A totally unimodular view of structured sparsity

Marwa El Halabi
marwa.elhalabi@epfl.ch

Laboratory for Information and Inference Systems (LIONS)
École Polytechnique Fédérale de Lausanne (EPFL)
Switzerland

SPARS 2015

Joint work with: Volkan Cevher

lions@epfl

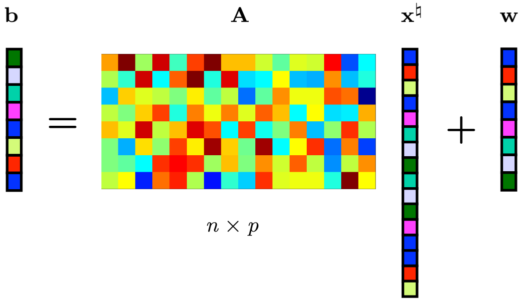
MARIE CURIE ACTIONS

FNSNF
FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

erc

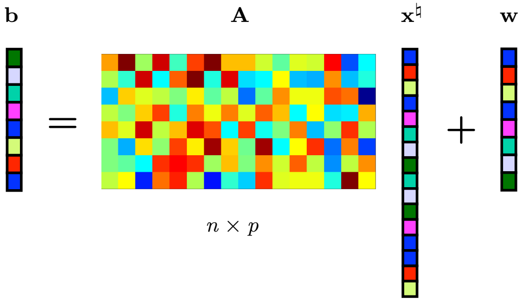
EPFL

Linear inverse problems

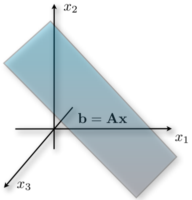


Applications: **Machine learning, signal processing, theoretical computer science...**

Linear inverse problems

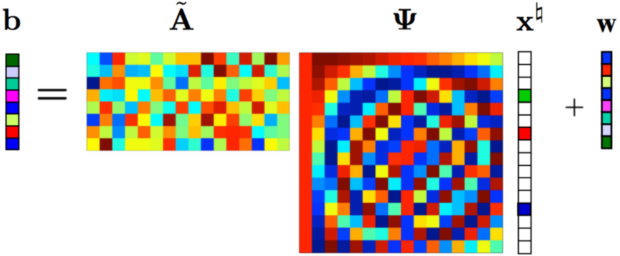


Nullspace (null) of $\mathbf{A}$: $\quad \mathbf{x}^\natural + \boldsymbol{\delta} \to \mathbf{b}, \quad \forall \boldsymbol{\delta} \in \text{null}(\mathbf{A})$



We need additional information on $\mathbf{x}^\natural$!

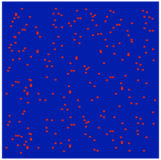
Sparsity to the rescue!



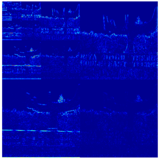
- $\mathbf{b} \in \mathbb{R}^n$, $\tilde{\mathbf{A}} \in \mathbb{R}^{n \times p}$, and $n < p$
- $\mathbf{\Psi} \in \mathbb{R}^{p \times p}$, $\mathbf{x}^\natural \in \Sigma_s$, and $s < n \ll p$
- Sparsity allows tractable recovery with $n = O(p \log(p/s))$

Beyond sparsity towards model-based or *structured* sparsity

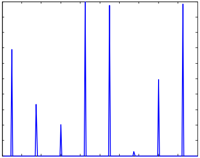
- The following signals can look the **same** from a **sparsity** perspective!



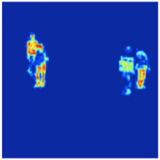
Sparse image



Wavelet coefficients of a natural image

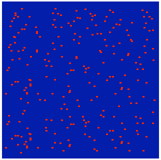


Spike train

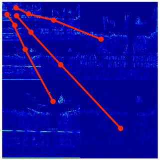


Background substracted image

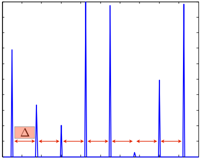
- In reality, these signals have additional structures beyond the simple sparsity



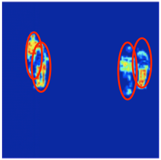
Sparse image



Wavelet coefficients of a natural image



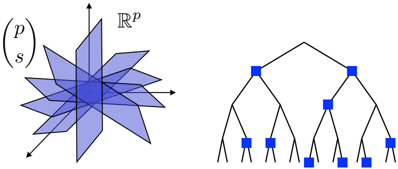
Spike train



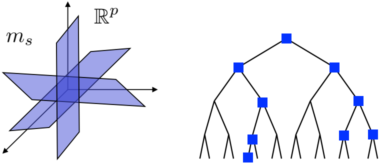
Background substracted image

Beyond sparsity towards model-based or ***structured*** sparsity

Sparsity model: Union of all s-dimensional canonical subspaces.



Structured sparsity model: A **particular** union of m_s s-dimensional canonical subspaces.



