Selective ℓ_1 minimization for sparse recovery

Van Luong Le, Fabien Lauer and Gérard Bloch

***Abstract*—Motivated by recent approaches to switched linear system identification based on sparse optimization, the paper deals with the recovery of sparse solutions of underdetermined systems of linear equations. More precisely, we focus on the associated convex relaxation where the ℓ_1-norm of the vector of variables is minimized and propose a new iteratively reweighted scheme in order to improve the conditions under which this relaxation provides the sparsest solution. We prove the convergence of the new scheme and derive sufficient conditions for the convergence towards the sparsest solution. Experiments show that the new scheme significantly improves upon the previous approaches for compressive sensing. Then, these results are applied to switched system identification.**

***Index Terms*—Convex relaxation, sparsity, system identification, hybrid systems, switched systems.**

I. Introduction

This paper considers a convex relaxation approach to recover sparse solutions of underdetermined systems of linear equations, with an application to hybrid dynamical system identification. Linear hybrid systems are dynamical systems switching between multiple linear subsystems. More precisely, we consider single-input single-output systems in switched autoregressive with external input form as

$$y_i = \boldsymbol{\theta}_{\lambda_i}^\top \boldsymbol{\varphi}_i + v_i, \tag{1}$$

where $\lambda_i \in \{1, \dots, s\}$ is the discrete state or mode with s the number of submodels, $\{\boldsymbol{\theta}_j\}_{j=1}^s$ are the parameter vectors of the submodels, $v_i \in \mathbb{R}$ is a noise term and $\boldsymbol{\varphi}_i = [y_{i-1}, \dots, y_{i-n_a},\ u_{i-1}, \dots, u_{i-n_b}]^\top \in \mathbb{R}^p$ is the regression vector with $p = n_a + n_b$, where n_a and n_b are the model orders. Then, the identification problem is, given a collection $\mathcal{S} = \{(\boldsymbol{\varphi}_i, y_i)\}_{i=1}^N$, to estimate: (i) the number of submodels s, (ii) the parameter vectors $\{\boldsymbol{\theta}_j\}_{j=1}^s$, and (iii) the switching sequence $\{\lambda_i\}_{i=1}^N$.

Related work. Many hybrid system identification approaches have been proposed over the last decade [1]. Here, we focus on methods based on convex optimization and which offer several guarantees, the first of which being that the estimates correspond to a global solution of the formulated optimization problem. In addition, when a convex relaxation of a nonconvex optimization problem is considered, theoretical guarantees of equivalence between the two formulations can be obtained. More particularly, we follow the approach of [2], which iteratively estimates each parameter vector individually by maximizing the sparsity of the error vector. This sparse optimization problem is solved via its ℓ_1-norm convex relaxation and the sparsity of the solution is improved by the iteratively reweighted ℓ_1 minimization scheme developed in the compressive sensing literature [3]. Note that many results on ℓ_1-norm relaxations were also developed in this field [4], [5], [6].

Contribution. This paper proposes in Sect. II a new iterative method based on ℓ_1-norm minimization for the recovery of sparse solutions. As in the method of [3] (recalled in Sect. II-A), we consider a weighted form of the ℓ_1 convex relaxation. But, instead of updating the weights in a soft manner, in Sect. II-B we explicitly set a weight to zero at each iteration. The proposed scheme offers three major advantages when compared with the one of [3]: (i) it converges in a finite number of steps, (ii) theoretical guarantees of convergence towards the sparsest solution can be obtained, and (iii) experiments in Sect. IV-A show that it allows the sparsest solution to be recovered in a larger range of sparsity level. The advantages of this new sparsity enhancing scheme are used in Sect. III to improve the approach of [2] for hybrid system identification, as witnessed by experiments in Sect. IV-B.

II. Sparse recovery

Consider an underdetermined system of linear equations, $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, with a full row rank matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and a non-zero vector $\boldsymbol{b} \in \mathbb{R}^m$, where $m \ll n$. We are interested in sparse solutions of this system, i.e., solutions with few nonzero components, which are specifically obtained by solving

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \|\boldsymbol{x}\|_0, \qquad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \tag{2}$$

where the ℓ_0-pseudo norm is defined as $\|\boldsymbol{x}\|_0 = |\{i : x_i \neq 0\}|$. More precisely we concentrate on instances of (2) with a unique minimizer and assume that the following assumption holds in the rest of the paper.

Assumption 1. *Problem* (2) *has a unique minimizer.*

The following theorem shows that Assumption 1 holds in many cases.

Theorem 1 (Uniqueness via the spark [7])**.** *If a system of linear equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ has a solution $\boldsymbol{x}$ obeying*

$$\|\boldsymbol{x}\|_0 < \frac{\mathit{spark}(\boldsymbol{A})}{2},$$

where spark$(\boldsymbol{A})$ is the smallest number of linearly dependent columns of $\boldsymbol{A}$, this solution is necessarily the sparsest possible.

Even when having a unique solution, (2) remains a nonconvex optimization problem which is intractable for large n due to its combinatorial search nature. Nonetheless, it has been the focus of many works over the last decade in various fields and particularly in the context of compressive sensing [4], [5], [6].

V.L. Le and G. Bloch are with the Centre de Recherche en Automatique de Nancy (CRAN), Université de Lorraine, CNRS, France, van-luong.le@univ-lorraine.fr, gerard.bloch@univ-lorraine.fr

F. Lauer is with the LORIA, Université de Lorraine, CNRS, Inria, France, fabien.lauer@loria.fr

As discussed in [7] and references therein, a common alternative for (2) is to consider the convex relaxation based on the ℓ_1-norm. This leads to

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \|\boldsymbol{x}\|_1, \qquad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \tag{3}$$

where $\|\boldsymbol{x}\|_1 = \sum_{i=1}^n |x_i|$ is the ℓ_1-norm of $\boldsymbol{x}$. This problem is convex and can typically be solved efficiently [8]. As shown in [7], for a column normalized matrix $\boldsymbol{A}$, problems (2) and (3) are equivalent if, for a solution of $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, the condition

$$\|\boldsymbol{x}\|_0 < \frac{1}{2}\left(1 + \frac{1}{\mu(\boldsymbol{A})}\right) \tag{4}$$

holds, where $\mu(\boldsymbol{A})$ is the mutual coherence of $\boldsymbol{A}$ [9] which is defined, for any matrix $\boldsymbol{A} = [\boldsymbol{A}_1, \dots, \boldsymbol{A}_n] \in \mathbb{R}^{m\times n}$, by

$$\mu = \mu(\boldsymbol{A}) = \max_{1\leq i,j\leq n, i\neq j} \frac{\left|\boldsymbol{A}_i^\top \boldsymbol{A}_j\right|}{\|\boldsymbol{A}_i\|_2 \|\boldsymbol{A}_j\|_2}. \tag{5}$$

In the case of an unnormalized matrix $\boldsymbol{A}$, a similar equivalence is obtained by considering a weighted version of (3), i.e., $\min_{\boldsymbol{x}\in\mathbb{R}^n} \|\boldsymbol{W}_A\boldsymbol{x}\|_1$, s.t. $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{W}_A = \operatorname{diag}(\|\boldsymbol{A}_1\|_2, \dots, \|\boldsymbol{A}_n\|_2)$.

