RECOVERY OF CORRELATED SPARSE SIGNALS FROM UNDER-SAMPLED MEASUREMENTS

Zhaofu Chen[1], Rafael Molina[2], Aggelos K. Katsaggelos[3]

[1,3] Electrical Engineering and Computer Science Dept., Northwestern University, Evanston, IL, USA
[2] Dept. de Ciencias de la Computación e I. A., Universidad de Granada, Granada, Spain

ABSTRACT

In this paper we consider the problem of recovering temporally smooth or correlated sparse signals from a set of under-sampled measurements. We propose two algorithmic solutions that exploit the signal temporal properties to improve the reconstruction accuracy. The effectiveness of the proposed algorithms is corroborated with experimental results.

Index Terms— Sparse signal recovery, multiple measurement, greedy algorithm, convex relaxation method

1. INTRODUCTION

Consider the measurement system expressed as

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{n}, \tag{1}$$

where the signal of interest $\mathbf{x} \in \mathbb{R}^{N\times 1}$ undergoes a linear transformation $\mathbf{\Phi} \in \mathbb{R}^{M\times N}$ and is corrupted by additive noise $\mathbf{n} \in \mathbb{R}^{M\times 1}$. Depending on the applications, the transformation $\mathbf{\Phi}$ can be a dictionary in sparse signal representation problems [1], a measurement system in compressive sensing [2], a degradation process in image processing [3], a steering matrix in source localization problems [4], etc.

In general $\mathbf{y}$ is an under-sampled measurement, i.e., $M \leq N$. In such cases the recovery of $\mathbf{x}$ from the noisy $\mathbf{y}$ is an ill-posed problem, i.e., there exist infinite number of solutions. A common approach to this problem is to constrain the solution space by incorporating prior knowledge about $\mathbf{x}$. As is well known, most natural signals are sparse either in its native domain (e.g., distributed sources in space) or in certain basis (e.g., image coefficients in the wavelet basis). Without loss of generality, we assume $\mathbf{x}$ is sparse, i.e., $||\mathbf{x}||_0 \ll N$, where $||\cdot||$ denotes the ℓ_0-(pseudo)norm.

Sparse signal recovery with single measurement as modeled by (1) is well-studied. Broadly, three categories of algorithms exist for solving such problems. Greedy algorithms, such as Matching Pursuit [5] and Orthogonal Matching Pursuit [6, 7] seek the "most representative" subset of columns in $\mathbf{\Phi}$ to approximate $\mathbf{y}$. Relaxation-based approaches, such as LASSO [8] and FOCUSS [9], replace the ℓ_0-(pseudo)norm with the sparsity-promoting ℓ_p-(pseudo)norm ($p \leq 1$) and solve a regularized fitting problem. Bayesian approaches adopt sparsity-promoting priors in hierarchical models and employ statistical inference to obtain estimates of $\mathbf{x}$ [10, 11].

When the measurement process occurs at T time instances, the data model in (1) is expanded as

$$\mathbf{Y} = \mathbf{\Phi}\mathbf{X} + \mathbf{N}, \tag{2}$$

where $\mathbf{Y} \in \mathbb{R}^{M\times T}$ contains the measurements $\{\mathbf{y}_{\cdot i}\}_{i=1}^{T}$ as its columns, and $\mathbf{X} \in \mathbb{R}^{N\times T}$ contains the signals $\{\mathbf{x}_{\cdot i}\}_{i=1}^{T}$ as its columns, respectively.

Given the T measurements, various strategies can be applied depending on the dynamics of $\mathbf{X} = [\mathbf{x}_{\cdot 1}, \cdots, \mathbf{x}_{\cdot T}]$. On one end, for static signals $\{\mathbf{x}_{\cdot i}\}$ and independent noise, averaging over the columns of $\mathbf{Y}$ yields $\bar{\mathbf{y}} = \sum_i \mathbf{y}_{\cdot i}/T$ with improved signal-to-noise ratio (SNR), which can then be considered as a single measurement. On the other end, if $\mathbf{x}_{\cdot i}$ is assumed independent of each other, the problem reduces to T single-measurement sparse recovery problems, each of which can be solved separately.

