Supplementary Material:
A totally unimodular view of structured sparsity

Marwa El Halabi Volkan Cevher

LIONS, EPFL

1 Numerical illustration of Sparse G-group cover's performance

In this section, we compare the performance of minimizing the TU relaxation $g^{**}_{\mathfrak{G},G}$ of the proposed Sparse G-group cover (c.f., Section 5.4) in problem (2), which we will call Sparse latent group lasso (SLGL), with Basis pursuit (BP) and Sparse group Lasso (SGL). Recall the SGL criteria is $(1-\alpha)\sum_{\mathcal{G}\in\mathfrak{G}}\sqrt{|\mathcal{G}|}\|\boldsymbol{x}_{\mathcal{G}}\|_q + \alpha\|\boldsymbol{x}_{\mathcal{G}}\|_1$, with $q=2$ in [3]. We compare also against SGL_∞ where we set $q=\infty$, which is better suited for signals with equal valued non-zero coefficients. We generate a sparse signal $\boldsymbol{x}^\natural$ in dimensions $p=200$, covered by $G=5$ groups, randomly chosen from the $M=29$ groups. The groups generated are interval groups, of equal size of 10 coefficients, and with an overlap of 3 coefficients between each two consecutive groups. The true signal $\boldsymbol{x}^\natural$ has 3 non-zero coefficients (all set to one) in each of its 5 active groups (cf., Figure 2). Note that these groups lead a TU group structure $\mathfrak{G}$, so the TU relaxation in this case is tight. We recover $\boldsymbol{x}^\natural$ from its compressive measurements $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}^\natural + \boldsymbol{w}$, where the noise $\boldsymbol{w}$ is a random Gaussian vector of variance $\sigma = 0.01$ and $\boldsymbol{A}$ is a random column normalized Gaussian matrix. We encode the data via $\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|_2 \le \|\boldsymbol{w}\|_2$ using the true ℓ_2-norm of the noise. We produce the data randomly 10 times and report the averaged results.

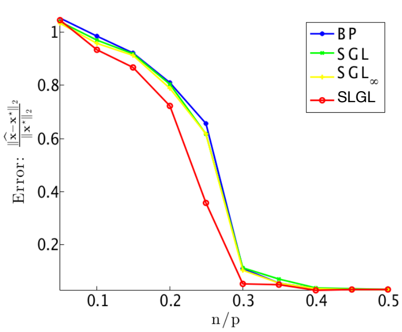

Figure 1: Recovery error of SLGL, SGL, and BP

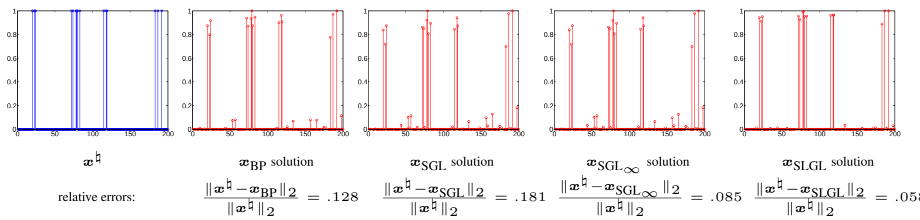

Figure 2: *Recovery for* $n = 0.25p, s = 15, p = 200, G = 5$ *out of* $M = 29$ *groups.*

Figure 1 measures the relative recovery error with $\frac{\|\boldsymbol{x}^\natural - \hat{\boldsymbol{x}}\|_2}{\|\boldsymbol{x}^\natural\|_2}$, as we vary the number of compressive measurements. Since the SLGL criteria uses the fact that $\boldsymbol{x}^\natural$ lies in the unit ℓ_∞-ball, we include this constraint in the all the other formulations for fairness. Since the true signal exhibit strong overall sparsity we use $\alpha = 0.95$ in SGL as suggested in [3] (we tried several values of α, and this seemed to give the best results for SGL). We use an interior point method to obtain high accuracy solutions to each formulation. Figure 8 shows that SLGL outperforms the other criterias as we vary the number of measurements.