However, in many applications, the matrix $\boldsymbol{A}$ cannot be freely chosen and the sufficient condition (4) might be violated. Yet, the problem defined in (2) may have a unique solution $\boldsymbol{x}_0$, as stated by Theorem 1. Thus, since $1 + \frac{1}{\mu(\boldsymbol{A})} \leq \operatorname{spark}(\boldsymbol{A})$ [7], there is a range of problems with

$$\frac{1}{2}\left(1 + \frac{1}{\mu(\boldsymbol{A})}\right) \leq \|\boldsymbol{x}_0\|_0 \leq \frac{\operatorname{spark}(\boldsymbol{A})}{2}.$$

For these problems, recovering the sparsest solution is therefore a well-defined problem, but not directly solvable through the ℓ_1 convex relaxation: a solution $\boldsymbol{x}_1$ to (3) may have more nonzero elements than $\boldsymbol{x}_0$.

A. The classical iteratively reweighted approach

In order to improve the sparsity of the solutions of (3), a reweighted ℓ_1 minimization scheme is proposed in [3]. At each iteration l, the following problem is solved:

$$\boldsymbol{x}_1^{(l)} \in \arg\min_{\boldsymbol{x}\in\mathbb{R}^n} \|\boldsymbol{W}_l\boldsymbol{x}\|_1, \qquad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \tag{6}$$

where $\boldsymbol{W}_l = \operatorname{diag}(w_1^{(l)}, \dots, w_n^{(l)})$ is a weighting diagonal matrix which penalizes differently the entries of $\boldsymbol{x}$. At the first iteration, the weights are equal, i.e., $\boldsymbol{W}_1 = \boldsymbol{I}_n$. Then, $\boldsymbol{W}_l$ is updated with

$$w_i^{(l+1)} = \frac{1}{\left|x_{1i}^{(l)}\right| + \epsilon},$$

where the $w_i^{(l+1)}$ are the weights at the $(l+1)$th iteration, $x_{1i}^{(l)}$ is the ith element of the solution of (6) at the lth iteration and $\epsilon > 0$ is a parameter preventing a division by zero. Note that the choice of ϵ has an influence on the convergence of $\boldsymbol{x}_1^{(l)}$.

Important open issues, highlighted in [3], regarding this scheme are: (i) *What are smart and robust rules for selecting the parameter ϵ?* (ii) *Under what conditions does the algorithm converge?* Though preliminary results on the convergence are given in [10], these rely on a condition involving both the matrix $\boldsymbol{A}$ and the solution to (2). On the contrary, the following presents another reweighting mechanism with a convergence analysis relying only on conditions on $\boldsymbol{A}$.

Note that other methods have been proposed to recover sparse solutions, including the greedy algorithm of [11] and the one in [12] that mixes ℓ_1 minimization with a greedy approach.

B. Selective ℓ_1 minimization

In this section, we propose a new method for updating the weighting matrix $\boldsymbol{W}_l$ in (6) in order to improve the sparsity of the solution. The new method, named Selective ℓ_1 Minimization (Sℓ_1M), is given in Algorithm 1, where $\|\boldsymbol{x}\|_\infty = \max_i |x_i|$ is the ℓ_∞-norm of $\boldsymbol{x}$.

Algorithm 1 Sℓ_1M

```
Require: $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$.
  - Initialize $l = 0$, $\boldsymbol{W}_1 = \boldsymbol{I}_n$.
  repeat
    - Set $l = l + 1$.
    - Get any $\boldsymbol{x}_1^{(l)}$ in the solution set of (6).
    - Select the smallest index $q^{(l)} \in \arg\max_{i=1,\dots,n} w_i^{(l)}|x_{1i}^{(l)}|$.
    - Calculate $\boldsymbol{W}_{l+1}$ with $w_i^{(l+1)} = \begin{cases} w_i^{(l)}, & \text{if } i \neq q^{(l)}, \\ 0, & \text{if } i = q^{(l)}. \end{cases}$
  until $\left\|\boldsymbol{W}_l\boldsymbol{x}_1^{(l)}\right\|_\infty = 0$ or $\left\|\boldsymbol{W}_{l+1}\boldsymbol{x}_1^{(l)}\right\|_\infty = 0$.
  return $\boldsymbol{x}_1^* = \boldsymbol{x}_1^{(l)}$.
```

Algorithm 1 relaxes the optimization of the nonzero variables by setting their weights w_i to 0 in the cost function of (6), thus putting more weight on the other variables that are pulled towards 0. When the stopping criterion is met, we have $\|\boldsymbol{W}_l\boldsymbol{x}_1^{(l)}\|_0 \leq 1$. Hence, if it returns at iteration $l < n$, the algorithm yields a sparse solution.

The following provides an analysis of the proposed iterative scheme. The convergence in a finite number of steps is proved and a condition on the matrix $\boldsymbol{A}$ and the sparsity level guaranteeing the convergence towards the desired solution is derived.

1) Convergence in a finite number of steps: The following theorem, based on Lemmas given in Appendix A, guarantees that the algorithm Sℓ_1M converges in a finite number of steps.

Theorem 2. *The solution $\boldsymbol{x}_1^*$, returned by Algorithm 1, is found in at most $m+1$ iterations and $\|\boldsymbol{x}_1^*\|_0 \leq m$, where m is the number of rows in $\boldsymbol{A}$.*

Proof: If Algorithm 1 does not converge after m iterations, Lemma 3 implies that $\left|x_{1q^{(l)}}^{(l)}\right| \neq 0, \forall l \leq m$. Then, according to Lemma 4, the columns $\boldsymbol{A}_{q^{(l)}}, l \in \{1, \dots, m\}$, of $\boldsymbol{A}$ are linearly independent and $\boldsymbol{b} \in \mathbb{R}^m$ can be expressed as a linear combination of these columns. Therefore, the minimum value of the sum $\sum_{i\notin\{q^{(l)}\}_{l=1}^m} |x_{1i}|$ is 0 and the solution $\boldsymbol{x}_1^{(m+1)}$ to (6) at iteration $m+1$ satisfies the stopping criterion of the algorithm as $\left\|\boldsymbol{W}_{(m+1)}\boldsymbol{x}_1^{(m+1)}\right\|_\infty = 0$.

