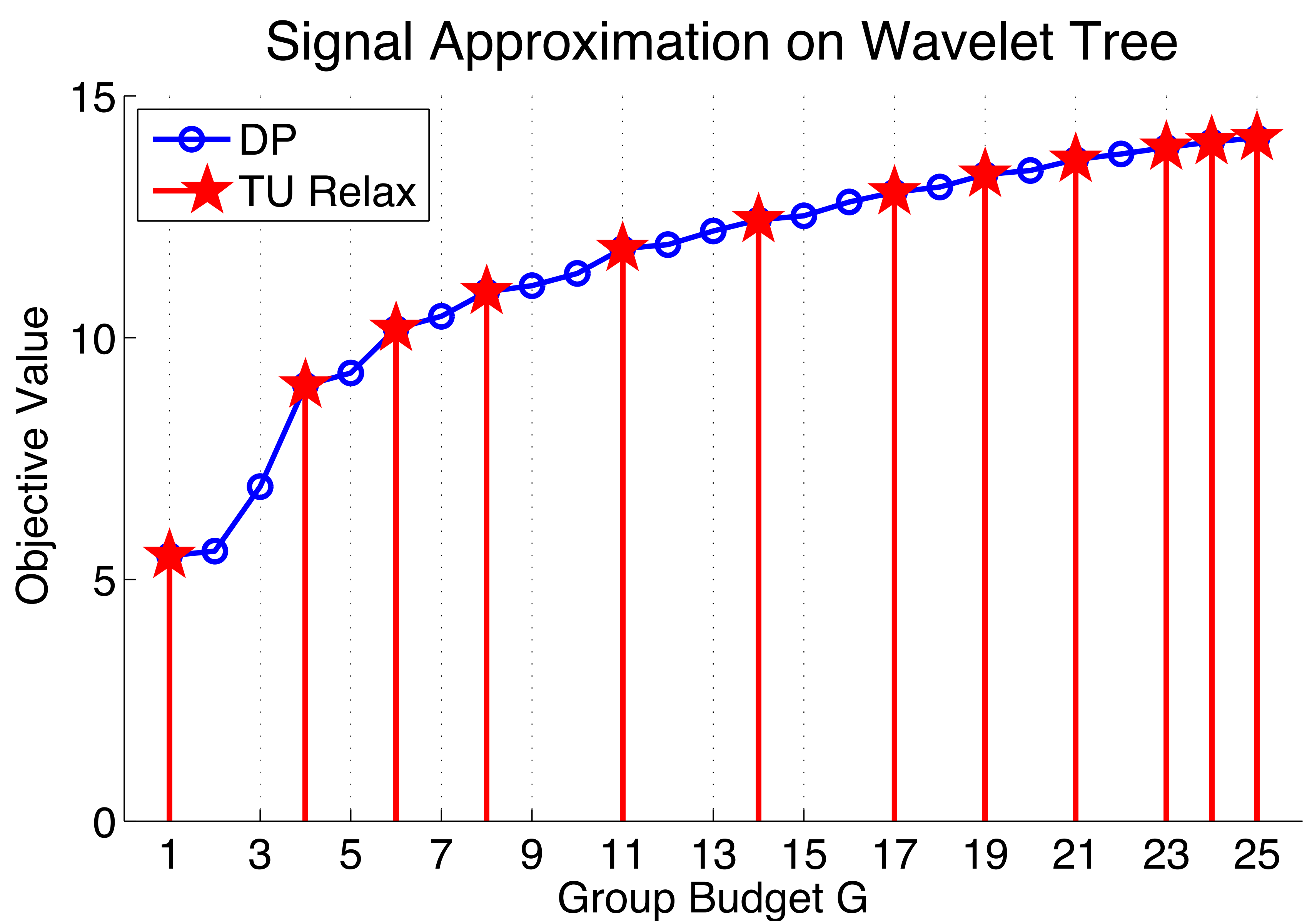
$$\mathcal{P} = \left\{ G, \; \sum_i x_i^2 y_i \right\}_{G=1}^{M}$$

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

Example

- 1D block signals have a sparse representation via the Haar wavelet coefficients.
- The coefficients are arranged hierarchically.
- This hierarchy is a special group structure, that we can handle **exactly**.

