An Adaptive Bayesian Framework for Recovery of Sources with Structured Sparsity

Ali Bereyhi and Ralf R. Müller
Friedrich-Alexander Universität Erlangen-Nürnberg, Germany
{ali.bereyhi, ralf.r.mueller}@fau.de

***Abstract*—In oversampled adaptive sensing (OAS), noisy measurements are collected in multiple subframes. The sensing basis in each subframe is adapted according to some posterior information exploited from previous measurements. The framework is shown to significantly outperform the classic non-adaptive compressive sensing approach.**

This paper extends the notion of OAS to signals with structured sparsity. We develop a low-complexity OAS algorithm based on structured orthogonal sensing. Our investigations depict that the proposed algorithm outperforms the conventional non-adaptive compressive sensing framework with group LASSO recovery via a rather small number of subframes.

***Index Terms*—Oversampled adaptive sensing, Bayesian estimation, structured sparsity, compressive sensing.**

I. Introduction

The recently proposed oversampled adaptive sensing (OAS) framework has shown privileged performance for *time-limited* sensing in noisy environments [1], [2]. Unlike earlier adaptive approaches, e.g., [3]–[5], this scheme allows for *oversampling*. In this scheme, the signal is sensed in multiple steps, referred to as *subframes*. The sensing matrix in each subframe is adapted based on some posterior information determined from the measurements of previous subframes. In [1], it has been demonstrated that OAS achieves a considerable performance gain, when some prior information on the signal is available. The most well-known form of such prior information is *sparsity* which was explicitly studied in [1], [2]. Investigations have depicted that even *suboptimal low-complexity* OAS algorithms outperform well-known non-adaptive compressive sensing techniques in time-limited scenarios. This is intuitively illustrated as follows: When the signal is sparse, zero samples are detected in initial subframes even by low-quality measurements. These samples are then excluded in next subframes, where we focus on sensing the non-zero samples.

The previous studies on the OAS framework model the samples of a sparse signal as an independent and identically distributed (i.i.d.) process which does not consider any structure on the sparsity. It is however known that in many applications with sparse signals, the samples have structural dependencies, e.g., [6], [7]. In such applications, the recovery performance of conventional compressive sensing techniques can be improved by taking into account the sparsity structure [7]–[10].

This work has been accepted for presentation in 2019 IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP). The link to the final version in the proceedings will be available later.

From Bayesian points of view, structured sparsity provides further prior information on the signal. This intuitively implies that OAS achieves higher performance gains when it is employed to sense signals with structured sparsity. In this work, we aim to study the performance of OAS in such scenarios. To this end, we develop a low-complexity OAS scheme based on structured orthogonal sensing. Our investigations show that the proposed adaptive scheme with few subframes significantly outperforms the non-adaptive state-of-the-art.

Notation: Scalars, vectors and matrices are shown with non-bold, bold lower case and bold upper case letters, respectively. $\mathbf{I}_K$ and $\mathbf{0}_{K\times N}$ are the $K \times K$ identity matrix and $K \times N$ all-zero matrix, respectively. $\mathbf{A}^{\mathsf{T}}$ denotes the transpose of $\mathbf{A}$. The set of real numbers is shown by $\mathbb{R}$. We use the shortened notation $[N]$ to represent $\{1, \dots, N\}$.

II. Problem Formulation

We consider a sensing setup in which a vector of N signal samples collected in $\boldsymbol{x} \in \mathbb{R}^N$ is to be sensed via K distinct sensors within a fixed time interval of duration T. The sensing process is assumed to be linear and noisy. Hence, the vector of measurements collected by the sensor network within $t \leq T$ seconds of sensing is represented as $\boldsymbol{y} = \mathbf{A}\,\boldsymbol{x} + \boldsymbol{z}$, where $\mathbf{A} \in \mathbb{R}^{K\times N}$ is the sensing matrix whose entries are tunable, and $\boldsymbol{z} \in \mathbb{R}^K$ denotes additive white Gaussian noise with zero mean and variance $\sigma^2(t)$. The dependency of the noise variance on the sensing time models the sensing quality.

