Convex block-sparse linear regression with expanders — provably

Anastasios Kyrillidis, Bubacarr Bah, Rouzbeh Hasheminezhad, Quoc Tran-Dinh, Luca Baldassarre, Volkan Cevher

The University of Texas at Austin

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

[Sparse matrices in compressed sensing]

Linear regression under the sparse setting, using convex optimization:

$$\text{(BP)} \qquad \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \|\boldsymbol{\beta}\|_1 \quad \text{subject to} \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta}$$

where

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}^\star, \qquad \mathbf{y} \in \mathbb{R}^n, \quad \mathbf{X} \in \mathbb{R}^{n \times p}$$

Main assumptions:

- $\boldsymbol{\beta}^\star$ is s-sparse or *approximately* s-sparse where $s \ll p$
- $\mathbf{X}$ has only a few non-zero elements per columns and satisfies the RIP-1, for all s-sparse vectors $\boldsymbol{\beta}$:

(RIP-1)

$$(1-\delta)\|\boldsymbol{\beta}\|_1 \leq \|\mathbf{X}\boldsymbol{\beta}\|_1 \leq \|\boldsymbol{\beta}\|_1, \quad \delta \in (0,1)$$

[What is known for this setting?]

$\mathbf{X}$ is an **expander matrix:**

Expander graph: a bipartite graph with p left nodes and n right nodes (with left degree d) is an *expander graph* if for any:

$$\mathcal{S} \subseteq [p], \ |\mathcal{S}| = s \ll p$$

and for constant $\epsilon_s \in (0, 1/2)$, the set of neighbors of $\mathcal{S}, \Gamma(\mathcal{S})$, satisfies:

$$|\Gamma(\mathcal{S})| \geq (1-\epsilon_s)d|\mathcal{S}|$$

Expander matrix $\mathbf{X}$: the adjacency matrix of an expander graph $\mathbf{X} \in \{0,1\}^{n \times p}$.

State-of-the-art result for BP: Berinde et al. 2008

Theorem: Assume $\mathbf{X} \in \{0,1\}^{n \times p}$ is an expander matrix — with degree d and expansion parameter $\epsilon_s \in (0,1/2)$ — such that it satisfies the RIP-1 condition as follows, for all s-sparse $\boldsymbol{\beta}$:

$$(1-\epsilon_s)d\|\boldsymbol{\beta}\|_1 \leq \|\mathbf{X}\boldsymbol{\beta}\|_1 \leq d\|\boldsymbol{\beta}\|_1.$$

Then, BP finds a solution $\widehat{\boldsymbol{\beta}}$ such that

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star\|_1 \leq \frac{2}{1-\frac{4\epsilon_s}{1-2\epsilon_s}} \cdot \|\boldsymbol{\beta}^\star - \boldsymbol{\beta}^\star_s\|_1$$

where $\boldsymbol{\beta}^\star_s$ is the s-sparse approximation of $\boldsymbol{\beta}^\star$.

[Structured sparsity]

Sparsity is merely a first-order description of $\boldsymbol{\beta}^\star$ in many applications - in practice, we know much more:

- (Overlapping) group sparsity: Genetic pathways in microarray data analysis, brain regions in neuro-imaging…
- Tree sparsity: Image processing and wavelet models…
- Graph sparsity…

For more information:

Yuan and Lin (2006), Baldassarre et al. (2013), Baraniuk et al. (2010), Asteris et al. (2014), Hegde et al. (2015)…

[In this work…]

Non-overlapping group sparsity:

Denote the block-sparse model:

$$\mathcal{M} := \{\mathcal{G}_1, \dots, \mathcal{G}_M\}$$

where

$$\mathcal{G}_i \subseteq [p], \ |\mathcal{G}_i| = g, \mathcal{G}_i \cap \mathcal{G}_j = \emptyset, \ i \neq j$$

and M is the total number of groups. The k-sparse block model is defined as a collection of sets $\mathcal{M}_k := \{\mathcal{G}_{i_1}, \dots, \mathcal{G}_{i_k}\}$. The best k-block sparse model satisfies:

$$\mathcal{M}_k^\star \in \arg\min_{\mathcal{M} \subseteq \mathcal{M}} \|\mathbf{w} - \mathbf{w}_{\mathcal{M}_k}\|_{2,1}$$

