Segmented Compressed Sampling for Analog-to-Information Conversion: Method and Performance Analysis

Omid Taheri, *Student Member, IEEE*, and Sergiy A. Vorobyov, *Senior Member, IEEE*

***Abstract*—A new segmented compressed sampling (CS) method for analog-to-information conversion (AIC) is proposed. An analog signal measured by a number of parallel branches of mixers and integrators (BMIs), each characterized by a specific random sampling waveform, is first segmented in time into M segments. Then the subsamples collected on M different segments and K different BMIs are reused so that a larger number of samples (at most K^2) than the number of BMIs is collected. This technique is shown to be equivalent to extending the measurement matrix, which consists of the BMI sampling waveforms, by adding new rows without actually increasing the number of BMIs. We prove that the extended measurement matrix satisfies the restricted isometry property with overwhelming probability if the original measurement matrix of BMI sampling waveforms satisfies it. We also prove that the signal recovery performance can be improved if our segmented CS-based AIC is used for sampling instead of the conventional AIC with the same number of BMIs. Therefore, the reconstruction quality can be improved by slightly increasing (by $M \leq K$ times) the sampling rate per each BMI. Simulation results verify the effectiveness of the proposed segmented CS method and the validity of our theoretical results. Particularly, our simulation results show significant signal recovery performance improvement when the segmented CS-based AIC is used instead of the conventional AIC with the same number of BMIs.**

***Index Terms*—Analog-to-information conversion (AIC), compressed sampling (CS), correlated random variables, Craig–Bernstein inequality, empirical risk minimization, segmented AIC, segmented CS.**

I. Introduction

ACCORDING to Shannon's sampling theorem, an analog band-limited signal can be recovered from its discrete-time samples if the sampling rate is at least twice the bandwidth of the signal. Recent theory of compressed sampling (CS), however, suggests that an analog signal can be recovered from fewer samples if it is sparse or compressible in some basis and not necessarily band-limited [1]–[4]. CS theory also suggests that a universal sampling matrix (for example, a random projection matrix) can be designed, and it can be used for all signals that are sparse or compressible in some basis regardless of their nature [2]. CS has already found a wide range of applications such as image acquisition [5], sensor networks [6], cognitive radios [7], communication channel estimation [8], etc.

The sampling process often used in the CS literature consists of two steps. First, an analog signal is sampled at the Nyquist rate and then a measurement matrix is applied to the time domain samples in order to collect the compressed samples. This sampling approach, however, defeats one of the primary purposes of CS, which is avoiding high rate sampling. A more practical approach for "direct" sampling and compression of analog signals belonging to the class of signals in a union of subspaces is taken in [9] and the follow up work [10]. Another practical approach to CS, which avoids high rate sampling, has been presented in [1], [11], and the name analog-to-information conversion (AIC) has been coined. The AIC device consists of several parallel branches of mixers and integrators (BMIs) in which the analog signal is measured against different random sampling waveforms. Therefore, for every collected compressed sample, there is a BMI that multiplies the signal to a sampling waveform and then integrates the result over the sampling period T.

In this paper, we propose a new segmented CS method and a new segmented CS-based AIC structure which is capable of collecting more samples than the number of BMIs. With more samples, the recovery performance can be improved as compared to the case when the AIC of [1] with the same number of BMIs is used for sampling. The specific contributions of this work are the following. i) A new segmented CS-based AIC structure is developed. Some preliminary results have been reported in [12]. In this structure, the integration period T is divided into M equal segments such that the sampling rate of the so-obtained segmented AIC is M times higher than the sampling rate of the AIC of [1]. Then the subsamples collected over M different segments and K different BMIs are reused so that a larger number of samples (at most K^2 correlated samples) than the number of BMIs is collected. We show that our segmented CS-based AIC technique is equivalent to extending the measurement matrix, which consists of the BMI sampling waveforms, by adding new

Manuscript received April 23, 2010; revised August 10, 2010, October 25, 2010; accepted October 29, 2010. Date of publication November 11, 2010; date of current version January 12, 2011. The associate editor coordinating the review of this paper and approving it for publication was Dr. Konstantinos I. Diamantaras. This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada and in part by the Alberta Innovates—Technology Futures, Alberta, Canada. Parts of this work were presented at the IEEE Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), Aruba, Dutch Antilles, 2009 and the Annual Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, California, November 2010.

The authors are with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, T6G 2V4 Canada (e-mail: otaheri@ece.ualberta.ca; vorobyov@ece.ualberta.ca).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org provided by the authors. This includes Matlab codes needed to generate the simulation results shown in the paper. This material is 20 KB in size.

Digital Object Identifier 10.1109/TSP.2010.2091411

rows without actually increasing the number of BMIs.[1] ii) We prove that the restricted isometry property (RIP), i.e., the sufficient condition for signal recovery based on compressed samples, is satisfied for the extended measurement matrix resulting from the segmented CS-based AIC structure with overwhelming probability if the original matrix of BMI sampling waveforms satisfies the RIP. Thus, our segmented AIC is a valid candidate for CS. iii) We also prove that the signal recovery performance based on the empirical risk minimization approach can be improved if our segmented AIC is used for sampling instead of the AIC of [1] with the same number of BMIs. Some preliminary results on this topic have been reported in [15]. The mathematical challenge in such a proof is that the samples collected by our segmented AIC are correlated, while all results on performance analysis of the signal recovery available in the literature are obtained for the case of uncorrelated samples.

The rest of this paper is organized as follows. The setup for CS of analog signals and background on CS signal recovery and AIC are given in Section II. The main idea of the paper, that is, the segmented CS, is explained in Section III. We prove in Section IV that the extended measurement matrix resulting from the proposed segmented CS satisfies the RIP and, therefore, the segmented CS is a legitimate CS method for AIC. The signal recovery performance analysis for our segmented CS is given in Section V. Section VI shows our simulation results and Section VII concludes the paper. This paper is reproducible research [16] and the software needed to generate the simulation results can be obtained from http://ieeexplore.ieee.org or http://www.ece. ualberta.ca/vorobyov/SegmentedCS.zip.

II. Setup and Background

Setup for CS of Analog Signals: CS deals with a low rate representation of sparse or compressible signals, i.e., such signals which have few nonzero or significantly different from zero projections on the vectors of an orthogonal basis (sparsity basis). It is assumed that the analog signal $f(t)$ can be represented or approximated as a linear combination of a finite number of N basis functions $\{\psi_n(t)\}_{n=1}^N$ defined over the time period $t \in [0,T]$. Hence, the signal $f(t)$ is also defined over the same time period and it can be mathematically expressed as

$$f(t) = \sum_{n=1}^{N} x_n \psi_n(t) = \boldsymbol{x}^T \boldsymbol{\Psi}(t) \tag{1}$$

where $\{x_n\}_{n=1}^N$ are some coefficients, $\boldsymbol{x} \triangleq (x_1, \ldots, x_N)^T$ is a vector of such (possibly complex) coefficients, $\boldsymbol{\Psi}(t) \triangleq (\psi_1(t), \ldots, \psi_N(t))^T$, and $(\cdot)^T$ stands for the transpose. If $f(t)$ is sparse or compressible, i.e., the vector $\boldsymbol{x}$ has a small number of nonzero or significantly different from zero elements, the basis $\{\psi_n(t)\}_{n=1}^N$ is called a sparsity basis and $\boldsymbol{\Psi}(t)$ maps the discrete vector of coefficients $\boldsymbol{x}$ onto a continuous signal $f(t)$. It is known that a universal CS method can be designed to effectively sample and recover S-sparse signals regardless of the specific sparsity domain [1], [2].

[1]In this respect, the works [13] and [14] also need to be mentioned. In [13], Toeplitz-structured measurement matrices are considered, while measurement matrix that is built based on only one random vector with shifts of $D \geq 1$ in between the rows appear in radar imaging application considered in [14].

The measurement operator $\boldsymbol{\Phi}(t)$ is the collection of $K < N$ sampling waveforms $\{\phi_k(t)\}_{k=1}^K$, i.e., $\boldsymbol{\Phi}(t) \triangleq (\phi_1(t), \ldots, \phi_K(t))^T$. One of the practical choices for the sampling waveforms is a pseudo random ± 1 chip sequence which alternates its value at a rate higher than, for example, the Nyquist rate for bandlimited signals [1], [11] or just the rate N/T in the traditional CS setup with the finite number of N possible projections. Let the chip duration T_c be set to T/N_c where N_c is the number of chips per signal period T. The discrete measurement y_k can be expressed as

$$y_k = \int_0^T f(t)\phi_k(t)dt. \tag{2}$$

Then the relationship between the $K \times 1$ vector of measurements $\boldsymbol{y} \triangleq (y_1, \ldots, y_K)^T$ and the sparse coefficient vector $\boldsymbol{x}$ can be explained in terms of the $K \times N$ matrix $\boldsymbol{\Phi}' = \boldsymbol{\Phi}(t)\boldsymbol{\Psi}(t)$ with its (k,n)th entry given as

$$[\boldsymbol{\Phi}']_{k,n} = \int_0^T \psi_n(t)\phi_k(t)dt. \tag{3}$$

Using the matrix $\boldsymbol{\Phi}'$, we can compactly represent the vector of discrete measurements as $\boldsymbol{y} = \boldsymbol{\Phi}'\boldsymbol{x}$. Then, the $K \times N_c$ measurement matrix $\boldsymbol{\Phi}$ and the $N \times N_c$ sparsity basis $\boldsymbol{\Psi}$ can be derived as the discrete equivalents of $\boldsymbol{\Phi}(t)$ and $\boldsymbol{\Psi}(t)$. Specifically, let the entries of $\boldsymbol{\Phi}$ be given as

$$[\boldsymbol{\Phi}]_{k,n} = \int_{(n-1)T_c}^{nT_c} \phi_k(t)dt \tag{4}$$

and the entries of $\boldsymbol{\Psi}$ as

$$[\boldsymbol{\Psi}]_{m,n} = \int_{(n-1)T_c}^{nT_c} \psi_m(t)dt. \tag{5}$$

Then it can be seen that $\boldsymbol{\Phi}' = \boldsymbol{\Phi}\boldsymbol{\Psi}^T$. Moreover, the discrete counterpart of the analog signal $f(t)$, denoted as vector $\boldsymbol{f}$, is given as $\boldsymbol{f} = \boldsymbol{\Psi}^T\boldsymbol{x}$. Using the measurement matrix $\boldsymbol{\Phi}$, the $K \times 1$ vector of compressed samples $\boldsymbol{y}$ can be equivalently written as $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{f} = \boldsymbol{\Phi}'\boldsymbol{x}$. In the noisy case, the sampling process can be expressed as

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{f} + \boldsymbol{w} = \boldsymbol{\Phi}'\boldsymbol{x} + \boldsymbol{w} \tag{6}$$

where $\boldsymbol{w}$ is a zero mean noise vector with identically and independently distributed (i.i.d.) entries of variance σ^2.

In the traditional CS setup for discrete signals, the sparsity basis matrix $\boldsymbol{\Psi}$ with entries given by (5) is considered to be an $N \times N$ orthonormal matrix. This corresponds to the case when $N_c = N$ for the sampling waveforms. However, there exist applications where this condition is not satisfied and N_c is larger than N. The mathematical analysis and the proves given in this paper consider the traditional CS setup where the matrix $\boldsymbol{\Psi}$ is square and orthonormal. However, we include some simulation results attesting to the fact that our segmented CS method also works when $N_c > N$.

Another important issue is the number of required compressed samples for successful signal recovery. Among various

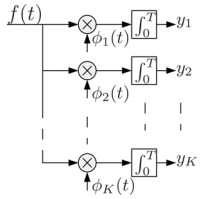


Fig. 1. The structure of the AIC based on RMPI.

bounds on the sufficient number of collected compressed samples[2] K $(S < K < N)$ required for recovering an S-sparse signal, the first and most popular one is given by the following inequality $S \leq CK/\log(N/K)$ where C is some constant [1]. This bound is derived based on the uniform uncertainty principle which states that $\mathbf{\Phi}$ must satisfy the following restricted isometry property (RIP) [1], [18]. Let $\mathbf{\Phi}_{\mathcal{T}}$ be a submatrix of $\mathbf{\Phi}$ retaining only the columns with their indexes in the set $\mathcal{T} \subset \{1, \ldots, N\}$. Then the S-restricted isometry constant δ_S is the smallest number satisfying the inequality

$$\frac{K}{N}(1-\delta_S)\|\boldsymbol{c}\|_{l_2}^2 \leq \|\mathbf{\Phi}_{\mathcal{T}}\boldsymbol{c}\|_{l_2}^2 \leq \frac{K}{N}(1+\delta_S)\|\boldsymbol{c}\|_{l_2}^2 \qquad (7)$$

for all sets $\mathcal{T}$ of cardinality less than or equal to S and all vectors $\boldsymbol{c}$. Here $\|\cdot\|_{l_2}$ denotes the Euclidean norm of a vector. As shown in [2], [19], if the entries of $\mathbf{\Phi}$ are, for example, independent zero mean Gaussian random variables with variance $1/N$, then $\mathbf{\Phi}$ satisfies the RIP for $S \leq CK/\log(N/K)$ with high probability.[3] It is known that the same holds when the entries of $\mathbf{\Phi}$ are independent zero mean Bernoulli variables with variance $1/N$ [19].