In addition, if Algorithm 1 converges at the lth iteration, $\left\|\boldsymbol{x}_1^{(l+1)}\right\|_0 \leq l-1$. Thus, $\|\boldsymbol{x}_1^*\|_0 \leq m$. ■

It is worth noting that if $\|\boldsymbol{x}_1^*\|_0 < \frac{\text{spark}(\boldsymbol{A})}{2}$, according to Theorem 1, $\boldsymbol{x}_1^*$ is the sparsest solution of (2).

2) Convergence towards the sparsest solution: In order for the iterative algorithm to converge to the solution $\boldsymbol{x}_0$ of (2) under Assumption 1, we need to ensure that the variables removed from the weighted sum in the cost function of (6) correspond to nonzeros in $\boldsymbol{x}_0$. The following proposition shows under which conditions the choice of the index $q^{(l)}$ in Algorithm 1 corresponds to a nonzero element in the solution of (2), $\boldsymbol{x}_0$, for which we define the two sets

$$I_0 = \{i : x_{0i} = 0\}, \quad I_1 = \{i : x_{0i} \neq 0\}. \tag{7}$$

Proposition 1. *If $\left\|\boldsymbol{x}_0 - \boldsymbol{x}_1^{(l)}\right\|_1 < \frac{\mu+1}{2\mu}\|\boldsymbol{W}_l\boldsymbol{x}_0\|_\infty$, where μ is defined as in* (5) *with a column normalized matrix $\boldsymbol{A}$, then $q^{(l)} \in I_1$ for all $q^{(l)} \in \arg\max_i w_i^{(l)}\left|x_{1i}^{(l)}\right|$.*

Proof: Assume $q^{(l)} \notin I_1$, $x_{0q^{(l)}} = 0$, then $\left|\left(\boldsymbol{x}_0 - \boldsymbol{x}_1^{(l)}\right)_{q^{(l)}}\right| = |x_{1q^{(l)}}^{(l)}| = \max_i w_i^{(l)}\left|x_{1i}^{(l)}\right|$. Let $j \in \arg\max_i w_i^{(l)}|x_{0i}|$ and $|x_{0j}| = \|\boldsymbol{W}_l\boldsymbol{x}_0\|_\infty > \frac{2\mu}{\mu+1}\left\|\boldsymbol{x}_0 - \boldsymbol{x}_1^{(l)}\right\|_1$. Then, since $\left|x_{1j}^{(l)}\right| \geq |x_{0j}| - \left|\left(\boldsymbol{x}_0 - \boldsymbol{x}_1^{(l)}\right)_j\right|$, Lemma 5 in Appendix applied to $\boldsymbol{\delta} = \boldsymbol{x}_0 - \boldsymbol{x}_1^{(l)}$ leads to $\left|x_{1j}^{(l)}\right| > \frac{\mu}{\mu+1}\left\|\boldsymbol{x}_0 - \boldsymbol{x}_1^{(l)}\right\|_1$. Thus

$$\left|x_{1q^{(l)}}^{(l)}\right| = \left|\left(\boldsymbol{x}_0 - \boldsymbol{x}_1^{(l)}\right)_{q^{(l)}}\right| \geq \left|x_{1j}^{(l)}\right| > \frac{\mu}{\mu+1}\left\|\boldsymbol{x}_0 - \boldsymbol{x}_1^{(l)}\right\|_1. \tag{8}$$

On the other hand, Lemma 5 implies that $\left|(\boldsymbol{x}_0 - \boldsymbol{x}_1^{(l)})_{q^{(l)}}\right| \leq \frac{\mu}{\mu+1}\left\|\boldsymbol{x}_0 - \boldsymbol{x}_1^{(l)}\right\|_1$, which contradicts (8). Thus, the assumption $q^{(l)} \notin I_1$ is wrong and we conclude that $q^{(l)} \in I_1$. ■

3) Influence of the weighting matrix on sparse recovery: The following lemma shows that, with a good choice of $\boldsymbol{W}$ such that $w_i = 0, i \in I_1$, the solution to (6) is exactly $\boldsymbol{x}_0$.

Lemma 1. *Given a diagonal matrix $\boldsymbol{W}$, with entries $w_i \geq 0$, if for all nonzero $\boldsymbol{\delta} \in Ker(\boldsymbol{A})$ the following condition holds*

$$\sum_{i \in I_1} w_i|\delta_i| < \sum_{i \in I_0} w_i|\delta_i|, \tag{9}$$

where I_0 and I_1 are defined by (7)*, then the solution $\boldsymbol{x}_0$ to* (2) *under Assumption 1 uniquely solves problem* (6)*, i.e., $\boldsymbol{x}_0 = \arg\min_{\boldsymbol{x} \in \mathbb{R}^n} \|\boldsymbol{W}\boldsymbol{x}\|_1 \quad s.t. \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$.*

Proof: $\boldsymbol{x}_0$ uniquely solves problem (6) if, for all nonzero $\boldsymbol{\delta} \in \text{Ker}(\boldsymbol{A})$, $\|\boldsymbol{W}\boldsymbol{x}_0\|_1 < \|\boldsymbol{W}(\boldsymbol{x}_0 + \boldsymbol{\delta})\|_1$, or equivalently if

$$\sum_{i \in I_1} w_i\left(|x_{0i}| - |x_{0i} + \delta_i|\right) < \sum_{i \in I_0} w_i\left|\delta_i\right|,$$

which is implied by the condition (9) since $\sum_{i \in I_1} w_i\left(|x_{0i}| - |x_{0i} + \delta_i|\right) \leq \sum_{i \in I_1} w_i|\delta_i|$. ■

If the condition in Proposition 1 is satisfied for h iterations, the condition (9) is relaxed to

$$\sum_{i \in I_1} |\delta_i| - \sum_{i \in \left\{q^{(l)}\right\}_{l=1}^{h}} |\delta_i| < \sum_{i \in I_0} |\delta_i|. \tag{10}$$

When this condition is satisfied, the solution of (6) with $\boldsymbol{W}_l = \boldsymbol{W}_{h+1}$ is the sparsest vector $\boldsymbol{x}_0$ thanks to Lemma 1. We see that the left-hand side (LHS) of (10) can decrease to zero after $|I_1|$ iterations. This also shows that the algorithm converges after at most $|I_1| + 1$ iterations if the indexes $q^{(l)}$ are all well-chosen in I_1.