If TU-relax obtains a G-group solution, it solves the original problem



Convex relaxations

Latent Group Lasso Norm:
promotes sparsity at the group level [Obozinski et al., 2011]

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

The group cover is defined by the non-zero terms in the optimal decomposition.

$\breve{\mathcal{S}}(x) := \{\mathcal{G}_j \in \mathfrak{G} : \exists\ \mathbf{v} \in \mathcal{V}(x) \text{ s.t. } \mathbf{v}_j \neq 0\}$ $\quad$ $\mathcal{V}(x)$ is the set of solutions of (*)

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

Convex relaxations

Latent Group Lasso Norm:

promotes sparsity at the group level [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 is defined by the non-zero terms in the optimal decomposition.

$\breve{\mathcal{S}}(x) := \{\mathcal{G}_j \in \mathfrak{G} : \exists\, \mathbf{v} \in \mathcal{V}(x) \text{ s.t. } \mathbf{v}_j \neq 0\}$ $\quad \mathcal{V}(x)$ is the set of solutions of (*)

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

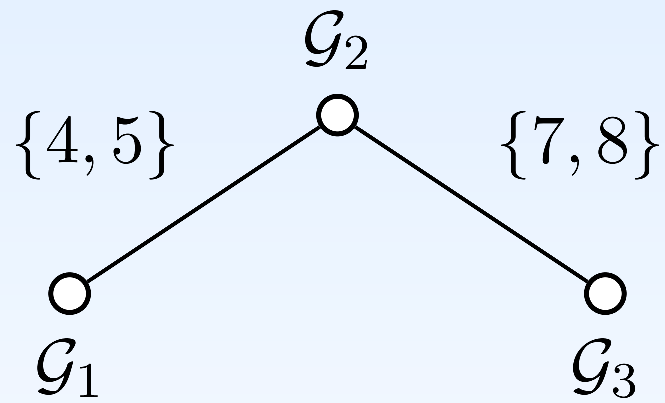
Group-sparse approximation: $\hat{x} = \operatorname*{argmin}_{z \in \mathbb{R}^N} \left\{ \|x - z\|_2^2 : \|z\|_{\mathfrak{G},\{1,2\}} \leq \lambda \right\}$

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

Example

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

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



Loopless pairwise overlapping groups

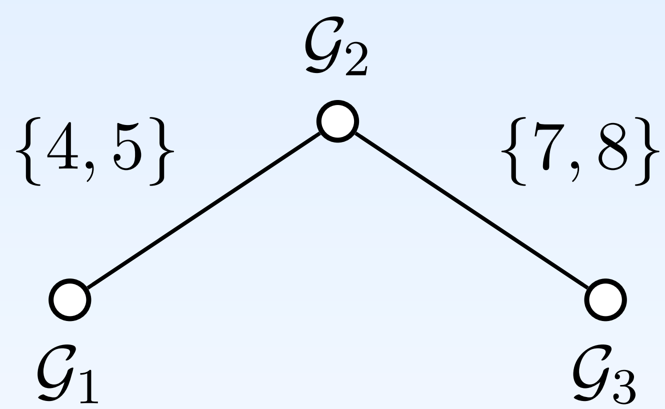
$$x = [\overbrace{0\ 0\ 1\ 1\ 1}^{\mathcal{G}_1}\ 0\ \overbrace{1\ 1\ 1\ 0\ 0}^{\mathcal{G}_3}]^\top$$

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

Example

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

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



Loopless pairwise overlapping groups

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

$\mathcal{G}_2$

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 $0 < \lambda \leq 2$

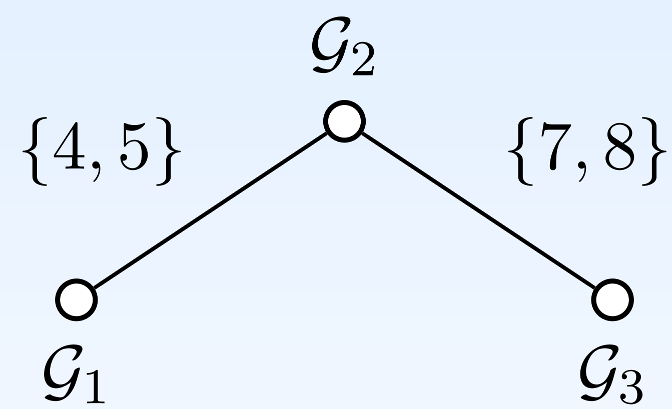
Convex relaxation solution: $\breve{\mathcal{S}}(x) = \mathfrak{G}$ with $d_j = \text{constant}$

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

Example

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

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



Loopless pairwise overlapping groups

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

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 $d_2 > \dfrac{2}{\sqrt{3}}$

Need to know beforehand which groups are irrelevant!

