PENALIZED L1 MINIMIZATION FOR RECONSTRUCTION OF TIME-VARYING SPARSE SIGNALS

Wei Chen†  Miguel R. D. Rodrigues⋆  Ian J. Wassell†

† Computer Laboratory, University of Cambridge, Cambridge CB3 0FD, UK
⋆ Instituto de Telecomunicações, Department of Computer Science, University of Porto

ABSTRACT

In this paper, we propose a penalized $\ell_1$ minimization algorithm for reconstructing a time-varying signal based on compressive sensing (CS) principles. The time-varying signal can be seen as a sequence of slow-changing frames. In the proposed algorithm, all frames of the sequence are sampled at an equal rate, which makes the encoder simpler than frame-categorized methods. We introduce a specialized Fréchet mean of the target frame and several adjacent frames as the penalty vector to make the algorithm close to $\ell_0$ minimization. We prove that the specialized Fréchet mean is a good approximation of the target frame for a sequence of slow time-varying signals. Experimental results demonstrate the superior reconstruction quality of the proposed algorithm.

1. INTRODUCTION

Most practical signals of interest have a sparse representation if they are transformed into a suitable basis. The transformed signal can then be compressed, but this wastes resources as most of the sampled information is discarded after compression. Compressive sensing (CS) leverages the compressibility of the signal by directly acquiring a smaller quantity of random linear measurements that contain a little redundancy in the information level. Thus, CS appears to be an excellent approach for applications in which data acquisition is expensive such as imaging at non-visible wavelengths and sampling made by wireless sensor nodes.

Recently, CS has been studied for recovering a sequence of time-varying sparse signals from linear measurement vectors. This problem arises in the application of environment monitoring by a wireless sensor network using CS video, where the images are taken by a CS camera, for example, a single pixel camera [1] or a random lens imager [2]. There are also related scenarios such as real-time magnetic resonance imaging and channel equalization in communications [3]. Several reconstruction approaches have been proposed for solving the problem of recovering time-varying signals. One natural method involves jointly sampling the sequence and then recovering the sequence at once using higher dimension transformation [4]. This method leads to a large increase in the computational complexity. An alternative method is sampling each frame independently and then recovering the sequence frame by frame. Side information obtained from the first frame, the prior frame or the key frame, can be used to decrease the number of measurements required for recovering the following frames [5–8] or reduce the reconstruction time [3]. However, this method requires the encoder to have the ability to differentiate the key frames in the sequence.

Motivated by the fact that fewer measurements are needed for exact reconstruction with $\ell_0$ minimization than with $\ell_1$ minimization [9], we propose a penalized CS framework where a prior estimation of the sparse representations is used to make $\ell_1$ minimization close to $\ell_0$ minimization. The proposed solution differs from the other side information based methods mentioned previously in two aspects. First, although all those methods use side information to improve the performance, the first frame and the key frame in our solution are not necessarily identified and sampled more than the other frames. Second, those methods update the side information using the knowledge of the key frame or the reconstruction result of the previous frame. However, we derive the side information from several adjacent frames. Our penalized CS framework can be viewed as being similar to the updating step of the iteratively reweighted $\ell_1$ minimization algorithm (IRL1) [9] where a weight vector is defined by the outcome of the previous iteration and is updated in each iteration. We propose to use a specialized Fréchet mean of several adjacent frames as the penalty vector, and our method only needs to solve the $\ell_1$ minimization once for each frame.