AIC: The random modulation preintegration (RMPI) structure is proposed for AIC in [1]. The RMPI multiplies the signal with the sampling waveforms in the analog domain and then integrates the product over the time period T to produce samples. It implies that the sampling device has a number of parallel BMIs in order to process the analog signal in real-time. The RMPI structure is shown in Fig. 1, where the previously introduced notations are used.

Recovery Methods: A sparse signal can be recovered from its noiseless sample vector $\boldsymbol{y}$ by solving the following convex optimization (linear programming) problem [2], [20]

$$\min \|\tilde{\boldsymbol{x}}\|_{l_1} \quad \text{subject to} \quad \mathbf{\Phi}^{'}\tilde{\boldsymbol{x}} = \boldsymbol{y} \qquad (8)$$

where $\|\cdot\|_{l_1}$ denotes the l_1-norm of a vector. In the noisy case, the recovery problem is modified as [21]

$$\min \|\tilde{\boldsymbol{x}}\|_{l_1} \quad \text{subject to} \quad \|\mathbf{\Phi}^{'}\tilde{\boldsymbol{x}} - \boldsymbol{y}\|_{l_2} \leq \gamma \qquad (9)$$

where γ is the bound on the square root of the noise energy. In order to reconstruct an analog signal, i.e., obtain the estimate $\hat{f}(t)$ from the measurement vector $\boldsymbol{y}$, one should first solve for $\tilde{\boldsymbol{x}}$ using (8) or (9) and then calculate $\hat{f}(t)$ based on (1).

Another technique for sparse signal recovery from noisy samples (see [4]) uses the empirical risk minimization method that was first developed in statistical learning theory for approximating an unknown function based on noisy measurements [22]. Note that the empirical risk minimization-based recovery method is of a particular interest since under some approximations (see [4, p. 4041]) it reduces to another well-known least absolute shrinkage and selection operator (LASSO) method [23]. Therefore, the risk minimization-based method provides the generality which we need in this paper.

In application to CS, the unknown function is the sparse signal and the collected data are the noisy compressed samples. Let the entries of the measurement matrix $\mathbf{\Phi}$ be selected with equal probability as $\pm 1/\sqrt{N}$, and the energy of the signal $\boldsymbol{f}$ be bounded so that $\|\boldsymbol{f}\|^2 \leq NB^2$. The risk $r(\hat{\boldsymbol{f}})$ of a candidate reconstruction $\hat{\boldsymbol{f}}$ and the empirical risk $\hat{r}(\hat{\boldsymbol{f}})$ are defined as [22]

$$r(\hat{\boldsymbol{f}}) \triangleq \frac{\|\hat{\boldsymbol{f}} - \boldsymbol{f}\|^2}{N} + \sigma^2, \quad \hat{r}(\hat{\boldsymbol{f}}) \triangleq \frac{1}{K}\sum_{j=1}^{K}(y_j - \boldsymbol{\phi}_j\hat{\boldsymbol{f}})^2. \quad (10)$$

Then the candidate reconstruction $\hat{\boldsymbol{f}}_K$ obtained based on K samples can be found as [4]

$$\hat{\boldsymbol{f}}_K = \arg \min_{\hat{\boldsymbol{f}} \in \mathcal{F}(B)} \left\{ \hat{r}(\hat{\boldsymbol{f}}) + \frac{c(\hat{\boldsymbol{f}})\log 2}{\epsilon K} \right\} \qquad (11)$$

where $\mathcal{F}(B) \triangleq \{\boldsymbol{f} : \|\boldsymbol{f}\|^2 \leq NB^2\}$, $c(\hat{\boldsymbol{f}})$ is a nonnegative number assigned to a candidate signal $\hat{\boldsymbol{f}}$, and $\epsilon = 1/(50(B+\sigma)^2)$. Moreover, $\hat{\boldsymbol{f}}_K$ given by (11) satisfies the following inequality [4]

$$E\left\{\frac{\|\hat{\boldsymbol{f}}_K - \boldsymbol{f}\|^2}{N}\right\} \leq C_1 \min_{\hat{\boldsymbol{f}} \in \mathcal{F}(B)} \left\{ \frac{\|\hat{\boldsymbol{f}} - \boldsymbol{f}\|^2}{N} + \frac{c(\hat{\boldsymbol{f}})\log 2 + 4}{\epsilon K} \right\} \qquad (12)$$

where $C_1 = [(27-4e)(B/\sigma)^2 + (50-4\sqrt{2})B/\sigma + 26]/[(23-4e)(B/\sigma)^2 + (50-4\sqrt{2})B/\sigma + 24]$, $e = 2.7183\ldots$, and $E\{\cdot\}$ stands for the expectation operation.

Let a compressible signal $\boldsymbol{f}$ be defined as a signal for which $\|\boldsymbol{f}^{(m)} - \boldsymbol{f}\|^2 \leq NC_A m^{-2\alpha}$, where $\boldsymbol{f}^{(m)}$ is the best m-term approximation of $\boldsymbol{f}$ which is obtained by retaining the m most significant coefficients of vector $\boldsymbol{x}$ and $C_A > 0$ and $\alpha \geq 0$ are some constants. Let also $\mathcal{F}_c(B, \alpha, C_A) \triangleq \{\boldsymbol{f} : \|\boldsymbol{f}\|^2 \leq NB^2, \|\boldsymbol{f}^{(m)} - \boldsymbol{f}\|^2 \leq NC_A m^{-2\alpha}\}$ be the set of compressible signals. Then based on the weight assignment $c(\hat{\boldsymbol{f}}) = 2\log(N)N_{\hat{\boldsymbol{x}}}$, where $N_{\hat{\boldsymbol{x}}}$ is the actual number of nonzero projections of $\hat{\boldsymbol{f}}$ onto the sparsity basis, the following inequality holds [4]

$$\sup_{\boldsymbol{f} \in \mathcal{F}_c(B,\alpha,C_A)} E\left\{\frac{\|\hat{\boldsymbol{f}}_K - \boldsymbol{f}\|^2}{N}\right\} \leq C_1 C_2 \left(\frac{K}{\log N}\right)^{-2\alpha/(2\alpha+1)} \qquad (13)$$

where $C_2 = C_2(B, \sigma, C_A) > 0$ is a constant.

If signal $\boldsymbol{f}$ is indeed sparse and belongs to $\mathcal{F}_s(B, S) \triangleq \{\boldsymbol{f} : \|\boldsymbol{f}\|^2 \leq NB^2, \|\boldsymbol{f}\|_{l_0} \leq S\}$, then there exists a constant $C_2' = C_2'(B, \sigma) > 0$ such that [4]

$$\sup_{\boldsymbol{f} \in \mathcal{F}_s(B,S)} E\left\{\frac{\|\hat{\boldsymbol{f}}_K - \boldsymbol{f}\|^2}{N}\right\} \leq C_1 C_2' \left(\frac{K}{S\log N}\right)^{-1}. \quad (14)$$

[2]See [17] for broader review.

[3]Note that in order to ensure consistency throughout the paper, the variance of the elements in $\mathbf{\Phi}$ is assumed to be $1/N$ instead of $1/K$ as, for example, in [2]. Thus, the multiplier K/N is added in the left- and right-hand sides of (7).

III. Segmented Compressed Sampling Method

A significant advantage of the AIC is that it removes the need for high speed sampling. The smaller the number of samples K being collected, the less number of BMIs is required, thus, the less complex the AIC hardware is. The minimum number of samples required for successful signal recovery is given by the bound based on the RIP (7). The practical rule of thumb for the noiseless case is that four incoherent measurements are required for successful recovery of each nonzero coefficient [1]. However, in the event that the measurements are noisy a larger number of samples allows for a better signal recovery. Indeed, the mean-square error (MSE) between the actual and recovered signals is bounded in the noisy case as given in [1, p. 27, Theorem 3] for compressible signals. Such a bound contains a coefficient which depends inversely on the number of available samples. Thus, the larger the number of samples, the better recovery performance can be achieved in the noisy case. Moreover, in practice when the signal sparsity level may not be exactly known, the number of BMIs may be insufficient to guarantee successful signal recovery. Therefore, we may need to collect a larger number of samples to enable recovery. In order to collect a larger number of compressed samples using the AIC structure in Fig. 1, we need to increase the hardware complexity by adding more BMIs. The latter makes the AIC device more complex although its sampling rate is much lower than that of the analog-to-digital converter (ADC). Therefore, it is desirable to reduce the number of parallel BMIs in the AIC without sacrificing significantly the signal recovery accuracy. It can be achieved by adding to the AIC the capability of sampling at a higher rate, which is, however, still significantly lower than the sampling rate required by the ADC. The latter can be achieved by splitting the integration period T in every BMI of the AIC in Fig. 1 into $M \leq K$ shorter subperiods (segments). Note that since the original integration period is divided into a number of smaller subperiods, the samples collected over all parallel BMIs during one subperiod do not have complete information about the signal. Therefore, they are called incomplete samples. Hereafter, the complete samples obtained over the whole period T are referred to as just samples, while the incomplete samples are referred to as subsamples.

A. The Basic Idea and the Model

The basic idea is to collect a number of subsamples by splitting the integration period into a number of subperiods and then reuse such subsamples in order to build additional samples. In this manner, a larger number of samples than the number of BMIs can be collected. It allows for a tradeoff between the AIC and ADC structures by allocating $M \leq K$ subsamples per time-unit T to K BMIs. Indeed, the signal is measured at a low rate by correlating it to a number of sampling waveforms just as in the AIC, while at the same time the integration period T is split into shorter subintervals, i.e., the sampling rate is slightly increased. However, such sampling rate is still significantly lower than that required by the ADC.

Let the integration period be split into M subintervals and $\boldsymbol{y}_k = (y_{k,1}, \ldots, y_{k,M})^T, k = 1, \ldots, K$ be the vectors of subsamples collected against the sampling waveforms $\{\boldsymbol{\phi}_k\}_{k=1}^K$. The subsample $y_{k,j}$ is given by

$$y_{k,j} = \int_{(j-1)T/M}^{jT/M} f(t)\phi_k(t)dt. \tag{15}$$

Then the total number of subsamples collected by all BMIs over all subperiods is MK. These subsamples can be gathered in the following $K \times M$ matrix

$$\boldsymbol{Y} = \begin{pmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,M} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,M} \\ \vdots & \vdots & \vdots & \vdots \\ y_{K,1} & y_{K,2} & \cdots & y_{K,M} \end{pmatrix} \tag{16}$$

where the kth row contains the subsamples obtained by correlating the measured signal with the waveform $\boldsymbol{\phi}_k$ over M subperiods each of length T/M.

The original K samples, i.e., the samples collected at BMIs over the whole time period T, can be obtained as

$$y_k = \sum_{m=1}^{M} [\boldsymbol{Y}]_{k,m} = \sum_{m=1}^{M} y_{k,m}, \quad k = 1, \ldots, K. \tag{17}$$

In order to construct additional samples, we consider column-wise permuted versions of $\boldsymbol{Y}$. The following definitions are then in order.

The *permutation* π is a one-to-one mapping of the elements of some set $\mathcal{D}$ to itself by simply changing the order of the elements. Then $\pi(k)$ stands for the index of the kth element in the permuted set. For example, let $\mathcal{D}$ consist of the elements of a $K \times 1$ vector $\boldsymbol{z}$, and the order of the elements in $\mathcal{D}$ is the same as in $\boldsymbol{z}$. After applying the permutation function π to $\boldsymbol{z}$, the permuted vector is $\boldsymbol{z}^{\pi} = (z_{\pi(1)}, \ldots, z_{\pi(k)}, \ldots, z_{\pi(K)})^T$. If vector $\boldsymbol{z}$ is itself the vector of indexes, i.e., $\boldsymbol{z} = (1, \ldots, K)^T$, then obviously $z_{\pi(k)} = \pi(k)$.