Between these two ends, it is common that $\{\mathbf{x}_{\cdot i}\}$, though non-static, share a fixed sparsity pattern, i.e., the locations of the nonzeros in $\mathbf{x}_{\cdot i}$ do not change. This is known as the Multiple Measurement Vector (MMV) problem, where the goal is to recover a row-wise sparse $\mathbf{X}$ from $\mathbf{Y}$. Algorithms for solving the MMV problem can also be broadly categorized as greedy [12], relaxation-based [13, 14], or Bayesian approaches [15].

While the existing algorithms induce row-sparsity in $\mathbf{X}$, they do not fully exploit the information in $\mathbf{X}$, and in particular, the correlation within the nonzero rows of $\mathbf{X}$. Utilizing such intra-row structure can potentially lead to improved recovery performance. In [16] a Bayesian model with a greedy inference procedure is proposed to achieve row-sparsity in $\mathbf{X}$ while promoting smoothness of the nonzero rows in $\mathbf{X}$. Experimental results have confirmed that such an approach can yield higher recovery accuracy, although at the cost of increased computation.

In this paper, we propose a greedy algorithm and a relaxation-based algorithm to find $\mathbf{X}$ that jointly meets the following criteria:

1. $\mathbf{X}$ fits the measurement $\mathbf{Y}$ well, i.e., $||\mathbf{Y}-\mathbf{\Phi}\mathbf{X}||_{\mathrm{F}}^2$ is small,

This work has been partially supported by a grant from the Department of Energy (DE-NA0000457), the Spanish Ministry of Economy and Competitiveness under project TIN2010-15137, the European Regional Development Fund (FEDER), and the CEI BioTic at the Universidad de Granada.

where $||\cdot||_{\rm F}$ denotes the matrix Frobenius norm.

2. $\mathbf{X}$ is row-wise sparse, i.e., the number of nonzero rows in $\mathbf{X}$ is much smaller than N.
3. Nonzero rows of $\mathbf{X}$ are smooth or correlated.

The rest of the paper is structured as follows. Section 2 presents the principles of the greedy algorithm. Section 3 formulates a least-squares fitting problem with relaxed regularization, and discusses an efficient algorithm to solve it. Numerical examples corroborating the effectiveness of the proposed algorithms are shown in Section 4. Finally, the paper is concluded in Section 5.

Notation: Matrices and vectors are denoted by uppercase and lowercase boldface letters, respectively. For a matrix $\mathbf{X}$, $\mathbf{x}_{i\cdot}$ denotes its $i^{\rm th}$ row, $\mathbf{x}_{\cdot j}$ denotes its $j^{\rm th}$ column, and X_{ij} denotes its $(i,j)^{\rm th}$ element, respectively.

2. GREEDY ALGORITHM

Consider the following optimization problem

$$\begin{array}{l}\min\limits_{\mathbf{X}} \left\{c||\mathbf{Y}-\mathbf{\Phi}\mathbf{X}||_{\rm F}^2+\sum_{i=1}^{N}\mathbf{x}_{i\cdot}\mathbf{P}\mathbf{x}_{i\cdot}^{\rm T}\right\} \\ \text{s.t.: number of nonzero rows in } \mathbf{X}\leq S\end{array}, \tag{3}$$

where $||\mathbf{Y}-\mathbf{\Phi}\mathbf{X}||_{\rm F}^2$ quantifies measurement fidelity, $\mathbf{x}_{i\cdot}\mathbf{P}\mathbf{x}_{i\cdot}^{\rm T}$ models our prior knowledge about the intra-row structure of $\mathbf{X}$, and c is a parameter balancing these two factors in the cost function. The user-defined parameter S determines the level of sparsity in $\mathbf{X}$, as well as serves as a convergence criterion as is presented shortly. The symmetric positive-semidefinite matrix $\mathbf{P}\in\mathbb{R}^{T\times T}$ can be used to impose a wide range of structural properties on the nonzero rows of $\mathbf{X}$. For example, if from prior knowledge $\mathbf{x}_{i\cdot}$ is smooth, $\mathbf{P}$ can be constructed as $\mathbf{P}=\mathbf{D}^{\rm T}\mathbf{D}$, where $\mathbf{D}\in\mathbb{R}^{T\times T}$ with

$$D_{ij}=\begin{cases}-2, & \text{if } i=j\\ 1, & \text{if } |i-j|=1\\ 0, & \text{else}\end{cases} \tag{4}$$