4) Sparse recovery condition: We now show, in Theorem 3, a condition on the matrix $\boldsymbol{A}$ and the sparsity level which can guarantee the convergence towards the desired solution. But this requires the following definition using the notation $\boldsymbol{\delta}_T$ for the subvector of $\boldsymbol{\delta}$ containing its components of indexes in T.

Definition 1 (Definition 1 in [5])**.** *A matrix $\boldsymbol{A} \in \mathbb{C}^{m \times n}$ is said to satisfy the null space property (NSP) of order k with constant $\gamma \in (0, 1)$ if*

$$\|\boldsymbol{\delta}_T\|_1 \leq \gamma\|\boldsymbol{\delta}_{T^c}\|_1,$$

for all sets $T \subset \{1, \ldots, n\}$ with $|T| \leq k$, $T^c = \{1, \ldots, n\} \backslash T$ and for all $\boldsymbol{\delta} \in Ker(\boldsymbol{A})$.

The following theorem extends Theorem 1 in [5] to cases where the sparsity level $\|\boldsymbol{x}_0\|_0$ equals $k + h$ with $h > 0$ and provides a sparse recovery condition for the iterative Algorithm 1 (proof given in Appendix B).

Theorem 3. *Given a matrix $\boldsymbol{A}$ that satisfies the NSP of order k with constant $\gamma \in (0, \frac{1}{2})$, if $\boldsymbol{x}_0$ is such that $\boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{b}$ and $\|\boldsymbol{x}_0\|_0 \leq k + h$ with the integer $h \in [1, k]$ satisfying*

$$\gamma < \frac{1 - (4h-1)\mu}{1 + (4h+1)\mu}, \tag{11}$$

where μ is defined as in (5)*, Algorithm 1 converges to $\boldsymbol{x}_0$ in at most $h + 1$ iterations.*

Note that the NSP of $\boldsymbol{A}$ and the value of γ can be difficult to determine directly, but can be related to the easier to handle restricted isometry property (see [5] for details).

III. Hybrid system identification

We now turn to the problem of hybrid system identification, i.e., of estimating the parameter vectors $\{\boldsymbol{\theta}_j\}_{j=1}^s$ in (1) from a data set $\mathcal{S} = \{(\boldsymbol{\varphi}_i, y_i)\}_{i=1}^N$.

We follow the approach of [2] in which we introduce the proposed reweighted scheme (Sℓ_1M). This yields Algorithm 2 for the estimation of a single parameter vector.

Algorithm 2 requires a number of iterations, N_s, in order to deal with noisy data (for which true sparsity cannot be obtained). However, with knowledge of the number of modes, s, we can set $N_s = \frac{s-1}{s}N$, since in this case the largest fraction of points of a mode in the data set is at least N/s.

As in [2], the identification procedure obtains the submodels one by one. After applying Algorithm 2 to estimate a parameter vector $\hat{\boldsymbol{\theta}}_j$, the data points verifying the error condition, $|y_i - \boldsymbol{\varphi}_i^\top\hat{\boldsymbol{\theta}}_j| \leq \delta$ where δ is a fixed threshold, are associated to this submodel and removed from the data set. Then, the next parameter vectors are iteratively estimated from reduced data sets until all data points are removed, at which point the estimated number of modes $\hat{s}$ is given by the number of submodels obtained. Note that, if the noise is unbounded,

```
**Algorithm 2** Estimation of a single parameter vector
Require: A data set $\mathcal{S} = \{(\boldsymbol{\varphi}_i, y_i)\}_{i=1}^N$ and a number of iterations $N_s$.
  - Initialize $l = 0$, $\boldsymbol{W} = \boldsymbol{I}_N$.
  while $l < N_s$ do
    - Set $l = l + 1$.
    - Obtain $\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \|\boldsymbol{W}\boldsymbol{e}(\boldsymbol{\theta})\|_1$, where $\boldsymbol{e}(\boldsymbol{\theta}) = \left[y_1 - \boldsymbol{\varphi}_1^\top\boldsymbol{\theta}, \dots, y_N - \boldsymbol{\varphi}_N^\top\boldsymbol{\theta}\right]^\top$ is the error vector.
    - Select an index in the maximal absolute error set (break the tie arbitrarily if necessary):
          $q \in \arg\max_i w_i \left|e_i\left(\boldsymbol{\theta}^*\right)\right|.$
    - Set the $q$th entry on the diagonal of $\boldsymbol{W}$ to $w_q = 0$.
  end while
  return $\boldsymbol{\theta}^*$ and $\boldsymbol{W}$.
```

e.g., Gaussian, the procedure should stop before that, i.e., when a small and predefined fraction of the data remains, to avoid creating irrelevant submodels for small groups of points corrupted by large noise terms.

IV. Numerical experiments

A. Compressive sensing example

We consider a classical example of sparse signal recovery used in many works to show the efficiency of the proposed method. The goal is to recover a sparse signal $\boldsymbol{x}$ of length n with $\|\boldsymbol{x}\|_0 = k$. The k nonzero positions are chosen randomly, and the nonzero values are randomly drawn according to a zero-mean unit-variance Gaussian distribution. The sensing matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is a Gaussian matrix, i.e., with entries following a zero-mean Gaussian distribution with variance $1/m$. The sparsity level k is increased from 10 to 60 to see the capacity in signal recovery.

To compare with the classical reweighting method, we set the experiment as in [3] with $n = 256, m = 100$. For each value of k, we run 500 trials to estimate the probability of perfect signal recovery (be successful if $\|\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_0\|_\infty \leq 10^{-3}$). Figure 1 reports the successful recovery probability, Pr(recovery), for the unweighted ℓ_1-norm minimization (Unweighted ℓ_1), the Orthogonal Matching Pursuit (OMP) [11], the reweighted scheme of [3] (Reweighted), and the proposed one (Sℓ_1M). We see that the requisite oversampling factor for perfect recovery [3], $\min_k m/k$ s.t. Pr(recovery)$= 1$, decreases from approximately 4 for Unweighted ℓ_1 or 3 for the method of [3] with $\epsilon = 1$ to $100/40 = 2.5$ for our method. Moreover, in our method there is no hyperparameter to tune, whereas ϵ can influence the results for the classical scheme [3].