A. Model for Time-Limited Sensing

As indicated, the sensing process is to be limited to a time duration of T. It can hence be performed either in one step for a duration of T or in M steps each lasting for T/M. In the former case, the sensing process ends with K noisy measurements; however, the latter scheme collects MK measurements in total. Intuitively, the quality of measurements obtained by the first approach is higher than those acquired via multiple sensing steps. We model this phenomenon by setting the noise variance reversely proportional to the sensing time, i.e., for sensing duration t, $\sigma^2(t) = \sigma_0^2/t$ with σ_0^2 denoting the variance of noise within a unit of time.

This model is straightforwardly justified for various types of sensing devices following the corresponding circuitry models; see [1] for some detailed discussions. From systematic viewpoint, this model agrees with the physical intuition, since the signal-to-noise ratio (SNR) at each sensor grows linearly with time. Considering this model, there is a trade-off between the number of total collected measurements and the sensing quality. More measurements are acquired at the expense of shorter sensing time which results in higher noise variance.

B. Bayesian OAS Framework

The Bayesian OAS framework, introduced and analyzed in [1], [2], refers to the following sequential sensing procedure:

(a) The sensing time T is divided to M *subframes*.

(b) In subframe $m \in [M]$, the sensors measure

$$\boldsymbol{y}_m = \mathbf{A}_m \, \boldsymbol{x} + \boldsymbol{z}_m \tag{1}$$

for some sensing matrix $\mathbf{A}_m$ and measuring noise $\boldsymbol{z}_m \sim \mathcal{N}\left(\mathbf{0}, \sigma_{\rm sub}^2 \mathbf{I}_K\right)$, where $\sigma_{\rm sub}^2 = \sigma^2 \left(T/M\right) = M\sigma^2\left(T\right)$.

(c) From the stacked measurements in subframe m, i.e.,

$$\mathbf{Y}_m \coloneqq \left[\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\right], \tag{2}$$

a *Bayesian estimation* of the samples is determined as

$$\hat{\boldsymbol{x}}_m = \mathbb{E}\left\{\boldsymbol{x} | \mathbf{Y}_m, \mathbb{A}_m\right\}, \tag{3}$$

where $\mathbb{A}_m = \{\mathbf{A}_1, \dots, \mathbf{A}_m\}$, and the expectation is taken with respect to some *postulated* prior distribution $q\left(\boldsymbol{x}\right)$.

(d) Given the estimation in subframe m, the vector of *posterior information* is determined as

$$\mathbf{d}_m = \mathbb{E}\left\{\mathrm{d}\left[\boldsymbol{x}; \hat{\boldsymbol{x}}_m\right] | \mathbf{Y}_m, \mathbb{A}_m\right\}, \tag{4}$$

for some distortion function $\mathrm{d}\left[\cdot;\cdot\right]$. In general, the dimension of the posterior information vector can be different from the signal dimension. We hence denote it by B, i.e., $\mathbf{d}_m \in \mathbb{R}^B$, to keep the formulation generic.

(e) The sensor network constructs the sensing matrix of the next subframe based on $\mathbf{d}_m$, i.e., $\mathbf{A}_{m+1} = f_{\rm Adp}\left(\mathbf{d}_m\right)$ for some adaptation function $f_{\rm Adp}\left(\cdot\right)$.

C. Signals with Structured Sparsity

We assume that the signal samples have a structured sparsity pattern. To model the signal, we follow the generic *structured sparsity* model introduced in [7, Definition 2]: For $L \leq N$, let $\mathbb{I} \subseteq [N]$ be a subset of L indices, i.e., $|\mathbb{I}| = L$. Define $\boldsymbol{x}_{\mathbb{I}} \in \mathbb{R}^L$ to be a vector constructed by collecting those entries in $\boldsymbol{x}$ whose indices are in $\mathbb{I}$. Then, $\mathbb{S}_{\mathbb{I}}$ is said to be a *canonical L-sparse subspace* corresponding to index subset $\mathbb{I}$, when

$$\mathbb{S}_{\mathbb{I}} = \left\{\boldsymbol{x} : \boldsymbol{x}_{\mathbb{I}} \in \mathbb{R}^L \text{ and } \boldsymbol{x}_{\mathbb{I}^C} = \mathbf{0}_{N-L}\right\}. \tag{5}$$