implements a second-order difference operator. Non-smooth $\mathbf{x}_{i\cdot}$ with abrupt variations will result in larger value of $\mathbf{x}_{i\cdot}\mathbf{P}\mathbf{x}_{i\cdot}^{\rm T}=||\mathbf{D}\mathbf{x}_{i\cdot}^{\rm T}||_2^2$ and hence be penalized more than smooth $\mathbf{x}_{i\cdot}$. As another example, if we know the statistical properties, e.g., the correlation among the elements of $\mathbf{x}_{i\cdot}$, $\mathbf{P}$ can be set as proportional to the precision (inverse of covariance) matrix of $\mathbf{x}_{i\cdot}$, which penalizes $\mathbf{x}_{i\cdot}$ that deviates much from the statistical model.

Let $\Omega\subseteq\{1,2,\ldots,N\}$ be a subset with cardinality S, i.e., $|\Omega|=S$. Denote by $\mathbf{\Phi}_{\cdot\Omega}$ the $M\times|\Omega|$ sub-matrix of $\mathbf{\Phi}$ consisting of its columns indexed by Ω. Solving (3) globally requires finding the optimal S columns of $\mathbf{\Phi}$. Since the number of ways to choose S out of N columns grows combinatorially with N, the problem in (3) is NP-hard. Therefore we can only find sub-optimal solutions to it. In this paper we present an iterative greedy algorithm that instead of finding S columns of $\mathbf{\Phi}$ all at once, finds one column at a time and incrementally populates Ω.

The algorithm starts with $\Omega=\emptyset$ and $\Omega^{\rm c}=\{1,2,\ldots,N\}$, and initializes the residual $\mathbf{R}=\mathbf{Y}$. In the iterations the algorithm fits the residual by incrementally selecting columns from $\mathbf{\Phi}$ and optimizing the corresponding coefficients in $\mathbf{X}$.

Specifically, at each iteration, in order to find a locally optimal column of $\mathbf{\Phi}$, the algorithm scans through $i\in\Omega^{\rm c}$ and solves a series of simple local optimization problems

$$p_i=\min_{\mathbf{x}\in\mathbb{R}^{1\times T}}\left\{c||\mathbf{R}-\boldsymbol{\phi}_{\cdot i}\mathbf{x}||_{\rm F}^2+\mathbf{x}\mathbf{P}\mathbf{x}^{\rm T}\right\},\quad i\in\Omega^{\rm c}, \tag{5}$$

where $\boldsymbol{\phi}_{\cdot i}$ is the $i^{\rm th}$ column of $\mathbf{\Phi}$. Setting the gradient of the quadratic cost function in (5) with respect to $\mathbf{x}$ to $\mathbf{0}$, the minimizer $\mathbf{x}^\star$ is found to be

$$\mathbf{x}^\star=\boldsymbol{\phi}_{\cdot i}^{\rm T}\mathbf{R}\left(||\boldsymbol{\phi}_{\cdot i}||_2^2\mathbf{I}+c^{-1}\mathbf{P}\right)^{-1}. \tag{6}$$

After obtaining $\{p_i\}_{i\in\Omega^{\rm c}}$ from (5), the index with the lowest cost is selected, and the sets Ω and $\Omega^{\rm c}$ are updated accordingly.

At the end of an iteration, the contribution from the currently selected $\mathbf{\Phi}_{\cdot\Omega}$ is removed from $\mathbf{R}$ via the following optimization

$$\mathbf{X}_\Omega=\mathop{\arg\min}_{\mathbf{X}\in\mathbb{R}^{|\Omega|\times T}}\left\{c||\mathbf{Y}-\mathbf{\Phi}_{\cdot\Omega}\mathbf{X}||_{\rm F}^2+\sum_{i=1}^{|\Omega|}\mathbf{x}_{i\cdot}\mathbf{P}\mathbf{x}_{i\cdot}^{\rm T}\right\}, \tag{7}$$

which, after the gradient of the cost function with respect to $\mathbf{X}$ is set to zero, leads to the following equation

$$\left(\mathbf{\Phi}_{\cdot\Omega}^{\rm T}\mathbf{\Phi}_{\cdot\Omega}\right)\mathbf{X}_\Omega+c^{-1}\mathbf{X}_\Omega\mathbf{P}=\mathbf{\Phi}_{\cdot\Omega}^{\rm T}\mathbf{Y}. \tag{8}$$