Different permuted versions of the subsample matrix $\boldsymbol{Y}$ can be obtained by applying different permutations to different columns of $\boldsymbol{Y}$. Specifically, let $\mathcal{P}^{(i)} = \{\pi_1^{(i)}, \ldots, \pi_j^{(i)}, \ldots, \pi_M^{(i)}\}$ be the ith set of column permutations with $\pi_j^{(i)}$ being the permutation function applied to the jth column of $\boldsymbol{Y}$, and let I stand for the number of such permutation sets. Then according to the above notations, the matrix resulting from applying the set of permutations $\mathcal{P}^{(i)}$ to the columns of $\boldsymbol{Y}$ can be expressed as $\boldsymbol{Y}^{\mathcal{P}^{(i)}} = (\boldsymbol{y}_1^{\pi_1^{(i)}}, \ldots, \boldsymbol{y}_j^{\pi_j^{(i)}}, \ldots, \boldsymbol{y}_M^{\pi_M^{(i)}})$ where $\boldsymbol{y}_j$ is the jth column of $\boldsymbol{Y}$.

Permutation sets $\mathcal{P}^{(i)}, i = 1, \ldots, I$ are chosen in such a way that all subsamples in a specific row of $\boldsymbol{Y}^{\mathcal{P}^{(i)}}$ come from different rows of the original subsample matrix $\boldsymbol{Y}$ as well as from different rows of other permuted matrices $\boldsymbol{Y}^{\mathcal{P}^{(1)}}, \ldots, \boldsymbol{Y}^{\mathcal{P}^{(i-1)}}$. For example, all subsamples in a specific row of $\boldsymbol{Y}^{\mathcal{P}^{(1)}}$ must come from different rows of the original matrix $\boldsymbol{Y}$ only, while the subsamples in a specific row of $\boldsymbol{Y}^{\mathcal{P}^{(2)}}$ come from different rows of $\boldsymbol{Y}$ and $\boldsymbol{Y}^{\mathcal{P}^{(1)}}$ and so on. This requirement is forced to make sure that any additional sample is correlated to any original or any other additional sample only over one segment. Then

the additional KI samples can be obtained based on the permuted matrices $\boldsymbol{Y}^{\mathcal{P}^{(i)}}, i=1,\ldots,I$ as

$$y_k^{\mathcal{P}^{(i)}} = \sum_{m=1}^{M} [\boldsymbol{Y}^{\mathcal{P}^{(i)}}]_{k,m}, \quad k=1,\ldots,K, \quad i=1,\ldots,I. \tag{18}$$

It is worth noting that in terms of the hardware structure, the subsamples used to generate additional samples must be chosen from different BMIs as well as different segments. This is equivalent to collecting additional samples by correlating the signal with additional sampling waveforms which are not present among the actual BMI sampling waveforms. Each of these additional sampling waveforms comprises the non-overlapping subperiods of M different original waveforms.

Now the question is how many permuted matrices, which satisfy the above conditions, can be generated based on $\boldsymbol{Y}$. Consider the following $K \times M$ matrix

$$\boldsymbol{Z} \triangleq \underbrace{(\boldsymbol{z}, \boldsymbol{z}, \ldots, \boldsymbol{z})}_{M \text{ times}} \tag{19}$$

where $\boldsymbol{z}$ is the vector of indexes. Applying the column permutation set $\mathcal{P}^{(i)}$ to the columns of $\boldsymbol{Z}$, we obtain a permuted matrix $\boldsymbol{Z}^{\mathcal{P}^{(i)}} = (\boldsymbol{z}^{\pi_1^{(i)}}, \ldots, \boldsymbol{z}^{\pi_j^{(i)}}, \ldots, \boldsymbol{z}^{\pi_M^{(i)}})$. Then the set of all permuted versions of $\boldsymbol{Z}$ can be denoted as $\mathcal{S}_{\boldsymbol{Z}} = \{\boldsymbol{Z}^{\mathcal{P}^{(1)}}, \ldots, \boldsymbol{Z}^{\mathcal{P}^{(I)}}\}$. With these notations, the following theorem is in order.

Theorem 1: The size of $\mathcal{S}_{\boldsymbol{Z}}$, i.e., the number I of permutation sets $\mathcal{P}^{(i)}, i=1,\ldots,I$ which satisfy the conditions

$$\left[\boldsymbol{Z}^{\mathcal{P}^{(i)}}\right]_{k,j} \neq \left[\boldsymbol{Z}^{\mathcal{P}^{(i)}}\right]_{k,r}, \ \forall \boldsymbol{Z}^{\mathcal{P}^{(i)}} \in \mathcal{S}_{\boldsymbol{Z}}, j \neq r, \ k \in \{1,\ldots,K\}, \ j,r \in \{1,\ldots,M\} \tag{20}$$

$$\exists! j \text{ or } \nexists j \text{ such that } \left[\boldsymbol{Z}^{\mathcal{P}^{(i)}}\right]_{k,j} = \left[\boldsymbol{Z}^{\mathcal{P}^{(l)}}\right]_{h,j}, \ \forall \boldsymbol{Z}^{\mathcal{P}^{(i)}}, \boldsymbol{Z}^{\mathcal{P}^{(l)}} \in \mathcal{S}_{\boldsymbol{Z}}, \ \boldsymbol{Z}^{\mathcal{P}^{(i)}} \neq \boldsymbol{Z}^{\mathcal{P}^{(l)}}, \forall j \in \{1,\ldots,M\} \ \forall k,h \in \{1,\ldots,K\} \tag{21}$$

is at most $K-1$.

Using the property that $z_{\pi(k)} = \pi(k)$ for the vector of indexes $\boldsymbol{z}$, the conditions (20) and (21) can also be expressed in terms of permutations as

$$\pi_j^{(i)}(k) \neq \pi_r^{(i)}(k) \quad \forall i \in \{1,\ldots,I\}, \ j \neq r, \ k \in \{1,\ldots,K\}, \ j,r \in \{1,\ldots,M\} \tag{22}$$

$$\exists! j \text{ or } \nexists j \text{ such that } \pi_j^{(i)}(k) = \pi_j^{(l)}(h) \quad \forall i,l \in \{1,\ldots,I\}, \ i \neq l, \ \forall j \in \{1,\ldots,M\}, \ \forall k,h \in \{1,\ldots,K\}. \tag{23}$$

Proof: See the Appendix. ∎

Example 1: Let the specific choice of index permutations be $\pi_s(k) = ((s+k-2) \bmod K)+1, s,k=1,\ldots,K$ with π_1 being the identity permutation and 'mod' standing for the modulo operation. For this specific choice, $\pi_j^{(i)} = \pi_{[i(j-1) \bmod K]+1}, \ i=1,\ldots,K-1, \ j=1,\ldots,M$. Consider the following matrix notation for the set $\mathcal{P}$ where the elements along the ith row are the permutations $\mathcal{P}^{(i)}, i=1,\ldots,I$

$$\mathcal{P} \triangleq \begin{pmatrix} \mathcal{P}^{(1)} \\ \mathcal{P}^{(2)} \\ \mathcal{P}^{(3)} \\ \vdots \\ \mathcal{P}^{(K-2)} \\ \mathcal{P}^{(K-1)} \end{pmatrix} = \begin{pmatrix} \pi_1^{(1)} & \pi_2^{(1)} & \pi_3^{(1)} & \cdots & \pi_M^{(1)} \\ \pi_1^{(2)} & \pi_2^{(2)} & \pi_3^{(2)} & \cdots & \pi_M^{(2)} \\ \pi_1^{(3)} & \pi_2^{(3)} & \pi_3^{(3)} & \cdots & \pi_M^{(3)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \pi_1^{(K-2)} & \pi_2^{(K-2)} & \pi_3^{(K-2)} & \cdots & \pi_M^{(K-2)} \\ \pi_1^{(K-1)} & \pi_2^{(K-1)} & \pi_3^{(K-1)} & \cdots & \pi_M^{(K-1)} \end{pmatrix} = \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 & \cdots & \pi_M \\ \pi_1 & \pi_3 & \pi_5 & \cdots & \pi_{[2(M-1) \bmod K]+1} \\ \pi_1 & \pi_4 & \pi_7 & \cdots & \pi_{[3(M-1) \bmod K]+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \pi_1 & \pi_{K-1} & \pi_{K-3} & \cdots & \pi_{[(K-2)(M-1) \bmod K]+1} \\ \pi_1 & \pi_K & \pi_{K-1} & \cdots & \pi_{[(K-1)(M-1) \bmod K]+1} \end{pmatrix}. \tag{24}$$

Note that not all permutations $\mathcal{P}^{(i)}, i=1,\ldots,I$ used in (24) may be permissible. In fact, the set of permutations $\mathcal{P}^{(i)}$ with $K/\gcd(i,K) < M$ has at least one repeated permutation that contradicts the condition (22). Here $\gcd(\cdot,\cdot)$ stands for the greatest common devisor of two numbers. For example, for $K=8$ and $M=4$, $K/\gcd(4,K) = 2 < M$ and $\mathcal{P}^{(4)}$ is impermissible. Therefore, instead of $K-1=7$, only the following 6 sets of permutations are allowed

$$\mathcal{P} = \begin{pmatrix} \pi_1^{(1)} & \pi_2^{(1)} & \pi_3^{(1)} & \pi_4^{(1)} \\ \pi_1^{(2)} & \pi_2^{(2)} & \pi_3^{(2)} & \pi_4^{(2)} \\ \pi_1^{(3)} & \pi_2^{(3)} & \pi_3^{(3)} & \pi_4^{(3)} \\ \pi_1^{(4)} & \pi_2^{(4)} & \pi_3^{(4)} & \pi_4^{(4)} \\ \pi_1^{(5)} & \pi_2^{(5)} & \pi_3^{(5)} & \pi_4^{(5)} \\ \pi_1^{(6)} & \pi_2^{(6)} & \pi_3^{(6)} & \pi_4^{(6)} \end{pmatrix} = \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 & \pi_4 \\ \pi_1 & \pi_3 & \pi_5 & \pi_7 \\ \pi_1 & \pi_4 & \pi_7 & \pi_2 \\ \pi_1 & \pi_6 & \pi_3 & \pi_8 \\ \pi_1 & \pi_7 & \pi_5 & \pi_3 \\ \pi_1 & \pi_8 & \pi_7 & \pi_6 \end{pmatrix}. \tag{25}$$

Theorem 1 shows how many different permuted versions of the original subsample matrix $\boldsymbol{Y}$ can be obtained such that the correlation between the original and additional samples would be minimal. Indeed, since the set of subsamples that are used to build additional samples is chosen in such a way that additional samples have at most one subsample in common with the previous samples, i.e., conditions (22) and (23) are satisfied, the set of permutations (24) is a valid candidate. The ith element of $\mathcal{P}$, i.e., the element $\mathcal{P}^{(i)} = (\pi_1^{(i)}, \ldots, \pi_M^{(i)})$, is the set of permutations applied to $\boldsymbol{Y}$ to obtain $\boldsymbol{Y}^{\mathcal{P}^{(i)}}$. Adding up the entries along the rows of $\boldsymbol{Y}^{\mathcal{P}^{(i)}}$, a set of K additional samples can be obtained.

Example 2: Let the number of additional samples K_a be at most K. This means that all permutations are given by only $\mathcal{P}^{(1)}$ in (24). In this special case, the subsample selection method can be given as follows. For constructing the $(K+1)$st sample, M subsamples on the main diagonal of $\boldsymbol{Y}$ are summed up together.

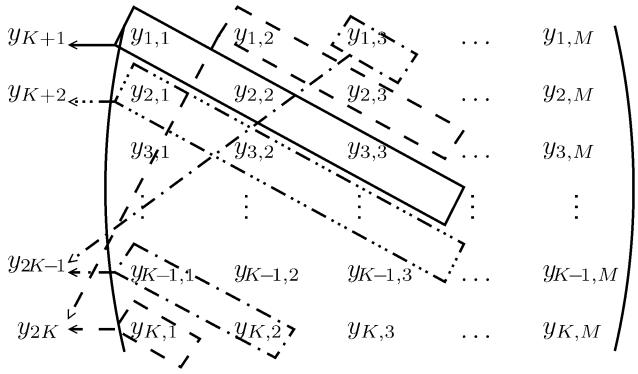


Fig. 2. Subsample selection principle for building additional samples in Example 2.

