COMPRESSED SENSING OF TIME-VARYING SIGNALS

D. Angelosante and G. B. Giannakis[*]

University of Minnesota
Dept. of ECE
Minneapolis, MN 55455, USA

E. Grossi

Università degli Studi di Cassino
DAEIMI
Cassino, IT 03043, Italy

ABSTRACT

Compressed sensing (CS) lowers the number of measurements required for reconstruction and estimation of signals that are sparse when expanded over a proper basis. Traditional CS approaches deal with time-invariant sparse signals, meaning that, during the measurement process, the signal of interest does not exhibit variations. However, many signals encountered in practice are varying with time as the observation window increases (e.g., video imaging, where the signal is sparse and varies between different frames). The present paper develops CS algorithms for time-varying signals, based on the least-absolute shrinkage and selection operator (Lasso) that has been popular for sparse regression problems. The Lasso here is tailored for smoothing time-varying signals, which are modeled as vector valued discrete time series. Two algorithms are proposed: the Group-Fused Lasso, when the unknown signal support is time-invariant but signal samples are allowed to vary with time; and the Dynamic Lasso, for the general class of signals with time-varying amplitudes and support. Performance of these algorithms is compared with a sparsity-unaware Kalman smoother, a support-aware Kalman smoother, and the standard Lasso which does not account for time variations. The numerical results amply demonstrate the practical merits of the novel CS algorithms.

***Index Terms*—** Compressed Sensing, Lasso, Group Lasso, Fused Lasso, Smoothing, Tracking.

1. INTRODUCTION

Since sparsity is a feature present in many natural phenomena as well as man made signals, estimation under sparsity constraints has been a topic of intense research in the last decades [1, 2]. The increasingly large amount of information to be processed in various applications urged researchers to focus on algorithms of manageable complexity that encourage sparse solutions. Among the various options, penalized regression has shown to be a widely accepted approach for sparse signal estimation [3]. Sparse regression has gained popularity in the engineering community thanks to the theory of compressed sensing (CS) [4, 5, 6]. CS promises to lower the Nyquist rate constraint for signal sampling whenever the signal is sparse or compressible over some basis by solving a convex optimization problem [4]. Since many signals in the real world have a parsimonious representation over some convenient basis, CS theory can be very promising.

In its noiseless formulation, sensing amounts to performing N inner products $y_n = \mathbf{m}'_n \mathbf{s}$ of the signal of interest $\mathbf{s} \in \mathbb{R}^K$, $(\cdot)'$ denotes transpose, with a collection of vectors $\{\mathbf{m}_n\}_{n=1}^N$. The signal vector $\mathbf{s}$ can be always represented as $\mathbf{s} = \mathbf{B}\mathbf{x}$ with respect to a given basis $\{\mathbf{b}_k\}_{k=1}^K$, where $\mathbf{B} := [\mathbf{b}_1 \cdots \mathbf{b}_K]$. Arranging the measurements y_n in a vector $\mathbf{y} \in \mathbb{R}^N$, and the projection vectors $\mathbf{m}'_n$ as rows in a matrix $\mathbf{M} \in \mathbf{R}^{N \times K}$, the observations can be written as $\mathbf{y} = \mathbf{M}\mathbf{s} = \mathbf{M}\mathbf{B}\mathbf{x} = \mathbf{H}\mathbf{x}$ where $\mathbf{H} := \mathbf{M}\mathbf{B}$. The main results of CS asserts that, if the signal has a *parsimonious* representation over the basis $\mathbf{B}$, i.e., the number of the non-zero entries of $\mathbf{x}$ is $S \ll K$, then $\mathbf{s}$ can be reconstructed (with high probability) from $N \geq CS \log (K/S)$ linear observations, where C is a fixed constant, if $\mathbf{H}$ is a random matrix (Gaussian, Bernoulli and Fourier cases have been considered). Equally interesting, this is possible by solving the following convex optimization problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s. t.} \quad \mathbf{y} = \mathbf{H}\mathbf{x}$$

where $\|\cdot\|_i$ denotes the ℓ_i norm. Thus, the number of needed measurements grows logarithmically with K rather than linearly [4, 5]. The results still holds if the signal is $R-$compressible, meaning that its entries decay exponentially with R, a behavior expected in real signals.

Another advantage of CS is its robustness to noisy observations [5]. Indeed, if the observation model is $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$, where $\mathbf{n}$ is a zero-mean noise vector with covariance matrix $\mathrm{E}[\mathbf{n}\mathbf{n}'] = \sigma_n^2 \mathbf{I}_N$, $\mathbf{x}$ can be estimated as the solution of the following problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s. t.} \quad \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 \leq \epsilon \tag{1}$$

where ϵ depends on the noise variance. It is to mention also

[*]Prepared through collaborative participation in the Communications and Networks Consortium sponsored by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

that the problem in (1) is very similar to the least-absolute shrinkage and selection operator (Lasso) [7, 8, 9], which amounts to

$$\min_{\mathbf{x}} \|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2 \quad \text{s. t.} \quad \|\mathbf{x}\|_1 \leq s$$

and both can be subsumed by the following unconstrained convex minimization problem

$$\min_{\mathbf{x}} \left\{ \frac{1}{2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \right\}. \tag{2}$$

The penalty parameter λ tunes the sparsity level, i.e., the larger λ is, the more coefficients will be shrunk to zero. Thus, the choice of λ affects the performance of the estimator, and two strategies are commonly adopted: if the noise variance is known, a reasonable choice is given in [10]; alternatively, the problem can be solved for various λ's, and the best one can be selected via cross-validation. Lasso has been studied for more than a dozen years and its properties explored, both asymptotically [11, 12, 13] and for finite sample sizes [8].