Assume $\mathbb{S}$ is a subspace which is partitioned into S canonical L-sparse subspaces, i.e., $\mathbb{S} = \cup_{s=1}^S \mathbb{S}_{\mathbb{I}_s}$ for some distinct index subsets $\mathbb{I}_1, \dots, \mathbb{I}_S$. In this case, $\mathbb{S}$ is said to represent a *structured sparsity model* with sparsity L on a union of S sparse subspaces. Examples of structured sparsity models are *tree-based* and *block sparse* signals [7], [9], [11].

An stochastic model for structured sparsity can be described by a prior distribution for which we have $\Pr\left\{\boldsymbol{x} \notin \mathbb{S}\right\} = 0$. In the sequel, we give a stochastic model for the specific example of block sparsity. We use this model later to investigate our approach. For sake of simplicity, we present the model for sparse signals whose blocks are of similar size. Extensions to signals consisting of blocks with various lengths is straightforward.

Definition 1 (Random block sparse model)**:** Let L be a divisor of N and define $B = N/L$. $\boldsymbol{x}$ is said to be block sparse with block length L and sparsity factor ξ, when for $b \in [B]$

$$\boldsymbol{x}_b = \left[x_{(b-1)L+1}, \dots, x_{bL}\right]^{\mathsf{T}}$$

reads $\boldsymbol{x}_b = \psi_b \boldsymbol{s}_b$ with $\boldsymbol{s}_b \in \mathbb{R}^L$ being a *continuous* random vector and ψ_b being a ξ-Bernoulli random variable, i.e.

$$\Pr\left\{\psi_b = 1\right\} = 1 - \Pr\left\{\psi_b = 0\right\} = \xi.$$

The above model consists of B blocks of length L, each of them being either a vector of all zeros or completely non-zero. Hence, knowing only one sample in each block, one can recover the *support* of $\boldsymbol{x}$. For large N, the fraction of non-zero blocks is ξ which equals the fraction non-zero samples.

D. Objectives and Performance Measure

The main objective of this study is to investigate the impact of sparsity structure on the performance of OAS. To this end, we consider the following metric to quantify the performance:

Definition 2 (Average distortion)**:** Let $\mathbf{Y} \in \mathbb{R}^{K \times M}$, for some integer M, contain all measurements of $\boldsymbol{x}$ collected within the restricted sensing time T. Assume $\mathbf{g}\left(\cdot\right) : \mathbb{R}^{K \times M} \mapsto \mathbb{R}^N$ denotes an algorithm recovering the samples in $\boldsymbol{x}$ as $\hat{\boldsymbol{x}} = \mathbf{g}\left(\mathbf{Y}\right)$. The average distortion D with respect to the distortion function $\Delta\left(\cdot;\cdot\right) : \mathbb{R}^N \times \mathbb{R}^N \mapsto \mathbb{R}^+$ is defined as $D = \mathbb{E}\left\{\Delta\left(\boldsymbol{x}; \hat{\boldsymbol{x}}\right)\right\}$.

III. Block-Wise OAS via Orthogonal Sensing

The complexity of the OAS framework mainly depends on two factors: the *ensemble* from which the sensing matrix is chosen, and the *postulated prior* which is used for estimation in each subframe. On one hand, one can set the postulated prior distribution to the true one and search in each subframe for the optimal sensing matrix for the next subframe. This approach results in optimal performance which is achieved at the expense of high computational complexity. On the other hand, one may restrict the ensemble of sensing matrices and/or postulate a different prior distribution, such that the estimation and sensing matrix construction is addressed in each subframe with low complexity. The investigations in [1] and [2] show that even by following the latter suboptimal approach, the OAS framework outperforms the benchmark.

