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

EPFL-Idiap-ETH Sparsity Workshop 2015

Joint work with: Volkan Cevher

lions@epfl

MARIE CURIE ACTIONS

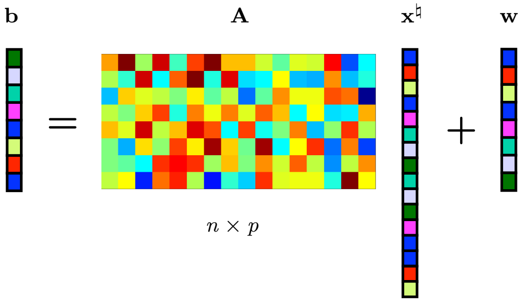
FNSNF
FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

erc

EPFL

Supervised learning and inverse problems

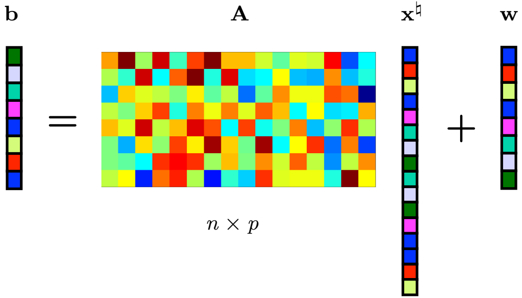
Running example:



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

Supervised learning and inverse problems

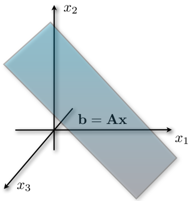
Running example:



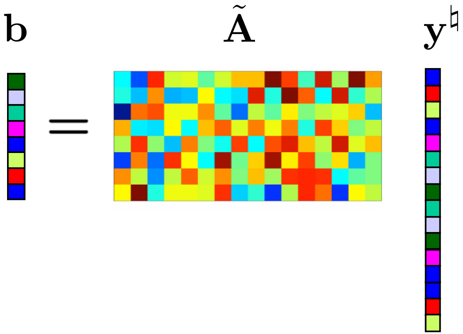
***A difficult estimation challenge when** $n < p$:*

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

- **Needle in a haystack:** ***We need additional information on $\mathbf{x}^\natural$!***

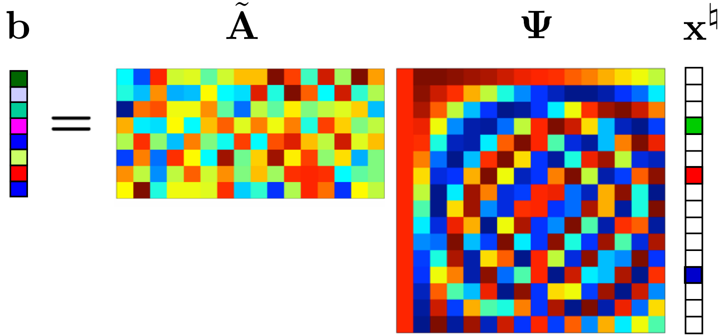


Sparsity to the rescue!



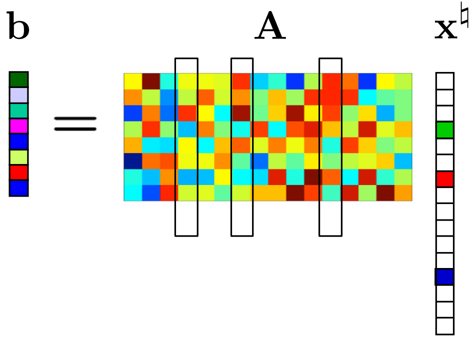
- $\mathbf{b} \in \mathbb{R}^n$, $\tilde{\mathbf{A}} \in \mathbb{R}^{n \times p}$, and $n < p$

Sparsity to the rescue!



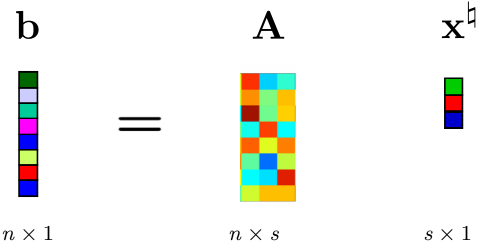
- $\mathbf{b} \in \mathbb{R}^n$, $\tilde{\mathbf{A}} \in \mathbb{R}^{n \times p}$, and $n < p$
- $\boldsymbol{\Psi} \in \mathbb{R}^{p \times p}$, $\mathbf{x}^\natural \in \Sigma_s$, and $s < n < p$

Sparsity to the rescue!