2. PENALIZED $\ell_1$ MINIMIZATION USING PRIOR ESTIMATION

Suppose that we observe $T$ measurement vectors of a sequence of time-varying sparse signals, which can be written as

$$y_t = \Phi_t f_t + e_t, \quad (1)$$

where $t = 1, \ldots, T$ is the time index of each frame, $f_t \in \mathbb{R}^N$ is the $t$th signal vector, $y_t \in \mathbb{R}^M (M \ll N)$ is the $t$th measurement vector, and $e_t$ is the noise. This work was supported by Fundação para a Ciência e a Tecnologia through the research project PDTC/EEA-TEL/100854/2008.
measurement vector, $\Phi_t \in \mathbb{R}^{M \times N}$ is a matrix and $e_t \in \mathbb{R}^M$ is a noise vector. We assume each signal can be transformed to a sparse version $x_t \in \mathbb{R}^N$ of the form $x_t = \Psi^H f_t$, where $\Psi \in \mathbb{R}^{N \times N}$ is the orthonormal sparsifying matrix. The vector $f_t$ can then be written in terms of a sparse representation $x_t$ as $f_t = \Psi x_t$. Consequently, the vector of observations $y_t$ can also be written as follows:

$$y_t = \Phi_t \Psi x_t + e_t = A_t x_t + e_t,$$

where $A_t = \Phi_t \Psi$. The classical CS approach to recover $x_t$ involves solving the following $\ell_1$ minimization:

$$\min_{x_t} \| \tilde{x}_t \|_{\ell_1} \quad \text{s.t.} \quad \| A_t \tilde{x}_t - y_t \|_{\ell_2} \leq \epsilon,$$  \hspace{1cm} (3)

where $\epsilon > 0$ relates to an estimate of the noise level.

We assume that the sparse representations change slowly over time during the observing time of the sequence. Reconstructing the sequence of signals independently is not an efficient way as inter-frame correlation is not exploited. With this idea in mind, we propose a penalized CS framework using a prior estimation $\hat{x}_t \in \mathbb{R}^N$ of the representation as a penalty vector in the $\ell_1$ minimization algorithm. Our solution can be expressed as:

$$\min_{x_t} \left\| \frac{\tilde{x}_t}{p_t} \right\|_{\ell_1} \quad \text{s.t.} \quad \| A_t \hat{x}_t - y_t \|_{\ell_2} \leq \epsilon,$$ \hspace{1cm} (4)

where $\tilde{}$ denotes array division and $p_t \in \mathbb{R}^N$ is the penalty of the $t$th frame with elements $p_{t,i} = |\tilde{x}_{t,i}| + \rho$ ($i = 1, \ldots, N$). Here, $\rho$ is a positive parameter to ensure that the algorithm is well-defined.

The motivation of the approach relates to the fact that the $\ell_1$-norm $\| \tilde{x}_t \|_{\ell_1}$ in (4) is a better approximation to the $\ell_0$-norm than the $\ell_1$-norm $\| x_t \|_{\ell_1}$ in (3) is because $\| \tilde{x}_t \|_{\ell_1} \approx \| x_t \|_{\ell_0}$ for a good prior estimate $\hat{x}_t$. Consequently, one expects the algorithm in (4) to outperform the algorithm in (3), in terms of higher estimation quality for a certain number of linear measurements, or, alternatively, a lower number of linear measurements for certain estimation quality. Successful implementation of the penalized CS algorithm depends on the quality of the estimation $\hat{x}_t$ of the actual sparse representation $x_t$.

3. PRIOR ESTIMATION OF TIME-VARYING SPARSE SIGNAL: A FRÉCHET MEAN APPROACH

As the sparse representation changes slowly over time, a natural idea is to use some form of the mean of several adjacent frames as an approximation. However, it appears that we cannot compute the mean vector directly without knowing explicitly the sparse representations. Instead, we propose to use the specialized Fréchet mean as a prior, which can be acquired by solving an ordinary least squares problem.