In the sequel, we develop a low complexity OAS algorithm for recovery of signals with structured sparsity. The algorithm selects the sensing matrix of each subframe from a certain class of *row-orthogonal* matrices. This restriction significantly simplifies the Bayesian estimation in each subframe. For sake of brevity, we restrict the derivations to signals with block sparsity whose non-zero samples are i.i.d. Gaussian: We assume that $\boldsymbol{x}$ is a random block sparse vector with $B = N/L$ blocks of length L in which $\boldsymbol{s}_b \sim \mathcal{N}\left(\mathbf{0}, \mathbf{I}_L\right)$ for $b \in [B]$. The framework is however extendable to other stochastic structured sparsity models with straightforward modifications.

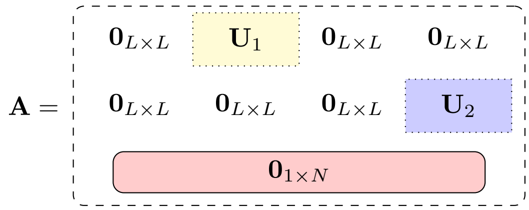

Fig. 1: An example of block-wise orthogonal matrices.

A. Block-wise Orthogonal Sensing Matrices

We start the derivations by defining a simple class of *block-wise orthogonal* matrices. This class comprises $F \coloneqq \lfloor K/L \rfloor$ orthogonal *principles*; namely, $\mathbf{U}_f \in \mathbb{R}^{L \times L}$ for $f \in [F]$ which satisfy $\mathbf{U}_f \mathbf{U}_f^{\mathsf{T}} = \mathbf{I}_L$. Let $\mathbb{F} = \{i_1, \dots, i_F\} \subset [B]$ be a subset with F distinct indices. The block-wise orthogonal matrix $\mathbf{A} \in \mathbb{R}^{K \times N}$ corresponding to $\mathbb{F}$ is then constructed by setting

$$\mathbf{A}\left(\left(f-1\right)L + \ell, \left(i_f - 1\right)L + \lambda\right) = \mathbf{U}_f\left(\ell, \lambda\right) \tag{6}$$

for $f \in [F]$ and $\ell, \lambda \in [L]$, and the other entries zero. An example of block-wise orthogonal matrices with $B = 4$ blocks of length L and $K = 2L + 1$ rows is given in Fig. 1. In this example, $F = 2$ and $\mathbb{F} = \{2, 4\}$. Hence, the entries in the last row are zero. $\mathbf{U}_1, \mathbf{U}_2 \in \mathbb{R}^{L \times L}$ denote the principles.

We define $\mathbb{O}_F$ to be the set of all possible block-wise matrices constructed by the principles $\mathbf{U}_f$ for $f \in [F]$ from F distinct indices. The main property of $\mathbf{A} \in \mathbb{O}_F$ is that the entries of $\mathbf{A}\boldsymbol{x}$ are partitioned into F blocks in which each block reads $\mathbf{U}_f \boldsymbol{x}_b$ for some $f \in [F]$ and $b \in [B]$. It is further straightforward to show that for $\mathbf{A} \in \mathbb{O}_F$ corresponding to $\mathbb{F}$

$$\boldsymbol{v} = \mathbf{A}^{\mathsf{T}} \mathbf{A} \boldsymbol{x} = \left[\boldsymbol{v}_1, \dots, \boldsymbol{v}_B\right]^{\mathsf{T}} \tag{7}$$

where $\boldsymbol{v}_b = \boldsymbol{x}_b \mathbf{1}\left\{b \in \mathbb{F}\right\}$.

By restricting the sensing matrices to be chosen from $\mathbb{O}_F$, Bayesian estimation and derivation of the posterior information become computationally tractable tasks. In the sequel, we derive these parameters for the given block sparse model.

B. Bayesian Estimator

Consider the Bayesian OAS framework, and let $\mathbf{A}_m \in \mathbb{O}_F$ be the sensing matrix in subframe $m \in [M]$. The vector of measurements in this subframe is therefore given by (1). With straightforward lines of derivations, it is shown that

$$\boldsymbol{w}_m = \mathbf{A}_m^{\mathsf{T}} \boldsymbol{y}_m = \mathbf{A}_m^{\mathsf{T}} \mathbf{A}_m \boldsymbol{x} + \mathbf{A}_m^{\mathsf{T}} \boldsymbol{z}_m \tag{8}$$

is a sufficient statistic. Hence,

$$\hat{\boldsymbol{x}}_m = \mathbb{E}\left\{\boldsymbol{x} | \mathbf{Y}_m, \mathbb{A}_m\right\} = \mathbb{E}\left\{\boldsymbol{x} | \boldsymbol{w}_1, \dots, \boldsymbol{w}_m\right\}. \tag{9}$$

