Detection of Row-Sparse Matrices with Row-Structure Constraints

David Gregoratti, Carlos Buelga, and Xavier Mestre

Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA)
Av. C. F. Gauss 7, 08860 Castelldefels, Barcelona (Spain)

Problem Description

In this work we consider an underdetermined multi-measurement vector (MMV) linear regression problem where the parameter matrix is row-sparse and where an additional constraint fixes the number of nonzero elements in the active rows (see also Fig. 1). Even if this additional constraint offers side structure information that could be exploited to improve the estimation accuracy, it is extremely nonconvex and must be dealt with with caution. A detection algorithm is proposed that capitalizes on compressed sensing results and on the generalized distributive law (message passing on factor graphs).

Details

We consider the most classic MMV framework where the observation matrix, $\mathbf{Y} \in \mathbb{R}^{N\times L}$, is modeled as the product of the sampling matrix, $\mathbf{S} \in \mathbb{R}^{N\times M}$, and the true parameter matrix, $\mathbf{X}^* \in \mathbb{R}^{M\times L}$, plus additive white Gaussian noise, $\mathbf{W} \in \mathbb{R}^{N\times L}$:

$$\mathbf{Y} = \mathbf{S}\mathbf{X}^* + \mathbf{W}.$$

As introduced before, we are interested in the underdetermined problem where $N < M$. Even if such problem is ill-conditioned, compressed sensing (CS) results show that $\mathbf{X}^*$ can be recovered from $\mathbf{Y}$ with high accuracy under some mild assumptions on $\mathbf{S}$ as long as $\mathbf{X}^*$ is sparse. One possible solution is to approximate $\mathbf{X}^*$ by

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{Y} - \mathbf{S}\mathbf{X}\|_{\mathrm{F}}^2 + \lambda\|\mathbf{X}\|_1 \qquad (1)$$

with λ a real positive constant.
However, the model of Fig. 1 is characterized by a distinguishing sparsity structure, which we would like to exploit to improve the estimate precision. Specifically, we consider the case where the nonzero elements of $\mathbf{X}^*$ are concentrated in few rows. Moreover, each of these active rows has a fixed number of nonzero elements, namely r. In other words, for each row $\mathbf{r}_m^* \in \mathbb{R}^L$ of $\mathbf{X}^*$, $m = 1, 2, \dots, M$, we have the additional constraint that either

$$\|\mathbf{r}_m^*\|_0 = 0 \qquad \text{or} \qquad \|\mathbf{r}_m^*\|_0 = r \qquad (2)$$

One readily sees that this constraint is nonconvex and should be handled with care.

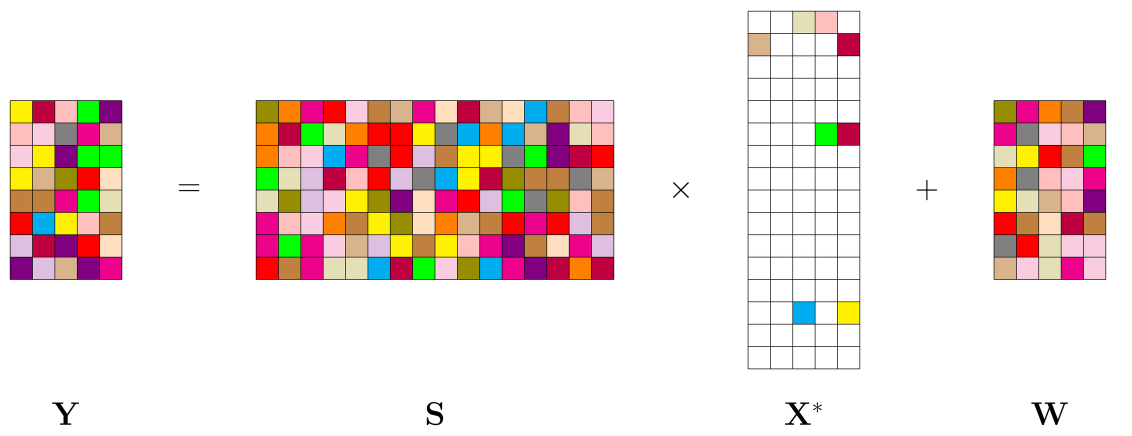


Fig. 1: Signal model of the considered MMV problem. Note that the parameter matrix is row-sparse and that active rows show a fixed number of nonzero elements.