From Theorem 1, we must have $k < \text{spark}(\boldsymbol{A})/2 \leq m/2 = 50$ to guarantee the uniqueness of the solution to (2). This explains why all methods have a small successful recovery probability with the sparsity level k close to 50. Nonetheless, our method shows a successful recovery probability greater than 0.9 at $k = 45$.

These improvements are paid for by the computational cost of the proposed method, which requires a larger number of iterations compared with [3]. This trade-off is consistent with the results of the OMP [11], which obtains a low probability of success but at a much lower computational cost.

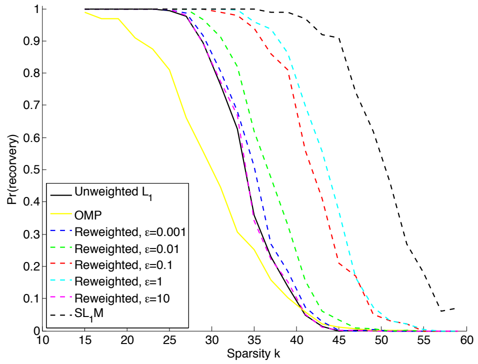

Fig. 1. Empirical probability of successful recovery versus the sparsity level k for the OMP method [11], the classical reweighting of [3] with various ϵ and the proposed one (Sℓ_1M).

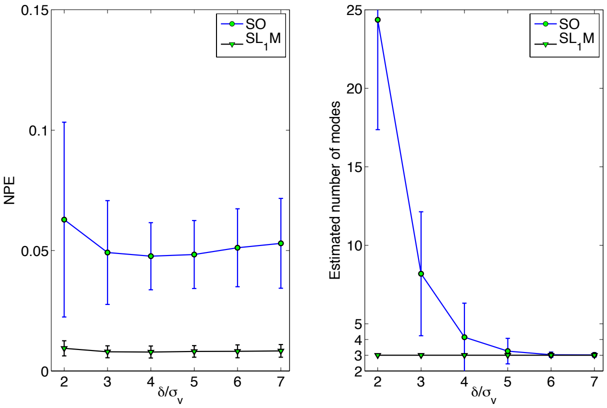

Fig. 2. Parametric error (NPE) and estimated number of modes for different values of the threshold δ used in the switched system identification example.

B. Hybrid system identification example

We now consider the switched linear system with 3 modes and $n_a = n_b = 2$ given as an example in Sect. 4 of [2] to test the identification procedure. Training sets of $N = 600$ points are generated with a uniformly distributed random sequence of $\lambda_i \in \{1, 2, 3\}$ and an additive Gaussian noise with $\sigma_v = 0.1$.

We compare the Normalized Parametric Error, $NPE = \sqrt{\sum_{j=1}^s \|\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j\|_2^2 / \sum_{j=1}^s \|\boldsymbol{\theta}_j\|_2^2}$, of the original method of [2] (SO) using the reweighting of [3] with the one of the same method based on Sℓ_1M and Algorithm 2 (labeled Sℓ_1M for short). More precisely, over 100 trials with different input, switching and noise sequences, we report the mean and standard deviation of the NPE. Since the methods estimate the parameter vectors one by one until the data set is empty, the number of modes cannot be fixed. If $\hat{s} > s$, the NPE is computed with the s best parameter vectors that yield the smallest NPE. The threshold δ used to assign data points to a submodel is varied in the range $[2\sigma_v, 7\sigma_v]$. Due to the unbounded Gaussian noise, both methods are stopped with 5% of the data left unassigned to a mode.

Figure 2 shows that the proposed method (with N_s set as suggested in Sect. III) yields a model with a smaller error and a better estimate of the number of modes than the SO method. Moreover, this is true over a large range of values for δ.

V. Conclusions

The paper proposed a new iterative algorithm to improve the sparsity of the solution of an ℓ_1-norm relaxation. Compared with the state-of-the-art scheme of [3], the proposed algorithm benefits from the absence of hyperparameters and a finite convergence in a number of steps at most equal to the number of linear equations. In addition, a sparse recovery condition, guaranteeing the convergence towards the sparsest solution, was proved and experiments showed that the new scheme can recover the sparsest solution in more difficult cases. Finally, we presented an application to hybrid system identification, where the increased sparse recovery capacity of the method translates into more accurate parameter estimates.

Appendix

Throughout the appendix, for a vector $\boldsymbol{x} \in \mathbb{R}^n$ and an index set $T \subset \{1, \dots, n\}$, we denote $\boldsymbol{x}_T$ the subvector of $\boldsymbol{x}$ containing its components of indexes in T and $\|\boldsymbol{x}_T\|_1 = \sum_{i \in T} |x_i|$.

A. Useful lemmas and definitions

Lemma 2. *Given a solution* $\boldsymbol{x}_1$ *of* $\min_{\boldsymbol{x} \in \mathbb{R}^n} \|\boldsymbol{x}_T\|_1$, *s.t.* $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, *if* $x_{1i} \neq 0$ *for some* $i \in T$, *then, for a solution,* $\tilde{\boldsymbol{x}}_1$, *of* $\min_{\boldsymbol{x} \in \mathbb{R}^n} \|\boldsymbol{x}_{\tilde{T}}\|_1$, *s.t.* $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, *where* $\tilde{T} = T \backslash i$, *we have* $\tilde{x}_{1i} \neq 0$.

Proof: We know that $\|\boldsymbol{x}_{1T}\|_1 \leq \|\tilde{\boldsymbol{x}}_{1T}\|_1$ or equivalently $\|\boldsymbol{x}_{1\tilde{T}}\|_1 + |x_{1i}| \leq \|\tilde{\boldsymbol{x}}_{1\tilde{T}}\|_1 + |\tilde{x}_{1i}|$ while $\|\boldsymbol{x}_{1\tilde{T}}\|_1 \geq \|\tilde{\boldsymbol{x}}_{1\tilde{T}}\|_1$. Therefore $|\tilde{x}_{1i}| \geq |x_{1i}| > 0$. ■

Lemma 3. *For* $\boldsymbol{b} \neq \boldsymbol{0}$, *after* l *iterations, if* $\left\|\boldsymbol{W}_l \boldsymbol{x}_1^{(l)}\right\|_\infty \neq 0$, *Algorithm 1 yields* $\boldsymbol{x}_1^{(l)}$ *such that* $\left|x_{1q^{(i)}}^{(l)}\right| \neq 0, \forall i = \{1, \dots, l\}$.