Since the blocks are independent, we have

$$\hat{\boldsymbol{x}}_{b,m} = \mathbb{E}\left\{\boldsymbol{x}_b | \boldsymbol{w}_{b,1}, \dots, \boldsymbol{w}_{b,m}\right\} \tag{10}$$

where $\hat{\boldsymbol{x}}_{b,m}, \boldsymbol{w}_{b,m} \in \mathbb{R}^L$ denote the b-th block of $\hat{\boldsymbol{x}}_m$ and $\boldsymbol{w}_m$ for $b \in [B]$, respectively. To continue with derivations, let us define the following two notations:

- $\mathbb{F}_m \subseteq [B]$ represents the index set corresponding to $\mathbf{A}_m$. This set contains indices of the blocks whose samples are sensed in subframe m.
- $\mathbb{M}_b\left(m\right) \subseteq [m]$ contains indices of all subframes at which block b is sensed, i.e. $\mathbb{M}_b\left(m\right) = \{i \in [m] : b \in \mathbb{F}_i\}$.

By these definition, the Bayesian estimator further reads

$$\hat{\boldsymbol{x}}_{b,m} = \mathbb{E}\left\{\boldsymbol{x}_b | \mathbb{W}_b\left(m\right)\right\} \tag{11}$$

where $\mathbb{W}_b\left(m\right) = \{\boldsymbol{w}_{b,i} : i \in \mathbb{M}_b\left(m\right)\}$.

Following the property of $\mathbb{O}_F$ given in (7), it is concluded that for $i \in \mathbb{M}_b\left(m\right)$, we have $\boldsymbol{w}_{b,i} = \boldsymbol{x}_b + \tilde{\boldsymbol{z}}_{b,i}$, where $\tilde{\boldsymbol{z}}_{b,i} \in \mathbb{R}^{L \times L}$ denotes the b-th block of $\mathbf{A}_i^{\mathsf{T}} \boldsymbol{z}_i$. This concludes

$$\bar{\boldsymbol{w}}_b\left(m\right) = \sum_{i \in \mathbb{M}_b(m)} \boldsymbol{w}_{b,i} = |\mathbb{M}_b\left(m\right)| \boldsymbol{x}_b + \sum_{i \in \mathbb{M}_b(m)} \tilde{\boldsymbol{z}}_{b,i} \tag{12}$$

is a sufficient static for estimating $\boldsymbol{x}_b$. Considering the structure of $\mathbf{A}_i$, one can conclude that for $i \in \mathbb{M}_b\left(m\right)$, the noise term reads $\tilde{\boldsymbol{z}}_{b,i} = \mathbf{U}_f^{\mathsf{T}} \boldsymbol{z}_i^0$ for some principle $\mathbf{U}_f$ and some $\boldsymbol{z}_i^0 \sim \mathcal{N}\left(\mathbf{0}, M\sigma^2\left(T\right)\mathbf{I}_L\right)$. Hence, we can write

$$\bar{\boldsymbol{w}}_b\left(m\right) = |\mathbb{M}_b\left(m\right)| \boldsymbol{x}_b + \bar{\boldsymbol{z}}_b \tag{13}$$

where $\bar{\boldsymbol{z}}_b \sim \mathcal{N}\left(\mathbf{0}, \sigma_{b,m}^2 \mathbf{I}_L\right)$ with $\sigma_{b,m}^2 \coloneqq |\mathbb{M}_b\left(m\right)| M\sigma^2\left(T\right)$.