Then the M subsamples on the second diagonal are used to construct the $(K+2)$nd sample, and so on up to the $(K+K_a)$th sample. Mathematically, the so constructed additional samples can be expressed in terms of the elements of $\boldsymbol{Y}$ as

$$y_{K+k} = \sum_{m=1}^{M} y_{l,m}, \quad k = 1 \ldots, K_a \tag{26}$$

where $l = [(k+m-2) \bmod K]+1$ and $K_a \leq K$. Fig. 2 shows schematically how the subsamples are selected in this example.

The proposed segmented sampling process can be equivalently expressed in terms of the measurement matrix. Let $\boldsymbol{\Phi}$ be the original $K \times N$ measurement matrix. Let the kth row of the matrix $\boldsymbol{\Phi}$ be $\boldsymbol{\phi}_k = (\boldsymbol{\phi}_{k,1}, \ldots, \boldsymbol{\phi}_{k,M})$ where $\boldsymbol{\phi}_{k,j}$, $j = 1, \ldots, M$ are some vectors. Let also for simplicity, the length of $\boldsymbol{\phi}_{k,j}$ be N/M where N/M be a positive integer number. The set of permutations applied to $\boldsymbol{Y}$ in order to obtain $\boldsymbol{Y}^{\mathcal{P}^{(i)}}$ is $\mathcal{P}^{(i)}$. Then the operation $\boldsymbol{\Phi}^{\mathcal{P}^{(i)}}$ can be expressed as follows. The first group of N/M columns of $\boldsymbol{\Phi}$, which are the vectors $\boldsymbol{\phi}_{k,1}$, $k = 1, \ldots, K$, are permuted with $\pi_1^{(i)}$. The second group of N/M columns of $\boldsymbol{\Phi}$ are permuted with $\pi_2^{(i)}$ and so on until the last group of N/M columns of $\boldsymbol{\Phi}$ which are permuted with $\pi_M^{(i)}$. Then the $K_e \times N$ extended measurement matrix which combines all possible permutations $\mathcal{P}^{(i)}, i = 1, \ldots, I$ can be expressed as

$$\boldsymbol{\Phi}_e = \left(\boldsymbol{\Phi}^T, \left(\boldsymbol{\Phi}^{\mathcal{P}^{(1)}} \right)^T, \ldots, \left(\boldsymbol{\Phi}^{\mathcal{P}^{(I)}} \right)^T \right)^T \tag{27}$$

where $K_e \triangleq K + K_a = K + KI$.

Example 3: Continuing with the set up used in Example 2, let $K_a \leq K$. Then the extended measurement matrix is

$$\begin{aligned} \boldsymbol{\Phi}_e &= \begin{pmatrix} \boldsymbol{\Phi} \\ \boldsymbol{\Phi}_1 \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{\phi}_{1,1} & \boldsymbol{\phi}_{1,2} & \cdots & \boldsymbol{\phi}_{1,M} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\phi}_{K,1} & \boldsymbol{\phi}_{K,2} & \cdots & \boldsymbol{\phi}_{K,M} \\ \boldsymbol{\phi}_{1,1} & \boldsymbol{\phi}_{2,2} & \cdots & \boldsymbol{\phi}_{M,M} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\phi}_{K_a,1} & \boldsymbol{\phi}_{\pi_2(K_a),M} & \cdots & \boldsymbol{\phi}_{\pi_M(K_a),M} \end{pmatrix} \end{aligned} \tag{28}$$

where $\boldsymbol{\Phi}_1$ contains only K_a rows of $\boldsymbol{\Phi}^{\mathcal{P}^{(1)}}$ and $\boldsymbol{\Phi}_1 = \boldsymbol{\Phi}^{\mathcal{P}^{(1)}}$ if $K_a = K$.

B. Implementation Issues and Discussion

Due to the special structure of the extended measurement matrix $\boldsymbol{\Phi}_e$, the sampling hardware needs only K parallel BMIs for collecting KI samples. These BMIs are essentially the same as those in Fig. 1. The only difference is that the integration period T is divided into M equal subperiods. At the end of every subperiod, each integrator's output is sampled and the integrator is reset. Therefore, some factors, which may influence the complexity of a hardware implementation of the proposed segmented AIC, are the following. Since the sampling rate of the segmented AIC is M times higher than that of the conventional AIC with the same number of BMIs, the segmented AIC complexity can slightly increase as compared to the conventional AIC with K BMIs. However, the rate increased in $M \leq K$ times is still by far less than, for example, the required Nyquist rate which depends on the signal bandwidth. As compared to the AIC with KI BMIs, the segmented AIC has only K BMIs, that makes the complexity of the segmented AIC for collecting KI samples significantly smaller than that of the conventional AIC with KI BMIs. In addition, a multiplexer which selects the subsamples for constructing additional samples is needed in the proposed segmented AIC. It is worth noting, however, that partial sums can be kept for constructing the samples (original and additional), that is, the results of the integration are updated and accumulated for each sample iteratively after each subperiod. In this way, there is no need of designing the circuitry for memorizing the matrix of subsamples $\boldsymbol{Y}$, but only the partial sums for each sample are memorized at any current subperiod. One more factor which may have an effect on the performance of the segmented AIC is the hardware sampling noise introduced at time instances mT/M, $m = 1, \ldots, M$ when the output of each BMI is sampled to collect a subsample. This sampling noise appears M times over a time period T for the segmented AIC while it appears once over T for the conventional AIC. However, the amount of the hardware sampling noise depends on the specific hardware implementation of the sampler and is out of the scope of this paper.

Finally, it is worth noting that the possibility of improving the signal recovery performance due to increasing the sampling rate in each BMI of the proposed segmented AIC agrees with the convention that the recovery performance cannot be improved only due to the post processing. Moreover, note that since the original random sampling waveforms are linearly independent with high probability, the additional sampling waveforms of our segmented CS method are also linearly independent with overwhelming probability. However, a sufficient condition that guarantees that the extended measurement matrix of the proposed segmented CS-based AIC scheme is a valid choice is the RIP. Therefore, the RIP for the proposed segmented CS method is analyzed in the next section.

IV. RIP For the Segmented Compressed Sampling Method

The purpose of this section is to show that the extended measurement matrix $\boldsymbol{\Phi}_e$ in (27) satisfies the RIP if the original measurement matrix $\boldsymbol{\Phi}$ satisfies it. The latter will also imply that $\boldsymbol{\Phi}_e$ can be used as a valid CS measurement matrix. In our setup it

is only assumed that the elements of the original measurement matrix are i.i.d. zero mean Gaussian variables and the measurement matrix is extended by constructing its permuted versions as described in the previous section.

Let us first consider the special case of Example 3. In this case, $\mathbf{\Phi}$, $\mathbf{\Phi}_1$, and $\mathbf{\Phi}_e$ are the original measurement matrix, the matrix of additional sampling waveforms, and the extended measurement matrix given by (28), respectively. Let the matrix $\mathbf{\Phi}$ satisfy the RIP with sufficiently high probability. For example, let the elements of $\mathbf{\Phi}$ be i.i.d. zero mean Gaussian random variables with variance $1/N$. Let $\mathcal{T}$ be any subset of size S of the set $\{1,\ldots,N\}$. Then for any $0 < \delta_S < 1$, the matrix $\mathbf{\Phi}_{\mathcal{T}}$, which is a submatrix of $\mathbf{\Phi}$ which consists of only the columns with their indexes in the set $\mathcal{T}$ satisfies (7) with the following probability [19]

$$\Pr\{\mathbf{\Phi}_{\mathcal{T}} \text{ satisfies (7)}\} \geq 1 - 2(12/\delta_S)^S e^{-C_0(\delta_S/2)K} \quad (29)$$

where $C_0(\delta_S/2) = \delta_S^2/16 - \delta_S^3/48$. Hereafter, the notation C_0 is used instead of $C_0(\delta_S/2)$ for brevity.

First, the following auxiliary result on the extended measurement matrix $\mathbf{\Phi}_e$ is of interest.

Lemma 1: Let the elements of the measurement matrix $\mathbf{\Phi}$ be i.i.d. zero mean Gaussian variables with variance $1/N$, $\mathbf{\Phi}_e$ be formed as shown in (28), and $\mathcal{T} \subset \{1,\ldots,N\}$ of size S. If K_a is chosen so that $\min\{K, K_a + M - 1\} \leq \lceil (K + K_a)/2 \rceil$, then for any $0 < \delta_S < 1$, the following inequality holds

$$\Pr\{(\mathbf{\Phi}_e)_{\mathcal{T}} \text{ satisfies (7)}\} \geq 1 - 4(12/\delta_S)^S e^{-C_0 \lfloor \frac{K+K_a}{2} \rfloor} \quad (30)$$

where $\lceil x \rceil$ and $\lfloor x \rfloor$ are the smallest integer larger than or equal to x and the largest integer smaller than or equal to x, respectively, and C_0 is a constant given after (29).

Proof: See the Appendix. ∎

Using the above lemma, the following main result, which states that the extended measurement matrix $\mathbf{\Phi}_e$ in (28) satisfies the RIP with overwhelming probability, is in order.

Theorem 2: Let $\mathbf{\Phi}_e$ be formed as in (28) and let the elements of $\mathbf{\Phi}$ be i.i.d. zero mean Gaussian variables with variance $1/N$. If $\min\{K, K_a + M - 1\} \leq \lceil (K + K_a)/2 \rceil$, then for any $0 < \delta_S < 1$, there exist constants C_3 and C_4, which depend only on δ_S, such that for $S \leq C_3 \lfloor (K + K_a)/2 \rfloor / \log(N/S)$ the inequality (7) holds for all S-sparse vectors with probability that satisfies the following inequality:

$$\Pr\{\mathbf{\Phi}_e \text{ satisfies RIP}\} \geq 1 - 4e^{-C_4 \lfloor (K+K_a)/2 \rfloor} \quad (31)$$

where $C_4 = C_0 - C_3[1 + (1 + \log(12/\delta_S))/\log(N/S)]$ and C_3 is small enough that guarantees that C_4 is positive.

Proof: See the Appendix. ∎

Let us consider now the general case when the number of additional samples K_a is larger than the number of BMIs K, i.e., $K_a > K$ and $K_e > 2K$, and the extended measurement matrix is given by (27). Note that while proving Lemma 1 for the special case of Example 3, we were able to split the rows of $\mathbf{\Phi}_e$ into two sets each consisting of independent entries. In the general case, some of the entries of the original measurement matrix appear more than twice in the extended measurement matrix $\mathbf{\Phi}_e$, and it is no longer possible to split the rows of $\mathbf{\Phi}_e$ into only two sets with independent entries. Because of the way the additional samples are built, the samples $y_{lK+1}, y_{lK+2}, \ldots, y_{(l+1)K}$ obtained based on the permuted matrix $\boldsymbol{Y}^{\mathcal{P}^{(l)}}$ are uncorrelated with each other, but they are correlated with every other set of samples obtained based on the original matrix $\boldsymbol{Y}$ and the permuted matrices $\boldsymbol{Y}^{\mathcal{P}^{(i)}}$, $\forall i, i \neq l$. Thus, the following principle can be used for partitioning the rows of $\mathbf{\Phi}_e$ into the sets with independent entries. First, the rows corresponding to the original samples form a single set with independent entries, then the rows corresponding to the first set of additional samples based on the matrix $\boldsymbol{Y}^{\mathcal{P}^{(1)}}$ form another set and so on. Then the number of such sets is $n_p = \lceil K_e/K \rceil$, while the size of each set is

$$K_i = \begin{cases} K, & 1 \leq i < \left\lceil \frac{K_e}{K} \right\rceil - 1 \\ K_e - \left(\left\lceil \frac{K_e}{K} \right\rceil - 1 \right) K, & i = \left\lceil \frac{K_e}{K} \right\rceil. \end{cases} \quad (32)$$

The extended measurement matrix (27) can be rewritten as

$$\mathbf{\Phi}_e = \left((\mathbf{\Phi}_e)_1^T, (\mathbf{\Phi}_e)_2^T, \ldots, (\mathbf{\Phi}_e)_{n_p}^T \right)^T \quad (33)$$

where $(\mathbf{\Phi}_e)_i$ is the ith partition of $\mathbf{\Phi}_e$ of size given by (32). Then the general form of Lemma 1 is as follows.