- $\mathbf{b} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{n \times p}$, $\mathbf{x}^\natural \in \Sigma_s$, and $s < n < p$

Sparsity to the rescue!

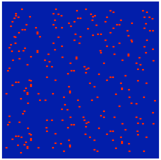


- $\mathbf{b} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{n \times p}$, $\mathbf{x}^\natural \in \Sigma_s$, and $s < n < p$

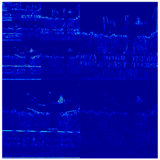
Impact: Support restricted columns of $\mathbf{A}$ leads to an *overcomplete* system.

Beyond sparsity towards model-based or *structured* sparsity

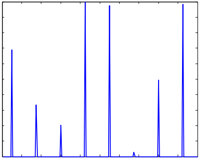
- The following signals can look the **same** from a **sparsity** perspective!



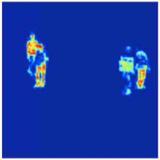
Sparse image



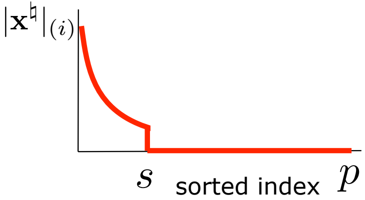
Wavelet coefficients of a natural image



Spike train

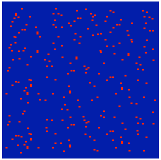


Background substracted image

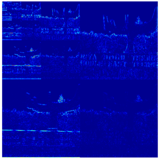


Beyond sparsity towards model-based or *structured* sparsity

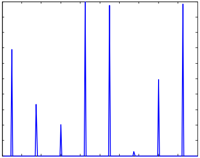
- The following signals can look the **same** from a **sparsity** perspective!



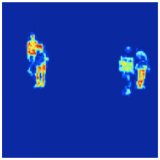
Sparse image



Wavelet coefficients of a natural image

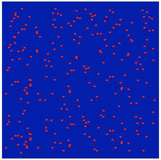


Spike train

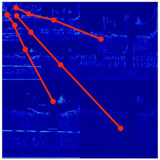


Background substracted image

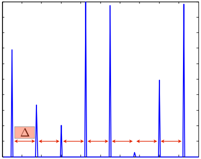
- In reality, these signals have additional structures beyond the simple sparsity



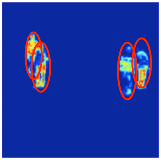
Sparse image



Wavelet coefficients of a natural image



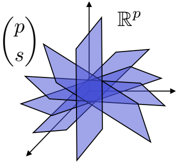
Spike train

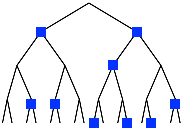


Background substracted image

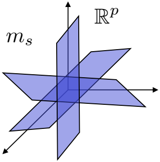
Beyond sparsity towards model-based or *structured* sparsity

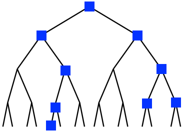
Sparsity model: Union of all s-dimensional canonical subspaces.





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





Three upshots of structured sparsity:

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

A simple template for linear inverse problems

> Find the "***sparsest***" $\mathbf{x}$ subject to ***structure*** and ***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

A combinatorial approach for estimating $\mathbf{x}^\natural$ from $\mathbf{b} = \mathbf{A}\mathbf{x}^\natural + \mathbf{w}$

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

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

Simple sparsity

A combinatorial approach for estimating $\mathbf{x}^\natural$ from $\mathbf{b} = \mathbf{A}\mathbf{x}^\natural + \mathbf{w}$

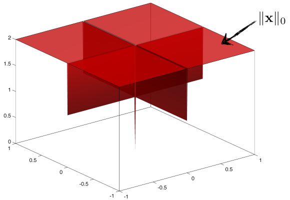
$$\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^p} \left\{ \|\mathbf{x}\|_0 : \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 \leq \|\mathbf{w}\|_2 \right\} \qquad (\mathcal{P}_0)$$

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

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



Simple sparsity

A combinatorial approach for estimating $\mathbf{x}^\natural$ from $\mathbf{b} = \mathbf{A}\mathbf{x}^\natural + \mathbf{w}$

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

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

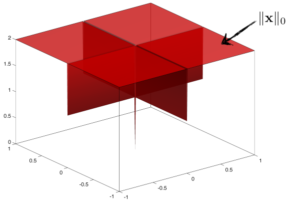
$\mathcal{P}_0$ has the following characteristics:

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

Convex relaxation:

Convex envelope is the largest convex lower bound.