In general, CS assumes that a certain number of random linear projections of the signal of interest can be obtained. In various applications however, it might not be possible to take sufficient linear measurements of the signal, and observations have to be acquired as time progresses. Clearly, the signal of interest may exhibit variations in time, rendering the standard CS processing ineffective. This happens, for example, in non-parametric spectrum analysis of time series, where the data is often partitioned into segments for statistical reliability, and the frequency amplitudes may fluctuate, or in brain imaging through magneto-encephalography, where several snapshots (measurement vectors) are obtained over a small time period in which there are variations in brain activity [14]. In this paper we adapt some sparse regression and CS notions to address processing of slow time-varying signals. Previous works dealing with CS of time-varying signals are [15, 16] for sequential CS, and [17, 18, 19] for smoothing CS in the presence of multiple measurement vectors (MMV). Here, the signal of interest is represented as a vector valued discrete time series, and, different from [17, 18, 19], advantage is taken from the additional assumption of its slow time-varying nature. To this end, we will use proper modifications of two recent algorithms for sparse regression, namely Group Lasso [20] and Fused Lasso [21], which are briefly described hereafter.

Except for sparsity, a limitation inherent to Lasso is that it does not exploit other properties of the signal. Many researchers have pursued variants of the Lasso to incorporate physical constraints on the signal. In particular, the Group Lasso has been introduced when variables can be a priori grouped, a property that arises in various problems, e.g., analysis of variance (ANOVA) [20]. Suppose that the variables can be arranged in G groups, each of size K_g, so that $\mathbf{x} = [\mathbf{x}_1' \cdots \mathbf{x}_G']'$, with $\sum_{g=1}^{G} K_g = K$. When group selection is needed, one should drive all the variables in one group to zero together. Group Lasso can achieve this objective by solving

$$\min_{\mathbf{x}_1,\ldots,\mathbf{x}_G} \left\| \mathbf{y} - \sum_{g=1}^{G} \mathbf{H}_g \mathbf{x}_g \right\|_2^2 \quad \text{s. t.} \quad \sum_{g=1}^{G} \sqrt{\mathbf{x}_g' \mathbf{P}_g \mathbf{x}_g} \leq s$$

where $\mathbf{H}_g$ is a matrix containing the K_g columns of $\mathbf{H}$ corresponding to the sub-vector $\mathbf{x}_g$, and $\mathbf{P}_g$ are kernel matrices. The simplest instance of the Group Lasso corresponds to the choice $\mathbf{P}_g = \mathbf{I}_{K_g}$, and it reduces to

$$\min_{\mathbf{x}_1,\ldots,\mathbf{x}_G} \left\| \mathbf{y} - \sum_{g=1}^{G} \mathbf{H}_g \mathbf{x}_g \right\|_2^2 \quad \text{s. t.} \quad \sum_{g=1}^{G} \|\mathbf{x}_g\|_2 \leq s.$$

Another characteristic that a sparse signal may posses is local constancy when the regressors are properly arranged. In many problems, e.g., gene classification using microarrays [21], the signal is sparse but exhibits one or more clusters of non-zero elements, and within each cluster the signal amplitudes are similar. However, the position, number, and size of the clusters are unknown. Fused Lasso precisely incorporates this a priori information [21]. Specifically, considering an additional penalty term which takes into account the ℓ_1 norm of the successive differences of the coefficients, the algorithm yields sparse and piecewise constant solutions. The Fused Lasso estimate is obtained as the solution of

$$\begin{aligned} &\min_{\mathbf{x}} \quad \|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2 \\ &\text{s. t.} \quad \|\mathbf{x}\|_1 \leq s_1 \text{ and } \sum_{k=2}^{K} |x_k - x_{k-1}| \leq s_2 \end{aligned}$$

which is a convex but non-differentiable problem. For a fixed pair (s_1, s_2), it can be solved through quadratic programming methods. Smoothness is a feature present in many signals. Since images exhibit spatial smoothness, Fused Lasso can have great potentials in image processing [21]. Indeed, the term $\sum_{k=2}^{K} |x_k - x_{k-1}|$ is often referred to as total variation (TV), and linear regression constraining the TV has been shown to be effective in image restoration [22, Cap. 6].

In what follows, we show how Fused Lasso and Group Lasso processing, if suitable modified and combined, can be used for CS of time-varying signals. The remainder of the paper is organized as follows: Sec. 2 deals with CS of time-varying signals that do not exhibit variations in the support, while in Sec. 3 an algorithm that does not rely on support constancy is proposed. Numerical simulations, testing the reconstruction capabilities of the proposed algorithms against traditional approaches, are presented in Sec. 4. Finally, some discussion and concluding remarks are given in Sec. 5.

2. CS OF TIME-VARYING SIGNALS WITH TIME-INVARIANT SUPPORT

In this section, we first provide a simple model for signal variations, where the signal entries are time-varying and the support is unknown but time-invariant. Subsequently, we introduce the Group-Fused Lasso (GF-Lasso) for CS of this type of signals. It is worth stressing that the dynamic model is introduced here for comparison purposes only, and will be used in the performance assessment: the GF-Lasso neither follows from a Bayesian approach nor it relies on any signal model.