Literature Overview

To the best of our knowledge, none of the available solutions for structured sparse problems captures the specificities of this model. Nevertheless, we briefly comment on how they can be employed to approximate the solution to the problem at hand.

Row Sparsity

The first approximation consists in solving the problem with tools designed for row-sparse parameter matrices, (see, e.g., [1] and references therein).

Cons: Algorithms for row-sparse matrices are indifferent to row structure and typically return rows with all active entries. An extra step is needed to enforce the required structure (e.g., ***hard thresholding*** to select the r entries with highest magnitude).

Composite Regularizer

A slightly more sophisticated solution consists in relaxing (1) subject to (2) into

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{Y} - \mathbf{S}\mathbf{X}\|_{\mathrm{F}}^2 + \lambda\|\mathbf{X}\|_1 + \mu\|\mathbf{X}\|_{2,1} \qquad (3)$$

with $\lambda, \mu > 0$. The purpose of the $\ell_{2,1}$ regularizer, namely

$$\|\mathbf{X}\|_{2,1} = \sum_{m=1}^{M} \sqrt{\sum_{l=1}^{L} x_{m,l}^2}$$

is to promote row sparsity. Together with the classic ℓ_1 regularizer, the resulting estimate, $\hat{\mathbf{X}}$, will show few active rows, each one with few active entries [2] (see also Fig. 2).

Pros: Problem (3) is convex and efficient; scalable algorithms exist for its solution.

Cons: The rows of the estimate, $\hat{\mathbf{X}}$, do not necessarily meet the structure constraints. Extra steps (hard decisions) may be required.

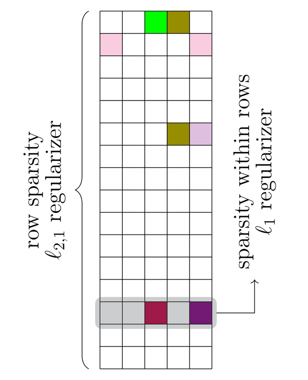


Fig. 2: Combined effects of the ℓ_1 and $\ell_{2,1}$ regularizers on the solution to (3).

Other Solutions

Other works in the literature allow for a more accurate characterization of the sparsity structure [3–5]. All these solutions, however, require an exhaustive search over the *atoms* of the sparsity model: This can be a severe limitation for the problem at hand where each row shows $\binom{L}{r}$ different activation patterns.

Proposed GDL-Based Approach

When dealing with sparsity structure, a number of works suggest that greedy algorithms are a better option. Indeed, convex continuous methods may not be able to induce an explicit distinction between active and inactive entries. Then, we replace (1) by

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{Y} - \mathbf{S}\mathbf{X}\|_{\mathrm{F}}^2 + \lambda\|\mathbf{X}\|_0 \quad \text{s. to (2).} \qquad (4)$$

By noting that the objective function can be decoupled along the columns of $\mathbf{X}$, that is

$$\frac{1}{2}\|\mathbf{Y} - \mathbf{S}\mathbf{X}\|_{\mathrm{F}}^2 + \lambda\|\mathbf{X}\|_0 = \frac{1}{2}\sum_{l=1}^{L}\|\mathbf{y}_l - \mathbf{S}\mathbf{x}_l\|_2^2 + \lambda\|\mathbf{x}_l\|_0$$

we see that each entry $x_{m,l}$ of $\mathbf{X}$ relates to only one of the column terms above and only one of the row structure constraints in (2), as depicted in the factor graph of Fig. 3. Then, we propose to solve problem (4) by means of an iterative *min-sum* message-passing algorithm that alternates between column problems and row problems. In other words, we apply the Generalized Distributive Law (GDL) [6, 7].
More specifically, the column-l-to-row-m message $\phi_{m,l}^{\mathrm{c}}(x_{m,l})$ is the column marginal

$$\phi_{m,l}^{\mathrm{c}}(x_{m,l}) = \min_{\{x_{i,l}\}_{i\neq m}} \frac{1}{2}\|\mathbf{y}_l - \mathbf{S}\mathbf{x}_l\|_2^2 + \lambda\|\mathbf{x}_l\|_0 + \sum_{i\neq m}\phi_{i,l}^{\mathrm{r}}(x_{i,l}) \qquad (5)$$

while

$$\phi_{m,l}^{\mathrm{r}}(x_{m,l}) = \min_{\{x_{m,j}\}_{j\neq l}} \sum_{j\neq l}\phi_{m,j}^{\mathrm{c}}(x_{m,j}) \quad \text{s. to (2)} \qquad (6)$$

is the row marginal that propagates from row m to column l.