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



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

Simple sparsity

A combinatorial approach for estimating $\mathbf{x}^\natural$ from $\mathbf{b} = \mathbf{A}\mathbf{x}^\natural + \mathbf{w}$

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

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

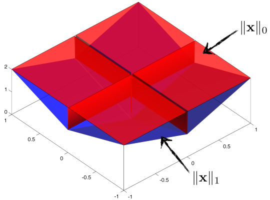
$\mathcal{P}_0$ has the following characteristics:

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

Convex relaxation:

Convex envelope is the largest convex lower bound.

$\|\mathbf{x}\|_1$ 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

A combinatorial approach for estimating $\mathbf{x}^\natural$ from $\mathbf{b} = \mathbf{A}\mathbf{x}^\natural + \mathbf{w}$

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

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

The role of convexity: Tractable & stable recovery

A combinatorial approach for estimating $\mathbf{x}^\natural$ from $\mathbf{b} = \mathbf{A}\mathbf{x}^\natural + \mathbf{w}$

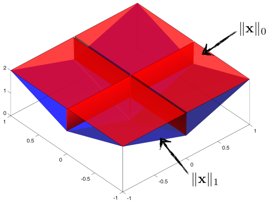
$$\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathbb{R}^p} \left\{ \|\mathbf{x}\|_1 : \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 \leq \|\mathbf{w}\|_2 , \|\mathbf{x}\|_\infty \leq 1 \right\} \qquad (BP)$$

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 [13]:

- 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 relaxations in general ?

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

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

How to compute the convex relaxation of g in general ?

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

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

Proposition (Hardness of conjugation)

The Fenchel conjugate of g results in the following combinatorial problem

$$g^*(\mathbf{y}) = \sup_{\boldsymbol{s}\in\{0,1\}^p} |\mathbf{y}|^T\boldsymbol{s} - F(\boldsymbol{s}).$$

which is NP-Hard in general.

Tractable convex relaxation

Prior work:

1. Monotone submodular penalties [1]
   - Tractable biconjugation via Lovász extension
   - Limited to certain structures
2. ℓ_q-regularized combinatorial functions [11] $(\mu F(\text{supp}(\mathbf{x})) + \nu\|\mathbf{x}\|_q)$
   - Tractable biconjugation even for some non-submodular functions
   - Not always tractable
   - May loose structure

Our work: New framework for tractable convex relaxations

- Easy to design
- Tractable biconjugation via linear programming (LP)
- Applicable to various submodular and non-submodular structures

Template for TU structures

Sparsity and structure together [5]

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

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

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

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.

Relaxation of ILP to LP [10]

When $\boldsymbol{M}$ is TU and $\boldsymbol{c}$ is integeral, then the LP

$$\max_{\boldsymbol{\beta}\in\mathbb{R}^m} \{\boldsymbol{\theta}^T\boldsymbol{\beta} : \boldsymbol{M}\boldsymbol{\beta} \leq \boldsymbol{c}, \boldsymbol{\beta} \geq 0\}$$

has integer optimal solutions (i.e., ILP $\equiv$ LP).

- "Exact convex relaxation" of: $g^*(\mathbf{y}) = \sup_{\boldsymbol{s}\in\{0,1\}^p} |\mathbf{y}|^T\boldsymbol{s} - F(\boldsymbol{s})$.
- Same idea behind the tractable biconjugation of submodular functions

Convexification of TU structures

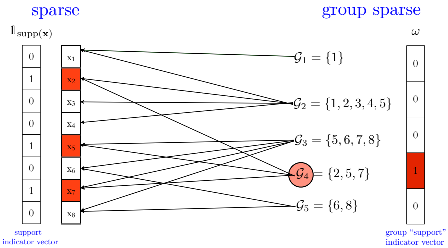
TU convex relaxation given by LP

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

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

- Special cases:
  - Rederive the convex envelope of several submodular models
  - Establish the tightness of some convex regularizers for non-submodular models
- Beyond linear objectives, some quadratic objectives can also be handled

Group cover sparsity: Minimal group cover [2, 12, 8]



Structure: *We seek the signal covered by a minimal number of groups.*

Objective: $\boldsymbol{d}^T\boldsymbol{\omega}$

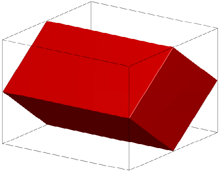
Linear description: For each non-zero coefficient, at least one group containing it is selected

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

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 it is TU.