2 Proof of Proposition 2

Proposition (Convexification)**.** *The convex envelope of* $g_{\mathfrak{G},\cap}(\boldsymbol{x})$ *over the unit* ℓ_∞*-ball is*

$$g^{**}_{\mathfrak{G},\cap}(\boldsymbol{x}) = \begin{cases} \sum_{\mathcal{G}_i \in \mathfrak{G}} d_i \|\boldsymbol{x}_{\mathcal{G}_i}\|_\infty & \textit{if } \boldsymbol{x} \in [-1,1]^p \\ \infty & \textit{otherwise} \end{cases}$$

Proof. Since $g_{\mathfrak{G},\cap}(\boldsymbol{x})$ is a TU-penalty, we can use Proposition 1 in the main text, to compute its convex envelope:

$$\begin{aligned} g^{**}_{\mathfrak{G},\cap}(\boldsymbol{x}) &= \min_{\boldsymbol{s}\in[0,1]^p, \boldsymbol{\omega}\in[0,1]^M} \{\boldsymbol{d}^T\boldsymbol{\omega} : \boldsymbol{H\beta} \leq 0, |\boldsymbol{x}| \leq \boldsymbol{s}\} \\ &= \min_{\boldsymbol{\omega}\in[0,1]^M} \{\boldsymbol{d}^T\boldsymbol{\omega} : \boldsymbol{H} \begin{bmatrix} \boldsymbol{\omega} \\ |\boldsymbol{x}| \end{bmatrix} \leq 0\} \\ &= \sum_{\mathcal{G}_i \in \mathfrak{G}} d_i \|\boldsymbol{x}_{\mathcal{G}_i}\|_\infty \qquad\qquad (\text{since } w_i^* = \|\boldsymbol{x}_{\mathcal{G}}\|_\infty) \end{aligned}$$

for $\boldsymbol{x} \in [-1,1]^p$, $g^{**}_{\mathfrak{G},\cap}(\boldsymbol{x}) = \infty$ otherwise. □

3 Proof of Proposition 3

Proposition (Convexification). *When the group structure leads to a TU biadjacency matrix $\boldsymbol{B}$, the convex envelope of the group ℓ_0-"norm" over the unit ℓ_∞-ball is*

$$g^{**}_{\mathfrak{G},0}(\boldsymbol{x}) = \begin{cases} \min_{\boldsymbol{\omega}\in[0,1]^M} \{\boldsymbol{d}^T\boldsymbol{\omega} : \boldsymbol{B}\boldsymbol{\omega} \geq |\boldsymbol{x}|\} & \text{if } \boldsymbol{x} \in [-1,1]^p \\ \infty & \text{otherwise} \end{cases}$$

Proof. Note that $g_{\mathfrak{G},0}(\boldsymbol{x})$ can be written in the form given in Definition 4 with $\boldsymbol{M} = [-\boldsymbol{B}, \boldsymbol{I}_p]$ and $\boldsymbol{c} = 0$. Thus, when $\boldsymbol{B}$ is TU, so is $\boldsymbol{M}$ [2, Proposition 2.1], and thus we can use Proposition 1 in the main text, to compute its convex envelope:

$$\begin{aligned} g^{**}_{\mathfrak{G},0}(\boldsymbol{x}) &= \min_{\substack{\boldsymbol{s}\in[0,1]^p \\ \boldsymbol{\omega}\in[0,1]^M}} \{\boldsymbol{d}^T\boldsymbol{\omega} : \boldsymbol{B}\boldsymbol{\omega} \geq \boldsymbol{s}, |\boldsymbol{x}| \leq \boldsymbol{s}\} \\ &= \min_{\boldsymbol{\omega}\in[0,1]^M} \{\boldsymbol{d}^T\boldsymbol{\omega} : \boldsymbol{B}\boldsymbol{\omega} \geq |\boldsymbol{x}|\} \end{aligned}$$

for $\boldsymbol{x} \in [-1,1]^p$, $g^{**}_{\mathfrak{G},0}(\boldsymbol{x}) = \infty$ otherwise. $\square$

4 Proof of Proposition 4

Proposition. *Given any group structure $\mathfrak{G}$, $g_{\mathfrak{G},s}(\boldsymbol{x})$ is not a TU penalty.*

Proof. Let $G(\mathfrak{G}\cup\mathcal{P}, \mathcal{E})$ denote the bipartite graph representation of the group structure $\mathfrak{G}$. We use the linearization trick employed in [1] to reduce $g_{\mathfrak{G},s}(\boldsymbol{x})$ to an integer program. For conciseness, we consider $g_{\mathfrak{G},s}(\boldsymbol{s})$ only for binary vectors $\boldsymbol{s} \in \{0,1\}^p$, since $g_{\mathfrak{G},s}(\boldsymbol{x}) = g_{\mathfrak{G},s}(\mathbb{1}_{\mathrm{supp}(\boldsymbol{x})})$.