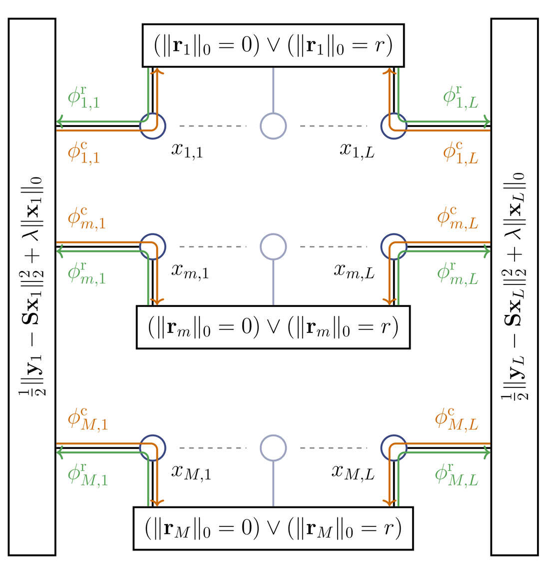


Fig. 3: Factor-graph representation of problem (4). Each row block can be further represente by the factor graph in Fig. 4.

Pros: This approach promotes the desired structure.

Cons: The number of cycles is huge, jeopardizing convergence. **Quick fix:** ignore rows where all column problems return zeros (which also helps complexity).

Row Marginals

For each row m, let us introduce the set of *hidden state variables* $\sigma_{m,l} \in \{0, 1, \dots, r\}$, with $l = 0, 1, \dots, L$, and define the state transition according to

$$\sigma_{m,0} = 0$$
$$\sigma_{m,l} = \sigma_{m,l-1} + \|x_{m,l}\|_0.$$

Moreover, the cost associated to the transition is $\phi_{m,l}^{\mathrm{c}}(x_{m,l})$.
Then, row problem (6) is equivalent to finding the most likely (minimum cost) *sequence of hidden states* $\sigma_{m,l}$ $(l = 0, 1, \dots, L)$ that leads to either $\sigma_{m,L} = 0$ or $\sigma_{m,L} = r$ from $\sigma_{m,0} = 0$.
The row marginals $\phi_{m,l}^{\mathrm{r}}(x_{m,l})$ are

$$\phi_{m,l}^{\mathrm{r}}(x_{m,l}) = \min_{\{x_{m,j}\}_{j\neq l}} \sum_{j\neq l}\phi_{m,j}^{\mathrm{c}}(x_{m,j})$$

subject to $x_{m,1}, x_{m,2}, \dots, x_{m,L}$ corresponding to a feasible sequence of hidden states.

Note that, since the state transition only depends on whether $x_{m,l} = 0$ or not (and not on the specific value taken by $x_{m,l}$), the row marginals can only take two values, namely

$$\phi_{m,l}^{\mathrm{r}}(x_{m,l}) = \begin{cases} \phi_{m,l}^{\mathrm{r}}(0) & \text{if } x_{m,l} = 0 \\ \phi_{m,l}^{\mathrm{r}}(*) & \text{if } x_{m,l} \neq 0. \end{cases}$$

This solution is, again, an application of the GDL and its factor-graph representation is depicted in Fig. 4. More specifically, it consists in running the Viterbi algorithm twice (left to right and right to left) on a trellis similar to the one in Fig. 5.

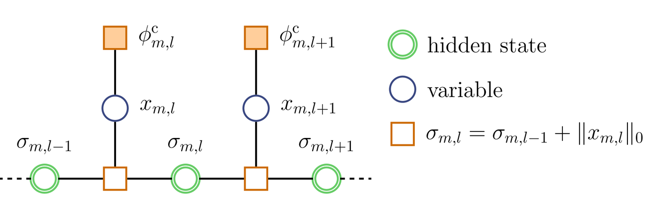


Fig. 4: Factor-graph representation of the row-wise minimization problem.