The matrix equation in (8) is known as the Sylvester equation in control theory, and can be solved using the Bartels-Stewart (B-S) algorithm. With $\mathbf{X}_\Omega$, the residual is updated as

$$\mathbf{R}=\mathbf{Y}-\mathbf{\Phi}_{\cdot\Omega}\mathbf{X}_\Omega. \tag{9}$$

This greedy algorithm terminates after S iterations, at which point the solution $\mathbf{X}$ is obtained by setting its Ω-indexed rows to $\mathbf{X}_\Omega$ and the other rows to zero.

The pseudocode is listed in Algorithm 1.

Algorithm 1 The greedy algorithm

Inputs: $\mathbf{Y}\in\mathbb{R}^{M\times T}$, $\mathbf{\Phi}\in\mathbb{R}^{M\times N}$, S, c
Outputs: Ω, $\mathbf{X}\in\mathbb{R}^{N\times T}$
Initialize: $\Omega=\emptyset$, $\Omega^{\rm c}=\{1,2,\ldots,N\}$, $\mathbf{R}=\mathbf{Y}$, $s=1$
while $s\leq S$ **do**
  $i^\star=\mathop{\arg\min}\limits_{i\in\Omega^{\rm c}}\min\limits_{\mathbf{x}\in\mathbb{R}^{1\times T}}\left\{c||\mathbf{R}-\boldsymbol{\phi}_{\cdot i}\mathbf{x}||_{\rm F}^2+\mathbf{x}\mathbf{P}\mathbf{x}^{\rm T}\right\}$
  $\Omega\leftarrow\Omega\cup\{i^\star\}$, $\Omega^{\rm c}\leftarrow\Omega^{\rm c}\setminus\{i^\star\}$
  $\mathbf{X}_\Omega=\mathop{\arg\min}\limits_{\mathbf{X}\in\mathbb{R}^{|\Omega|\times T}}\left\{c||\mathbf{Y}-\mathbf{\Phi}_{\cdot\Omega}\mathbf{X}||_{\rm F}^2+\sum_{i=1}^{|\Omega|}\mathbf{x}_{i\cdot}\mathbf{P}\mathbf{x}_{i\cdot}^{\rm T}\right\}$
  $\mathbf{R}=\mathbf{Y}-\mathbf{\Phi}_{\cdot\Omega}\mathbf{X}_\Omega$
end while
Let $\mathbf{X}$ be an $N\times T$ all-zero matrix.
Set Ω-indexed rows of $\mathbf{X}$ to $\mathbf{X}_\Omega$.

3. RELAXATION-BASED ALGORITHM

In this section we present a relaxation-based approach to find a solution $\mathbf{X}$ that satisfies the criteria set forth in Section 1. Specifically, this approach formulates a regularized

fitting problem, where the regularization term incorporates the properties of $\mathbf{X}$. We solve this regularized fitting problem by adopting the Alternating Directions Method of Multipliers (ADMM) framework.

Consider the following unconstrained optimization

$$\min_{\mathbf{X}\in\mathbb{R}^{N\times T}} \left\{ \frac{c}{2}||\mathbf{Y}-\mathbf{\Phi X}||_{\mathrm{F}}^2 + \sum_{i=1}^{N} ||\mathbf{x}_{i\cdot}||_{\mathbf{P}} \right\}, \qquad (10)$$

where

$$||\mathbf{x}_{i\cdot}||_{\mathbf{P}} = (\mathbf{x}_{i\cdot}\mathbf{P}\mathbf{x}_{i\cdot}^{\mathrm{T}})^{1/2} \qquad (11)$$

denotes the $\mathbf{P}$-weighted norm. Note when $\mathbf{P}=\mathbf{I}$, (11) reduces to the conventional ℓ_2-norm, and the regularization term in (10) reduces to the $\ell_1\ell_2$-norm of $\mathbf{X}$.