Lemma 2: Let the elements of the measurement matrix $\mathbf{\Phi}$ be i.i.d. zero mean Gaussian variables with variance $1/N$, $\mathbf{\Phi}_e$ be the extended measurement matrix (27), and $\mathcal{T} \subset \{1,\ldots,N\}$ be of size S. Let also $K_a > K$ and $n_p = \lceil K_e/K \rceil$. Then, for any $0 < \delta_S < 1$, the following inequality holds

$$\Pr\{(\mathbf{\Phi}_e)_{\mathcal{T}} \text{ satisfies (7)}\} \geq 1 - 2(n_p - 1)(12/\delta_S)^S \times (e^{-C_0 K}) - 2(12/\delta_S)^S (e^{-C_0 K_{n_p}}) \quad (34)$$

where $K_{n_p} = K_e - (\lceil (K_e/K) \rceil - 1)K$ and C_0 is a constant given after (29).

Proof: See the Appendix. ∎

Lemma 2 is needed to prove that the extended measurement matrix (33) satisfies the RIP with overwhelming probability. Therefore, the general version of Theorem 2 is as follows.

Theorem 3: Let the elements of $\mathbf{\Phi}$ be i.i.d. zero mean Gaussian variables with variance $1/N$ and $\mathbf{\Phi}_e$ be formed as in (27). If $K_a > K$, then there exist constants C_3, C_4, and C_4' for any $0 < \delta_S < 1$, such that for $S \leq C_3 K_{n_p}/\log(N/S)$ the inequality (7) holds for all S-sparse vectors with probability that satisfies the following inequality:

$$\Pr\{\mathbf{\Phi}_e \text{ satisfies RIP}\} \geq 1 - 2(n_p - 1)e^{-C_4' K} - 2e^{-C_4 K_{n_p}} \quad (35)$$

where $C_4' = C_0 - (C_3 K_{n_p}/K)[1 + (1 + \log(12/\delta_S))/\log(N/S)]$, C_4 is given after (31), and C_3 is small enough to guarantee that C_4 and C_4' are both positive.

Proof: See Appendix. ∎

When splitting the rows of $\mathbf{\Phi}_e$ in a number of sets as described before Lemma 2, it may happen that the last subset $(\mathbf{\Phi}_e)_{n_p}$ has the smallest size K_{n_p}. As a result, the dominant term in (35) will likely be the term $2e^{-C_4 K_{n_p}}$. It may lead to a more stringent sparsity condition, that is, $S \leq C_3 K_{n_p}/\log(N/S)$. To improve the lower bound in (35), we can move some of the rows from $(\mathbf{\Phi}_e)_{n_p - 1}$ to $(\mathbf{\Phi}_e)_{n_p}$ in order to make the last

two partitions of almost the same size. Then the requirement on the sparsity level will become $S \leq C_3 K'/\log(N/S)$ where $K' = \lfloor (K + K_{n_p})/2 \rfloor$. Therefore, the lower bound on the probability calculated in (35) improves.

V. Performance Analysis of the Recovery

In this section, we aim at answering the question whether signal recovery also improves if the proposed segmented CS method, i.e., the extended measurement matrix $\boldsymbol{\Phi}_e$ (27), is used instead of the original matrix $\boldsymbol{\Phi}$. The study is performed based on the empirical risk minimization method for signal recovery from noisy random projections [4]. As mentioned in Section II, the LASSO method can be viewed as one of the possible implementations of the empirical risk minimization method.

We first consider the special case of Example 3 when the extended measurement matrix is given by (28). Let the entries of the measurement matrix $\boldsymbol{\Phi}$ be selected with equal probability as $\pm 1/\sqrt{N}$, i.e., be i.i.d. Bernoulli distributed with variance $1/N$. The Bernoulli case is used here in order to keep our derivations short by only emphasizing the differences caused by our construction of matrix $\boldsymbol{\Phi}_e$ with correlated rows as compared to the case analyzed in [4], where the measurement matrix consists of all i.i.d. entries. Moreover, the Bernoulli case is the one which is practically appealing. Note that our results can be easily applied to the case of Gaussian distributed entries of $\boldsymbol{\Phi}$ by only changing the moments of Bernoulli distribution to the moments of Gaussian distribution.

Let $r(\hat{\boldsymbol{f}}, \boldsymbol{f}) \triangleq r(\hat{\boldsymbol{f}}) - r(\boldsymbol{f})$ be the "excess risk" between the candidate reconstruction $\hat{\boldsymbol{f}}$ of the signal sampled using the extended measurement matrix $\boldsymbol{\Phi}_e$ and the actual signal $\boldsymbol{f}$, and $\hat{r}(\hat{\boldsymbol{f}}, \boldsymbol{f}) \triangleq \hat{r}(\hat{\boldsymbol{f}}) - \hat{r}(\boldsymbol{f})$ be the "empirical excess risk" between the candidate signal reconstruction and the actual signal. Then the difference between the "excess risk" and the "empirical excess risk" can be found as

$$r(\hat{\boldsymbol{f}}, \boldsymbol{f}) - \hat{r}(\hat{\boldsymbol{f}}, \boldsymbol{f}) = \frac{1}{K_e} \sum_{j=1}^{K_e} (U_j - E[U_j]) \tag{36}$$

where $U_j \triangleq (y_j - \boldsymbol{\phi}_j \boldsymbol{f})^2 - (y_j - \boldsymbol{\phi}_j \hat{\boldsymbol{f}})^2$.

The MSE between the candidate reconstruction and the actual signal can be expressed as [22]

$$\text{MSE} \triangleq E\{\|\boldsymbol{g}\|^2\} = N r(\hat{\boldsymbol{f}}, \boldsymbol{f}) \tag{37}$$

where $\boldsymbol{g} \triangleq \hat{\boldsymbol{f}} - \boldsymbol{f}$. Therefore, if we know an upper bound on the right-hand side of (36), denoted hereafter as U, we can immediately find an upper bound on the MSE in the form $\text{MSE} \leq N\hat{r}(\hat{\boldsymbol{f}}, \boldsymbol{f}) + NU$. In other words, to find the candidate reconstruction $\hat{\boldsymbol{f}}$, one can minimize $\hat{r}(\hat{\boldsymbol{f}}, \boldsymbol{f}) + U$. This minimization will result in a bound on the MSE as in (12).

The Craig–Bernstein inequality [4], [24] can be used in order to find an upper bound U on the right-hand side of (36). In our notations the Craig–Bernstein inequality states that the probability of the following event:

$$\frac{1}{K_e} \sum_{j=1}^{K_e} (U_j - E\{U_j\}) \leq \frac{\log\left(\frac{1}{\delta}\right)}{K_e \epsilon} + \frac{\epsilon \operatorname{var}\left\{\sum_{j=1}^{K_e} U_j\right\}}{2K_e(1-\zeta)} \tag{38}$$

is greater than or equal to $1 - \delta$ for $0 < \epsilon h \leq \zeta < 1$, if the random variables U_j satisfy the following moment condition for some $h > 0$ and for all $k \geq 2$

$$E\left\{|U_j - E\{U_j\}|^k\right\} \leq \frac{k! \operatorname{var}\{U_j\} h^{k-2}}{2}. \tag{39}$$

The second term in the right-hand side of (38) contains the variance $\operatorname{var}\{\sum_{j=1}^{K_e} U_j\}$, which we need to calculate or at least find an upper bound on it.

In the case of the extended measurement matrix, the random variables U_j, $j = 1, \ldots, K_e$ all satisfy the moment condition for the Craig-Bernstein inequality [24] with the same coefficient $h = 16B^2 e + 8\sqrt{2}B\sigma$ where σ^2 is the variance of the Gaussian noise.[4] Moreover, it is easy to show that the following bound on the variance of U_j is valid for the extended measurement matrix

$$\operatorname{var}\{U_j\} \leq \left(2\frac{\|\boldsymbol{g}\|^2}{N} + 4\sigma^2\right) \frac{\|\boldsymbol{g}\|^2}{N} \leq (8B^2 + 4\sigma^2) r(\hat{\boldsymbol{f}}, \boldsymbol{f}). \tag{40}$$

However, unlike [4], in the case of the extended measurement matrix, the variables U_j are not independent from each other. Thus, we can not simply replace the term $\operatorname{var}\{\sum_{j=1}^{K_e} U_j\}$ with the sum of the variances for U_j, $j = 1, \ldots, K_e$. Using the definition of the variance, we can write that

$$\begin{aligned}
&\operatorname{var}\left\{\sum_{j=1}^{K_e} U_j\right\} \\
&\quad \triangleq E\left\{\left(\sum_{j=1}^{K_e} U_j\right)^2\right\} - \left(E\left\{\sum_{j=1}^{K_e} U_j\right\}\right)^2 \\
&\quad = \sum_{j=1}^{K_e} E\{U_j^2\} + 2\sum_{i=1}^{K_e-1} \sum_{j=i+1}^{K_e} E\{U_i U_j\} - K_e^2 \left(\frac{\|\boldsymbol{g}\|^2}{N}\right)^2 \\
&\quad = \sum_{j=1}^{K_e} \left(E\left\{U_j^2\right\} - \left(\frac{\|\boldsymbol{g}\|^2}{N}\right)^2\right) \\
&\qquad + 2\sum_{i=1}^{K_e-1} \sum_{j=i+1}^{K_e} \left(E\{U_i U_j\} - \left(\frac{\|\boldsymbol{g}\|^2}{N}\right)^2\right) \\
&\quad = \sum_{j=1}^{K_e} \operatorname{var}\{U_j\} + 2\sum_{i=1}^{K_e-1} \sum_{j=i+1}^{K_e} \left(E\{U_i U_j\} - \left(\frac{\|\boldsymbol{g}\|^2}{N}\right)^2\right)
\end{aligned} \tag{41}$$

where the upper bound on $\operatorname{var}\{U_j\}$ is given by (40). Using the fact following form the noisy model (6) that the random noise components w_i and w_j are independent from $\boldsymbol{\phi}_i \boldsymbol{g}$ and $\boldsymbol{\phi}_j \boldsymbol{g}$, respectively, $E\{U_i U_j\}$ can be expressed as

$$\begin{aligned}
&E\{U_i U_j\} \\
&\quad = E\{[2w_i \boldsymbol{\phi}_i \boldsymbol{g} - (\boldsymbol{\phi}_i \boldsymbol{g})^2][2w_j \boldsymbol{\phi}_j \boldsymbol{g} - (\boldsymbol{\phi}_j \boldsymbol{g})^2]\} \\
&\quad = 4E\{w_i w_j\} E\{\boldsymbol{\phi}_i \boldsymbol{g} \boldsymbol{\phi}_j \boldsymbol{g}\} - 2E\{w_i\} E\{\boldsymbol{\phi}_i \boldsymbol{g} (\boldsymbol{\phi}_j \boldsymbol{g})^2\} \\
&\qquad - 2E\{w_j\} E\{\boldsymbol{\phi}_j \boldsymbol{g} (\boldsymbol{\phi}_i \boldsymbol{g})^2\} + E\{(\boldsymbol{\phi}_i \boldsymbol{g})^2 (\boldsymbol{\phi}_j \boldsymbol{g})^2\}.
\end{aligned} \tag{42}$$

[4]The derivation of the coefficient h coincides with a similar derivation in [4], and therefore, is omitted.

The latter expression can be further simplified using the fact that $E\{w_i\} = E\{w_j\} = 0$. Thus, we obtain that

$$E\{U_iU_j\} = 4E\{w_iw_j\}E\{(\boldsymbol{\phi}_i\boldsymbol{g})(\boldsymbol{\phi}_j\boldsymbol{g})\} + E\{(\boldsymbol{\phi}_i\boldsymbol{g})^2(\boldsymbol{\phi}_j\boldsymbol{g})^2\}. \quad (43)$$

It is easy to verify that if $\boldsymbol{\phi}_i$ and $\boldsymbol{\phi}_j$ are independent, then $E(U_iU_j) = E\{(\boldsymbol{\phi}_i\boldsymbol{g})^2\}E\{(\boldsymbol{\phi}_j\boldsymbol{g})^2\} = (\|\boldsymbol{g}\|^2/\ N)^2$ as in [4]. However, in our case, $\boldsymbol{\phi}_i$ and $\boldsymbol{\phi}_j$ may depend on each other. If they indeed depend on each other, they have $L = N/M$ common entries, while the rest of the entries are independent. In addition, the additive noise terms w_i and w_j are no longer independent random variables as well and, thus, $E\{w_iw_j\} = \sigma^2/M$. Without loss of generality, let the first L entries of $\boldsymbol{\phi}_i$ and $\boldsymbol{\phi}_j$ be the same, that is,

$$\boldsymbol{\phi}_i\boldsymbol{g} = \overbrace{g_1a_1 + \ldots + g_La_L}^{A} + \overbrace{g_{L+1}\phi_{i,L+1} + \cdots + g_N\phi_{i,N}}^{P_i} \quad (44)$$

$$\boldsymbol{\phi}_j\boldsymbol{g} = \overbrace{g_1a_1 + \cdots + g_La_L}^{A} + \overbrace{g_{L+1}\phi_{j,L+1} + \cdots + g_N\phi_{j,N}}^{P_j} \quad (45)$$

with $a_1, \ldots, a_L$ being the common part between $\boldsymbol{\phi}_i$ and $\boldsymbol{\phi}_j$.