Group sparse norm:

$$\|\boldsymbol{\beta}\|_{2,1} := \sum_{\mathcal{G} \in \mathcal{M}} w_i \|\boldsymbol{\beta}_{\mathcal{G}}\|_2, \ w_i > 0.$$

[Problem statement]

Consider linear model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}^\star, \quad \text{where} \quad \boldsymbol{\beta}^\star \in \mathcal{M} \quad \text{is k-block sparse}$$

Consider optimization criterion:

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \|\boldsymbol{\beta}\|_{2,1} \quad \text{subject to} \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} \qquad (\star)$$

can we obtain guarantees of the form

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star\|_{2,1} \leq C \cdot \|\boldsymbol{\beta}^\star - \boldsymbol{\beta}^\star_{\mathcal{M}_k^\star}\|_{2,1} \ ?$$

Observe that the distance is measured using a more appropriate norm, given model assumptions!

[Our results]

Theorem: Assume $\mathbf{X} \in \{0,1\}^{n \times p}$ is an expander matrix — with degree d and expansion parameter $\epsilon_{\mathcal{M}_k}$ — such that it satisfies the **model based** RIP-1 condition as follows, for all k-block sparse $\boldsymbol{\beta}$:

$$(1-\epsilon_{\mathcal{M}_k})d\|\boldsymbol{\beta}\|_1 \leq \|\mathbf{X}\boldsymbol{\beta}\|_1 \leq d\|\boldsymbol{\beta}\|_1$$

Then, (*) finds a solution such that

$$\|\widehat{\boldsymbol{\beta}}\|_{2,1} \leq \|\boldsymbol{\beta}^\star\|_{2,1}$$

and

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star\|_{2,1} \leq \frac{2}{1-\frac{4\epsilon_{\mathcal{M}_k} g}{1-2\epsilon_{\mathcal{M}_k}}} \cdot \|\boldsymbol{\beta}^\star - \boldsymbol{\beta}^\star_{\mathcal{M}_k^\star}\|_{2,1}$$

where

$$\epsilon_{\mathcal{M}_k} \in \left(0, \frac{1}{2(1+2g)}\right)$$

Corollary: Assuming noise such that $\|\mathbf{X}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})\|_1 = \gamma \geq 0$, then:

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star\|_{2,1} \leq \frac{2}{1-\frac{4\epsilon_{\mathcal{M}_k} g}{1-2\epsilon_{\mathcal{M}_k}}} \cdot \|\boldsymbol{\beta}^\star - \boldsymbol{\beta}^\star_{\mathcal{M}_k^\star}\|_{2,1} + \frac{\gamma}{1-\frac{4\epsilon_{\mathcal{M}_k} g}{1-2\epsilon_{\mathcal{M}_k}}}$$