2.1. Signal model

Let $\mathbf{x}_t := [x_{1;t}, \ldots, x_{K;t}]' \in \mathbb{R}^K$ collect the coordinates of the signal $\mathbf{s}_t$ at epoch t, for $t = 1, \ldots, T$. Let $\mathscr{S}_t = \big\{k \in \{1, \ldots, K\} : x_{k;t} \neq 0\big\}$ denote the support of $\mathbf{x}_t$, with cardinality S_t. We assume that the signal is sparse, i.e., $S_t \ll K$; the support is time-invariant, i.e., $\mathscr{S}_t = \mathscr{S}\ \forall t$; and the signal entries are time-varying. The signal is observed through a set of noisy linear measurements $\mathbf{y}_t \in \mathbb{R}^N$ for $t = 1, \ldots, T$,

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{n}_t \tag{3}$$

where $\mathbf{H}_t \in \mathbb{R}^{N \times K}$ are known regression matrices, and $\mathbf{n}_t \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 \mathbf{I}_N)$.

Given the observations $\{\mathbf{y}_t\}_{t=1}^T$, the matrices $\{\mathbf{H}_t\}_{t=1}^T$, and assuming that the signal is sparse and slowly time-varying, with unknown time-invariant support, the problem is to estimate $\{\mathbf{x}_t\}_{t=1}^T$.

2.2. Optimal estimator: Support-aware Kalman smoother

Suppose that a dynamic model is assigned for the evolution of the non-zero signal coefficients. In particular, for a given cardinality S of $\mathscr{S}$, let $\mathbf{C}_t, \mathbf{D}_t \in \mathbb{R}^{S \times S}$, and let $\mathbf{x}_t^{(\mathscr{S})} \in \mathbb{R}^S$ be the vector obtained by selecting the non-zero entries of $\mathbf{x}_t$. Suppose that $\mathbf{x}_t^{(\mathscr{S})}$ evolves according to the following equation

$$\mathbf{x}_t^{(\mathscr{S})} = \mathbf{C}_t \mathbf{x}_{t-1}^{(\mathscr{S})} + \mathbf{D}_t \mathbf{z}_t \tag{4}$$

where $\mathbf{x}_t^{(\mathscr{S})} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\mathbf{z}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_S)$, and $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

In this case, since (3) and (4) describe a Gauss-Markov state-space model [23], the Bayes solution is known to be the support-aware Kalman smoother (SKS), which determines the limiting mean-square error performance of any estimation procedure. Let $\mathbf{H}_t^{(\mathscr{S})} \in \mathbb{R}^{N \times S}$ denote the matrix obtained by selecting the S columns of $\mathbf{H}_t$ dictated by the support of the signal $\mathscr{S}$. Given $\mathscr{S}$, $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\mathbf{C}_t$, $\mathbf{D}_t$, $\mathbf{H}_t^{(\mathscr{S})}$, and σ_n^2, for $t = 1, \ldots, T$, the SKS first performs standard Kalman filtering, evaluating $\widehat{\mathbf{x}}_{t|t} = \mathrm{E}[\mathbf{x}_t | \mathbf{y}_1, \ldots, \mathbf{y}_t]$, followed by a smoothing phase to obtain $\widehat{\mathbf{x}}_{t|T} = \mathrm{E}[\mathbf{x}_t | \mathbf{y}_1, \ldots, \mathbf{y}_T]$; see, e.g., [23].

The SKS requires knowledge of the matrices involved in the dynamic model, and the support of the signal. While the former is sometimes available from the physics of the problem, the latter makes SKS inapplicable. When the support is unknown, one can simply neglect the sparsity, consider $\mathscr{S} = \{1, \ldots, K\}$, and use a (sparsity-unaware) Kalman smoother (UKS).

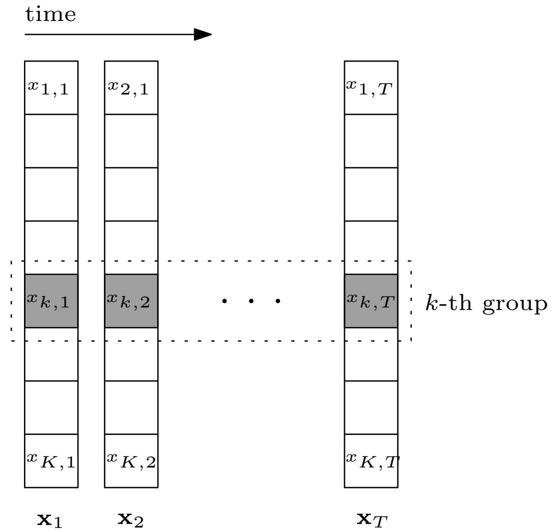

Fig. 1: The time evolution of each entry forms a group if the support is time invariant.

2.3. The GF-Lasso

Consider the coefficients $x_{k;1}, \ldots, x_{k;T}$, which represent the time evolution of the k-th entry of the vectors $\{\mathbf{x}_t\}_{t=1}^T$; see also Fig. 1. It is clear that they form a group, and since the signal support is time-invariant, one should require group selection instead of variable selection, that is drive all the weights in one group to zero together. Moreover, since the signal is supposed to be slowly time-varying, one would expect that the entries of each group do not exhibit abrupt variations during the observation time, i.e., the signal amplitude profile is expected to be smooth. An effective CS algorithm, then, should take into account both of these properties.