Generalization

- Introduce sparsity budget K.
- Allow individual element within a group to be selected.
- Discrete problem still NP-HARD.

Generalized Weighted Maximum Coverage

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

- Hierarchical constraints can be encoded in the Generalized WMC.
- We developed a polynomial-time Dynamic Program to solve the GWMC for hierarchical constraints.
- Hierarchical constraints are TU: the discrete relaxation (in both group and sparsity budget) allows polynomial solvers.

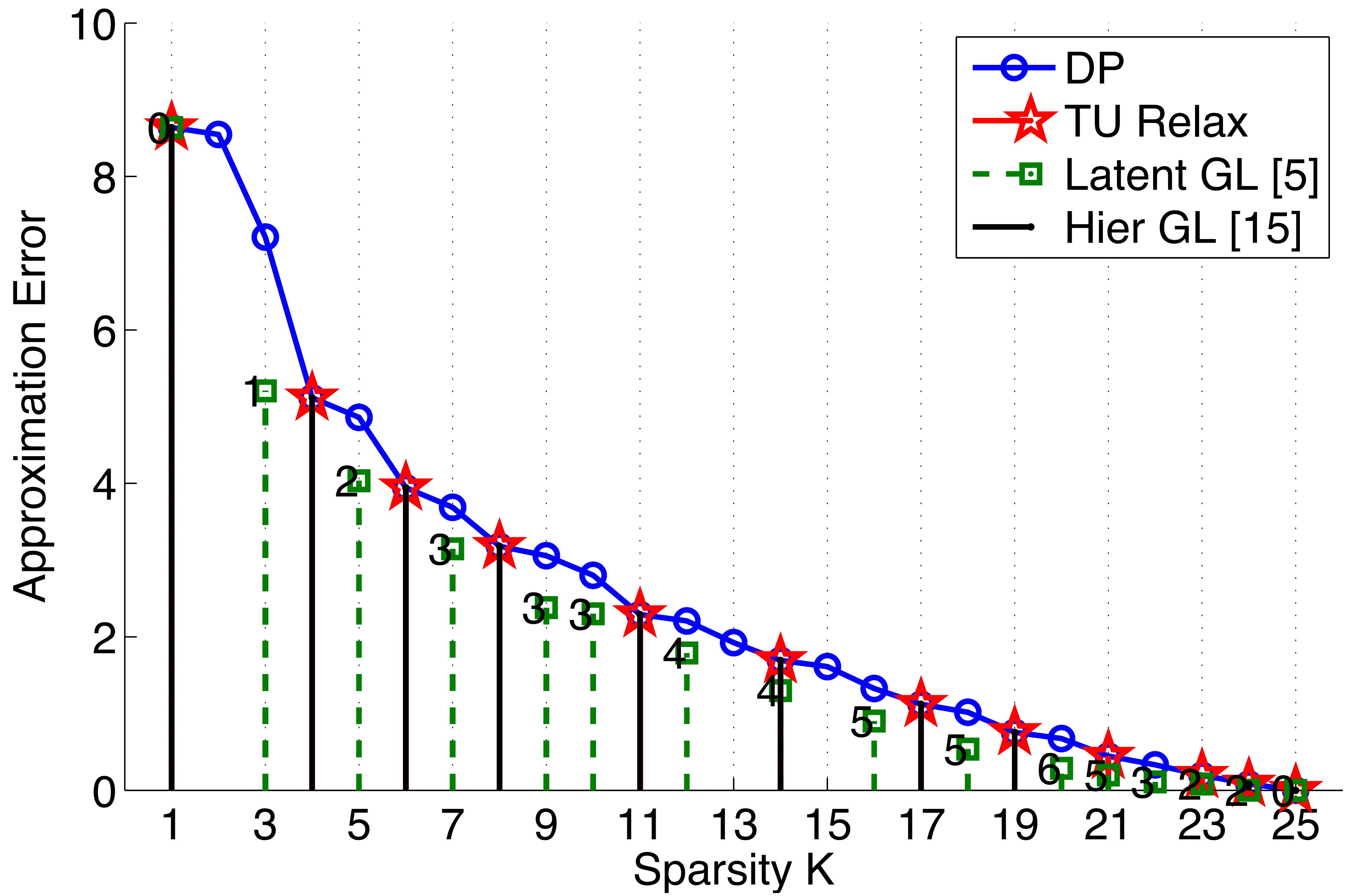
$$\max_{\omega \in \mathbb{B}^M,\ y \in \mathbb{B}^N} \left\{ \sum_{i=1}^N y_i x_i^2 - \lambda_G \sum_{j=1}^M \omega_j - \lambda_K \sum_{i=1}^N y_i : A^{\mathfrak{G}} \omega \geq y \right\}$$