As is known, the $\ell_1\ell_2$-norm can be used to find row-wise sparse $\mathbf{X}$ for the under-determined system in (2). However, the $\ell_1\ell_2$-norm does not assume any structural property of the nonzero rows in $\mathbf{X}$. In contrast, the regularization term in (10), i.e., the ℓ_1-norm of the $\mathbf{P}$-weighted ℓ_2-norms, both promotes row-wise sparsity in $\mathbf{X}$ and encourages the nonzero rows of $\mathbf{X}$ to comply with the smoothness or structural prior knowledge.

In order to solve the problem in (10), we employ the ADMM framework. ADMM is a generic primal-dual procedure based on the augmented Lagrangian, and has been successfully applied to a wide range of problems [17].

Since the cost function in (10) is convex in $\mathbf{X}$, in principle we can apply standard algorithms, such as subgradient method, to find a global minimum. Alternatively, we can decouple the two terms in the cost function by introducing an auxiliary primal variable $\mathbf{Z}\in\mathbb{R}^{N\times T}$ and the associated constraint as follows

$$\begin{aligned}\min_{\mathbf{X},\mathbf{Z}} &\left\{ \tfrac{c}{2}||\mathbf{Y}-\mathbf{\Phi X}||_{\mathrm{F}}^2 + \sum_{i=1}^{N} ||\mathbf{z}_{i\cdot}||_2 \right\} \\ &\text{s.t.: } \mathbf{Z}=\mathbf{X}\mathbf{P}^{1/2},\end{aligned} \qquad (12)$$

where $\mathbf{P}^{1/2}$ is the positive-semidefinite square root of $\mathbf{P}$. It is clear that (12) is equivalent to (10). This decoupling of variable yields simpler and more tractable iterations.

To solve the constrained problem in (12), we associate a dual variable $\mathbf{G}\in\mathbb{R}^{N\times T}$ with the constraint. The augmented Lagrangian function is then defined as

$$\begin{aligned}L_\rho(\mathbf{X},\mathbf{Z},\mathbf{G}) = &\tfrac{c}{2}||\mathbf{Y}-\mathbf{\Phi X}||_{\mathrm{F}}^2 + \sum_{i=1}^{N}||\mathbf{z}_{i\cdot}||_2 \\ &+ \langle \mathbf{G}, \mathbf{Z}-\mathbf{X}\mathbf{P}^{1/2}\rangle + \tfrac{\rho}{2}||\mathbf{Z}-\mathbf{X}\mathbf{P}^{1/2}||_{\mathrm{F}}^2,\end{aligned} \qquad (13)$$

where $\langle , \rangle$ denotes the inner product of two matrices with the same dimension, and $\rho>0$ is called the penalty parameter.

The ADMM procedure consists of the following primal-dual updates

$$\mathbf{X}^{k+1} = \arg\min_{\mathbf{X}} L_\rho(\mathbf{X},\mathbf{Z}^k,\mathbf{G}^k) \qquad (14)$$

$$\mathbf{Z}^{k+1} = \arg\min_{\mathbf{Z}} L_\rho(\mathbf{X}^{k+1},\mathbf{Z},\mathbf{G}^k) \qquad (15)$$

$$\mathbf{G}^{k+1} = \mathbf{G}^k + \rho(\mathbf{Z}^{k+1}-\mathbf{X}^{k+1}\mathbf{P}^{1/2}), \qquad (16)$$

where k is the iteration index. It is clear that ADMM implements block descent with respect to the primal variables $\mathbf{X}$ and $\mathbf{Z}$, with updates performed on the dual variable $\mathbf{G}$. By separating the updates of $\mathbf{X}$ and $\mathbf{Z}$, the per iteration updates are simpler and analytically more tractable.

The minimization (14) can be carried out by setting the gradient of $L_\rho(\mathbf{X},\mathbf{Z}^k,\mathbf{G}^k)$ with respect to $\mathbf{X}$ to zero, which results in

$$c\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\mathbf{X}^{k+1}+\rho\mathbf{X}^{k+1}\mathbf{P} = (\mathbf{G}^k+\rho\mathbf{Z}^k)\mathbf{P}^{1/2}+c\mathbf{\Phi}^{\mathrm{T}}\mathbf{Y}. \qquad (17)$$

Again, this is a Sylvester equation that can be solved for $\mathbf{X}^{k+1}$ using the B-S algorithm.