Let $\boldsymbol{g}_A$ be a subvector of $\boldsymbol{g}$ containing the L elements of $\boldsymbol{g}$ corresponding to the common part between $\boldsymbol{\phi}_i$ and $\boldsymbol{\phi}_j$, and $\boldsymbol{g}_{A'}$ be the subvector comprising the rest of the elements. Then using the fact that A, P_i, and P_j are all zero mean independent random variables, we can express $E\{(\boldsymbol{\phi}_i\boldsymbol{g})(\boldsymbol{\phi}_j\boldsymbol{g})\}$ from the first term on the right-hand side of (43) as

$$\begin{aligned} &E\{(\boldsymbol{\phi}_i\boldsymbol{g})(\boldsymbol{\phi}_j\boldsymbol{g})\} \\ &\quad = E\{(A+P_i)(A+P_j)\} \\ &\quad = E\{A^2\} + E\{AP_i\} + E\{AP_j\} + E\{P_iP_j\} \\ &\quad = E\{A^2\} = \frac{\left(\sum_{k=1}^{L} g_k^2\right)^2}{N} = \frac{\|\boldsymbol{g}_A\|^2}{N}. \end{aligned} \quad (46)$$

Similar, the second term on the right-hand side of (43) can be expressed as

$$\begin{aligned} &E\{(\boldsymbol{\phi}_i\boldsymbol{g})^2(\boldsymbol{\phi}_j\boldsymbol{g})^2\} \\ &\qquad = E\{(A^2 + P_i^2 + 2AP_i)(A^2 + P_j^2 + 2AP_j)\}. \end{aligned} \quad (47)$$

Using the facts that $4E\{w_iw_j\} = 4\sigma^2/M$, $E\{A^2\} = \|\boldsymbol{g}_A\|^2/N$, and $E\{P_i^2\} = \|\boldsymbol{g}_{A'}\|^2/N$, the expression (47) can be rewritten as

$$\begin{aligned} &E\{(\boldsymbol{\phi}_i\boldsymbol{g})^2(\boldsymbol{\phi}_j\boldsymbol{g})^2\} \\ &\quad = E\left\{A^4 + A^2P_i^2 + A^2P_j^2 + P_i^2P_j^2\right\} \\ &\quad = E\{A^4\} + 2\frac{\|\boldsymbol{g}_A\|^2}{N}\cdot\frac{\|\boldsymbol{g}_{A'}\|^2}{N} + \left(\frac{\|\boldsymbol{g}_{A'}\|^2}{N}\right)^2 \\ &\quad = E\{A^4\} + \left(\frac{\|\boldsymbol{g}\|^2}{N}\right)^2 - \left(\frac{\|\boldsymbol{g}_A\|^2}{N}\right)^2. \end{aligned} \quad (48)$$

Substituting (46) and (48) into (43), we obtain that

$$E\{U_iU_j\} = \frac{4\sigma^2}{M}\cdot\frac{\|\boldsymbol{g}_A\|^2}{N} + E\{A^4\} + \left(\frac{\|\boldsymbol{g}\|^2}{N}\right)^2 - \left(\frac{\|\boldsymbol{g}_A\|^2}{N}\right)^2. \quad (49)$$

Moreover, substituting (49) into (41), we find that

$$\begin{aligned} \operatorname{var}\left\{\sum_{j=1}^{K_e} U_j\right\} &= \sum_{j=1}^{K_e}\operatorname{var}\{U_j\} + 2 \\ &\times \sum_{\boldsymbol{\phi}_i,\boldsymbol{\phi}_j\,\text{dependent}} \left(E\{A^4\} - \left(\frac{\|\boldsymbol{g}_A\|^2}{N}\right)^2 + \frac{4\sigma^2}{M}\cdot\frac{\|\boldsymbol{g}_A\|^2}{N}\right). \end{aligned} \quad (50)$$

Since the extended measurement matrix is constructed so that the waveforms $\{\boldsymbol{\phi}_i\}_{i=K+1}^{K_e}$ are built upon M rows of the original matrix and using then the inequality[5] $E\{A^4\} - (\|\boldsymbol{g}_A\|^2/N)^2 \leq 2(\|\boldsymbol{g}_A\|^2/N)^2$ for all these M rows, we obtain for every additional $\boldsymbol{\phi}_i$ that

$$\begin{aligned} &\sum_{k=1}^{M}\left(E\{A^4\} - \left(\frac{\|\boldsymbol{g}_A\|^2}{N}\right)^2 + \frac{4\sigma^2}{M}\cdot\frac{\|\boldsymbol{g}_A\|^2}{N}\right) \\ &\qquad \leq \sum_{k=1}^{M}\left(2\left(\frac{\|\boldsymbol{g}_A\|^2}{N}\right)^2 + \frac{4\sigma^2}{M}\cdot\frac{\|\boldsymbol{g}_A\|^2}{N}\right) \end{aligned} \quad (51)$$

where $\boldsymbol{g}_A$ corresponds to the first L entries of $\boldsymbol{g}$ for $k = 1$, to the entries from $L+1$ to $2L$ for $k = 2$ and so on. Applying also the triangle inequality, we find that

$$\begin{aligned} &\sum_{k=1}^{M}\left(2\left(\frac{\|\boldsymbol{g}_A\|^2}{N}\right)^2 + \frac{4\sigma^2}{M}\cdot\frac{\|\boldsymbol{g}_A\|^2}{N}\right) \\ &\qquad \leq 2\left(\frac{\|\boldsymbol{g}\|^2}{N}\right)^2 + \frac{4\sigma^2}{M}\cdot\frac{\|\boldsymbol{g}\|^2}{N}. \end{aligned} \quad (52)$$

Combining (51) and (52) and using the fact that there are K_a additional rows in the extended measurement matrix, we obtain that

$$\begin{aligned} &2\sum_{\boldsymbol{\phi}_i,\boldsymbol{\phi}_j\,\text{dependent}}\left(E\{A^4\} - \left(\frac{\|\boldsymbol{g}_A\|^2}{N}\right)^2 + \frac{4\sigma^2}{M}\cdot\frac{\|\boldsymbol{g}_A\|^2}{N}\right) \\ &\qquad \leq 4K_a\left(\frac{\|\boldsymbol{g}\|^2}{N}\right)^2 + \frac{8\sigma^2K_a}{M}\cdot\frac{\|\boldsymbol{g}\|^2}{N}. \end{aligned} \quad (53)$$

Noticing that $\|\boldsymbol{g}\|^2/N = r(\hat{\boldsymbol{f}}, \boldsymbol{f})$ and $\|\boldsymbol{g}\|^2 \leq 4NB^2$, the right-hand side of the inequality (53) can be further upper bounded as

$$\begin{aligned} &4K_a\left(\frac{\|\boldsymbol{g}\|^2}{N}\right)^2 + \frac{8\sigma^2K_a}{M}\cdot\frac{\|\boldsymbol{g}\|^2}{N} \\ &\qquad \leq 16K_aB^2\,r(\hat{\boldsymbol{f}}, \boldsymbol{f}) + \frac{8\sigma^2K_a}{M}\,r(\hat{\boldsymbol{f}}, \boldsymbol{f}). \end{aligned} \quad (54)$$

[5]We skip the derivation of this inequality since it is relatively well known and can be found, for example, in [4, p. 4039].

Using the upper bound (54) for the second term in (50) and the upper bound (40) for the first term in (50), we finally find the upper bound for $\mathrm{var}\{\sum_{j=1}^{K_e} U_j\}$ as

$$\mathrm{var}\left\{\sum_{j=1}^{K_e} U_j\right\} \leq K_e\left(8B^2\left(1+\frac{2K_a}{K_e}\right) +4\sigma^2\left(1+\frac{2K_a}{MK_e}\right)\right) r(\hat{\boldsymbol{f}},\boldsymbol{f}). \quad (55)$$

Therefore, based on the Craig–Bernstein inequality, the probability that for a given candidate signal $\hat{\boldsymbol{f}}$ the following inequality holds

$$r(\hat{\boldsymbol{f}},\boldsymbol{f}) - \hat{r}(\hat{\boldsymbol{f}},\boldsymbol{f}) \leq \frac{\log(\frac{1}{\delta})}{K_e\epsilon} + \frac{\left(8B^2\left(1+\frac{2K_a}{K_e}\right)+4\sigma^2\left(1+\frac{2K_a}{MK_e}\right)\right) r(\hat{\boldsymbol{f}},\boldsymbol{f})\epsilon}{2(1-\zeta)} \quad (56)$$

is greater than or equal to $1-\delta$.

Let $c(\hat{\boldsymbol{f}})$ be chosen such that the Kraft inequality $\sum_{\hat{\boldsymbol{f}}\in\mathcal{F}(B)} 2^{c(\hat{\boldsymbol{f}})} \leq 1$ is satisfied (see also [4]), and let $\delta(\hat{\boldsymbol{f}}) = 2^{-c(\hat{\boldsymbol{f}})}\,\delta$. Applying the union bound to (56), it can be shown that for all $\hat{\boldsymbol{f}} \in \mathcal{F}(B)$ and for all $\delta > 0$, the following inequality holds with probability of at least $1-\delta$

$$r(\hat{\boldsymbol{f}},\boldsymbol{f}) - \hat{r}(\hat{\boldsymbol{f}},\boldsymbol{f}) \leq \frac{c(\hat{\boldsymbol{f}})\log 2 + \log(\frac{1}{\delta})}{K_e\epsilon} + \frac{\left(8B^2\left(1+\frac{2K_a}{K_e}\right)+4\sigma^2\left(1+\frac{2K_a}{MK_e}\right)\right) r(\hat{\boldsymbol{f}},\boldsymbol{f})\epsilon}{2(1-\zeta)}. \quad (57)$$

Finally, setting $\zeta = \epsilon h$ and

$$a = \frac{\left(8B^2\left(1+\frac{2K_a}{K_e}\right)+4\sigma^2\left(1+\frac{2K_a}{MK_e}\right)\right)\epsilon}{2(1-\zeta)} \quad (58)$$

$$\epsilon < \frac{1}{\left(4\left(1+\frac{2K_a}{K_e}\right)+16e\right)B^2 + 8\sqrt{B}\sigma + 2\sigma^2\left(1+\frac{2K_a}{MK_e}\right)} \quad (59)$$

where $0 < \epsilon h \leq \zeta < 1$ as required by the Craig-Bernstein inequality, the following inequality holds with probability of at least $1-\delta$ for all $\hat{\boldsymbol{f}} \in \mathcal{F}(B)$:

$$(1-a)r(\hat{\boldsymbol{f}},\boldsymbol{f}) \leq \hat{r}(\hat{\boldsymbol{f}},\boldsymbol{f}) + \frac{c(\hat{\boldsymbol{f}})\log 2 + \log(\frac{1}{\delta})}{K_e\epsilon}. \quad (60)$$

The following result on the recovery performance of the empirical risk minimization method is in order.

Theorem 4: Let ϵ be chosen as

$$\epsilon = \frac{1}{(60(B+\sigma)^2)} \quad (61)$$

which satisfies the inequality (59), then the signal reconstruction $\hat{\boldsymbol{f}}_{K_e}$ given by

$$\hat{\boldsymbol{f}}_{K_e} = \arg\min_{\hat{\boldsymbol{f}}\in\mathcal{F}(B)}\left\{\hat{r}(\hat{\boldsymbol{f}}) + \frac{c(\hat{\boldsymbol{f}})\log 2}{\epsilon K_e}\right\} \quad (62)$$

satisfies the following inequality

$$E\left\{\frac{\|\hat{\boldsymbol{f}}_{K_e}-\boldsymbol{f}\|^2}{N}\right\} \leq C_{1e}\min_{\hat{\boldsymbol{f}}\in\mathcal{F}(B)}\left\{\frac{\|\hat{\boldsymbol{f}}-\boldsymbol{f}\|^2}{N} + \frac{c(\hat{\boldsymbol{f}})\log 2 + 4}{\epsilon K_e}\right\} \quad (63)$$

where C_{1e} is the constant given as

$$C_{1e} = \frac{1+a}{1-a},$$
$$a = \frac{2\left(1+\frac{2K_a}{K_e}\right)\left(\frac{B}{\sigma}\right)^2 + \left(1+\frac{2K_a}{MK_e}\right)}{(30-8e)\left(\frac{B}{\sigma}\right)^2 + (60-4\sqrt{2})\left(\frac{B}{\sigma}\right) + 30} \quad (64)$$

with the coefficient a obtained from (58) for the specific choice of ϵ in (61).