$$\begin{aligned} g_{\mathfrak{G},s}(\boldsymbol{s}) &= \min_{\boldsymbol{\omega}\in\{0,1\}^M} \{\sum_{i=1}^M \omega_i \|\boldsymbol{s}_{\mathcal{G}_i}\|_0 : \boldsymbol{M}\boldsymbol{\beta} \leq 0\} \\ &= \min_{\boldsymbol{\omega}\in\{0,1\}^M} \{\sum_{(i,j)\in\mathcal{E}} \omega_i s_j : \boldsymbol{M}\boldsymbol{\beta} \leq 0\} \\ &= \min_{\substack{\boldsymbol{\omega}\in\{0,1\}^M \\ \boldsymbol{z}\in\{0,1\}^{|\mathcal{E}|}}} \{\sum_{(i,j)\in\mathcal{E}} z_{ij} : \boldsymbol{M}\boldsymbol{\beta} \leq 0, \boldsymbol{E}\boldsymbol{\beta} \leq \boldsymbol{z} + \mathbb{1}\} \end{aligned}$$

Recall that $\boldsymbol{E}$ is the edge-node incidence matrix of $G(\mathfrak{G}\cup\mathcal{P}, \mathcal{E})$. The constraint $\boldsymbol{E}\boldsymbol{\beta} \leq \boldsymbol{z} - \mathbb{1}$ corresponds to $z_{ij} \geq \omega_i + s_j - 1, \forall (i,j) \in \mathcal{E}$. Although both matrices $\boldsymbol{M}$ and $\boldsymbol{E}$ are TU, their concatenation $\widetilde{\boldsymbol{M}} = \begin{bmatrix} \boldsymbol{M} \\ \boldsymbol{E} \end{bmatrix}$ is not TU. To see this, let us first focus on the case where $\boldsymbol{M} = [-\boldsymbol{B}, \boldsymbol{I}_p]$.

Given any coefficient $i \in \mathcal{P}$ covered by at least one group $\mathcal{G}_i$, we denote the corresponding edge in the bipartite graph by $e_j = (i, M+i)$, which corresponds to the j^{th} row of $\boldsymbol{E}$. This translates into having the entries $\widetilde{\boldsymbol{M}}_{i,i} = -1, \widetilde{\boldsymbol{M}}_{i,M+i} = 1, \widetilde{\boldsymbol{M}}_{p+j,i} = 1$, and $\widetilde{\boldsymbol{M}}_{p+j,M+i} = 1$. The determinant of the submatrix resulting from these entries is -2, which contradicts the definition of TU (cf., Def. 4). It follows then that $\widetilde{\boldsymbol{M}}$ is TU iff $\mathfrak{G} = \{\emptyset\}$.

A similar argument holds for $\boldsymbol{M} = \boldsymbol{H}$. $\square$

5 Proof of Proposition 5

Proposition (Convexification)**.** *The convex surrogate via Proposition 1 in the main text, for $g_{\mathfrak{G},s}(\boldsymbol{x})$ with $\boldsymbol{M} = \boldsymbol{H}$ (i.e., the group intersection model with sparse groups) is given by*

$$\Omega_{\mathfrak{G},s}(\boldsymbol{x}) := \sum_{(i,j)\in\mathcal{E}} (\|\boldsymbol{x}_{\mathcal{G}_i}\|_\infty + |x_j| - 1)_+$$

for $\boldsymbol{x} \in [-1,1]^p$, and, $\Omega_{\mathfrak{G},s}(\boldsymbol{x}) := \infty$ otherwise. Note that $\Omega_{\mathfrak{G},s}(\boldsymbol{x}) \leq g^{**}_{\mathfrak{G},s}(\boldsymbol{x})$.

Proof. For $\boldsymbol{x} \in [-1,1]^p$,

$$\begin{aligned}\Omega_{\mathfrak{G},s}(\boldsymbol{x}) &= \min_{\substack{\boldsymbol{\omega}\in[0,1]^M \\ \boldsymbol{z}\in[0,1]^{|\mathcal{E}|}}} \{ \sum_{(i,j)\in\mathcal{E}} z_{ij} : \boldsymbol{H\beta} \leq 0, \boldsymbol{E\beta} \leq \boldsymbol{z} + \mathbb{1}, |\boldsymbol{x}| \leq \boldsymbol{s}\} \\ &= \sum_{(i,j)\in\mathcal{E}} (\|\boldsymbol{x}_{\mathcal{G}_i}\|_\infty + |x_j| - 1)_+\end{aligned}$$

since $\omega_i^* = \|\boldsymbol{x}_{\mathcal{G}_i}\|_\infty, \boldsymbol{s}^* = |\boldsymbol{x}|$, and $z_{ij}^* = (\omega_i^* + s_j^* - 1)_+$. □