The Fréchet mean $\bar{x}$ of $K$ adjacent frames $y_k$, $k = 1, \ldots, K$, is defined as follows:

$$\bar{x} = \arg \min_{\hat{x}} \sum_{k=1}^{K} \lambda_k d^2(\hat{x}, y_k, A_k),$$ \hspace{1cm} (5)

where $\lambda_k > 0$ is the contribution weight of the $k$th frame and

$$d(\hat{x}, y_k, A_k) = \| A_k \hat{x} - y_k \|_{\ell_2},$$ \hspace{1cm} (6)

is a Euclidean distance function. In view of equation (2), we can also rewrite the distance function as $d(\hat{x}, y_k, A_k) = \| A_k \hat{x} - A_k x_k - e_k \|_{\ell_2}$. The Fréchet mean minimizes the sum of weighted squared distances between the observations and prediction. The weights allow for the possibility of some frames contributes more than other to the value of the Fréchet mean. For example, consider the reconstruction of some target frame by using a higher weight on the target frame than on other previous frames we expect to acquire a more accurate estimation of the sparse representation of the target frame. This is then used as prior estimate in the previous $\ell_1$ minimization algorithm to improve performance.

We compute the specialized Fréchet mean defined in (5) and (6) by solving the following optimization problem:

$$\hat{x} = \arg \min_{\hat{x}} \left\| \tilde{A} \hat{x} - \tilde{y} \right\|_{\ell_2}^2,$$ \hspace{1cm} (7)

where the extended sensing matrix $\tilde{A}$ and the extended measurement vector are given by $\tilde{A} = [\sqrt{\lambda_1} A_1^H \cdots \sqrt{\lambda_K} A_K^H]^H$ and $\tilde{y} = [\sqrt{\lambda_1} y_1^H \cdots \sqrt{\lambda_K} y_K^H]^H$. If $\text{rank}(\tilde{A}) = N$, (7) turns out to be an ordinary least squares problem, where the solution can be written explicitly as $\hat{x} = (\tilde{A}^H \tilde{A})^{-1} \tilde{A}^H \tilde{y}$ or computed by a conjugate gradient (CG) algorithm [10]. Consequently, for a sequence of independently generated random matrices $A_k$, at least $K = \left\lceil \frac{\delta S}{M} \right\rceil$ frames are needed in the calculation of the prior estimation of a target frame.

3.1. Estimation error

We aim to investigate the error between the target frame estimate given by the specialized Fréchet mean and the real target frame. Before presenting our results, we first give the definition of the restricted isometry property (RIP) [11] which is a widely used tool for analyzing random projections. Formally, a matrix $A$ of size $M \times N$ is said to satisfy the RIP of order $S$ with restricted isometry constant (RIC) $\delta_S \in (0, 1)$ as the smallest number such that

$$(1 - \delta_S)\|x\|_{\ell_2}^2 \leq \|A x\|_{\ell_2}^2 \leq (1 + \delta_S)\|x\|_{\ell_2}^2$$ \hspace{1cm} (8)

holds for all $x$ with $\|x\|_{\ell_0} \leq S$.

Theorem 1 Consider the measurement model $y_k = A_k x_k + e_k$, $k = 1, \ldots, K$, where the matrices $A_k$ have
RIC δS,k. Fix the integers S₁, S₂ and the positive numbers λ₁, k = 1, . . . , K. If ∥xₖ∥ℓ₀ ≤ S₁, k = 1, . . . , K, ∥xₖ₁ − xₖ₂∥ℓ₀ ≤ S₂, k₁, k₂ ∈ {1, . . . , K} and ∥eₖ∥ℓ₂ ≤ ϵ², k ∈ {1, . . . , K}, Then

\[ \| \hat{x} - x_t \|^2_{\ell_2} \leq \sum_{k=1}^{K} C_{1,k} \| x_t - x_k \|^2_{\ell_2} + C_{2,k} \epsilon^2, \; t = 1, . . . , K \]

where Q = S₁ + (K − 1)S₂, C_{1,k} = \frac{\lambda_k(1 + \delta_{S,k})}{\lambda_k(1 - \delta_{Q,t})} and C_{2,k} = \frac{\lambda_k^2}{\lambda_k^2(1 - \delta_{Q,t})}.