Three upshots of structured sparsity: [Baraniuk et al., 2010]

1. Reduced sample complexity
2. Better noise robustness
3. Better interpretability

A simple template for structured sparsity recovery

$\boxed{\text{Find the "}\textit{sparsest}\text{" } \mathbf{x} \text{ subject to } \textit{structure} \text{ and } \textit{data}.}$

- **Sparsity**

  We can generalize this desideratum to other notions of simplicity
- **Structure**

  We only allow certain sparsity patterns
- **Data fidelity**

  We have many choices of convex constraints & losses to represent data; e.g.,

$$\|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 \leq \kappa$$

Simple sparsity

Sparse estimator

$$\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^p} \left\{ \|\mathbf{x}\|_0 : \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 \leq \|\mathbf{w}\|_2 \right\} \qquad (\mathcal{P}_0)$$

where $\|\mathbf{x}\|_0 := \mathbb{1}^T \boldsymbol{s}$, $\boldsymbol{s} = \mathbb{1}_{\mathrm{supp}(\mathbf{x})}, \mathrm{supp}(\mathbf{x}) = \{i | x_i \neq 0\}$

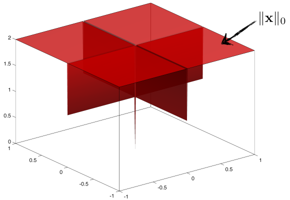
Simple sparsity

Sparse estimator

$$\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^p} \left\{ \|\mathbf{x}\|_0 : \|\mathbf{b}-\mathbf{A}\mathbf{x}\|_2 \le \|\mathbf{w}\|_2 \right\} \qquad (\mathcal{P}_0)$$

where $\|\mathbf{x}\|_0 := \mathbb{1}^T \boldsymbol{s}$, $\boldsymbol{s} = \mathbb{1}_{\mathrm{supp}(\mathbf{x})}, \mathrm{supp}(\mathbf{x}) = \{i | x_i \neq 0\}$

$\|\mathbf{x}\|_0$ over the unit ℓ_∞-ball



$\mathcal{P}_0$ has the following characteristics:

- sample complexity: $\mathcal{O}(s)$
- computational effort: NP-Hard
- stability: No

Simple sparsity

Sparse estimator

$$\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^p} \left\{ \|\mathbf{x}\|_0 : \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 \le \|\mathbf{w}\|_2 \right\} \qquad (\mathcal{P}_0)$$

where $\|\mathbf{x}\|_0 := \mathbb{1}^T \boldsymbol{s}$, $\boldsymbol{s} = \mathbb{1}_{\mathrm{supp}(\mathbf{x})}$, $\mathrm{supp}(\mathbf{x}) = \{i | x_i \neq 0\}$

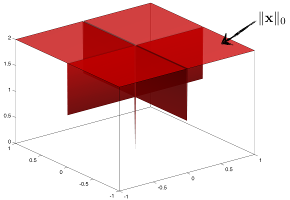
$\mathcal{P}_0$ has the following characteristics:

- sample complexity: $\mathcal{O}(s)$
- computational effort: NP-Hard
- stability: No

Convex relaxation:

Convex envelope is the largest convex lower bound.

$\|\mathbf{x}\|_0$ over the unit ℓ_∞-ball



A technicality: Restrict $\mathbf{x}^\natural \in [-1, 1]^p$.

Simple sparsity

Sparse estimator

$$\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^p} \left\{ \|\mathbf{x}\|_0 : \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 \leq \|\mathbf{w}\|_2 \right\} \qquad (\mathcal{P}_0)$$

where $\|\mathbf{x}\|_0 := \mathbb{1}^T \boldsymbol{s}, \boldsymbol{s} = \mathbb{1}_{\mathrm{supp}(\mathbf{x})}, \mathrm{supp}(\mathbf{x}) = \{i | x_i \neq 0\}$

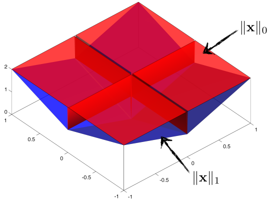
$\mathcal{P}_0$ has the following characteristics:

- sample complexity: $\mathcal{O}(s)$
- computational effort: NP-Hard
- stability: No

Convex relaxation:

Convex envelope is the largest convex lower bound.

$\|\mathbf{x}\|_1$ is the convex envelope of $\|\mathbf{x}\|_0$



A technicality: Restrict $\mathbf{x}^\natural \in [-1, 1]^p$.