Proof: The proof follows the same steps as the proof of the related result for the uncorrelated case [4, p. 4039–4040] with the exception of using, in our correlated case, the above calculated values for ϵ (61) and a (64) instead of ϵ and a for the uncorrelated case. ■

Example 4: Let one set of samples be obtained based on the measurement matrix $\boldsymbol{\Phi}_e$ with $K_a = K$, $K_e = 2K$, and $M = 8$, and let another set of samples be obtained using a $2K \times N$ measurement matrix with all i.i.d. (Bernoulli) elements. Let also ϵ be selected as given by (61). Then the MSE error bounds for these two cases differ from each other only by a constant factor given for the former case by C_{1e} in (64) and in the latter case by C_1 (see (12) and the row after). Considering the two limiting cases when $B/\sigma \to 0$ and $B/\sigma \to \infty$, the intervals of change for the corresponding coefficients can be obtained as $1.08 \leq C_{1e} \leq 2.88$ and $1.06 \leq C_1 \leq 1.63$, respectively.

The following result on the achievable recovery performance for a sparse or compressible signal sampled based on the extended measurement matrix $\boldsymbol{\Phi}_e$ is of importance.

Theorem 5: For a sparse signal $\boldsymbol{f} \in \mathcal{F}_s(B,S) \triangleq \{\boldsymbol{f} : \|\boldsymbol{f}\|^2 \leq NB^2, \|\boldsymbol{f}\|_{l_0} \leq S\}$ and corresponding reconstructed signal $\hat{\boldsymbol{f}}_{K_e}$ obtained according to (62), there exists a constant $C'_{2e} = C'_{2e}(B,\sigma) > 0$, such that

$$\sup_{\boldsymbol{f}\in\mathcal{F}_s(B,S)} E\left\{\frac{\|\hat{\boldsymbol{f}}_{K_e}-\boldsymbol{f}\|^2}{N}\right\} \leq C_{1e}C'_{2e}\left(\frac{K_e}{S\log N}\right)^{-1}. \quad (65)$$

Similar, for a compressible signal $\boldsymbol{f} \in \mathcal{F}_c(B,\alpha,C_A) \triangleq \{\boldsymbol{f} : \|\boldsymbol{f}\|^2 \leq NB^2, \|\boldsymbol{f}^{(m)}-\boldsymbol{f}\|^2 \leq NC_A m^{-2\alpha}\}$ and corresponding reconstructed signal $\hat{\boldsymbol{f}}_{K_e}$ obtained according to (62), there exists a constant $C_{2e} = C_{2e}(B,\sigma,C_A) > 0$, such that

$$\sup_{\boldsymbol{f}\in\mathcal{F}_c(B,\alpha,C_A)} E\left\{\frac{\|\hat{\boldsymbol{f}}_{K_e}-\boldsymbol{f}\|^2}{N}\right\} \leq C_{1e}C_{2e}\left(\frac{K_e}{\log N}\right)^{-2\alpha/(2\alpha+1)}. \quad (66)$$

Proof: The proof follows the same steps as the proofs of the related results for the uncorrelated case [4, p. 4040–4041] with the exception of using, in our correlated case, the above calculated values for ϵ (61) and a (64) instead of ϵ and a for the uncorrelated case. ■

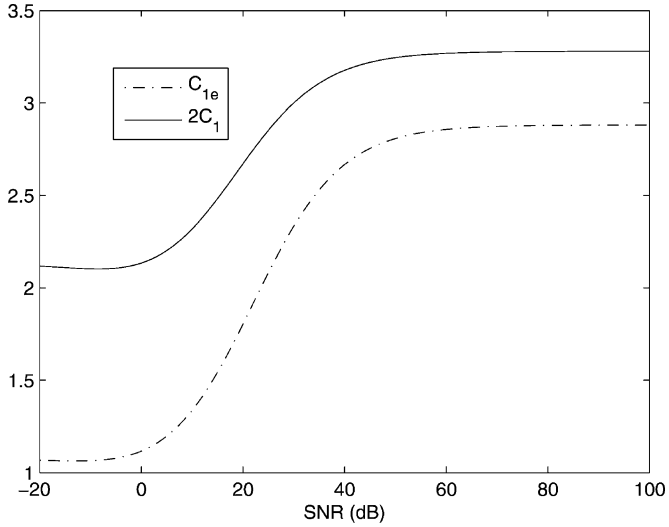


Fig. 3. This figure corresponds to Example 5 and shows the coefficients C_{1e} and $2C_1$ versus SNR. Since $C_{1e} < 2C_1$ for all values of SNR, one can conclude that the MSE corresponding to the empirical risk minimization-based recovery for the proposed segmented CS-based AIC must be lower than that for the conventional AIC.

Example 5: Let one set of samples be obtained based on the extended measurement matrix $\mathbf{\Phi}_e$ with $K_a = K$, $K_e = 2K$, and $M = 8$ and let another set of samples be obtained using the $K \times N$ measurement matrix with all i.i.d. (Bernoulli) elements. The error bounds corresponding to the case of K uncorrelated samples and our case of K_e correlated samples are (14) and (65), respectively. The comparison between these two error bounds boils down in this example to comparing $2C_1C_2'$ and $C_{1e}C_{2e}'$. Assuming the same ϵ as (61) for both methods, the following holds true $C_{2e}' = C_2'$. Fig. 3 compares C_{1e} and $2C_1$ versus the signal-to-noise ratio (SNR) B^2/σ^2. Since $C_{1e} < 2C_1$ for all values of SNR, the quality of the signal recovery, i.e., the corresponding MSE, for the case of $2K \times N$ extended measurement matrix is expected to be better than the quality of the signal recovery for the case of $K \times N$ measurement matrix of all i.i.d. entries.

The above results can be easily generalized for the case when $K_a > K$. Indeed, we only need to recalculate $\mathrm{var}\{\sum_{j=1}^{K_e} U_j\}$ for $K_e > 2K$. The only difference with the previous case of $K_a \leq K$ is the increased number of pairs of dependent rows in the extended measurement matrix $\mathbf{\Phi}_e$, which has a larger size now. The latter affects only the second term in (50). In particular, every row in $\mathbf{\Phi}^{\mathcal{P}^{(1)}}$ depends on M rows of the original measurement matrix $\mathbf{\Phi}$. Moreover, the term $\sum_{i=1}^{2K-1}\sum_{j=i+1}^{2K} E\{U_iU_j\}$ over all these M rows is bounded as in (52). Then considering all KM pairs of dependent rows from $\mathbf{\Phi}$ and $\mathbf{\Phi}^{\mathcal{P}^{(1)}}$, we have

$$2\sum_{\boldsymbol{\phi}_i,\boldsymbol{\phi}_j \text{dependent}} \left(E\{A^4\} - \left(\frac{\|\boldsymbol{g}_A\|^2}{N}\right)^2 + \frac{4\sigma^2}{M}\cdot\frac{\|\boldsymbol{g}_A\|^2}{N}\right) \leq 4K\left(\frac{\|\boldsymbol{g}\|^2}{N}\right)^2 + \frac{8\sigma^2 K}{M}\cdot\frac{\|\boldsymbol{g}\|^2}{N}. \quad (67)$$

Similar, every row of $\mathbf{\Phi}^{\mathcal{P}^{(2)}}$ depends on M rows of $\mathbf{\Phi}^{\mathcal{P}^{(1)}}$ and M rows of $\mathbf{\Phi}$. Considering all these $2KM$ pairs of dependent rows, we have

$$2\sum_{\boldsymbol{\phi}_i,\boldsymbol{\phi}_j \text{dependent}} \left(E\{A^4\} - \left(\frac{\|\boldsymbol{g}_A\|^2}{N}\right)^2 + \frac{4\sigma^2}{M}\cdot\frac{\|\boldsymbol{g}_A\|^2}{N}\right) \leq 4(2K)\left(\frac{\|\boldsymbol{g}\|^2}{N}\right)^2 + \frac{8\sigma^2(2K)}{M}\cdot\frac{\|\boldsymbol{g}\|^2}{N}. \quad (68)$$

Finally, the number of rows in the last matrix $(\mathbf{\Phi}_e)_{n_p}$ is K_{n_p} (see (32) and (33)). Every row of $(\mathbf{\Phi}_e)_{n_p}$ depends on M rows of each of the previous $n_p - 1$ matrices $\mathbf{\Phi}^{\mathcal{P}^{(i)}}$, $i = 1, \ldots, n_p - 1$. Considering all $(n_p - 1)K_{n_p}M$ pairs of dependent rows, we have

$$2\sum_{\boldsymbol{\phi}_i,\boldsymbol{\phi}_j \text{dependent}} \left(E\{A^4\} - \left(\frac{\|\boldsymbol{g}_A\|^2}{N}\right)^2 + \frac{4\sigma^2}{M}\cdot\frac{\|\boldsymbol{g}_A\|^2}{N}\right) \leq 4(n_p-1)K_{n_p}\left(\frac{\|\boldsymbol{g}\|^2}{N}\right)^2 + \frac{8\sigma^2(n_p-1)K_{n_p}}{M}\cdot\frac{\|\boldsymbol{g}\|^2}{N}. \quad (69)$$

Using (41) and the inequalities (67)–(69), we can find the following bound

$$\mathrm{var}\left\{\sum_{j=1}^{K_e} U_j\right\} \leq K_e\left(8B^2\left(1+\frac{D}{K_e}\right) + 4\sigma^2\left(1+\frac{D}{MK_e}\right)\right) r(\hat{\boldsymbol{f}}, \boldsymbol{f}) \quad (70)$$

where $D = 2K\sum_{i=1}^{n_p-2} i + 2K_{n_p}(n_p - 1)$. Note that in the case when $K_e = n_pK$, we have $D/K_e = n_p - 1$.

Therefore, it can be shown for the general extended matrix (27) that the inequality (60) holds for the following values of a and ϵ:

$$a = \frac{\left(8B^2\left(1+\frac{D}{K_e}\right) + 4\sigma^2\left(1+\frac{D}{MK_e}\right)\right)\epsilon}{2(1-\zeta)} \quad (71)$$

$$\epsilon < \frac{1}{\left(4\left(1+\frac{D}{K_e}\right)+16e\right)B^2 + 8\sqrt{B}\sigma + 2\sigma^2\left(1+\frac{D}{MK_e}\right)}. \quad (72)$$

Moreover, the theorems similar to Theorems 4 and 5 follow straightforwardly with the corrections to a and ϵ which are given now by (71) and (72), respectively.

We finally make some remarks on *non-RIP* conditions for l_1-norm-minimization-based recovery. Since the extended measurement matrix of the proposed segmented CS method satisfies the RIP, the results of [21] on recoverability and stability of the l_1-norm minimization straightforwardly apply. A different non-RIP-based approach for studying the recoverability and stability of the l_1-norm minimization, which uses some properties of the null space of the measurement matrix, is used in [25]. Then the non-RIP sufficient condition for recoverability of a

sparse signal from its noiseless compressed samples with the algorithm (8) is [25]

$$\sqrt{S} < \min\left\{0.5\frac{\|\boldsymbol{v}\|_{l_1}}{\|\boldsymbol{v}\|_{l_2}} : \boldsymbol{v} \in \{\mathcal{N}(\boldsymbol{\Phi}) \setminus \{0\}\}\right\} \tag{73}$$

where $\mathcal{N}(\boldsymbol{\Phi})$ denotes the null space of the measurement matrix $\boldsymbol{\Phi}$.