6 Proof of Proposition 6

Proposition (Convexification)**.** *The convex surrogate given by Proposition 1 in the main text, for $g_{\mathfrak{G},s}(\boldsymbol{x})$ with $\boldsymbol{M} = [-\boldsymbol{B}, \boldsymbol{I}_p]$ (i.e., the group ℓ_0-"norm" with sparse groups) is given by*

$$\Omega_{\mathfrak{G},s}(\boldsymbol{x}) := \min_{\boldsymbol{\omega}\in[0,1]^M} \{ \sum_{(i,j)\in\mathcal{E}} (\omega_i + |x_j| - 1)_+ : \boldsymbol{B\omega} \geq |\boldsymbol{x}|\}$$

for $\boldsymbol{x} \in [-1,1]^p$, $\Omega_{\mathfrak{G},s}(\boldsymbol{x}) = \infty$ otherwise.

Proof. For $\boldsymbol{x} \in [-1,1]^p$,

$$\begin{aligned}\Omega_{\mathfrak{G},s}(\boldsymbol{x}) &= \min_{\substack{\boldsymbol{\omega}\in[0,1]^M \\ \boldsymbol{z}\in[0,1]^{|\mathcal{E}|}}} \{ \sum_{(i,j)\in\mathcal{E}} z_{ij} : \boldsymbol{B\omega} \geq \boldsymbol{s}, \boldsymbol{E\beta} \leq \boldsymbol{z} + \mathbb{1}, |\boldsymbol{x}| \leq \boldsymbol{s}\} \\ &= \min_{\boldsymbol{\omega}\in[0,1]^M} \{ \sum_{(i,j)\in\mathcal{E}} (\omega_i + |x_j| - 1)_+ : \boldsymbol{B\omega} \geq |\boldsymbol{x}|\}\end{aligned}$$

since $\boldsymbol{s}^* = |\boldsymbol{x}|$, and $z_{ij}^* = (\omega_i + s_j^* - 1)_+$. □

7 Proof of Proposition 8

Proposition. *(Convexification) The convexification of the tree ℓ_0-"norm" over the unit ℓ_∞-ball is given by*

$$g^{**}_{T,0}(\boldsymbol{x}) = \begin{cases} \sum_{\mathcal{G}\in\mathfrak{G}_H} \|x_{\mathcal{G}}\|_\infty & \text{if } \boldsymbol{x} \in [-1,1]^p \\ \infty & \text{otherwise} \end{cases}$$

Proof. Since this is a TU-penalty we can use Proposition 1 in the main text, to compute its convex envelope:

$$\begin{aligned} g_{T,0}^{**}(\boldsymbol{x}) &= \min_{\boldsymbol{s}\in[0,1]^p} \{\mathbb{1}^T\boldsymbol{s} : \boldsymbol{T}\boldsymbol{s} \geq 0, |\boldsymbol{x}| \leq \boldsymbol{s}\} \\ &\stackrel{\star}{=} \sum_{\mathcal{G}\in\mathfrak{G}_H} \|x_{\mathcal{G}}\|_\infty \end{aligned}$$

for $\boldsymbol{x} \in [-1,1]^p$, ∞ otherwise, and where the groups $\mathcal{G} \in \mathfrak{G}_H$ are defined as each node and all its descendants. ($\star$) holds since any feasible $\boldsymbol{s}$ should satisfy $\boldsymbol{s} \geq |\boldsymbol{x}|$ and $s_{\text{parent}} \geq s_{\text{child}}$, so starting from the leaves, each leaf satisfies $s_i \geq |x_i|$, and since we are looking to minimize the sum of s_i's, we simply set $s_i = x_i$. For a node i with two children j, k as leaves, it will satisfy $s_i \geq |x_i|, |s_j|, |s_k|$, thus $s_i = \max\{|x_i|, |x_j|, |x_k|\}$, and so on. Thus, $s_i = \max_{\{k \text{ is a descendant of } i \text{ or } i \text{ itself}\}} |x_k|$ □