Proof: Lemma 3 is a consequence of Lemma 2 and the fact that $\left\|\boldsymbol{W}_1 \boldsymbol{x}_1^{(1)}\right\|_\infty = \left|x_{1q^{(1)}}^{(1)}\right| \neq 0$ if $\boldsymbol{b} \neq \boldsymbol{0}$. ■

Lemma 4. *In Algorithm 1, if,* $\forall l \leq m$, $\left|x_{1q^{(l)}}^{(l)}\right| \neq 0$, *then the columns* $\boldsymbol{A}_{q^{(i)}}$, $i = 1, \dots, l$, *of the full row rank matrix* $\boldsymbol{A}$ *are linearly independent.*

Proof: To prove that, we show that at the jth iteration, $\forall j \geq 2$, if $\left|x_{1q^{(l)}}^{(l)}\right| \neq 0, \forall l \leq j$, then $\boldsymbol{A}_{q^{(j)}}$ is linearly independent of $\left\{\boldsymbol{A}_{q^{(l)}}\right\}_{l=1}^{j-1}$. Assume that this is not true, i.e., $\boldsymbol{A}_{q^{(j)}} = \sum_{l=1}^{(j-1)} \beta_l \boldsymbol{A}_{q^{(l)}}$. Then,

$$\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}_1^{(j)} = \sum_{l=1}^{j-1} \left(x_{1q^{(l)}}^{(j)} + x_{1q^{(j)}}^{(j)} \beta_l\right) \boldsymbol{A}_{q^{(l)}} + \sum_{i \notin \left\{q^{(l)}\right\}_{l=1}^{j}} x_{1i}^{(j)} \boldsymbol{A}_i,$$

and we can get another solution $\boldsymbol{x}^{*(j)}$ of $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}$ whose elements are given as the coefficients of the columns of $\boldsymbol{A}$ in the above. In particular, $x_{q^{(j)}}^{*(j)} = 0$ and $\left\|\boldsymbol{W}_j \boldsymbol{x}^{*(j)}\right\|_1 = \sum_{i \notin \left\{q^{(l)}\right\}_{l=1}^{j}} \left|x_{1i}^{(j)}\right| < |x_{1q^{(j)}}^{(j)}| + \sum_{i \notin \left\{q^{(l)}\right\}_{l=1}^{j}} \left|x_{1i}^{(j)}\right| = \left\|\boldsymbol{W}_j \boldsymbol{x}_1^{(j)}\right\|_1$. But this contradicts the definition of $\boldsymbol{x}_1^{(j)}$ by (6) and we conclude that $\boldsymbol{A}_{q^{(j)}}$ is linearly independent of $\left\{\boldsymbol{A}_{q^{(l)}}\right\}_{l=1}^{j-1}$. ■

Lemma 5. *For any* $\boldsymbol{\delta} \in \mathbb{R}^n$ *and* $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ *such that* $\boldsymbol{A}\boldsymbol{\delta} = \boldsymbol{0}$ *and all columns of* $\boldsymbol{A}$ *are unit vectors, the bound*

$$|\delta_i| \leq \frac{\mu}{\mu + 1} \|\boldsymbol{\delta}\|_1,$$

where μ *is the mutual coherence of* $\boldsymbol{A}$, *holds.*

Proof: If $\boldsymbol{A}\boldsymbol{\delta} = \boldsymbol{0}$, then $\left(\boldsymbol{A}^\top \boldsymbol{A} - \boldsymbol{I}_n\right) \boldsymbol{\delta} = -\boldsymbol{\delta}$. Thus, $|\delta_i| \leq \sum_{j=1}^n |\delta_j| \left|\left(\boldsymbol{A}^\top \boldsymbol{A} - \boldsymbol{I}_n\right)_{i,j}\right| \leq \sum_{j=1}^n \mu |\delta_j| - \mu |\delta_i|$. Rearranging the terms yields the sought statement. ■

Definition 2 (taken from [5]). *Let* $\Sigma_k = \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\|_0 \leq k\}$. *The best* k-term approximation error *in terms of* ℓ_p*-norm of a vector* $\boldsymbol{x}$ *is defined as*

$$\sigma_k(\boldsymbol{x})_p = \inf_{\boldsymbol{z} \in \Sigma_k} \|\boldsymbol{x} - \boldsymbol{z}\|_p.$$

B. Proof of Theorem 3

We now give a condensed proof of Theorem 3, while full details can be found in chapter 4 of [13]. The proof is decomposed in two main steps: (i) showing that $q^{(l)} \in I_1$, $\forall l \leq h$, (ii) showing that, after Step (i), the algorithm converges to the unique solution $\boldsymbol{x}_0$ at iteration $h + 1$.

Step (i). First, we will prove by induction that $\forall l \leq h$,

$$\left\|\boldsymbol{x}_0 - \boldsymbol{x}_1^{(l)}\right\|_1 < \frac{\mu + 1}{2\mu} \|\boldsymbol{W}_l \boldsymbol{x}_0\|_\infty, \tag{12}$$

from which we will use Proposition 1 to conclude that $q^{(l)} \in I_1$, $\forall l \leq h$. To prove (12), we follow a path similar to that of the proof of Theorem 1 in [5]. We define the sets $T_q^{(j)} = \left\{q^{(i)}\right\}_{i=1}^{j-1}$, $T_s^{(j)} = \{1, \dots, n\} \backslash T_q^{(j)}$, $T^{(j)}$ as the set of index of k entries of $\boldsymbol{x}_0$ with largest magnitude such that $T^{(j)} \cap T_q^{(j)} = \emptyset$ and $T_r^{(j)} = \{1, \dots, n\} \setminus \left\{T^{(j)} \cup T_q^{(j)}\right\}$. Let $\boldsymbol{\delta}^{(j)} = \boldsymbol{x}_0 - \boldsymbol{x}_1^{(j)}$. We have $\left\|\boldsymbol{\delta}^{(l)}\right\|_1 = \left\|\boldsymbol{\delta}_{T^{(l)}}^{(l)}\right\|_1 + \left\|\boldsymbol{\delta}_{T_q^{(l)}}^{(l)}\right\|_1 + \left\|\boldsymbol{\delta}_{T_r^{(l)}}^{(l)}\right\|_1$, where the first term on the right-hand side (RHS) can be bounded as follows. By the fact that $\boldsymbol{A}$ satisfies the NSP (Definition 1) and that $\boldsymbol{\delta}^{(l)} \in \mathrm{Ker}(\boldsymbol{A})$, the definitions of $T^{(l)}$ and $T_r^{(l)}$ imply

$$\left\|\boldsymbol{\delta}_{T^{(l)}}^{(l)}\right\|_1 \leq \gamma \left(\left\|\boldsymbol{\delta}_{T_q^{(l)}}^{(l)}\right\|_1 + \left\|\boldsymbol{\delta}_{T_r^{(l)}}^{(l)}\right\|_1\right). \tag{13}$$