To solve (15), we note that

$$\begin{aligned}&L_\rho(\mathbf{X}^{k+1},\mathbf{Z},\mathbf{G}^k)\\ &= \textstyle\sum_{i=1}^{N}||\mathbf{z}_{i\cdot}||_2 + \langle \mathbf{G}^k,\mathbf{Z}\rangle + \frac{\rho}{2}||\mathbf{Z}-\mathbf{X}^{k+1}\mathbf{P}^{1/2}||_{\mathrm{F}}^2 + C\\ &= \textstyle\sum_{i=1}^{N}||\mathbf{z}_{i\cdot}||_2 + \frac{\rho}{2}||\mathbf{Z}-(\mathbf{X}^{k+1}\mathbf{P}^{1/2}-\rho^{-1}\mathbf{G}^k)||_{\mathrm{F}}^2 + C\\ &= \textstyle\sum_{i=1}^{N}||\mathbf{z}_{i\cdot}||_2 + \frac{\rho}{2}||\mathbf{Z}-\mathbf{H}||_{\mathrm{F}}^2 + C\\ &= \textstyle\sum_{i=1}^{N}\left\{||\mathbf{z}_{i\cdot}||_2 + \frac{\rho}{2}||\mathbf{z}_{i\cdot}-\mathbf{h}_{i\cdot}||_2^2\right\} + C,\end{aligned} \qquad (18)$$

where we have defined $\mathbf{H}=\mathbf{X}^{k+1}\mathbf{P}^{1/2}-\rho^{-1}\mathbf{G}^k$ for notational clarity, and C is a constant independent of $\mathbf{Z}$.

Two observations follow from (18). First, the cost function $L_\rho(\mathbf{X}^{k+1},\mathbf{Z},\mathbf{G}^k)$ is separable across the rows of $\mathbf{Z}$, and therefore the minimization reduces to N separate minimization problems in $\{\mathbf{z}_{i\cdot}\}$. Second, the functional form on the last line of (18) is smooth in $\mathbf{z}_{i\cdot}$ except at the origin $\mathbf{z}_{i\cdot}=\mathbf{0}$. With some algebra, the minimizer can be found to be

$$\mathbf{z}_{i\cdot} = \begin{cases} \mathbf{h}_{i\cdot}\left(1-\frac{1}{\rho||\mathbf{h}_{i\cdot}||_2}\right), & ||\mathbf{h}_{i\cdot}||_2 > \rho^{-1} \\ 0, & \text{else}\end{cases}. \qquad (19)$$

From (19) we note that the update rule for $\mathbf{Z}$ is simply soft-thresholding applied on its rows. This is analogous to the element-wise soft-thresholding performed when solving the ℓ_1-norm regularized sparse vector recovery problem.

The pseudocode for the relaxation-based algorithm is listed in Algorithm 2.

Algorithm 2 ADMM solver for the regularized fitting problem

Inputs: $\mathbf{Y}\in\mathbb{R}^{M\times T}$, $\mathbf{\Phi}\in\mathbb{R}^{M\times N}$, ρ, c
Outputs: $\mathbf{X}\in\mathbb{R}^{N\times T}$
Initialize: $\mathbf{X}^0=\mathbf{Z}^0=\mathbf{G}^0=\mathbf{0}$, $k=0$
while not converged **do**
    Update primal variable $\mathbf{X}^{k+1}$ by solving (17)
    Update primal variable $\mathbf{Z}^{k+1}$ using (19)
    Update dual variable $\mathbf{G}^{k+1}$ using (16)
    $k\leftarrow k+1$
end while
Set $\mathbf{X}=\mathbf{X}^k$

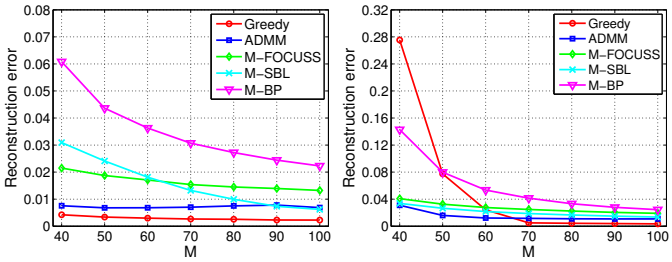
4. NUMERICAL EXAMPLES

In this section we demonstrate the performance of the proposed algorithms (denoted by "Greedy" and "ADMM" respectively) with experimental results. For validation and comparison, we also include in the examples the follow-