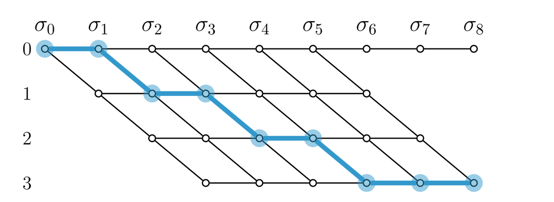


Fig. 5: Trellis representation of the row-wise minimization problem.

Future Work

The GDL-based approach for this trivial structure (exactly r active elements per active row) suggests that other more complex row structures can be investigated: For instance, the sequence of active and inactive entries of an active row can be mapped onto a codeword of a given binary code.

Column Marginals

Since the row marginals only take two values (for $x_{m,l} = 0$ and for $x_{m,l} \neq 0$), we see that problem (5) associates a cost $\phi_{i,l}^{\mathrm{r}}(0)$ to all entries $x_{i,l} = 0$ and a cost $\lambda + \phi_{i,l}^{\mathrm{r}}(*)$ to all entries $x_{i,l} \neq 0$.
Also, as far as the row problems (and our detection problem, in general) are concerned, we are not interested in characterizing the entire marginal $\phi_{m,l}^{\mathrm{c}}(x_{m,l})$ but we only need the values $\phi_{m,l}^{\mathrm{c}}(0)$ and $\phi_{m,l}^{\mathrm{c}}(*) = \min_{x_{m,l}\neq 0}\phi_{m,l}^{\mathrm{c}}(x_{m,l})$, which can be computed by solving the following equivalent problems (with $\mathbb{I}(\cdot)$ the indicator function)

$$\phi_{m,l}^{\mathrm{c}}(0) = \min_{\{x_{i,l}\}_{i=1}^{M}} \frac{1}{2}\|\mathbf{y}_l - \mathbf{S}\mathbf{x}_l\|_2^2 + \lambda\|\mathbf{x}_l\|_0 + \sum_{i\neq m}\phi_{i,l}^{\mathrm{r}}(x_{i,l}) + \mathbb{I}(x_{m,l} = 0)$$

$$\phi_{m,l}^{\mathrm{c}}(*) = \min_{\{x_{i,l}\}_{i=1}^{M}} \frac{1}{2}\|\mathbf{y}_l - \mathbf{S}\mathbf{x}_l\|_2^2 + \lambda\|\mathbf{x}_l\|_0 + \sum_{i\neq m}\phi_{i,l}^{\mathrm{r}}(x_{i,l}) + \mathbb{I}(x_{m,l} \neq 0).$$

The solution can be computed by means of Algorithm 1.