8 Proof of Proposition 9

Proposition (Convexification)**.** *The convex envelope of $g_{\boldsymbol{D}}(\boldsymbol{x})$ over the unit ℓ_∞-ball when $\boldsymbol{B}^T$ is a TU matrix is given by*

$$g_{\boldsymbol{D}}^{**}(\boldsymbol{x}) = \begin{cases} \max_{\mathcal{G}\in\mathfrak{G}} \|\boldsymbol{x}_{\mathcal{G}}\|_1 & \text{if } \boldsymbol{x} \in [-1,1]^p, \boldsymbol{B}^T|\boldsymbol{x}| \leq \mathbb{1} \\ \infty & \text{otherwise} \end{cases}$$

Proof. Since this is a TU penalty we can use Proposition 1 in the main text, to compute its convex envelope:

$$\begin{aligned} g_{\boldsymbol{D}}^{**}(\boldsymbol{x}) &= \begin{cases} \min_{\substack{\omega\in[0,1] \\ \boldsymbol{s}\in[0,1]^p}} \{\omega : \boldsymbol{B}^T\boldsymbol{s} \leq \omega\mathbb{1}, |\boldsymbol{x}| \leq \boldsymbol{s}\} & \text{if } \boldsymbol{x} \text{ feasible} \\ \infty & \text{otherwise} \end{cases} \\ &= \begin{cases} \|\boldsymbol{B}^T|\boldsymbol{x}|\|_\infty & \text{if } \boldsymbol{x} \in [-1,1]^p, \boldsymbol{B}^T|\boldsymbol{x}| \leq \mathbb{1} \\ \infty & \text{otherwise} \end{cases} \end{aligned}$$

□

9 Proof of Proposition 11

Proposition (Convexification)**.** *The convex envelope of $g_{\mathcal{G},\mathcal{D}}(\boldsymbol{x})$ over the unit ℓ_∞-ball is*

$$g_{\mathcal{G},\mathcal{D}}^{**}(\boldsymbol{x}) = \begin{cases} \sum_{(i,j)\in\mathcal{E}} (|x_i| + |x_j| - 1)_+ & \text{if } \boldsymbol{x} \in [-1,1]^p \\ \infty & \text{otherwise} \end{cases}$$

Proof. We use the linearization trick employed in [1] to reduce $g_{\mathcal{G},\mathcal{D}}(\boldsymbol{x})$ to a TU penalty. Let $\boldsymbol{s} = \mathbb{1}_{\mathrm{supp}(\boldsymbol{x})}$,

$$\begin{aligned} g_{\mathcal{G},\mathcal{D}}(\boldsymbol{x}) &= \sum_{(i,j)\in\mathcal{E}} s_i s_j \\ &= \min_{\boldsymbol{z}\in\{0,1\}^{|\mathcal{E}|}} \{ \sum_{(i,j)\in\mathcal{E}} z_{ij} : z_{ij} \geq s_i + s_j - 1\} \\ &= \min_{\boldsymbol{z}\in\{0,1\}^{|\mathcal{E}|}} \{ \sum_{(i,j)\in\mathcal{E}} z_{ij} : \boldsymbol{E}_G \boldsymbol{s} \leq \boldsymbol{z} - \mathbb{1}\} \end{aligned}$$

Now we can apply Proposition 1 in the main text, to compute the convex envelope:

$$\begin{aligned} g^{**}_{\mathcal{G},\mathcal{D}}(\boldsymbol{x}) &= \min_{\boldsymbol{s}\in[0,1]^p, \boldsymbol{z}\in[0,1]^{|\mathcal{E}|}} \{ \sum_{(i,j)\in\mathcal{E}} z_{ij} : \boldsymbol{E}_{\mathcal{G}} \boldsymbol{s} \leq \boldsymbol{z} - \mathbb{1}, |\boldsymbol{x}| \leq \boldsymbol{s}\} \\ &= \sum_{(i,j)\in\mathcal{E}} (|x_i| + |x_j| - 1)_+ \qquad\qquad (\boldsymbol{s}^* = \boldsymbol{x}, z^*_{ij} = (s^*_i + s^*_j - 1)_+) \end{aligned}$$

for $\boldsymbol{x} \in [-1,1]^p$, $g^{**}_{\mathcal{G},\mathcal{D}}(\boldsymbol{x}) = \infty$ otherwise. $\square$

References

[1] Marcin Kamiński. Quadratic programming on graphs without long odd cycles. 2008.

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

[3] Noah Simon, Jerome Friedman, Trevor Hastie, and Robert Tibshirani. A sparse-group lasso. *Journal of Computational and Graphical Statistics*, 22(2):231–245, 2013.