Bearing this in mind, we define the order two Group-Fused Lasso (GF2-Lasso) as the solution of the following minimization problem

$$\min_{\mathbf{x}_1, \ldots, \mathbf{x}_T} \quad \sum_{t=1}^{T} \|\boldsymbol{y}_t - \mathbf{H}_t \mathbf{x}_t\|_2^2$$

$$\text{s. t.} \quad \sum_{k=1}^{K} \sqrt{\sum_{t=1}^{T} |x_{k;t}|^2} \leq s_1 \text{ and } \sum_{k=1}^{K} \sqrt{\sum_{t=2}^{T} |x_{k;t} - x_{k;t-1}|^2} \leq s_2.$$

This algorithm penalizes the least-squares error with two terms:

- $\sum_{k=1}^{K} \sqrt{\sum_{t=1}^{T} |x_{k;t}|^2}$, which encourages grouping; the smaller s_1 is, the larger the number of groups (i.e., signal components) shrunk to zero will be; and

- $\sum_{k=1}^{K}\sqrt{\sum_{t=2}^{T}|x_{k;t}-x_{k;t-1}|^2}$, which promotes smoothness for each group of variables; the smaller s_2 is, the smoother the solution will be.

Notice that the second penalty term contains a quadratic smoothing function, which works well when the original signal is very smooth. If some occasional rapid variations in the original signal are to be preserved, then a TV smoothing function, similar to that in the Fused Lasso, can be used. In this case the minimization problem becomes

$$\begin{aligned}&\min_{\mathbf{x}_1,\ldots,\mathbf{x}_T} \quad \sum_{t=1}^{T}\|\boldsymbol{y}_t-\mathbf{H}_t\mathbf{x}_t\|_2^2\\ &\text{s. t.} \quad \sum_{k=1}^{K}\sqrt{\sum_{t=1}^{T}|x_{k;t}|^2}\le s_1 \text{ and } \sum_{k=1}^{K}\sum_{t=2}^{T}|x_{k;t}-x_{k;t-1}|\le s_2\end{aligned}$$

and its solution will be henceforth referred to as order one Group-Fused Lasso (GF1-Lasso).

The constraints in both GF-Lasso are convex and non-differentiable. Nonetheless, for fixed (s_1, s_2), these criteria lead to a quadratic programming problem and algorithms for its solution are available. Finally, we point out that GF-Lasso algorithms are non recursive but treat consecutive observations in a batch form. As a result, the complexity of GF-Lasso increases with time, and if estimation of long time series is sought, GF-Lasso should split observations in batches. When causal estimation of the signal is needed, a competitive recursive algorithm for estimating sparse signal is the R-Lasso introduced in [16].

Remark 2.1. It is worthwhile re-iterating that the GF-Lasso does not rely on any statistical model of the signal dynamics: the parameter s_2 alone controls the time-variations of its estimates. If $s_2 = 0$, the signal is constant (i.e., $\mathbf{x}_1 = \cdots = \mathbf{x}_T$), and GF-Lasso reduces to the standard Lasso, while large values of s_2 allow GF-Lasso to follow fast amplitude variations. In the limiting case $s_2 = \infty$, GF-Lasso coincides with the Group Lasso

$$\min_{\mathbf{x}_1,\ldots,\mathbf{x}_T}\sum_{t=1}^{T}\|\boldsymbol{y}_t-\mathbf{H}_t\mathbf{x}_t\|_2^2 \quad \text{s. t.} \quad \sum_{k=1}^{K}\sqrt{\sum_{t=1}^{T}|x_{k;t}|^2}\le s_1,$$

which is also known in the literature on MMV [18, 19], and will be hereafter referred to as MMV-12. Unfortunately, optimal selection of the pair (s_1, s_2) is still an open issue; in any case, cross-validation techniques can be employed for selecting (s_1, s_2) from a pre-specified grid of values.

3. CS OF GENERAL TIME-VARYING SIGNALS: THE DYNAMIC LASSO

In the previous section, estimation of sparse dynamical processes has been addressed for constant-support signals. Even if this assumption is realistic in certain applications, we propose here an algorithm, referred to as Dynamic Lasso (D-Lasso), that is robust to variations in the signal support.

Paralleling GF2-Lasso, the order 2 Dynamic Lasso (D2-Lasso) estimate is obtained as the solution of

$$\begin{aligned}&\min_{\mathbf{x}_1,\ldots,\mathbf{x}_T} \quad \sum_{t=1}^{T}\|\boldsymbol{y}_t-\mathbf{H}_t\mathbf{x}_t\|_2^2\\ &\text{s. t.} \quad \sum_{k=1}^{K}\sum_{t=1}^{T}|x_{k;t}|\le s_1 \text{ and } \sum_{k=1}^{K}\sqrt{\sum_{t=2}^{T}|x_{k;t}-x_{k;t-1}|^2}\le s_2.\end{aligned}$$

Similar to GF2-Lasso, D2-Lasso penalizes the least-squares error with two terms:

- $\sum_{k=1}^{K}\sum_{t=1}^{T}|x_{k;t}|$, which encourages sparsity, in both time and space; and
- $\sum_{k=1}^{K}\sqrt{\sum_{t=2}^{T}|x_{k;t}-x_{k;t-1}|^2}$, which promotes smoothness across time, for each signal component.

The difference between D2-Lasso and GF2-Lasso resides in the first penalty term, where the ℓ_2 norm of the vectors $[x_{k;1}\cdots x_{k;T}]$ is replaced by the ℓ_1 norm. Indeed, since the sparsity pattern is unknown and time-varying, no group can be a priori defined. D2-Lasso, then, seeks an unstructured sparse pattern with entries smoothly varying across time.