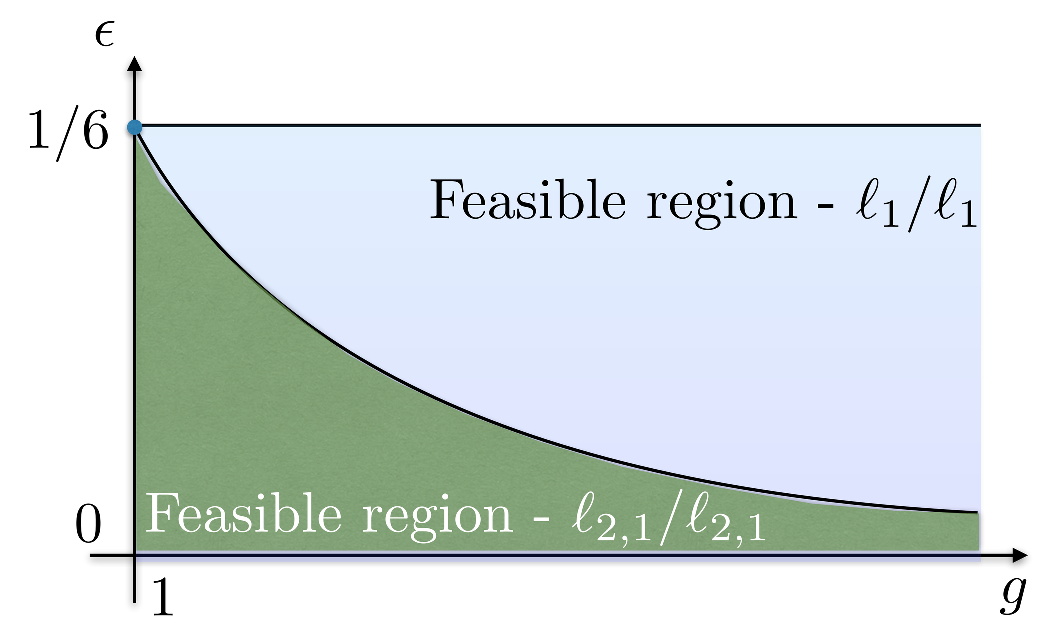
[Discussion]

Remark: In the extreme case where $g = 1$, our result is analogous to that of Berinde et al. (2008).

- When g grows, feasible values of $\epsilon_{\mathcal{M}_k} \to 0$, i.e., we require more rows in order to construct an expander matrix with the desirable expansion property.