Therefore, $\left\|\boldsymbol{\delta}^{(l)}\right\|_1$ is bounded by

$$\left\|\boldsymbol{\delta}^{(l)}\right\|_1 \leq (1 + \gamma) \left(\left\|\boldsymbol{\delta}_{T_q^{(l)}}^{(l)}\right\|_1 + \left\|\boldsymbol{\delta}_{T_r^{(l)}}^{(l)}\right\|_1\right). \tag{14}$$

Now, we bound the second term of the RHS of (14) as follows. For $l \geq 2$, we have

$$\begin{aligned}
&\|\boldsymbol{x}_{0T^{(l)}}\|_1 + \left\|\boldsymbol{x}_{0T_q^{(l)}}\right\|_1 + \left\|\boldsymbol{x}_{0T_r^{(l)}}\right\|_1 = \|\boldsymbol{x}_0\|_1 \geq \left\|\boldsymbol{x}_1^{(1)}\right\|_1 \\
&\geq \left\|\boldsymbol{x}_{1T_s^{(2)}}^{(2)}\right\|_1 + \left|x_{1q^{(1)}}^{(1)}\right| \geq \left\|\boldsymbol{x}_{1T_s^{(3)}}^{(3)}\right\|_1 + \left|x_{1q^{(1)}}^{(1)}\right| + \left|x_{1q^{(2)}}^{(2)}\right| \geq \cdots \\
&\geq \left\|\boldsymbol{x}_{1T_s^{(l)}}^{(l)}\right\|_1 + \sum_{i=1}^{l-1} \left|x_{1q^{(i)}}^{(i)}\right| = \left\|\boldsymbol{x}_{1T^{(l)}}^{(l)}\right\|_1 + \left\|\boldsymbol{x}_{1T_r^{(l)}}^{(l)}\right\|_1 + \sum_{i=1}^{l-1} \left|x_{1q^{(i)}}^{(i)}\right|.
\end{aligned}$$

Then, by using the triangle inequalities, $\left\|\boldsymbol{x}_{1T^{(l)}}^{(l)}\right\|_1 \geq \|\boldsymbol{x}_{0T^{(l)}}\|_1 - \left\|\boldsymbol{\delta}_{T^{(l)}}^{(l)}\right\|_1$ and $\left\|\boldsymbol{x}_{1T_r^{(l)}}^{(l)}\right\|_1 \geq \left\|\boldsymbol{\delta}_{T_r^{(l)}}^{(l)}\right\|_1 - \left\|\boldsymbol{x}_{0T_r^{(l)}}\right\|_1$, and by keeping the term with $\boldsymbol{\delta}_{T_r^{(l)}}^{(l)}$ on the LHS, we have

$$\left\|\boldsymbol{\delta}_{T_r^{(l)}}^{(l)}\right\|_1 \leq \left\|\boldsymbol{\delta}_{T^{(l)}}^{(l)}\right\|_1 + 2\left\|\boldsymbol{x}_{0T_r^{(l)}}\right\|_1 + \left\|\boldsymbol{x}_{0T_q^{(l)}}\right\|_1 - \sum_{i=1}^{l-1}\left|x_{1q^{(i)}}^{(i)}\right|. \tag{15}$$

The second term of the RHS of (15) can be computed as a best k-term approximation error (see Definition 2), i.e.,

$$\left\|\boldsymbol{x}_{0T_r^{(l)}}\right\|_1 = \sigma_k\left(\boldsymbol{W}_l\boldsymbol{x}_0\right)_1. \tag{16}$$

For the two last terms on the RHS of (15), the triangle inequality yields $\left\|\boldsymbol{x}_{0T_q^{(l)}}\right\|_1 - \sum_{i=1}^{l-1}\left|x_{1q^{(i)}}^{(i)}\right| \leq \sum_{i=1}^{l-1}\left|\delta_{q^{(i)}}^{(i)}\right|$. Then, introducing this inequality with (13) and (16) in (15) gives, by keeping all the terms with $\boldsymbol{\delta}_{T_r^{(l)}}^{(l)}$ on the LHS,

$$\left\|\boldsymbol{\delta}_{T_r^{(l)}}^{(l)}\right\|_1 \leq \frac{1}{1-\gamma}\left(\gamma\left\|\boldsymbol{\delta}_{T_q^{(l)}}^{(l)}\right\|_1 + 2\sigma_k\left(\boldsymbol{W}_l\boldsymbol{x}_0\right)_1 + \sum_{i=1}^{l-1}\left|\delta_{q^{(i)}}^{(i)}\right|\right). \tag{17}$$

Thus, introducing (17) in (14) leads to

$$\left\|\boldsymbol{\delta}^{(l)}\right\|_1 \leq \frac{1+\gamma}{1-\gamma}\left(2\sigma_k\left(\boldsymbol{W}_l\boldsymbol{x}_0\right)_1 + \sum_{i=1}^{l-1}\left|\delta_{q^{(i)}}^{(i)}\right| + \left\|\boldsymbol{\delta}_{T_q^{(l)}}^{(l)}\right\|_1\right). \tag{18}$$