Similarly to GF-Lasso, the quadratic smoothing function will attenuate or remove fast signal amplitude variations. However, since a dynamic support is being considered here, abrupt signal variations should be accounted for, such as the "birth" and "death" of signal entries. For this reason, we introduce the following alternative estimation procedure, referred to as order one Dynamic Lasso (D1-Lasso), as the solution of

$$\begin{aligned}&\min_{\mathbf{x}_1,\ldots,\mathbf{x}_T} \quad \sum_{t=1}^{T}\|\boldsymbol{y}_t-\mathbf{H}_t\mathbf{x}_t\|_2^2\\ &\text{s. t.} \quad \sum_{k=1}^{K}\sum_{t=1}^{T}|x_{k;t}|\le s_1 \text{ and } \sum_{k=1}^{K}\sum_{t=2}^{T}|x_{k;t}-x_{k;t-1}|\le s_2\end{aligned}$$

Although D1-Lasso is reminiscent of Fused Lasso, smoothing here is effected only over time amplitude variations. Since Fused Lasso can produce piecewise constant solutions [21], D1-Lasso is expected to be able to follow abrupt signal variations, such as birth or death of signal components.

Analogous to GF-Lasso, for a given pair (s_1, s_2) D-Lasso is a convex non-differentiable minimization problem, and can be solved through quadratic programming methods. Notice that D-Lasso does not rely on grouping across time, and can be adopted under very general signal conditions. Nevertheless, numerical results will demonstrate that it is an effective technique also for constant-support signals.

Table 1: TSE of GF1-Lasso (constant support).

$s_1 \backslash s_2$	**10**	**11**	**12**	**13**	**14**
32	19.481	21.254	23.254	24.570	24.372
34	9.6286	9.6509	9.6668	9.6958	9.7443
36	3.6652	3.6621	3.6809	3.6986	3.7215
38	2.7733	2.7680	**2.7635**	2.7680	2.7703
40	3.5619	3.5140	3.4735	3.4399	3.4045
42	3.7919	3.7670	3.7436	3.7268	3.7017
44	3.8457	3.8400	3.8278	3.8184	3.8085

Table 2: TSE of GF2-Lasso (constant support).

$s_1 \backslash s_2$	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
30	17.464	17.637	17.937	18.345	18.827
32	9.6780	9.6540	9.7780	9.9960	10.251
34	4.5976	4.4181	4.3647	4.3503	4.5703
36	2.5244	2.3356	**2.2859**	2.3240	2.4425
38	3.2772	2.9508	2.7757	2.6894	2.6561
40	3.3187	2.9849	2.8207	2.8128	2.8911
42	3.3187	2.9849	2.8207	2.8128	2.8910

Table 3: TSE of D1-Lasso (constant support).

$s_1 \backslash s_2$	**1**	**2**	**3**	**4**	**5**
246	4.9006	3.9303	4.4268	4.7773	5.3748
248	4.6834	3.6091	3.9473	4.6170	5.0399
250	4.4868	3.2836	3.4827	4.2424	4.4482
252	4.3427	**3.0864**	3.3051	3.5260	4.0742
254	4.2886	3.1789	3.6505	3.6609	3.8388
256	4.2719	3.1981	3.5805	3.4350	3.6648
258	4.3081	3.2349	3.4391	3.2455	3.5206

Table 4: TSE of D2-Lasso (constant support).

$s_1 \backslash s_2$	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
250	2.8066	2.6261	2.6212	2.6607	2.8212
252	2.6806	2.4946	2.4370	2.4932	2.6511
254	2.6194	2.4309	2.3396	2.4033	2.5391
256	2.6253	2.3793	**2.3029**	2.3557	2.4727
258	2.7000	2.4150	2.3264	2.3539	2.4438
260	2.7881	2.5097	2.4007	2.3953	2.4597
262	2.9150	2.6480	2.5116	2.4813	2.5266

Remark 3.1. The proposed algorithms can perform CS under very general conditions. In particular, setting $s_2 = 0$, corresponds to forcing time-invariance, and thus performing CS with classical Lasso, as in the GF-Lasso case. On the other hand, $s_2 = \infty$, implies complete time independence in the amplitude variation, and D-Lasso reduces again to a standard Lasso problem

$$\min_{\mathbf{x}_1,\dots,\mathbf{x}_T} \sum_{t=1}^{T} \|\boldsymbol{y}_t - \mathbf{H}_t\mathbf{x}_t\|_2^2 \quad \text{s. t.} \quad \sum_{k=1}^{K}\sum_{t=1}^{T} |x_{k;t}| \leq s_1.$$

This recovery algorithm has been proposed also in the MMV literature [17], and will be hereafter referred to as MMV-11. Selection of s_2 plays a critical role in tracking abrupt changes. If cross-validation is performed, one can, in principle, reconstruct the signal without any information about its time-variations. On the other hand, if approximate models are given, these can be included and help selecting proper values for (s_1, s_2).

4. NUMERICAL SIMULATIONS

The proposed algorithms are tested here under two scenarios, namely time-varying and time-invariant signal support, in terms of the total square-error, TSE $:= \sum_{t=1}^{T} \|\mathbf{x}_t - \widehat{\mathbf{x}}_t\|_2^2$.

For comparison purposes, the standard Lasso (that neglects the time variation of the signal) has been also considered, i.e.

$$\widehat{\mathbf{x}}_1 = \cdots = \widehat{\mathbf{x}}_T = \arg\min_{\mathbf{x}} \left\{ \frac{1}{2} \sum_{t=1}^{T} \|\mathbf{y}_t - \mathbf{H}_t\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1 \right\}$$

with λ chosen within a grid around $\sqrt{2\sigma_n^2 \log(NT)}$, value suggested in [10]. As pointed out in Remarks 2.1 and 3.1, this solution corresponds to both GF-Lasso and D-Lasso with $s_2 = 0$.[1] All convex optimization problems have been solved by SeDuMi [24] interfaced with Yalmip [25].