The role of convexity: Tractable & stable recovery

Basis pursuit estimator

$$\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^p} \left\{ \|\mathbf{x}\|_1 : \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 \le \|\mathbf{w}\|_2 \,, \|\mathbf{x}\|_\infty \le 1 \right\} \qquad (BP)$$

where $\|\mathbf{x}\|_1 := \mathbb{1}^T|\mathbf{x}|$

The role of convexity: Tractable & stable recovery

Basis pursuit estimator

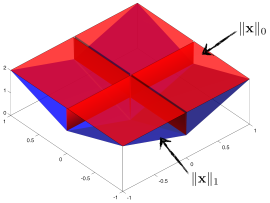
$$\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^p} \left\{ \|\mathbf{x}\|_1 : \|\mathbf{b}-\mathbf{A}\mathbf{x}\|_2 \leq \|\mathbf{w}\|_2 , \|\mathbf{x}\|_\infty \leq 1 \right\} \qquad (BP)$$

where $\|\mathbf{x}\|_1 := \mathbb{1}^T|\mathbf{x}|$

$\|\mathbf{x}\|_1$ is the convex envelope of $\|\mathbf{x}\|_0$

BP has the following characteristics:

- sample complexity: $\mathcal{O}(s\log(\frac{p}{s}))$
- computational effort: Tractable; $\mathcal{O}(n^2p^{1.5}\log(\frac{1}{\epsilon}))$ via IPM (for $w=0$)
- stability: Robust to noise



A technicality: Restrict $\mathbf{x}^\natural \in [-1,1]^p$.

Convex relaxation in general?

We encode the structure over the support by $g(\mathbf{x}) = F(\text{supp}(\mathbf{x}))$

- $\text{supp}(\mathbf{x}) = \{i | x_i \neq 0\}$
- $F(\boldsymbol{s}) : \{0,1\}^p \to \mathbb{R} \cup \{+\infty\}$

How does the current literature compute the convex relaxation of g ?

1. Case by case heuristics
2. Biconjugation ($\equiv$ convex envelope): Fenchel conjugate of Fenchel conjugate.

Fenchel conjugate: $g^*(\mathbf{y}) := \sup_{\mathbf{x}:\text{dom}(g)} \mathbf{x}^T\mathbf{y} - g(\mathbf{x})$.

Challenge: Fenchel conjugate of a general g is NP-Hard.

Tractable convex relaxation

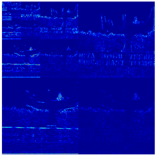
Prior work:

1. Monotone submodular penalties [Bach, 2010]
   - Tractable biconjugation via Lovász extension
   - Limited to certain structures
2. ℓ_q-regularized combinatorial functions [Obozinski and Bach, 2012] $(\mu F(\mathrm{supp}(\mathbf{x})) + \nu\|\mathbf{x}\|_q)$
   - Homogeneous convex envelope
   - Tractable biconjugation even for some non-submodular functions
   - Not always tractable
   - May lose structure [El Halabi and Cevher, 2014]

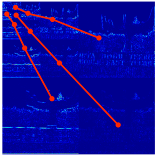
Our goal:

- Easy to design priors
- Tractable biconjugation
- Applies to as many structures as possible

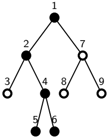
Tree sparsity



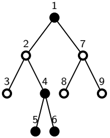
Wavelet coefficients



Wavelet tree



Valid selection of nodes



Invalid selection of nodes

We seek the sparsest signal with a rooted connected tree support. [Baraniuk et al., 2010]

Objective: $\|\mathbf{x}\|_0 \equiv \mathbb{1}^T \boldsymbol{s}$ s.t. $\mathbb{1}_{\mathrm{supp}(\mathbf{x})} = \boldsymbol{s}$

Linear constraint: A valid support satisfy $s_{\text{parent}} \geq s_{\text{child}}$ over tree $\mathcal{T}$

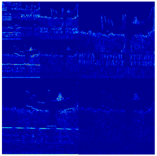
$$\boxed{\boldsymbol{T}\mathbb{1}_{\mathrm{supp}(\mathbf{x})} := \boldsymbol{T}\boldsymbol{s} \geq 0}$$

where $\boldsymbol{T}$ is the directed edge-node incidence matrix.

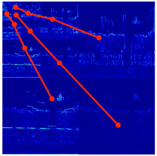
$$\boldsymbol{T} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \end{bmatrix}$$

Key Observation: Matrix $\boldsymbol{T}$ is totally unimodular (TU).

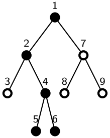
Tree sparsity



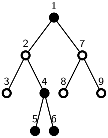
Wavelet coefficients



Wavelet tree



Valid selection of nodes



Invalid selection of nodes

We seek the sparsest signal with a rooted connected tree support. [Baraniuk et al., 2010]

Objective: $\|\mathbf{x}\|_0 \equiv \mathbb{1}^T \boldsymbol{s}$ s.t. $\mathbb{1}_{\text{supp}(\mathbf{x})} = \boldsymbol{s}$

Linear constraint: A valid support satisfy $s_{\text{parent}} \geq s_{\text{child}}$ over tree $\mathcal{T}$

$$\boxed{\boldsymbol{T}\mathbb{1}_{\text{supp}(\mathbf{x})} := \boldsymbol{T}\boldsymbol{s} \geq 0}$$

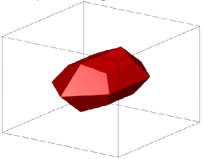
where $\boldsymbol{T}$ is the directed edge-node incidence matrix.