**Proof**: As the specialized Fréchet mean \( \hat{x} \) is the point that minimizes the sum of squared distances to all the other points, we have

\[ \sum_{k=1}^{K} \lambda_k \| A_k \hat{x} - y_k \|^2_{\ell_2} \leq \sum_{k=1}^{K} \lambda_k \| A_k x_t - y_k \|^2_{\ell_2}. \]  

Then, according to the RIP (8) and the inequality (10), we can derive

\[
\begin{align*}
\| \hat{x} - x_t \|^2_{\ell_2} &\leq \frac{1}{1 - \delta_{Q,t}} \| A_t (\hat{x} - x_t) \|^2_{\ell_2} \\
&\leq \frac{1}{1 - \delta_{Q,t}} \| A_t \hat{x} - y_t \|^2_{\ell_2} + \frac{\| e_t \|^2_{\ell_2}}{1 - \delta_{Q,t}} \\
&\leq \frac{1}{1 - \delta_{Q,t}} \sum_{k=1}^{K} \lambda_k \| A_k \hat{x} - y_k \|^2_{\ell_2} + \frac{\epsilon^2}{1 - \delta_{Q,t}},
\end{align*}
\]

and

\[
\begin{align*}
\| A_k x_t - y_k \|^2_{\ell_2} &\leq \| A_k (x_t - x_k) \|^2_{\ell_2} + \| e_k \|^2_{\ell_2} \\
&\leq (1 + \delta_{S,k}) \| x_t - x_k \|^2_{\ell_2} + \epsilon^2.
\end{align*}
\]

Then, from the inequality (10), (11) and (12), we can deduce that

\[
\begin{align*}
\| \hat{x} - x_t \|^2_{\ell_2} &\leq \sum_{k=1}^{K} \lambda_k(1 + \delta_{S,k}) \| x_t - x_k \|^2_{\ell_2} + \frac{(\lambda_k + \lambda_k^2) \epsilon^2}{\lambda_k(1 - \delta_{Q,t})} + \frac{\lambda_k^2}{\lambda_k^2(1 - \delta_{Q,t})} \\
&= \sum_{k=1}^{K} C_{1,k} \| x_t - x_k \|^2_{\ell_2} + C_{2,k} \epsilon^2.
\end{align*}
\]

**Remark**: Theorem 1 gives the upper bound of the squared pairwise Euclidean distance between the specialized Fréchet mean and a target frame. The bound is tight in the noiseless case \( \epsilon = 0 \) and all frames considered are same, i.e. \( x_k = x_t \). In this case, the specialized Fréchet mean is equal to the target frame. If the signal changes slowly over time, then the squared estimation error bound is equal to a sum of weighted distances between the target frame and other frames plus a noise term. For example, suppose \( \delta_{S_2} = 0.4, \delta_{Q_2} = 0.5, \lambda_k = 1, \lambda_k = 2, \epsilon^2 \leq 0.001 \| x_k \|^2_{\ell_2} \) and the Euclidean distance between the target frame \( x_k \) and any of the \( K - 1 = 4 \) adjacent frame \( x_k \) satisfies \( \| x_t - x_k \|^2_{\ell_2} \leq 0.01 \| x_k \|^2_{\ell_2}, \) then the Euclidean distance between the the Fréchet Mean and the target frame satisfies \( \| x - x_t \|^2_{\ell_2} \leq 0.06 \| x_t \|^2_{\ell_2} \). Note that the real error will be smaller than the bound given in Theorem 1.

4. ALGORITHM

The pseudo-code for penalized \( \ell_1 \) minimization for a sequence of signals is described as follows:

**Algorithm 1** Reconstruction procedure

**Input**: A time sequence of measurement matrices \( A_t \) (\( t = 1, 2, \ldots \)), a time sequence of measurement vectors \( y_t \) (\( t = 1, 2, \ldots \)), \( K \) positive weights \( \lambda_k \) (\( k = 1, \ldots, K \)), an estimate of the noise level \( \epsilon \) and a positive parameter \( \rho \);

**Output**: An estimate time sequence of signal vector \( x_t \) (\( t = 1, 2, \ldots \)).

**Process**: For \( t > 0 \), do

1. Compute the specialized Fréchet mean \( \hat{x}_t \) by considering \( y_t \) and \( K - 1 \) adjacent measurement vectors with weights \( \lambda_k \);
2. Compute the penalty \( p_t = \| \hat{x}_t \| + \rho \);
3. Compute \( x_t = \arg \min_{\tilde{x}_t} \| \tilde{x}_t \|_{\ell_1} \) s.t. \( \| A_t \tilde{x}_t - y_t \|^2_{\ell_2} \leq \epsilon \).