ing existing approaches: M-FOCUSS [13] (with $p = 0.8$), M-BP [14], and M-SBL [15]. For M-FOCUSS and M-BP, implementations from the Multiple-Spars Toolbox [18] were used, and for M-SBL the implementation obtained from the authors' website was used. For all algorithms considered herein, the parameters were empirically tuned to yield the best results. Denote by $\mathbf{X}$ and $\hat{\mathbf{X}}$ the ground truth and the reconstructed signal, respectively. The reconstruction error is quantified as $\epsilon = ||\mathbf{X} - \hat{\mathbf{X}}||_{\mathrm{F}}^2 / ||\mathbf{X}||_{\mathrm{F}}^2$.

We first demonstrate the capability of the proposed algorithms in reconstructing smooth signals from noisy measurements. We considered problems with measurement size M varying from 40 to 100, while the dimensions of the latent signal $\mathbf{X}$ were fixed at $N = 200$ and $T = 50$. Two sparsity levels of $\mathbf{X}$ were considered, e.g., either 5 or 10 out of the 200 rows of $\mathbf{X}$ were nonzeros. Each nonzero row of $\mathbf{X}$ was generated as a Hanning-window tapered sinusoid, where the number of periods was uniformly drawn between 1 and 3, and the phase was uniformly distributed between 0 and π; see Figure 2(a) for an example. The transformation $\mathbf{\Phi}$ was generated according to a uniform spherical ensemble, i.e., each $\boldsymbol{\phi}_{\cdot i}$ was independently drawn from a uniform distribution on the M-sphere with radius 1. Independent and identically distributed Gaussian noise $\mathbf{N}$ was added to the measurement, resulting an SNR at 10 dB.

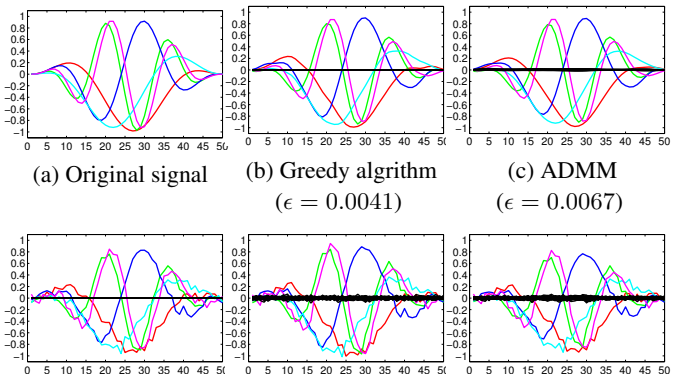
To model the smoothness of the nonzero rows of $\mathbf{X}$, we used $\mathbf{P} = \mathbf{D}^{\mathrm{T}}\mathbf{D}$ with $\mathbf{D}$ defined in (4). Reconstruction errors averaged over 100 random runs are shown in Figure 1. As we see, the proposed methods outperform the existing approaches under a wide range of experimental conditions. Comparing the two proposed algorithms, we see that ADMM is more robust to the lack of measurements than the greedy algorithm, while the latter yields lower reconstruction error when sufficient number of measurements are available.



(a) 5 nonzero rows in $\mathbf{X}$ (b) 10 nonzero rows in $\mathbf{X}$

Fig. 1: Reconstruction error v.s. number of measurements

The benefit of incorporating intra-row smoothness to signal recovery is clear when we examine the waveforms in Figure 2. In Figures 2(b)-(f), we use colored curves to denote the rows of $\hat{\mathbf{X}}$ corresponding to the true nonzero indices, and use black curves to denote the remaining rows. As is evidenced by the figures, both of the proposed algorithms and M-FOCUSS can accurately identify the nonzero rows of $\mathbf{X}$, while the reconstructions by M-SBL and M-BP have a number of spurious signals with very small but nonzero amplitudes. In addition, it is clear that by using smoothness-promoting regularizations in (3) and (10), the proposed methods yield more accurate estimates of the latent signals, while the other approaches suffer from over-fitting the noisy measurements.