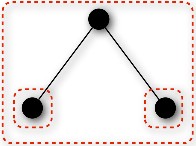
$$\boldsymbol{T} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \end{bmatrix}$$

Key Observation: Matrix $\boldsymbol{T}$ is totally unimodular (TU).

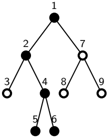
Total unimodular (TU): $\boldsymbol{M} \in \mathbb{R}^{l \times m}$ is TU iff the determinant of every square submatrix of $\boldsymbol{M}$ is 0, or ± 1.

Tree sparsity





$\mathfrak{G}_H = \{\{1,2,3\},\{2\},\{3\}\}$



valid selection of nodes

We seek the sparsest signal with a rooted connected tree support. [Baraniuk et al., 2010]

Objective: $\|\mathbf{x}\|_0 \equiv \mathbb{1}^T \boldsymbol{s}$ s.t. $\mathbb{1}_{\text{supp}(\mathbf{x})} = \boldsymbol{s}$

Linear constraint: A valid support satisfy $s_{\text{parent}} \geq s_{\text{child}}$ over tree $\mathcal{T}$

$$\boxed{\boldsymbol{T}\mathbb{1}_{\text{supp}(\mathbf{x})} := \boldsymbol{T}\boldsymbol{s} \geq 0}$$

where $\boldsymbol{T}$ is the directed edge-node incidence matrix.

$$\boldsymbol{T} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \end{bmatrix}$$

Key Observation: Matrix $\boldsymbol{T}$ is totally unimodular (TU).

Biconjugate: Tractable! $\sum_{\mathcal{G} \in \mathfrak{G}_H} \|x_{\mathcal{G}}\|_\infty$

This is known as the hierarchical group lasso [Zhao et al., 2006, Jenatton et al., 2011b].

How is TU property helpful?

Recall Fenchel conjugate: $g^*(\mathbf{y}) := \sup_{\mathbf{x}:\mathsf{dom}(g)} \mathbf{x}^T\mathbf{y} - g(\mathbf{x})$.

How is TU property helpful?

Recall Fenchel conjugate: $g^*(\mathbf{y}) := \sup_{\mathbf{x}:\mathrm{dom}(g)} \mathbf{x}^T\mathbf{y} - g(\mathbf{x})$.

Integer program

$$\boldsymbol{\beta}^\natural \in \arg\max_{\boldsymbol{\beta}\in\mathbb{Z}^m} \{\boldsymbol{\theta}^T\boldsymbol{\beta} : \boldsymbol{M}\boldsymbol{\beta} \leq \boldsymbol{c}, \boldsymbol{\beta} \geq 0\} \quad \text{(IP)}$$

NP-Hard (in general)

How is TU property helpful?

Recall Fenchel conjugate: $g^*(\mathbf{y}) := \sup_{\mathbf{x}:\mathrm{dom}(g)} \mathbf{x}^T\mathbf{y} - g(\mathbf{x})$.

Integer program

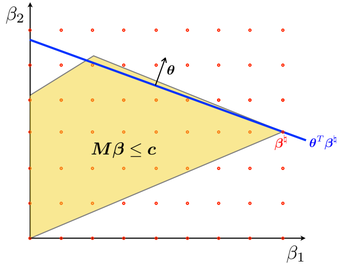
$$\boldsymbol{\beta}^\natural \in \arg\max_{\boldsymbol{\beta}\in\mathbb{Z}^m} \{\boldsymbol{\theta}^T\boldsymbol{\beta} : \boldsymbol{M}\boldsymbol{\beta} \leq \boldsymbol{c}, \boldsymbol{\beta} \geq 0\} \quad \text{(IP)}$$

NP-Hard (in general)

Convex relaxation

$$\boldsymbol{\beta}^\star \in \arg\max_{\boldsymbol{\beta}\in\mathbb{R}^m} \{\boldsymbol{\theta}^T\boldsymbol{\beta} : \boldsymbol{M}\boldsymbol{\beta} \leq \boldsymbol{c}, \boldsymbol{\beta} \geq 0\} \quad \text{(LP)}$$

Obtains an upperbound



How is TU property helpful?

Recall Fenchel conjugate: $g^*(\mathbf{y}) := \sup_{\mathbf{x}:\mathrm{dom}(g)} \mathbf{x}^T\mathbf{y} - g(\mathbf{x})$.