- When we are *oblivious* to any, a priori sparsity model, we obtain from Berinde et al. (2008):

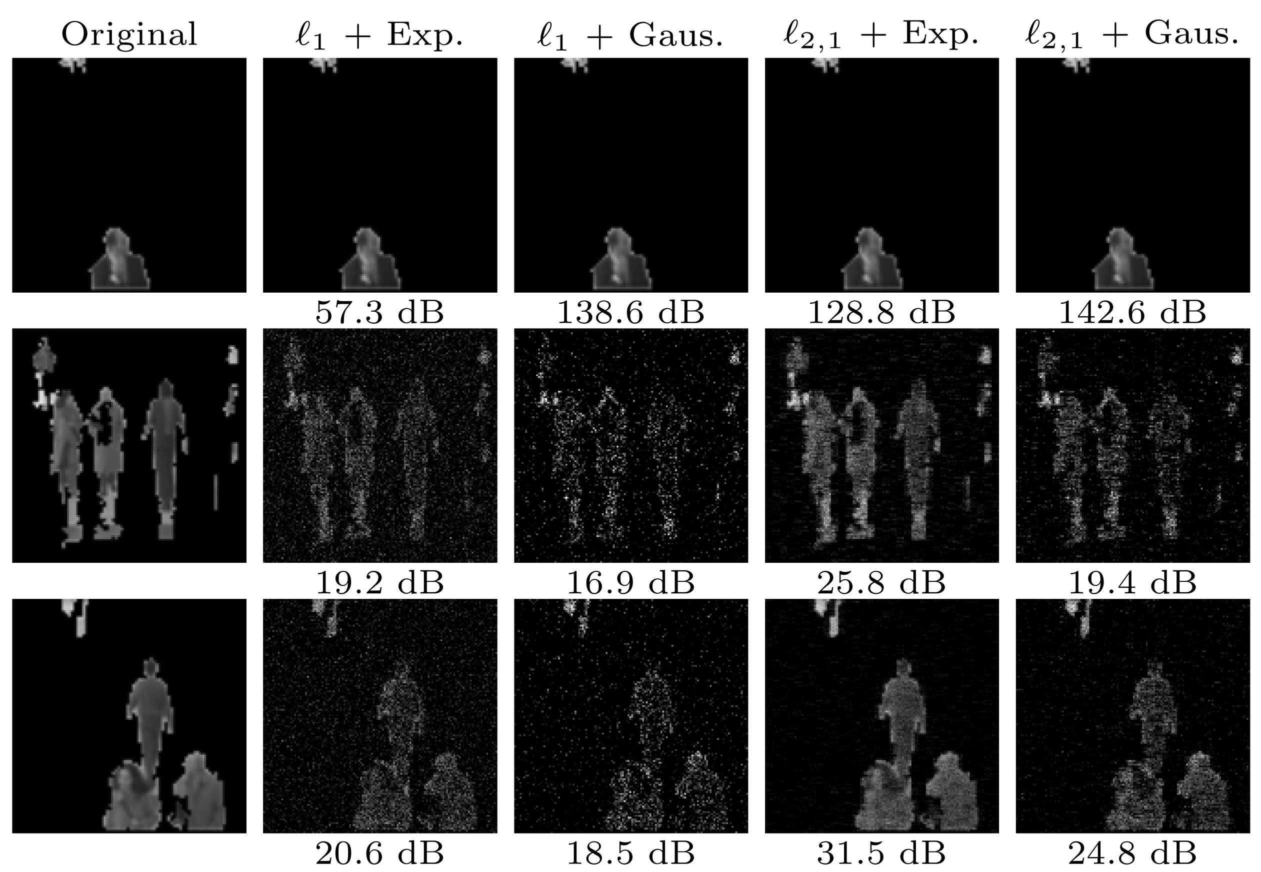
$$\|\boldsymbol{\beta}^\star - \widehat{\boldsymbol{\beta}}\|_{2,1} \leq \frac{2\sqrt{g}}{1-4\cdot\frac{\epsilon}{1-2\epsilon}} \cdot \|\boldsymbol{\beta}^\star - \boldsymbol{\beta}^\star_{\mathcal{M}_k^\star}\|_{2,1}.$$