By substituting the true prior of the block sparse signal, the Bayesian estimator in subframe m reduces to

$$\hat{\boldsymbol{x}}_{b,m} = \mathbb{E}\left\{\boldsymbol{x}_b | \bar{\boldsymbol{w}}_b\left(m\right)\right\} = |\mathbb{M}_b\left(m\right)| \frac{\bar{\boldsymbol{w}}_b\left(m\right)}{C\left(\bar{\boldsymbol{w}}_b\left(m\right)\right)} \tag{14}$$

where function $C\left(\cdot\right) : \mathbb{R}^L \mapsto \mathbb{R}^+$ reads

$$C\left(\boldsymbol{y}\right) \coloneqq V_b\left(m\right)\left(1 + \frac{\left(1-\xi\right)\phi\left(\boldsymbol{y} | \sigma_{b,m}^2\right)}{\xi \phi\left(\boldsymbol{y} | V_b\left(m\right)\right)}\right) \tag{15}$$

with $V_b\left(m\right) \coloneqq |\mathbb{M}_b\left(m\right)|^2 + \sigma_{b,m}^2$, and $\phi\left(\boldsymbol{y} | \sigma^2\right)$ denoting the distribution of a zero-mean Gaussian random vector with covariance matrix $\sigma^2 \mathbf{I}_L$.

C. Posterior Information and Adaptation

The sensing matrix of each subframe is adapted via an adaption function based on the posterior information obtained in the previous subframe. A common choice for the posterior information in Bayesian OAS is the *posterior mean squared error (MSE)* which in the most basic case is determined for each sample of the signal. In order to exploit the sparsity structure of block sparse signals, we set the posterior information to be a B-dimensional vector, i.e., $\mathbf{d}_m = \left[d_{1,m}, \dots, d_{B,m}\right]^{\mathsf{T}}$, whose b-th entry is the posterior MSE of block b in subframe m, i.e.,

$$d_{b,m} = \mathbb{E}\left\{\|\boldsymbol{x}_b - \hat{\boldsymbol{x}}_{b,m}\|^2 | \bar{\boldsymbol{w}}_b\left(m\right)\right\} \tag{16}$$

By substituting (14) into the definition, we have

$$d_{b,m} = \frac{1}{C\left(\bar{\boldsymbol{w}}_b\left(m\right)\right)}\left(\sigma_{b,m}^2 - \frac{|\mathbb{M}_b\left(m\right)|^2 \|\bar{\boldsymbol{w}}_b\left(m\right)\|^2}{C\left(\bar{\boldsymbol{w}}_b\left(m\right)\right)}\right). \tag{17}$$

The posterior information $\mathbf{d}_m$ is given to an adaption function which constructs the sensing matrix in the next subframe, i.e., $\mathbf{A}_{m+1}$. Note that in our simplified framework, $\mathbf{A}_{m+1}$ is

Algorithm 1 Block-Wise OAS via Orthogonal Sensing

Initiate Set $d_{b,0} = +\infty$, $\bar{\boldsymbol{w}}_b(0) = \mathbf{0}_{L\times 1}$ and $\mathbb{M}_b(0) = \emptyset$.

for $m \in [M]$ **do**

1) Determine $\mathbb{F}_m$ by worst-case adaptation on $\mathbf{d}_{m-1}$.
2) Update $\mathbb{M}_b(m) = \mathbb{M}_b(m-1) \cup \{m\}$ for all $b \in \mathbb{F}_m$.
3) Select $\mathbf{A}_m \in \mathbb{O}_F$ which corresponds to $\mathbb{F}_m$.
4) Sense the samples for duration T/M via $\mathbf{A}_m$.
5) Determine sufficient statistic $\boldsymbol{w}_m = \mathbf{A}_m^\mathsf{T} \boldsymbol{y}_m$ from $\boldsymbol{y}_m$.
6) Update $\bar{\boldsymbol{w}}_b(m) = \bar{\boldsymbol{w}}_b(m-1) + \boldsymbol{w}_{b,m}$ for all $b \in \mathbb{F}_m$.
7) Update $\hat{\boldsymbol{x}}_{b,m}$ and $d_{b,m}$ using (14) and (17) respectively.