Integer program

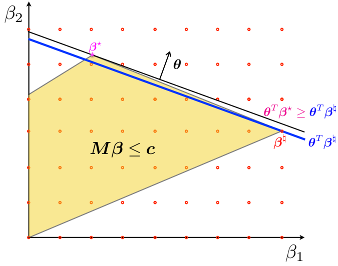
$$\boldsymbol{\beta}^\natural \in \arg\max_{\boldsymbol{\beta}\in\mathbb{Z}^m} \{\boldsymbol{\theta}^T\boldsymbol{\beta} : \boldsymbol{M}\boldsymbol{\beta} \le \boldsymbol{c}, \boldsymbol{\beta} \ge 0\} \quad \text{(IP)}$$

NP-Hard (in general)

Convex relaxation

$$\boldsymbol{\beta}^\star \in \arg\max_{\boldsymbol{\beta}\in\mathbb{R}^m} \{\boldsymbol{\theta}^T\boldsymbol{\beta} : \boldsymbol{M}\boldsymbol{\beta} \le \boldsymbol{c}, \boldsymbol{\beta} \ge 0\} \quad \text{(LP)}$$

Obtains an upperbound



How is TU property helpful?

Recall Fenchel conjugate: $g^*(\mathbf{y}) := \sup_{\mathbf{x}:\mathrm{dom}(g)} \mathbf{x}^T\mathbf{y} - g(\mathbf{x})$.

Integer program

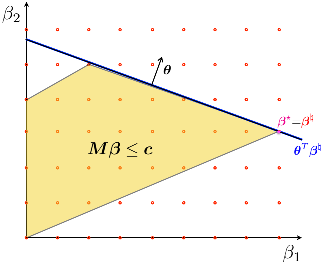
$$\boldsymbol{\beta}^\natural \in \arg\max_{\boldsymbol{\beta}\in\mathbb{Z}^m} \{\boldsymbol{\theta}^T\boldsymbol{\beta} : \boldsymbol{M}\boldsymbol{\beta} \leq \boldsymbol{c}, \boldsymbol{\beta} \geq 0\} \quad \text{(IP)}$$

NP-Hard (in general)

Convex relaxation

$$\boldsymbol{\beta}^\star \in \arg\max_{\boldsymbol{\beta}\in\mathbb{R}^m} \{\boldsymbol{\theta}^T\boldsymbol{\beta} : \boldsymbol{M}\boldsymbol{\beta} \leq \boldsymbol{c}, \boldsymbol{\beta} \geq 0\} \quad \text{(LP)}$$

Obtains an upperbound



Integral polytopes

$$\mathcal{P} = \{\boldsymbol{\beta} | \boldsymbol{M}\boldsymbol{\beta} \leq \boldsymbol{c}, \boldsymbol{\beta} \geq 0\}$$

When every vertex of $\mathcal{P}$ is integer, **LP is an "exact" relaxation.**

How is TU property helpful?

Recall Fenchel conjugate: $g^*(\mathbf{y}) := \sup_{\mathbf{x}:\mathrm{dom}(g)} \mathbf{x}^T\mathbf{y} - g(\mathbf{x})$.

Integer program

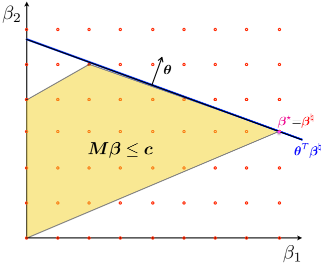
$$\boldsymbol{\beta}^\natural \in \arg\max_{\boldsymbol{\beta}\in\mathbb{Z}^m} \{\boldsymbol{\theta}^T\boldsymbol{\beta} : \boldsymbol{M}\boldsymbol{\beta} \le \boldsymbol{c}, \boldsymbol{\beta} \ge 0\} \quad \text{(IP)}$$

NP-Hard (in general)

Convex relaxation

$$\boldsymbol{\beta}^\star \in \arg\max_{\boldsymbol{\beta}\in\mathbb{R}^m} \{\boldsymbol{\theta}^T\boldsymbol{\beta} : \boldsymbol{M}\boldsymbol{\beta} \le \boldsymbol{c}, \boldsymbol{\beta} \ge 0\} \quad \text{(LP)}$$

Obtains an upperbound



Integral polytopes

$$\mathcal{P} = \{\boldsymbol{\beta} | \boldsymbol{M}\boldsymbol{\beta} \le \boldsymbol{c}, \boldsymbol{\beta} \ge 0\}$$

When every vertex of $\mathcal{P}$ is integer, **LP is an "exact" relaxation.**

Polyhedra $\mathcal{P} = \{\boldsymbol{M}\boldsymbol{\beta} \le \boldsymbol{c}, \boldsymbol{\beta} \ge 0\}$ has integer vertices when $\boldsymbol{M}$ is TU and $\boldsymbol{c}$ is integer

TU structures

TU modeling framework offers:

- ✓ General easy to design structure sparsity template,
- ✓ Tight convexifications for submodular/non-submodular examples,
- ✓ Efficient evaluation of biconjugate via LP.