For the first frame, the adjacent frames could be the next \( K - 1 \) frames. For other frames, their adjacent frames could be former frames or later ones. The weights \( \lambda_k \) (\( k = 1, \ldots, K \)) are required as the input to the algorithm, which can be simply given the same value. An interesting approach is to use different weights to take account of the time-varying nature of the sequence of signals.

![A sequence of crack images](a.png)

(b) Small varying of wavelet coefficients

Fig. 1. A sequence of crack images and the normalized squared Euclidean distances of wavelet coefficients between adjacent frames.
Fig. 2. Reconstructed result of the 8th crack image. (a) Traditional \( \ell_1 \) minimization with \( \frac{M}{N} = 0.3 \); (b) Proposed algorithm with \( \frac{M}{N} = 0.2 \); (c) Proposed algorithm with \( \frac{M}{N} = 0.15 \)

5. EXPERIMENTAL RESULTS

In this section, we test the reconstruction performance of the proposed algorithm with a sequence of crack images, which are taken by a sensor camera for the purpose of monitoring bridge condition. As shown in Fig 1(a), there are 8 frames, each of size \( 32 \times 32 \) pixels. We calculate the Daubechies wavelet "db1" coefficients \( x_t \) of each frame at level 2, and then plot the normalized squared Euclidean distance \( \frac{\| x_t - x_{t+1} \|^2}{\| x_t \|^2} \) between any pair of adjacent frames in Fig 1(b). We note that the sparse representations of the sequence are quite close in the Euclidean space, which verifies the assumption of a slow time-varying sparse signal.

The \( M \times N \) sensing matrices \( \Phi_t \) \( (t = 1, \ldots, 8) \) are generated independently with i.i.d. Gaussian entries \( \mathcal{N}(0, \frac{1}{M}) \). In Fig 2, we compare the proposed algorithm with traditional \( \ell_1 \) minimization. For the reconstruction of the 4th image of the sequence, we let \( K = 8, \lambda_t = 7 \) and \( \lambda_k = 1 \) \( (k = 1, \ldots, 8; k \neq t) \), which makes the target frame have a higher impact than other frames in the calculation of the Fréchet mean. We note in Fig 2 that the proposed penalized \( \ell_1 \) minimization outperforms the traditional CS algorithm by visual comparison. We achieve similar performance for the other images of the sequence which are not shown here due to the limited space.

Fig 3 shows the accuracy of reconstruction using the signal-to-noise ratio (SNR) \( \frac{\| x_t \|_2}{\| x_t - \hat{x}_t \|_2} \), where \( \hat{x}_t \) is the reconstructed vector. In this experiment, the number of measurements are reduced to 15% of the original except for the first frame of the Modified-CS [6], which has 50% measurements. We calculate the SNR by averaging over 200 experiment trials. The proposed method and the Modified-CS have similar performance except for the first frame. The Modified-CS algorithm requires specific notification of the key frames while the proposed method does not. Thus, the proposed method can be applied in situations where the encoder is ignorant concerning the locations of the key frames.

6. CONCLUSIONS

In this paper, we study the problem of reconstructing a time-varying signal with a small number of linear incoherent measurements. We proposed a penalized \( \ell_1 \) minimization algorithm for reconstruction that even uses fewer measurements than does the traditional CS method. We use a specialized Fréchet mean as the penalty vector, which is shown to be close to the original signal. Experiments demonstrate the advantage of our method in the reconstruction of a sequence of images.

7. REFERENCES