Group cover sparsity: Minimal group cover [2, 12, 8]



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

Structure: *We seek the signal covered by a minimal number of groups.*

$$\text{Objective: } \boldsymbol{d}^T \boldsymbol{\omega}$$

Linear description: For each non-zero coefficient, at least one group containing it is selected

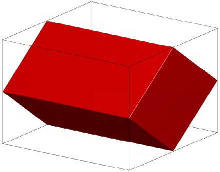
$$\boxed{\boldsymbol{B}\boldsymbol{\omega} \geq \boldsymbol{s}}$$

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 it is TU.

Biconjugate: $g_{TU}^{**}(\mathbf{x}) = \min_{\boldsymbol{\omega} \in [0,1]^M} \{\boldsymbol{d}^T \boldsymbol{\omega} : \boldsymbol{B}\boldsymbol{\omega} \geq |\mathbf{x}|\}$ for $\mathbf{x} \in [-1, 1]^p, \infty$ otherwise

Group cover sparsity: Minimal group cover [2, 12, 8]



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

Structure: *We seek the signal covered by a minimal number of groups.*

Objective: $\boldsymbol{d}^T\boldsymbol{\omega}$

Linear description: For each non-zero coefficient, at least one group containing it is selected

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

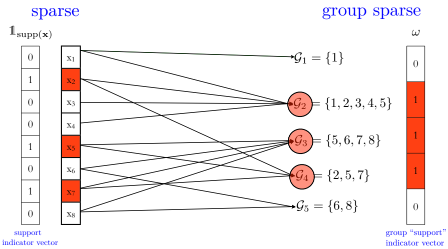
where $\boldsymbol{B}$ is the biadjacency matrix of $\mathfrak{G}$, i.e., $\boldsymbol{B}_{ij} = 1$ iff i-th coefficient is in $\mathcal{G}_j$.

When $\boldsymbol{B}$ is an interval matrix, or $\mathfrak{G}$ has a *loopless* group intersection graph it is TU.

Biconjugate: $g^{**}_{TU}(\mathbf{x}) = \min_{\boldsymbol{\omega}\in[0,1]^M}\{\boldsymbol{d}^T\boldsymbol{\omega} : \boldsymbol{B}\boldsymbol{\omega} \geq |\mathbf{x}|\}$ for $\mathbf{x} \in [-1,1]^p, \infty$ otherwise

$$\overset{\star}{=} \min_{\mathbf{v}_i\in\mathbb{R}^p}\{\textstyle\sum_{i=1}^M d_i\|\mathbf{v}_i\|_\infty : \mathbf{x} = \sum_{i=1}^M \mathbf{v}_i, \forall \mathrm{supp}(\mathbf{v}_i) \subseteq \mathcal{G}_i\},$$

Group intersection sparsity [9, 14, 1]



Structure: *We seek the signal intersecting with minimal number of groups.*

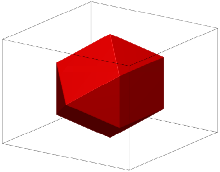
Objective: $\boldsymbol{d}^T\boldsymbol{\omega}$ (*submodular:* $F(\mathcal{S}) = \sum_{\mathcal{G}_i \in \mathfrak{G}, \mathcal{S}\cap\mathcal{G}_i \neq \emptyset} d_i$)

Linear description: All groups containing a non-zero coefficient are selected

$$\boxed{\boldsymbol{H}_k \boldsymbol{s} \leq \boldsymbol{\omega}, \forall k \in \{0, \cdots, p\}}$$

where $\boldsymbol{H}_k(i,j) = \begin{cases} 1 & \text{if } j = k, j \in \mathcal{G}_i \\ 0 & \text{otherwise} \end{cases}$, which is TU.