Model	Convex envelope	Submodular structure	ℓ_q-regularized structure	TU structure
Minimal Group cover	Latent group lasso	✗	✓	✓*
Group intersection sparsity	Group lasso	✓	✓	✓
Rooted connected tree	Hierarchical group lasso	✗	?	✓
Sparse group knapsack	ℓ_1 with knapsack constraint	✗	≠	✓*
Relaxed group knapsack	Exclusive lasso	✗	✓**	✓*

$\star$ this is shown only for group structures that lead to TU constraints.

$\star\star$ this is shown only for partition groups.

Template for TU structures

Sparsity and structure together [El Halabi and Cevher, 2014]

Given some weights $\boldsymbol{d} \in \mathbb{R}^d$, $\boldsymbol{e} \in \mathbb{R}^p$ and an integeral vector $\boldsymbol{c} \in \mathbb{Z}^l$, we define

$$g_{TU}(\mathbf{x}) := \boldsymbol{e}^T \boldsymbol{s} + \min_{\boldsymbol{\omega}} \{ \boldsymbol{d}^T \boldsymbol{\omega} : \boldsymbol{M} \begin{bmatrix} \boldsymbol{\omega} \\ \boldsymbol{s} \end{bmatrix} \leq \boldsymbol{c}, \mathbb{1}_{\mathrm{supp}(\mathbf{x})} = \boldsymbol{s}, \boldsymbol{\omega} \in \{0,1\}^d \}$$

for all feasible $\mathbf{x}$, ∞ otherwise. The parameter $\boldsymbol{\omega}$ is useful for latent modeling.

- $\boldsymbol{M}$ is TU.
- "Exact convex relaxation" of: $g^*(\mathbf{y}) = \sup_{\boldsymbol{s} \in \{0,1\}^p} |\mathbf{y}|^T \boldsymbol{s} - F(\boldsymbol{s})$.

Convexification of TU structures

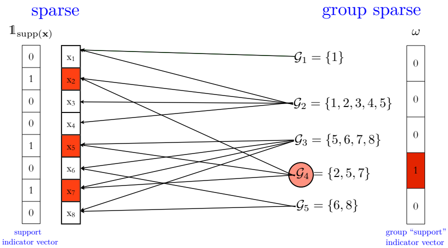
TU convex relaxation given by LP

$$g^{**}_{TU}(\mathbf{x}) := \min_{\boldsymbol{\omega},\boldsymbol{s}}\{\boldsymbol{d}^T\boldsymbol{\omega} + \boldsymbol{e}^T\boldsymbol{s} : \boldsymbol{M}\begin{bmatrix}\boldsymbol{\omega}\\ \boldsymbol{s}\end{bmatrix} \le \boldsymbol{c}, |\mathbf{x}| \le \boldsymbol{s}, \boldsymbol{\omega} \in [0,1]^d, \boldsymbol{s} \in [0,1]^p\}$$

for all feasible $\mathbf{x}$, ∞ otherwise.

- Usually g^{**}_{TU} have closed form solution.

Group cover sparsity: Minimal group cover



We seek the signal covered by a minimal number of groups.
[Baldassarre et al., 2013, Obozinski et al., 2011, Huang et al., 2011]

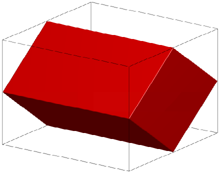
Objective: $\boldsymbol{d}^T\boldsymbol{\omega}$ (Group sparsity)
Linear constraint: For each non-zero coefficient, at least one group containing it is selected

$$\boxed{\boldsymbol{B}\boldsymbol{\omega} \geq \boldsymbol{s}}$$

where $\boldsymbol{B}$ is the biadjacency matrix of $\mathfrak{G}$, i.e., $\boldsymbol{B}_{ij} = 1$ iff i-th coefficient is in $\mathcal{G}_j$.

When $\boldsymbol{B}$ is an interval matrix, or $\mathfrak{G}$ has a loopless group intersection graph [Baldassarre et al., 2013], it is TU.

Group cover sparsity: Minimal group cover



$\mathfrak{G} = \{\{1,2\},\{2,3\}\}$, unit group weights $\boldsymbol{d} = \mathbb{1}$.

We seek the signal covered by a minimal number of groups.
[Baldassarre et al., 2013, Obozinski et al., 2011, Huang et al., 2011]

Objective: $\boldsymbol{d}^T\boldsymbol{\omega}$ (Group sparsity)
Linear constraint: For each non-zero coefficient, at least one group containing it is selected

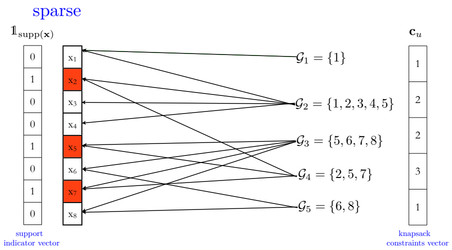
$$\boxed{\boldsymbol{B}\boldsymbol{\omega} \geq \boldsymbol{s}}$$

where $\boldsymbol{B}$ is the biadjacency matrix of $\mathfrak{G}$, i.e., $\boldsymbol{B}_{ij} = 1$ iff i-th coefficient is in $\mathcal{G}_j$.