Example: Approximation using wavelet tree

- 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

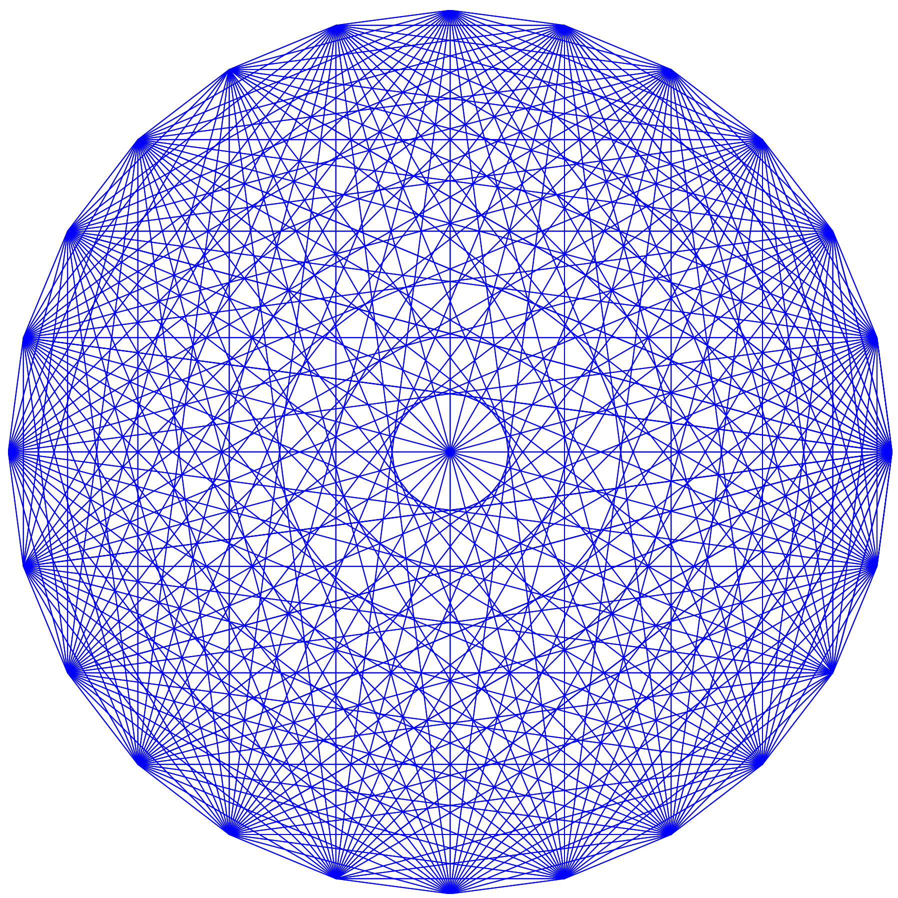
- Convex relaxations **cannot solve** the original discrete NP-hard problem and give only a subset of the solutions.

- We identified group structures that lead to **tractable group-based interpretations.**

- Totally Unimodular group structures allow **efficient relaxations.**

- However, some relevant application do not have tractable group structures:

  ➡ e.g. Breast cancer dataset: group-graph of the top 25 pathways from the Molecular Signature Database [Subramanian et al., 2005]



References

- 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.