- Our analysis provides weaker conditions bounds w.r.t. to range of values of $\epsilon_{\mathcal{M}_k}$



- However, the solution returned by standard BP solvers *does not necessarily belong to the model* $\mathcal{M}$.

[Block sparsity in image processing]



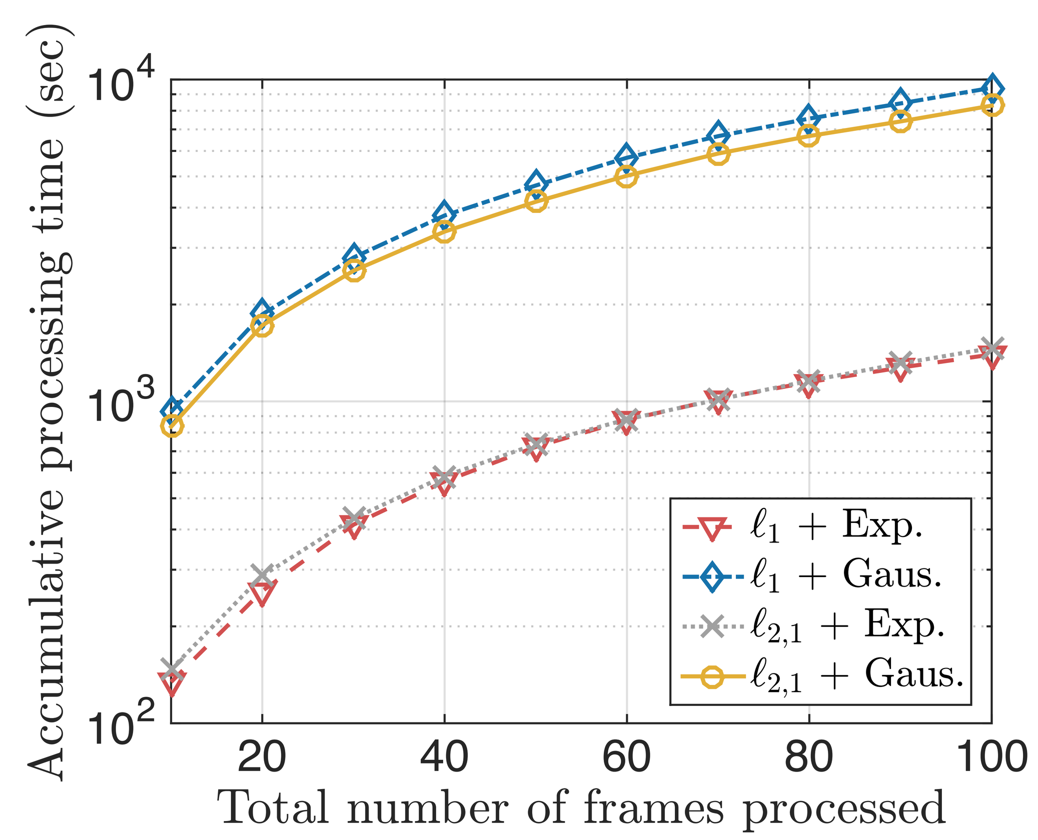


Figure *Left panel*: representative examples of subtracted frame recovery from compressed measurements. Here, $n = \lceil 0.3 \cdot p \rceil$ measurements are observed for $p = 2^{16}$. Block sparse model $\mathcal{M}$ contains groups of consecutive indices where $g = 4$. *Right panel*: Accumulative computational time required to process 100 frames. Overall, using Gaussian matrices in the $\ell_{2,1}$-norm case, DecOpt required almost 2.8 hours (upper bound), as compared to 0.55 hours when $\mathbf{X}$ is a sparse expander matrix. Thus, while Gaussian matrices is known to lead to better recovery results if no time restrictions apply, sparse sensing matrices constitute an appealing choice in practice.

[Algorithm evaluation]

Setting:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}^\star, \qquad \mathbf{y} \in \mathbb{R}^n, \quad \mathbf{X} \in \mathbb{R}^{n \times p}$$

- Expander matrix is designed with degree:

$$d = \left\lceil 22 \cdot \frac{\log(M)}{g} \right\rceil$$

- Last case: we force smaller degree ($d = 7$) - leads to faster convergence

Model		$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star\|_2$		Time (sec)	
p	$k \cdot g$	Gaus.	Exp.	Gaus.	Exp.
10000	300	8.6e-07	3.3e-06	24.3	**5.8**
	400	8.2e-06	3.4e-06	27.5	**6.2**
	500	8.6e-06	3.2e-06	27.8	**6.0**
	600	8.6e-06	3.4e-06	31.2	**7.6**
20000	300	8.1e-07	3.4e-06	95.5	**18.5**
	400	8.1e-06	3.3e-06	79.4	**16.4**
	500	8.5e-06	3.4e-06	83.9	**15.8**
	600	8.5e-06	3.5e-06	91.3	**18.9**
50000	300	8.2e-06	3.3e-06	419.3	**49.6**
	400	8.1e-06	3.4e-06	432.8	**45.8**
	500	8.5e-06	3.6e-06	436.0	**52.9**
	600	8.4e-06	3.5e-06	435.4	**51.1**
100000	600	8.1e-06	9.4e-06	1585.5	**55.1**
	800	8.1e-06	9.5e-06	1598.2	**54.5**
	1000	8.4e-06	9.4e-06	1600.6	**56.2**
	1200	8.1e-06	9.3e-06	1648.0	**55.6**

Table 1: Summary of comparison results for reconstruction and efficiency. Median values are reported. As a stopping criterion, we used $\|\boldsymbol{\beta}_{i+1} - \boldsymbol{\beta}_i\|_2 / \|\boldsymbol{\beta}_{i+1}\|_2 \leq 10^{-6}$, where $\boldsymbol{\beta}_i$ is the estimate at the i-th iteration. In all cases, $n = \lceil 0.4 \cdot p \rceil$.