Group intersection sparsity [9, 14, 1]



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

Structure: *We seek the signal intersecting with minimal number of groups.*

Objective: $\boldsymbol{d}^T\boldsymbol{\omega}$ (*submodular:* $F(\mathcal{S}) = \sum_{\mathcal{G}_i \in \mathfrak{G}, \mathcal{S}\cap\mathcal{G}_i \neq \emptyset} d_i$)

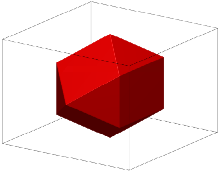
Linear description: All groups containing a non-zero coefficient are selected

$$\boldsymbol{H}_k \boldsymbol{s} \leq \boldsymbol{\omega}, \forall k \in \{0, \cdots, p\}$$

where $\boldsymbol{H}_k(i,j) = \begin{cases} 1 & \text{if } j = k, j \in \mathcal{G}_i \\ 0 & \text{otherwise} \end{cases}$, which is TU.

Biconjugate: $g_{TU}^{**}(\mathbf{x}) = \min_{\boldsymbol{\omega}\in[0,1]^M}\{\boldsymbol{d}^T\boldsymbol{\omega} : \boldsymbol{H}_k|\mathbf{x}| \leq \boldsymbol{\omega}, \forall k \in \mathfrak{P}\}$
for $\mathbf{x} \in [-1,1]^p$, ∞ otherwise.

Group intersection sparsity [9, 14, 1]



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

Structure: *We seek the signal intersecting with minimal number of groups.*

Objective: $\boldsymbol{d}^T\boldsymbol{\omega}$ (*submodular:* $F(\mathcal{S}) = \sum_{\mathcal{G}_i \in \mathfrak{G}, \mathcal{S} \cap \mathcal{G}_i \neq \emptyset} d_i$)

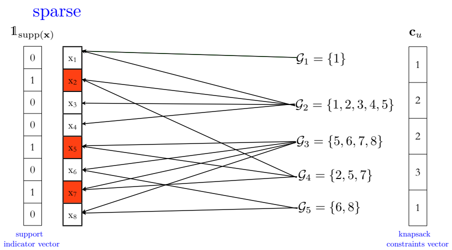
Linear description: All groups containing a non-zero coefficient are selected

$$\boldsymbol{H}_k \boldsymbol{s} \leq \boldsymbol{\omega}, \forall k \in \{0, \cdots, p\}$$

where $\boldsymbol{H}_k(i,j) = \begin{cases} 1 & \text{if } j = k, j \in \mathcal{G}_i \\ 0 & \text{otherwise} \end{cases}$, which is TU.

Biconjugate: $g_{TU}^{**}(\mathbf{x}) = \min_{\boldsymbol{\omega} \in [0,1]^M} \{\boldsymbol{d}^T\boldsymbol{\omega} : \boldsymbol{H}_k|\mathbf{x}| \leq \boldsymbol{\omega}, \forall k \in \mathfrak{P}\} \overset{\star}{=} \sum_{\mathcal{G} \in \mathfrak{G}} \|x_{\mathcal{G}}\|_\infty$ for $\mathbf{x} \in [-1,1]^p$, ∞ otherwise.

Group knapsack sparsity [15, 7, 6]



Structure: *We seek the sparsest signal with group allocation constraints.*

Objective: $\mathbb{1}^T \boldsymbol{s}$

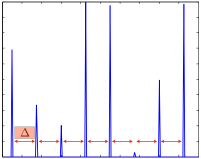
Linear description: A valid support obeys budget constraints over $\mathfrak{G}$

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

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, it is TU.

Group knapsack sparsity [15, 7, 6]



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

Structure: *We seek the sparsest signal with group allocation constraints.*

Objective: $\mathbb{1}^T \boldsymbol{s}$

Linear description: A valid support obeys budget constraints over $\mathfrak{G}$

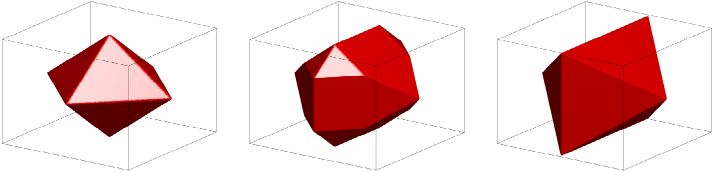
$$\boxed{\boldsymbol{B}^T \boldsymbol{s} \leq \boldsymbol{c}_u}$$

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, it is TU.