Test Case 1. The signal of interest is assumed to be a constant-support vector with time-varying entries, with $K = 10$ and $S = 3$ (without loss of generality, $\mathscr{S} = \{1, 2, 3\}$). The signal model is that of Eqs. (3) and (4): a Gauss-Markov process with parameters $\boldsymbol{\mu} = \mathbf{0}_S$, $\mathbf{\Sigma} = \mathbf{I}_S$, $\mathbf{C}_t = \beta\mathbf{I}_S$, $\mathbf{D}_t = \sqrt{1-\beta^2}\mathbf{I}_S$, and $\beta = 0.99$ is used to capture the signal variations; regarding the measurement process, a single linear observation of the signal is taken per time ($N = 1$), $\sigma_n^2 = 10^{-2}$, and the entries of $\mathbf{H}_t$ are i.i.d. zero-mean unit-variance Gaussian variates. The observation period is $T = 50$.

[1] In principle, T independent batch-Lassos can be used as a term of comparison, i.e. $\widehat{x}_t = \arg\min_{\mathbf{x}} \left\{ \frac{1}{2}\|\mathbf{y}_t - \mathbf{H}_t\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1 \right\}$. However, in the model further considered, only one measurement per time is taken ($N = 1$), and this alternative solution would perform very poorly.

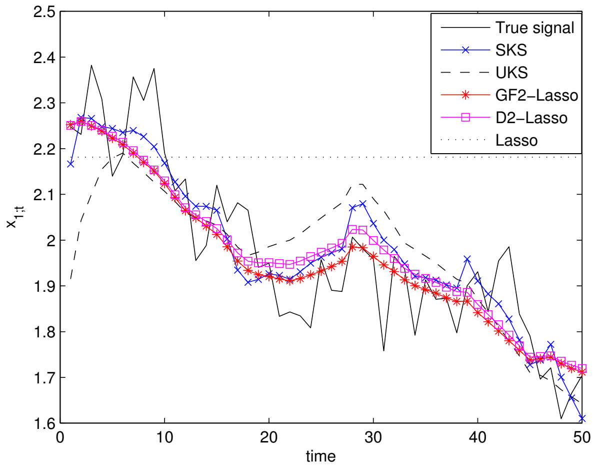

Fig. 2: First active component (constant support).

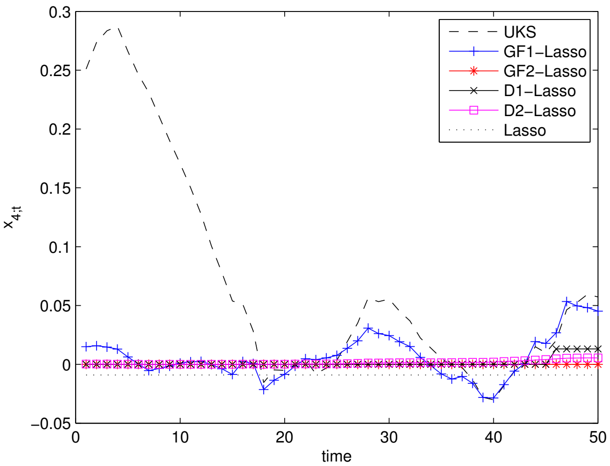

Fig. 3: Fourth component assumed inactive.

Tables 1 and 2 show the TSE of GF-Lasso for various (s_1, s_2) pair, while Tables 3 and 4 that of the D-Lasso (due to space constraints, only a subset of values of (s_1, s_2) around the minimum TSE has been reported in each table). For this model, the TSE of SKS, which is Bayes optimum, is 1.6763, and that of UKS 8.1373. Inspecting Table 1 and 2, it can be seen the minimum TSE of GF1-Lasso is 2.7635, while for GF2-Lasso 2.2859. As concerns D-Lasso, from Tables 3 and 4, the minimum TSE is 3.0864 and 2.3029 for D1-Lasso and D2-Lasso, respectively. As expected, GF-Lasso, which takes into account support constancy, outperforms D-Lasso, and both surpass UKS for a wide range of penalty parameters. The standard Lasso, on the other hand, exhibits disappointing performance, its minimum TSE being 10.289. Notice also that, since the signal amplitude is slowly time-varying, the order two strategies (GF2-Lasso and D2-Lasso) are superior. The MMV-11 and MMV-12 algorithms have been also implemented, and the minimum TSE achieved is 406.64 and 395.37, respectively. These values are comparable with that of the plain LS estimator, which is 407.66.[2] The poor performance of MMV-11 and MMV-12 comes from not exploiting the slowly time-varying nature of the signal.

Fig. 2 depicts the GF2-Lasso, D2-Lasso, SKS, UKS, and standard Lasso estimates of the first (active) signal component. The pair (s_1, s_2) achieving minimum TSE (taken from Tables 1 and 3) have been used for GF2-Lasso and D2-Lasso (these values can be considered close to those given by cross validation). The advantage of GF2 and D2-Lasso over competing alternatives is evident. In particular, the standard Lasso algorithm neglects the signal variation, and its estimates are simply time-averages of the true signal. Fig. 3 shows the estimates of the fourth signal component, which is inactive. All the proposed algorithms except GF1-Lasso estimate very low signal components. But even the GF1-Lasso estimates are smaller than those of the UKS. Finally, Fig. 4 compares GF-Lassos and D-Lassos. Notice that, as expected, the quadratic smoothing penalty produces wiggler solutions than the TV. Furthermore D1-Lasso provides a piecewise constant approximation of the signal of interest. This simple simulated example, then, demonstrates that if GF-Lasso or D-Lasso are used, only $TN = 50$ random linear projections suffice to effectively estimate the variations of a $K = 10$ dimensional signal over $T = 50$ time instants.