Algorithm 1: Modified OMP

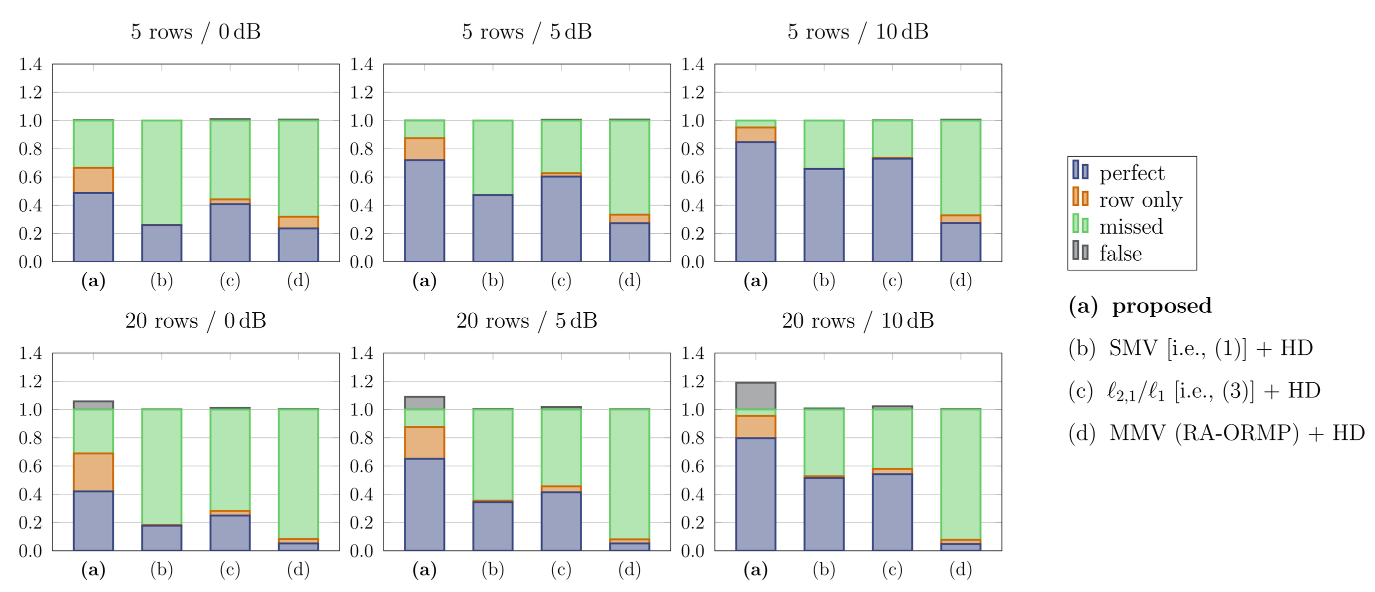
$\boldsymbol{\rho}_0 \leftarrow \mathbf{y}_l$, $i \leftarrow 0$, $\Omega_0 \leftarrow \emptyset$
repeat
  $i \leftarrow i + 1$
  for all $j \notin \Omega_{i-1}$ **do**
    $\gamma_j \leftarrow \frac{1}{2}(\mathbf{s}_j^{\mathrm{T}}\boldsymbol{\rho}_{i-1})^2 + \phi_{j,l}^{\mathrm{r}}(0) - \phi_{j,l}^{\mathrm{r}}(*) - \lambda$
  end for
  $k_i \leftarrow \arg\max_j \gamma_j$
  $\Omega_i \leftarrow \Omega_{i-1} \cup \{k_i\}$
  $\boldsymbol{\rho}_i \leftarrow (\mathbf{I}_N - \mathbf{S}_{\Omega_i}\mathbf{S}_{\Omega_i}^{\dagger})\mathbf{y}_l$
until $\gamma_{k_i} < 0$
$\Omega_* \leftarrow \Omega_i \setminus \{k_i\}$

Numerical Results

We test our algorithms with the following setup.

- Matrix $\mathbf{S}$ has size $N = 40$ and $M = 200$, and independent Gaussian entries with zero mean and unitary variance.
- The rows of the parameter matrix, $\mathbf{X}^*$, have either zero or $r = 2$ active entries. Active entries are drawn from a zero-mean, unitary-variance Gaussian distribution.
- AWGN matrix $\mathbf{W}$ has Gaussian i.i.d. entries with zero mean and variance SNR^{-1}.
- The regularizer parameters, λ and μ, have been chosen to minimize support errors.
- We run the algorithms 1000 times (all with different $\mathbf{S}$, $\mathbf{X}^*$ and $\mathbf{W}$) and we track *perfect* detection (i.e., the algorithm activates the right rows and the right entries), *row-only* detection (i.e., the algorithm selects the correct row but not the correct entries), *missed* detection (i.e., the algorithm fails to detect a row) and *false* detection (i.e., the algorithm selects an inactive row), all normalized w.r.t. the total number of active rows.
- For algorithms marked with HD, constraint (2) is forced by a *hard decision* based on the highest entry magnitudes.

The proposed GDL-based approach outperforms all other strategies by up to 25%!



References

[1] M. E. Davies *et al.*, "Rank awarness in joint sparse recovery," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1135–1146, Feb. 2012.

[2] A. Gramfort *et al.*, "Time-frequency mixed-norm estimates: Sparse M/EEG imaging with non-stationary source activations," *Neuroimage*, vol. 70, pp. 410–422, Apr. 2013.

[3] R. Jenatton *et al.*, "Proximal methods for hierarchical sparse coding," *J. Mach. Learn. Res.*, vol. 12, pp. 2297–2334, Jul. 2011.

[4] R. G. Baraniuk *et al.*, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.

[5] J. Huang *et al.*, "Learning with structured sparsity," *J. Mach. Learn. Res.*, vol. 12, pp. 3371–3412, Nov. 2011.

[6] S. M. Aji *et al.*, "The generalized distributive law," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 325–343, Mar. 2000.

[7] F. R. Kschischang *et al.*, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.