When $\boldsymbol{B}$ is an interval matrix, or $\mathfrak{G}$ has a loopless group intersection graph [Baldassarre et al., 2013], it is TU.

Biconjugate: $g_{TU}^{**}(\mathbf{x}) = \min_{\mathbf{v}_i \in \mathbb{R}^p}\{\sum_{i=1}^M d_i\|\mathbf{v}_i\|_\infty : \mathbf{x} = \sum_{i=1}^M \mathbf{v}_i, \forall \mathrm{supp}(\mathbf{v}_i) \subseteq \mathcal{G}_i\}$
This is known as the Latent group lasso [Obozinski et al., 2011].

Group knapsack sparsity



We seek the sparsest signal with group allocation constraints.
[Hegde et al., 2009, Gerstner and Kistler, 2002]

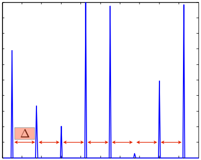
Objective: $\mathbb{1}^T \boldsymbol{s}$
Linear constraints: A valid support obeys budget constraints over $\mathfrak{G}$

$$\boxed{\boldsymbol{B}^T \boldsymbol{s} \leq \boldsymbol{c}_u}$$

where $\boldsymbol{B}$ is the biadjacency matrix of $\mathfrak{G}$, i.e., $\boldsymbol{B}_{ij} = 1$ iff i-th coefficient is in $\mathcal{G}_j$.

When $\boldsymbol{B}$ is an interval matrix or $\mathfrak{G}$ has a loopless group intersection graph [Baldassarre et al., 2013], it is TU.

Group knapsack sparsity



$$\boldsymbol{B}^T = \begin{bmatrix} 1 & 1 & \cdots & 1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 1 & \cdots & 1 & 1 & 0 & \cdots & 0 \\ & & & & \ddots & & & & \\ 0 & \cdots & 0 & 0 & 1 & 1 & \cdots & 1 & 1 \end{bmatrix}_{(p-\Delta+1)\times p}$$

We seek the sparsest signal with group allocation constraints.
[Hegde et al., 2009, Gerstner and Kistler, 2002]

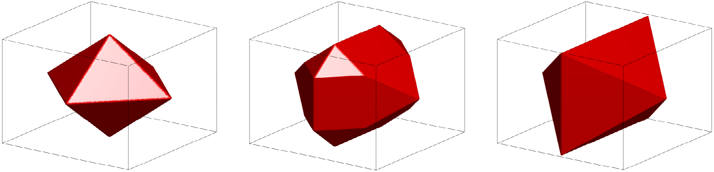
Objective: $\mathbb{1}^T\boldsymbol{s}$
Linear constraints: A valid support obeys budget constraints over $\mathfrak{G}$

$$\boxed{\boldsymbol{B}^T\boldsymbol{s} \leq \boldsymbol{c}_u}$$

where $\boldsymbol{B}$ is the biadjacency matrix of $\mathfrak{G}$, i.e., $\boldsymbol{B}_{ij} = 1$ iff i-th coefficient is in $\mathcal{G}_j$.

When $\boldsymbol{B}$ is an interval matrix or $\mathfrak{G}$ has a loopless group intersection graph [Baldassarre et al., 2013], it is TU.

Group knapsack sparsity



(left) $g_{TU}^{**}(\mathbf{x}) \leq 1$ (middle) $g_{TU}^{**}(\mathbf{x}) \leq 1.5$ (right) $g_{TU}^{**}(\mathbf{x}) \leq 2$ for $\mathfrak{G} = \{\{1,2\},\{2,3\}\}$

We seek the sparsest signal with group allocation constraints.
[Hegde et al., 2009, Gerstner and Kistler, 2002]

Objective: $\mathbb{1}^T \boldsymbol{s}$
Linear constraints: A valid support obeys budget constraints over $\mathfrak{G}$

$$\boxed{\boldsymbol{B}^T \boldsymbol{s} \leq \boldsymbol{c}_u}$$

where $\boldsymbol{B}$ is the biadjacency matrix of $\mathfrak{G}$, i.e., $\boldsymbol{B}_{ij} = 1$ iff i-th coefficient is in $\mathcal{G}_j$.

When $\boldsymbol{B}$ is an interval matrix or $\mathfrak{G}$ has a loopless group intersection graph [Baldassarre et al., 2013], it is TU.