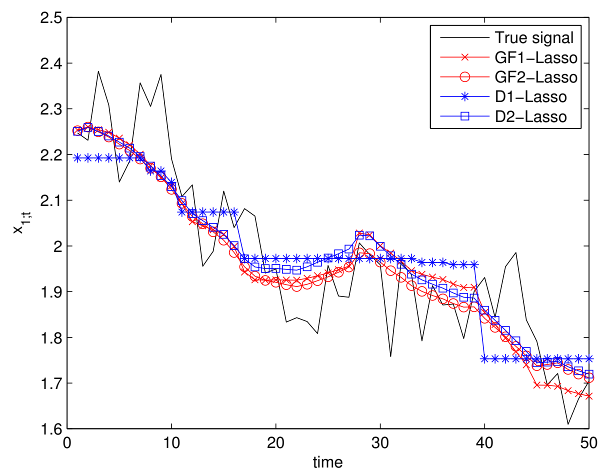

Fig. 4: First active component (constant support).

Test Case 2. We assume a dynamic signal model similar to the Test Case 1, allowing this time the birth and death of some signal components. In particular, the first component starts

[2]The LS estimator is $[\hat{\mathbf{x}}'_1 \cdots \hat{\mathbf{x}}'_T]' := \boldsymbol{H}^{\dagger}\boldsymbol{y}$, where $(\cdot)^{\dagger}$ denotes pseudo-inverse, $\boldsymbol{y} := [\mathbf{y}'_1 \cdots \mathbf{y}'_T]'$, and $\boldsymbol{H}$ is a block diagonal matrix built from $\mathbf{H}_1, \ldots, \mathbf{H}_T$.

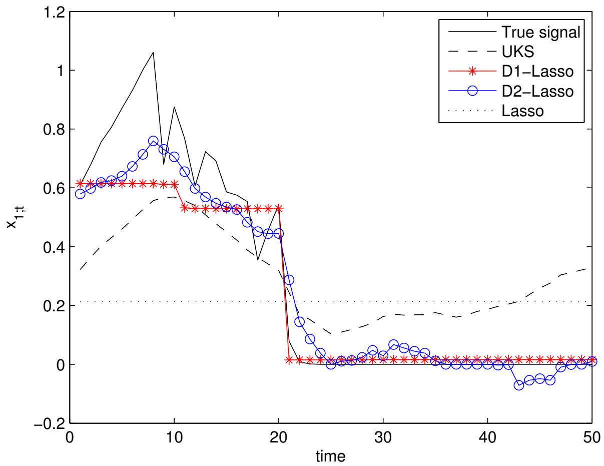

Fig. 5: First active component (time-varying support).

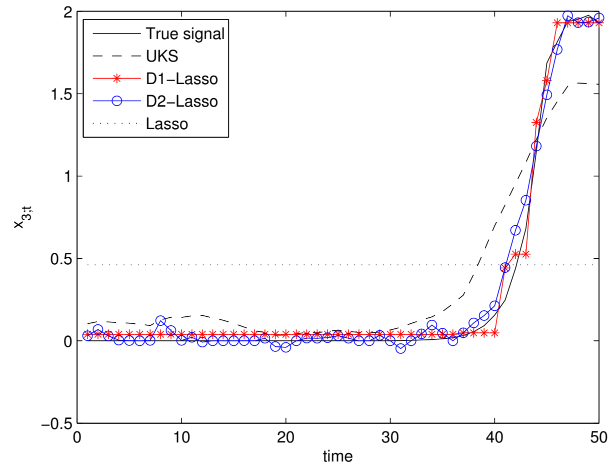

Fig. 6: Third active component (time-varying support).

decaying at time $t = 20$ and, for $t \geq 25$, it is equal to zero; the second component is active for the entire time interval; and the third one becomes active from time $t = 30$ on. Since the support is time varying, SKS and GF-Lasso are not applicable. The TSE of UKS is 17.1624 while that of the standard Lasso is 44.144. The tables listing the TSE for the D-Lassos have not been reported due to lack of space, however, they illustrate that D1 and D2-Lasso exhibit a much lower TSE for a wide range of (s_1, s_2), the minimum being 2.4656 and 1.7844 for D1-Lasso and D2-Lasso, respectively. Notice that, again, D2-Lasso outperforms D1-Lasso.

Figs. 5 and 6 show the estimates of the first and third components of the signal, respectively. Inspecting Fig. 5, it can be seen that UKS has an inertia in setting to zero the component that becomes inactive, while D-Lasso is able to follow abrupt changes, D1-Lasso being faster than D2-Lasso. Thus, D1-Lasso should be preferred over D2-Lasso when abrupt changes are known to occur. Fig. 6, on the other hand, shows that UKS exhibits a hysteresis when inactive components suddenly become active. This inertia causes UKS estimates to have large amplitude even when the signal is actually zero. Finally, Fig. 7 shows the fourth estimated component, which is inactive. While D-Lassos are able to set this component to zero, the standard Lasso and UKS are not.