end for

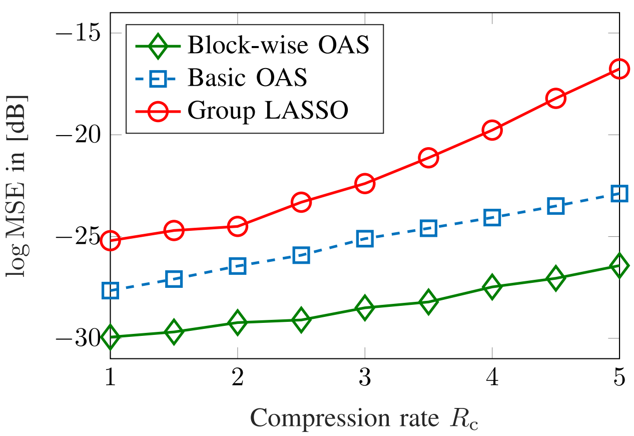

Fig. 2: MSE vs. compression rate for $B = 100$ and $L = 4$.

restricted to be chosen from $\mathbb{O}_F$. We hence employ the *worst-case adaptation* strategy proposed in [1] and utilized in [2]: In subframe m, the adaptation function finds the permutation

$$\Pi_m([B]) = \{i_1, \dots, i_B\}, \tag{18}$$

such that $d_{i_1,m} \geq \dots \geq d_{i_B,m}$. It then sets the sensing matrix of the next subframe to $\mathbf{A}_{m+1} \in \mathbb{O}_F$ whose corresponding index set is $\mathbb{F}_{m+1} = \{i_1, \dots, i_F\}$. The proposed OAS approach is summarized in Algorithm 1.

IV. Numerical Investigations

We investigate the proposed framework by conducting some numerical experiments. To this end, we consider the following *time-limited* sensing scenario:

- $T = 1$ and $\sigma^2(t) = 0.01/t$.
- The vector of signal samples consists of B blocks of L samples with sparsity factor $\xi = 0.1$.
- The compression rate is defined as $R_c = N/K = BL/K$.
- The performance is quantified via the MSE which is given by the average distortion when $\Delta(\boldsymbol{x}; \hat{\boldsymbol{x}}) = \|\hat{\boldsymbol{x}} - \boldsymbol{x}\|^2/N$.

We study three different signal recovery schemes:

1) Algorithm 1 with $M = 8$ subframes.
2) The *basic* OAS algorithm with orthogonal measurements which does not take the sparsity structure into account and treats samples as an i.i.d. sparse Gaussian sequence [2, Algorithm 1]. Similar to Scheme 1, we set $M = 8$.

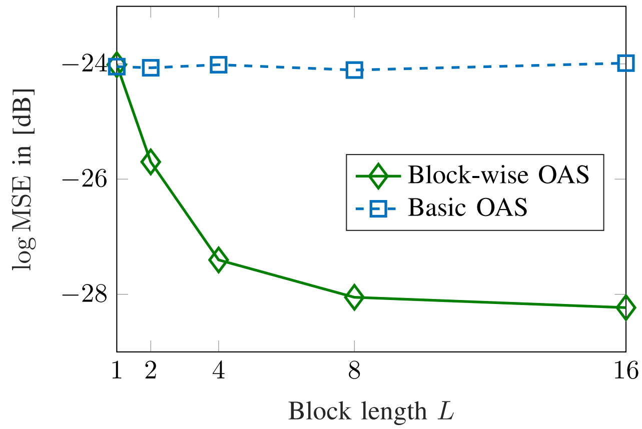

Fig. 3: MSE vs. block-length for $N = 1600$ and $R_c = 4$.

3) The *benchmark* in which the sensors measure the samples via sensing matrix $\mathbf{A}$ in a single subframe. The samples are recovered from measurement vector $\boldsymbol{y}$ via the *group LASSO* algorithm which for some λ reads [12]

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{v}\in\mathbb{R}^N}{\operatorname{argmin}} \|\boldsymbol{y} - \mathbf{A}\boldsymbol{v}\|^2 + \lambda \sum_{b=1}^{B} \|\boldsymbol{v}_b\| \tag{19}$$

We assume $\mathbf{A}$ is an i.i.d. matrix whose entries are zero-mean with variance $1/K$. This is a conventional setting in classic compressive sensing; see for example [13], [14].