Biconjugate: $g_{TU}^{**}(\mathbf{x}) = \begin{cases} \|\mathbf{x}\|_1 & \text{if } \boldsymbol{B}^T|\mathbf{x}| \leq \boldsymbol{c}_u, \mathbf{x} \in [-1,1]^p \\ \infty & \text{otherwise} \end{cases}$

Group knapsack sparsity example: A stylized spike train

- Basis pursuit (BP): $\|\mathbf{x}\|_1$
- TU-relax (TU):

$$g_{TU}^{**}(\mathbf{x}) = \begin{cases} \|\mathbf{x}\|_1 & \text{if } \mathbf{x} \in [-1,1]^p, \boldsymbol{B}^T|\mathbf{x}| \leq \boldsymbol{c}_u, \\ \infty & \text{otherwise} \end{cases}$$

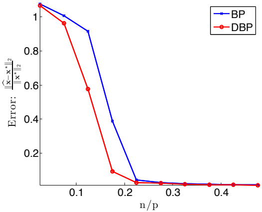
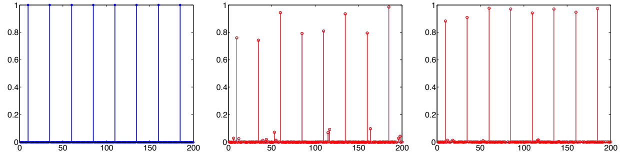


Figure: *Recovery for $n = 0.18p$.*



	$\mathbf{x}^\natural$	$\mathbf{x}_{\text{BP}}$ solution	$\mathbf{x}_{\text{TU}}$ solution
relative errors:		$\frac{\|\mathbf{x}^\natural - \mathbf{x}_{\text{BP}}\|_2}{\|\mathbf{x}^\natural\|_2} = .200$	$\frac{\|\mathbf{x}^\natural - \mathbf{x}_{\text{TU}}\|_2}{\|\mathbf{x}^\natural\|_2} = .067$

Conclusions

TU modeling framework offers:

- ✓ General easy to design structure sparsity template,
- ✓ Tight convexifications for submodular/non-submodular examples,
- ✓ Efficient evaluation of biconjugate via LP.

References I

- Bach, F. (2010).
  Structured sparsity-inducing norms through submodular functions.
  Adv. Neur. Inf. Proc. Sys. (NIPS), pages 118–126.
- Baldassarre, L., Bhan, N., Cevher, V., and Kyrillidis, A. (2013).
  Group-sparse model selection: Hardness and relaxations.
  arXiv preprint arXiv:1303.3207.
- Baraniuk, R., Cevher, V., Duarte, M., and Hegde, C. (2010).
  Model-based compressive sensing.
  IEEE Trans. Inf. Theory, 56(4):1982–2001.
- El Halabi, M. and Cevher, V. (2014).
  A totally unimodular view of structured sparsity.
  preprint.
  arXiv:1411.1990v1 [cs.LG].
- Gerstner, W. and Kistler, W. (2002).
  Spiking neuron models: Single neurons, populations, plasticity.
  Cambridge university press.
- Hegde, C., Duarte, M., and Cevher, V. (2009).
  Compressive sensing recovery of spike trains using a structured sparsity model.
  In *Sig. Proc. with Adapative Sparse Struct. Rep. (SPARS)*.

References II

- Huang, J., Zhang, T., and Metaxas, D. (2011).
  Learning with structured sparsity.
  J. Mach. Learn. Res., 12:3371–3412.
- Jenatton, R., Gramfort, A., Michel, V., Obozinski, G., Bach, F., and Thirion, B. (2011a).
  Multi-scale mining of fmri data with hierarchical structured sparsity.
  In *Pattern Recognition in NeuroImaging (PRNI)*.
- Jenatton, R., Mairal, J., Obozinski, G., and Bach, F. (2011b).
  Proximal methods for hierarchical sparse coding.
  J. Mach. Learn. Res., 12:2297–2334.
- Nemhauser, G. L. and Wolsey, L. A. (1999).
  Integer and combinatorial optimization, volume 18.
  Wiley New York.
- Obozinski, G. and Bach, F. (2012).
  Convex relaxation for combinatorial penalties.
  arXiv preprint arXiv:1205.1240.
- Obozinski, G., Jacob, L., and Vert, J. (2011).
  Group lasso with overlaps: The latent group lasso approach.
  arXiv preprint arXiv:1110.0413.

References III

- Oymak, S., Thrampoulidis, C., and Hassibi, B. (2013).
  Simple bounds for noisy linear inverse problems with exact side information.
  arXiv:1312.0641v2 [cs.IT].

- Zhao, P., Rocha, G., and Yu, B. (2006).
  Grouped and hierarchical model selection through composite absolute penalties.
  Department of Statistics, UC Berkeley, Tech. Rep, 703.

- Zhao, P. and Yu, B. (2006).
  On model selection consistency of Lasso.
  J. Mach. Learn. Res., 7:2541–2563.

- Zhou, H., Sehl, M., Sinsheimer, J., and Lange, K. (2010).
  Association screening of common and rare genetic variants by penalized regression.
  Bioinformatics, 26(19):2375.