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

For the neuronal spike example, we have $\boldsymbol{c}_u = \mathbb{1}$.

Group knapsack sparsity [15, 7, 6]



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

Structure: *We seek the sparsest signal with group allocation constraints.*

Objective: $\mathbb{1}^T \boldsymbol{s}$

Linear description: 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, it is TU.

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

For the neuronal spike example, we have $\boldsymbol{c}_u = \mathbb{1}$.

Group knapsack sparsity example: A stylized spike train

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

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

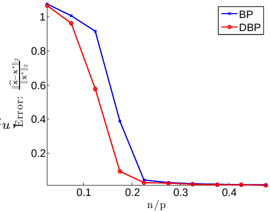
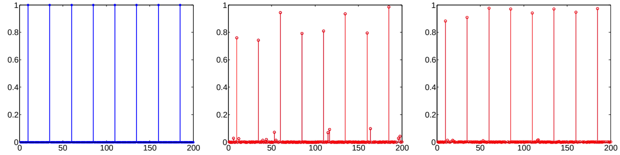


Figure: *Recovery for $n = 0.18p$.*



	$\mathbf{x}^\natural$	$\mathbf{x}_{\text{BP}}$ 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

Our work: TU modeling framework

- Complement previous approaches
- Convex programs (not necessarily norms)
- Tight convexifications, non-submodular examples
- Easy to design and "usually" efficient via an LP

References I

[1] Francis Bach.
Structured sparsity-inducing norms through submodular functions.
Adv. Neur. Inf. Proc. Sys. (NIPS), pages 118–126, 2010.

[2] L. Baldassarre, N. Bhan, V. Cevher, and A. Kyrillidis.
Group-sparse model selection: Hardness and relaxations.
arXiv preprint arXiv:1303.3207, 2013.

[3] Venkat Chandrasekaran, Benjamin Recht, Pablo A. Parrilo, and Alan S. Willsky.
The convex geometry of linear inverse problems.
Found. Comp. Math., 12:805–849, 2012.

[4] S. Chen, D. Donoho, and M. Saunders.
Atomic decomposition by basis pursuit.
SIAM J. Sci. Comp., 20(1):33–61, 1998.

[5] Marwa El Halabi and Volkan Cevher.
A totally unimodular view of structured sparsity.
preprint, 2014.
arXiv:1411.1990v1 [cs.LG].

[6] W Gerstner and W. Kistler.
Spiking neuron models: Single neurons, populations, plasticity.
Cambridge university press, 2002.

References II

[7] C. Hegde, M. Duarte, and V. Cevher.
Compressive sensing recovery of spike trains using a structured sparsity model.
In *Sig. Proc. with Adapative Sparse Struct. Rep. (SPARS)*, 2009.

[8] J. Huang, T. Zhang, and D. Metaxas.
Learning with structured sparsity.
J. Mach. Learn. Res., 12:3371–3412, 2011.

[9] R. Jenatton, A. Gramfort, V. Michel, G. Obozinski, F. Bach, and B. Thirion.
Multi-scale mining of fmri data with hierarchical structured sparsity.
In *Pattern Recognition in NeuroImaging (PRNI)*, 2011.

[10] George L Nemhauser and Laurence A Wolsey.
Integer and combinatorial optimization, volume 18.
Wiley New York, 1999.

[11] G. Obozinski and F. Bach.
Convex relaxation for combinatorial penalties.
arXiv preprint arXiv:1205.1240, 2012.

[12] G. Obozinski, L. Jacob, and J.P. Vert.
Group lasso with overlaps: The latent group lasso approach.
arXiv preprint arXiv:1110.0413, 2011.

References III

[13] Samet Oymak, Christos Thrampoulidis, and Babak Hassibi.
Simple bounds for noisy linear inverse problems with exact side information.
2013.
arXiv:1312.0641v2 [cs.IT].

[14] Peng Zhao and Bin Yu.
On model selection consistency of Lasso.
J. Mach. Learn. Res., 7:2541–2563, 2006.

[15] H. Zhou, M.E. Sehl, J.S. Sinsheimer, and K. Lange.
Association screening of common and rare genetic variants by penalized regression.

Bioinformatics, 26(19):2375, 2010.