Fig. 2 shows the MSE against the compression rate R_c for all the schemes when $B = 100$ and $L = 4$. For Scheme 3, the results are given by minimizing the MSE with respect to λ numerically. As the figure shows, the block-wise OAS scheme with $M = 8$ subframes outperforms the benchmark for a large range of compression rates. This observation indicates that *even by suboptimal adaptation* the sequential approach of OAS improves the recovery performance which is intuitive: The proposed algorithm recovers the zero blocks from the low-quality measurements of first few subframes. It then excludes these blocks in next subframes and only measures the non-zero blocks. Due to the longer sensing time, the latter measurements are of higher quality resulting in a good recovery.

It is further observed in Fig. 2 that the proposed scheme outperforms the *basic* OAS algorithm. Such an observation is due to the fact that in *basic* OAS the sparsity structure does not play any role in the recovery and adaptation. To further illustrate this latter observation, we sketch the MSE against the block length for Scheme 1 and Scheme 2 in Fig. 3 when $R_c = 4$. To fair comparison, at block length L the number of blocks B is chosen such that $N = 1600$. As the figure depicts, the MSE achieved by block-wise OAS reduces as the block-length L increases. This follows the fact that the number of canonical sparse subspaces in the block sparse model reduces with the block length which improves the recovery performance. Such a behavior is however not observed in basic OAS following the fact that this scheme ignores the sparsity structure.

V. Conclusions

A low-complexity OAS framework has been developed to sequentially measure signals with structured sparsity. The proposed scheme exploits the sparsity pattern of the signal to improve adaptation and recovery. Our numerical investigations demonstrated that this scheme outperforms the classic non-adaptive compressive sensing framework with the well-known group LASSO recovery algorithm, as well as the basic OAS framework previously developed for the i.i.d. sparsity model.

References

[1] R. R. Müller, A. Bereyhi, and C. Mecklenbräuker, "Oversampled adaptive sensing," *Proc. Information Theory and Applications Workshop (ITA)*, pp. 1–7, February 2018, San Diego, CA, USA.

[2] R. R. Müller, A. Bereyhi, and C. F. Mecklenbräuker, "Oversampled adaptive sensing with random projections: Analysis and algorithmic approaches," *Proc. IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pp. 336–341, December 2018, Louisville, KY, USA.

[3] J. D. Haupt, R. G. Baraniuk, R. M. Castro, and R. D. Nowak, "Compressive distilled sensing: Sparse recovery using adaptivity in compressive measurements," *Proc. Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*, pp. 1551–1555, November 2009, Pacific Grove, CA, USA.

[4] J. Haupt, R. Nowak, and R. Castro, "Adaptive sensing for sparse signal recovery," *Proc. 13th Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop*, pp. 702–707, January 2009, Marco Island, FL, USA.

[5] M. L. Malloy and R. D. Nowak, "Near-optimal adaptive compressed sensing," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 4001–4012, July 2014.

[6] L. He and L. Carin, "Exploiting structure in wavelet-based Bayesian compressive sensing," *IEEE Transactions on Signal Processing*, vol. 57, no. 9, pp. 3488–3497, September 2009.

[7] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1982–2001, April 2010.

[8] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5302–5316, November 2009.

[9] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4053–4085, September 2011.

[10] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3042–3054, June 2010.

[11] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, February 2006.

[12] W. Deng, W. Yin, and Y. Zhang, "Group sparse optimization by alternating direction method," *Proc. SPIE Wavelets and Sparsity XV*, vol. 8858, p. 88580R, September 2013, San Diego, CA, USA.

[13] A. Bereyhi, S. Haghighatshoar, and R. R. Müller, "Theoretical bounds on map estimation in distributed sensing networks," *Proc. IEEE International Symposium on Information Theory (ISIT)*, pp. 1385–1389, June 2018, Vail, CO, USA.

[14] Y. Shiraki and Y. Kabashima, "Typical reconstruction performance for distributed compressed sensing based on $\ell_{2,1}$-norm regularized least square and Bayesian optimal reconstruction: Influences of noise," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2016, no. 6, p. 063304, May 2015.