(a) Original signal (b) Greedy algrithm ($\epsilon = 0.0041$) (c) ADMM ($\epsilon = 0.0067$)

(d) M-FOCUSS ($\epsilon = 0.019$) (e) M-SBL ($\epsilon = 0.033$) (f) M-BP ($\epsilon = 0.051$)

Fig. 2: Examples of reconstructed smooth signals with reconstruction errors.

Since over-fitting is an important issue when noise is present, it is interesting to examine how noise level affects the algorithmic performance. In this experiment, we set $M = 40$ and the number of nonzero rows in $\mathbf{X}$ to be 5. Noise level was varied to yield a range of SNR values from 5 dB to 40 dB, which is of practical interest for applications such as EEG/MEG signal processing. As can be seen in Figure 3, both of the proposed methods give significantly lower reconstruction errors than the alternative approaches, especially at low SNR conditions. The robustness to noise is attributed to the incorporation of prior smoothness knowledge.

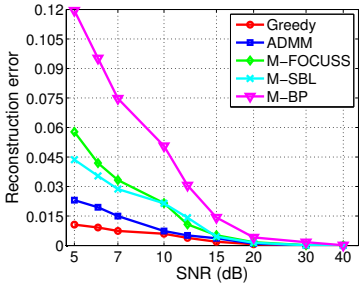


Fig. 3: Reconstruction error v.s. SNR

In addition to the sinusoid signals, we have considered $\mathbf{X}$ generated via first-order auto-regressive (AR) processes, i.e., $X_{ij} = aX_{ij-1} + E_{ij}$, where $a \in [0, 1)$ controls the correlation between adjacent samples in $\mathbf{x}_{i\cdot}$ and E_{ij} is a sample from the standard Gaussian distribution. The AR signals, though not smooth, have well defined correlation structure, which can

be exploited by setting $\mathbf{P}$ to be proportional to the precision matrix of $\mathbf{x}_i$..

In this experiment, 5 out of the 200 rows of $\mathbf{X}$ were generated as realizations of 50-sample AR processes, where the parameter a was varied from 0.1 to 0.8. The number of measurement M was kept at 40. Figure 4 shows the reconstruction errors averaged over 100 experimental runs. As can be seen, the proposed approaches are among the best in terms of reconstruction performance.

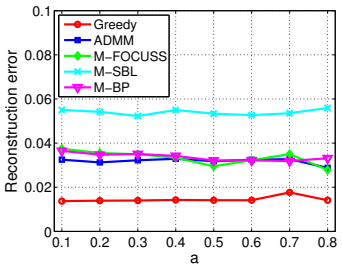


Fig. 4: Reconstruction error v.s. correlation

5. CONCLUSIONS

In this paper we considered the sparse signal recovery problem from a set of under-sampled measurements, where the sparse signals of interest are temporally smooth or correlated. We proposed a greedy algorithm and a relaxation-based algorithm that both incorporate prior knowledge about the signal properties. Effectiveness of the proposed algorithms have been demonstrated with numerical examples.

REFERENCES

[1] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 6–18, Jan 2006.

[2] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[3] A. K. Katsaggelos, *Digital Image Restoration*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1991.

[4] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, Aug 2005.

[5] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Proc.*, vol. 41, no. 12, pp. 3397–3415, Dec 1993.

[6] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Conf. Rec. of The 27th Asilomar Conf. on Signals, Systems and Computers*, Nov 1993, vol. 1, pp. 40–44.

[7] 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, Dec 2007.

[8] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. pp. 267–288, 1996.

[9] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 600–616, Mar 1997.

[10] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, September 2001.

[11] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Bayesian compressive sensing using Laplace priors," *IEEE Transactions on Image Processing*, vol. 19, no. 1, pp. 53–63, 2010.

[12] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Processing*, vol. 86, no. 3, pp. 572 – 588, 2006.

[13] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2477–2488, July 2005.

[14] J. A. Tropp, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Processing*, vol. 86, no. 3, pp. 589 – 602, 2006.

[15] D. P. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3704–3716, July 2007.

[16] M. Luessi, D. S. Babacan, R. Molina, and A. K. Katsaggelos, "Bayesian simultaneous sparse approximation with smooth signals," *IEEE Transactions on Signal Processing*, vol. 61, no. 22, pp. 5716–5729, Nov 2013.

[17] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[18] "Multiple-Spars Toolbox," 2014.