By applying Lemma 5 to $\boldsymbol{\delta}^{(j)}, \forall j \in \{1, \ldots, l\}$, we have $\forall i \in \{1, \ldots, n\}$, $\left|\delta_i^{(j)}\right| \leq \frac{\mu}{\mu+1}\|\boldsymbol{\delta}^{(j)}\|_1$. With $j = i, i = 1 \ldots l-1$, this gives $\sum_{i=1}^{l-1}\left|\delta_{q^{(i)}}^{(i)}\right| \leq \frac{\mu}{\mu+1}\sum_{i=1}^{l-1}\left\|\boldsymbol{\delta}^{(i)}\right\|_1$ and, with $j = l$, $\left\|\boldsymbol{\delta}_{T_q^{(l)}}^{(l)}\right\|_1 \leq (l-1)\frac{\mu}{\mu+1}\left\|\boldsymbol{\delta}^{(l)}\right\|_1$. On the other hand, the condition (11) leads to

$$\frac{1+\gamma}{1-\gamma} < \frac{\mu+1}{4h\mu}. \tag{19}$$

Then, using these bounds in (18) leads to

$$\left\|\boldsymbol{\delta}^{(l)}\right\|_1 \leq \frac{2(\mu+1)}{\mu\left(4h-l+1\right)}\sigma_k\left(\boldsymbol{W}_l\boldsymbol{x}_0\right)_1 + \frac{1}{4h-l+1}\sum_{i=1}^{l-1}\left\|\boldsymbol{\delta}^{(i)}\right\|_1. \tag{20}$$

Now, we prove by induction that the inequality (12) holds $\forall l \in \{1, \ldots, h\}$. Let denote $\bar{x}_{0j}(i)$ the ith largest absolute value of $\boldsymbol{W}_j\boldsymbol{x}_0$. For $l = 1$, Theorem 1 in [5] yields

$$\left\|\boldsymbol{x}_0 - \boldsymbol{x}_1^{(1)}\right\|_1 \leq \frac{2(1+\gamma)}{1-\gamma}\sigma_k(\boldsymbol{x}_0)_1. \tag{21}$$

Since $\sigma_k(\boldsymbol{x}_0)_1 = \sigma_k(\boldsymbol{W}_1\boldsymbol{x}_0)_1 \leq h\bar{x}_{01}(k+1) \leq h\bar{x}_{01}(1) = h\left\|\boldsymbol{W}_1\boldsymbol{x}_0\right\|_\infty$, and (19), the result (21) leads to (12) with $l = 1$.

Now, assume that (12) is true until $l-1$ with $l \geq 2$. To prove that it is true for l, we need to bound the sum $\sum_{i=1}^{l-1}\left\|\boldsymbol{\delta}^{(i)}\right\|_1$ involved in (20). We can show, as detailed in [13], that each term in the sum is bounded by

$$\left\|\boldsymbol{\delta}^{(j)}\right\|_1 < \frac{\mu+1}{2\mu}\left\|\boldsymbol{W}_l\boldsymbol{x}_0\right\|_\infty, \forall j \in \{1, \ldots, l-1\} \tag{22}$$

via (20) and the fact that, for $j < l \leq h \leq k$,

$$\sigma_k\left(\boldsymbol{W}_j\boldsymbol{x}_0\right)_1 \leq (h-j+1)\left\|\boldsymbol{W}_l\boldsymbol{x}_0\right\|_\infty. \tag{23}$$

In addition, similar steps can be used to complete the induction on l and prove that (12) holds $\forall l \in \{1, \ldots, h\}$. Then, by using Proposition 1 we conclude that $q^{(l)} \in I_1, \forall l \leq h$ with $h \leq k$.

Step 2. Now, we prove that in the $(h+1)$th iteration, solving (6) yields $\boldsymbol{x}_0$. By using Lemma 1, $\boldsymbol{x}_0$ uniquely solves problem (6) if for all nonzero $\boldsymbol{\delta} \in \mathrm{Ker}(\boldsymbol{A})$, the following condition holds:

$$\sum_{i\in I_1} w_i^{(h+1)}|\delta_i| < \sum_{i\in I_0} w_i^{(h+1)}|\delta_i|. \tag{24}$$

Indeed, the LHS of (24) can be rewritten as

$$\sum_{i\in I_1\setminus T_q^{(h+1)}} |\delta_i| < \gamma\sum_{i\in I_0}|\delta_i| + \gamma\sum_{i\in T_q^{(h+1)}}|\delta_i| \tag{25}$$

since $w_i^{(h+1)} = 0, \forall i \in T_q^{(h+1)}$, and the NSP of order k (Definition 1) is applied with $T = I_1 \setminus T_q^{(h+1)}$ and $|T| = k$. Then, we apply once more the NSP of order k with $T = T_q^{(h+1)}$ and $|T| = h \leq k$ for the RHS of (25) to get $\sum_{i\in I_1\setminus T_q^{(h+1)}}|\delta_i| < \gamma\sum_{i\in I_0}|\delta_i| + \gamma^2\sum_{i\in I_0}|\delta_i| + \gamma^2\sum_{i\in I_1\setminus T_q^{(h+1)}}|\delta_i|$, which we rewrite as

$$\sum_{i\in I_1\setminus T_q^{(h+1)}}|\delta_i| < \frac{\gamma}{1-\gamma}\sum_{i\in I_0}|\delta_i|. \tag{26}$$

Thus, the assumption $\gamma < \frac{1}{2}$ guarantees that $\sum_{i\in I_1\setminus T_q^{(h+1)}}|\delta_i| < \sum_{i\in I_0}|\delta_i|$ and that (24) holds. Hence, we obtain the unique solution $\boldsymbol{x}_0$ in $h+1$ iterations.

REFERENCES

[1] A. Garulli, S. Paoletti, and A. Vicino, "A survey on switched and piecewise affine system identification," in *16th IFAC Symp. on System Identification, Brussels, Belgium*, 2012, pp. 344–355.

[2] L. Bako, "Identification of switched linear systems via sparse optimization," *Automatica*, vol. 47, no. 4, pp. 668–677, 2011.

[3] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, 2008.

[4] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, "Introduction to compressed sensing," in *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.

[5] M. Fornasier and H. Rauhut, "Compressive sensing," in *Handbook of Mathematical Methods in Imaging*, O. Scherzer, Ed., 2011, pp. 187–229.

[6] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[7] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[8] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[9] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.

[10] W. Xu, M. A. Khajehnejad, A. S. Avestimehr, and B. Hassibi, "Breaking through the thresholds: an analysis for iterative reweighted ℓ_1 minimization via the Grassmann angle framework," in *IEEE Int. Conf. on Acoustics Speech and Signal Processing*, 2010, pp. 5498–5501.

[11] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[12] C. Novara, "Sparse identification of nonlinear functions and parametric set membership optimality analysis," *IEEE Transactions on Automatic Control*, vol. 57, no. 12, pp. 3236–3241, 2012.

[13] V. L. Le, "Hybrid dynamical system identification: geometry, sparsity and nonlinearities," Ph.D. dissertation, Université de Lorraine, 2013, http://tel.archives-ouvertes.fr/tel-00874283.