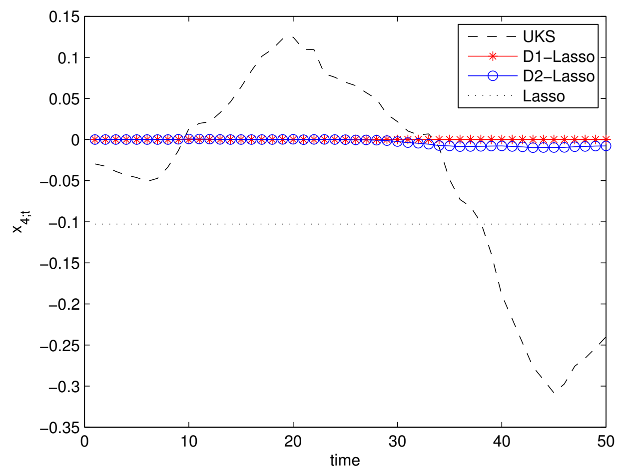

Fig. 7: Fourth component assumed inactive.

5. DISCUSSION AND CONCLUDING REMARKS

In this paper, the CS of time-varying signals has been considered and analyzed. Unlike existing studies that have dealt with reconstruction of static signals, here we have proposed two algorithms for estimating a sparse time-varying signal from its linear and noisy projections: the GF-Lasso for signals with constant support and time-varying amplitudes; and the D-Lasso for problems where support constancy cannot be granted. These algorithms have been tested, and simulation results have unquestionably demonstrated that they outperform sparsity-unaware competitors, and approach the performance of support-aware estimators. In particular, D-Lasso was found to be robust to abrupt changes in the support, where other sparsity unaware estimators exhibited an hysteresis phenomenon. The main difference/advantage of these algorithms is that they do not rely on any information on signal statistics, as other procedures do (e.g., [15]). Furthermore, the additional penalty term on the amplitude profile demonstrated to be very effective for CS of slowly time-varying signals. In particular, for this type of signals, the proposed algorithms perform significantly better than those proposed for sparse recovery with MMVs [17, 18, 19], which do not take into account the slow amplitude variations. CS for time-varying signals is a fertile ground for both analytical and application-oriented future research directions. GF-Lasso and D-Lasso have shown to be effective for a given choice of tuning parameters, whose selection is non-trivial and may require computationally intensive cross-validations. Algorithms that approximate their solution path may also have considerable impact in reducing the computational requirements for signal reconstruction.

6. REFERENCES

[1] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. 19, no. 6, pp. 716–723, Dec. 1974.

[2] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, 1978.

[3] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *J. Amer. Stat. Ass.*, vol. 96, no. 456, pp. 1348–1360, Dec. 2001.

[4] E. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[5] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math*, vol. 59, no. 8, pp. 1207–1223, 2006.

[6] E. Candès and T. Tao, "The Dantzig selector: statistical estimation when p is much larger than n," *Ann. Statist.*, vol. 35, no. 6, pp. 2313–2351, 2007.

[7] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal Statist. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.

[8] E. Candès and Y. Plan, "Near-ideal model selection by ℓ_1 minimization," *Ann. Statist.*, to be published.

[9] Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Mag.*, vol. 25, pp. 21–30, Mar. 2008.

[10] S. S.Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[11] K. Knight and W. J. Fu, "Asymptotics for lasso-type estimators," *Ann. Statist.*, vol. 28, no. 5, pp. 1356–1378, 2000.

[12] P. Zhao and B. Yu, "On model selection consistency of lasso," *J. Mach. Learn. Res.*, vol. 7, pp. 2541–2563, 2006.

[13] H. Zou, "The adaptive lasso and its oracle properties," *J. Amer. Stat. Ass.*, vol. 101, pp. 1418–1429, Dec. 2006.

[14] J. W. Phillips, R. M. Leahy, and J. C. Mosher, "Meg-based imaging of focal neuronal current sources," *IEEE Trans. Med. Imag.*, vol. 16, no. 3, pp. 338–348, Mar. 1997.

[15] N. Vaswani, "Kalman filtered compressed sensing," in *Proc. IEEE Intl. Conf. Image Proc.*, San Diego, CA, USA, Oct. 2008.

[16] D. Angelosante and G. B. Giannakis, "Rls-weighted lasso for adaptive estimation of sparse signals," in *Proc. IEEE Intl. Conf. Acous. Speech. Sig. Proc.*, Taipei, Taiwan, Apr. 2009.

[17] J. Troop, "Algorithms for simultaneous sparse approximation. part II: Convex relaxation," *Signal Processing*, vol. 86, pp. 589–602, Mar. 2006.

[18] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Processing*, vol. 54, pp. 4634–4643, Dec. 2006.

[19] S. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Processing*, vol. 53, pp. 2477–2488, July 2005.

[20] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Royal Statist. Soc. Ser. B*, vol. 68, no. 1, pp. 49–67, 2006.

[21] R. Tibshirani, M. Saunders, S. Rosset, and K. Knight J. Zhu, "Sparsity and smoothness via the fused lasso," *J. Royal Statist. Soc. Ser. B*, vol. 61, no. 1, pp. 91–108, 2005.

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

[23] Y. Bar-Shalom and Xiao-Rong Li, *Estimation and Tracking: Principles, Techniques and Software*, Artech House, 1993.

[24] J. Sturm, "Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones," *Optim. Meth. Softw.*, vol. 11, no. 12, pp. 625–653, 1999.

[25] J. Löfberg, "Yalmip: Software for solving convex (and nonconvex) optimization problems," in *Proc. Amer. Contr. Conf.*, Minneapolis, MN, USA, June 2006.