Let us show that the condition (73) is also satisfied for the extended measurement matrix $\boldsymbol{\Phi}_e$. Let $\boldsymbol{d}$ be any vector in the null space of $\boldsymbol{\Phi}_e$, i.e., $\boldsymbol{d} \in \mathcal{N}(\boldsymbol{\Phi}_e)$. Therefore, $[\boldsymbol{\Phi}_e]_i\boldsymbol{d} = 0,\ i = 1,\ldots,K_e$ where $[\boldsymbol{\Phi}_e]_i$ is the ith $1 \times N$ row-vector of $\boldsymbol{\Phi}_e$. Since the first K rows of $\boldsymbol{\Phi}_e$ are exactly the same as the K rows of $\boldsymbol{\Phi}$, we have $[\boldsymbol{\Phi}]_i\boldsymbol{d} = 0,\ i = 1,\ldots,K$. Hence, $\boldsymbol{d} \in \mathcal{N}(\boldsymbol{\Phi})$ and we can conclude that $\mathcal{N}(\boldsymbol{\Phi}_e) \subset \mathcal{N}(\boldsymbol{\Phi})$. Due to this property, we have $\min\{0.5\|\boldsymbol{v}\|_{l_1}/\|\boldsymbol{v}\|_{l_2} : \boldsymbol{v} \in \mathcal{N}(\boldsymbol{\Phi})\} \leq \min\{0.5\|\boldsymbol{v}\|_{l_1}/\|\boldsymbol{v}\|_{l_2} : \boldsymbol{v} \in \mathcal{N}(\boldsymbol{\Phi}_e)\}$. Therefore, if the original measurement matrix $\boldsymbol{\Phi}$ satisfies (73), so does the extended measurement matrix $\boldsymbol{\Phi}_e$, and the signal is recoverable from the samples taken by $\boldsymbol{\Phi}_e$.

Moreover, the necessary and sufficient condition for all signals with $\|\boldsymbol{x}\|_{l_0} < S$ to be recoverable from noiseless compressed samples using the l_1-norm minimization (8) is that [25]

$$\|\boldsymbol{v}\|_{l_1} > 2\|\boldsymbol{v}_{\mathcal{T}}\|_{l_1}, \quad \forall \boldsymbol{v} \in \{\mathcal{N}(\boldsymbol{\Phi}) \setminus \{0\}\} \tag{74}$$

where $\mathcal{T}$ is the set of indexes corresponding to the nonzero coefficients of $\boldsymbol{x}$. It is easy to see that since $\mathcal{N}(\boldsymbol{\Phi}_e) \subset \mathcal{N}(\boldsymbol{\Phi})$, the condition (74) also holds for the extended measurement matrix if the original measurement matrix satisfies it.

VI. Simulation Results

Throughout our simulations, three different measurement matrices (sampling schemes) are used: i) the $K \times N_c$ measurement matrix $\boldsymbol{\Phi}$ with i.i.d. entries referred to as the original measurement matrix; ii) the extended $K_e \times N_c$ measurement matrix $\boldsymbol{\Phi}_e$ obtained using the proposed segmented CS method and referred to as the extended measurement matrix; and iii) the $K_e \times N_c$ measurement matrix with all i.i.d entries referred to as the enlarged measurement matrix. This last measurement matrix corresponds to the sampling scheme with K_e independent BMIs in the AIC in Fig. 1. The number of segments in the proposed segmented CS method M is set to 8.

The noisy case corresponding to the model (6) is always considered. Then in order to make sure that the measurement noise for additional samples obtained based on the proposed extended measurement matrix is correlated with the measurement noise of the original samples, the $K \times M$ matrix of noisy subsamples with the noise variance σ^2/M is first generated. Then the permutations are applied to this matrix and the subsamples along each row of the original and permuted matrices are added up together to build noisy samples.

The recovery performance for three aforementioned sampling schemes is measured using the MSE between the recovered and original signals. In all examples, MSE values are computed based on 1000 independent simulation runs for all sampling schemes tested.

A. Simulation Example 1: Time-Sparse Signal With l_1 Norm Minimization-Based Recovery

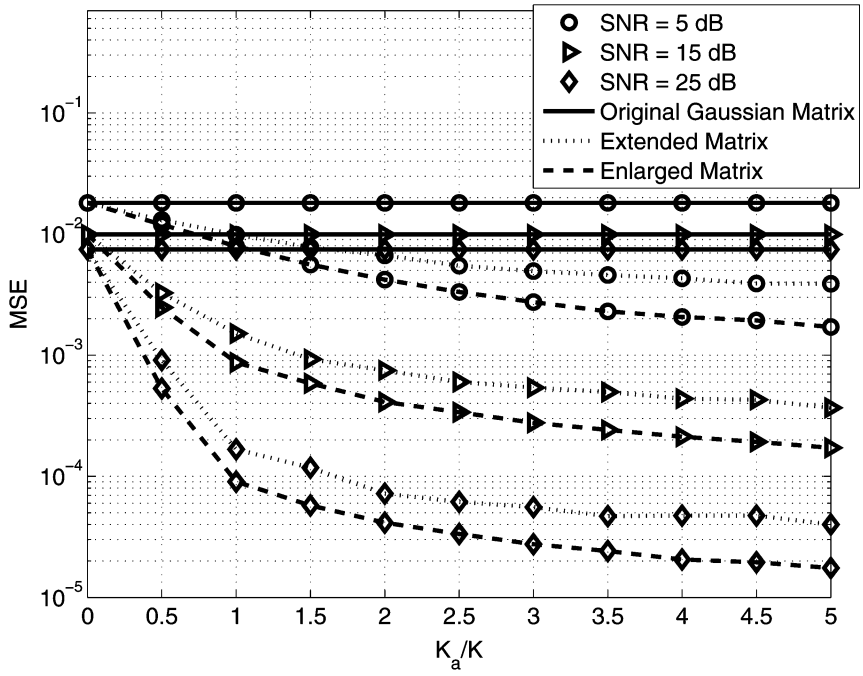
In our first example, the signal is assumed to be sparse in the time domain. Particularly, let $f(t)$ be the continuous signal as in (1) with $N = 128$ basis functions $\{\psi_n(t)\}_{n=1}^{N}$ of the type

$$\psi_n(t) = \begin{cases} \frac{N}{T}, & t \in [(n-1)T/N,\, nT/N] \\ 0, & \text{otherwise} \end{cases}, \quad n = 1,\ldots,N. \tag{75}$$

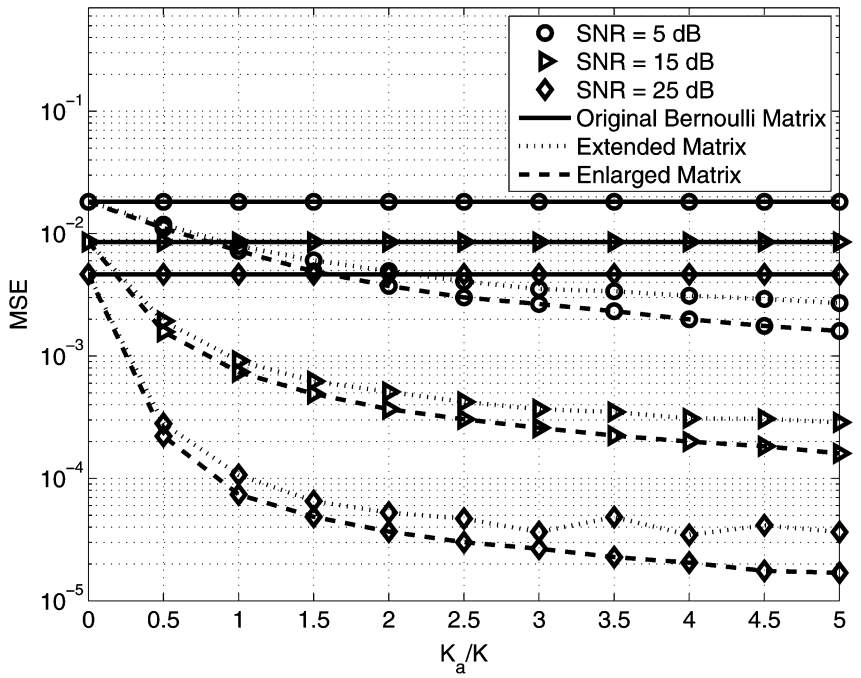
These basis functions form the sparsity basis $\boldsymbol{\Psi}(t)$. Choosing N_c to be equal to N, we obtain based on (5) that $\boldsymbol{\Psi} = \boldsymbol{I}$. Over one time period T only 3 projections of the signal onto the sparsity basis are nonzero and are set to $+1$ or -1 with equal probabilities.

The l_1-norm minimization algorithm (9) is used to recover the signal sampled using the three aforementioned sampling schemes. Since $\boldsymbol{\Psi} = \boldsymbol{I}$ for the considered time-sparse signal, then $\boldsymbol{\Phi}' = \boldsymbol{\Phi}$ in (9). The number of BMIs in the sampling device is $K = 16$, while γ in (9), which is the bound on the root square of the noise energy, is set to $\sqrt{K'}\sigma$. Here $K' = K$ for the sampling scheme based on the original measurement matrix, while $K' = K_e$ in the other two schemes. The entries of the original and enlarged measurement matrices are generated as i.i.d. Gaussian or i.i.d. Bernoulli distributed random variables with zero mean and variance $1/N$. This corresponds to the case of sampling waveforms with chip duration T/N and i.i.d. Gaussian or i.i.d. Bernoulli distributed chip amplitudes, respectively. The SNR is defined as $\|\boldsymbol{\Phi f}\|_{l_2}^2/\|\boldsymbol{w}\|_{l_2}^2$. Approximating $\|\boldsymbol{\Phi f}\|_{l_2}^2$ by $(K'/N)\|\boldsymbol{f}\|_{l_2}^2$, which is valid because of (7), the corresponding noise variance σ^2 can be calculated when SNR is given and vice versa. For example, the approximate SNR in decibels can be calculated as $10\log_{10}(3/N\sigma^2)$.

Fig. 4(a) and (b) shows the MSEs corresponding to all three aforementioned measurement matrices versus the ratio of the number of additional samples to the number of original samples K_a/K for the Gaussian and Bernoulli cases, respectively. The results are shown for three different SNR values of 5, 15, and 25 dB. It can be seen from the figures that better recovery quality is achieved by using the extended measurement matrix as compared to the original measurement matrix. The corresponding MSE curves in Fig. 4(a) and (b) are similar to each other which confirms the fact that both Gaussian and Bernoulli measurement matrices are good candidates, although Bernoulli is practically preferable. As expected, the recovery performance in the case of the extended measurement matrix is slightly worse than that in the case of the enlarged measurement matrix. This difference, however, is small as compared to the performance improvement provided by the extended measurement matrix over the original measurement matrix. Note also that in the case of the enlarged measurement matrix, the AIC in Fig. 1 consists of K_e BMIs, while only K BMIs are required in the case of the extended measurement matrix. For example, the number of such BMIs halves for the proposed segmented AIC if $K_a/K = 1$. Additionally, it can be seen that the rate of MSE improvement decreases as the number of collected samples increases. The latter can be observed for both the extended and enlarged measurement matrices and for all three values of SNR.



(a)



(b)

Fig. 4. Recovery of the time-sparse signal based on the l_1-norm minimization algorithm: MSEs versus K_a/K. (a) Measurement matrix with Gaussian distributed entries, (b) Measurement matrix with Bernoulli distributed entries.

B. Simulation Example 2: Time-Sparse Signal With Empirical Risk Minimization-Based Recovery

In our second simulation example, the empirical risk minimization method is used to recover the same time-sparse signal as in our first simulation example. The signal is sampled using the three sampling schemes tested with $K = 24$. The minimization problem (11) is solved to obtain a candidate reconstruction $\hat{\boldsymbol{f}}_{K'}$ of the original sparse signal $\boldsymbol{f}$. Considering $\hat{\boldsymbol{f}}_{K'} = \boldsymbol{\Psi}^T \hat{\boldsymbol{x}}_{K'}$, the problem (11) can be rewritten in terms of $\hat{\boldsymbol{x}}_{K'}$ as

$$\begin{aligned}\hat{\boldsymbol{x}}_{K'} &= \arg\min_{\hat{\boldsymbol{x}}\in\mathcal{X}} \left\{ \hat{r}(\boldsymbol{\Psi}^T\hat{\boldsymbol{x}}) + \frac{c(\hat{\boldsymbol{x}})\log 2}{\epsilon K'} \right\} \\ &= \arg\min_{\hat{\boldsymbol{x}}\in\mathcal{X}} \left\{ \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}^T\hat{\boldsymbol{x}}\|_{l_2}^2 + \frac{2\log 2\log N}{\epsilon}\|\hat{\boldsymbol{x}}\|_{l_0} \right\} \quad (76)\end{aligned}$$

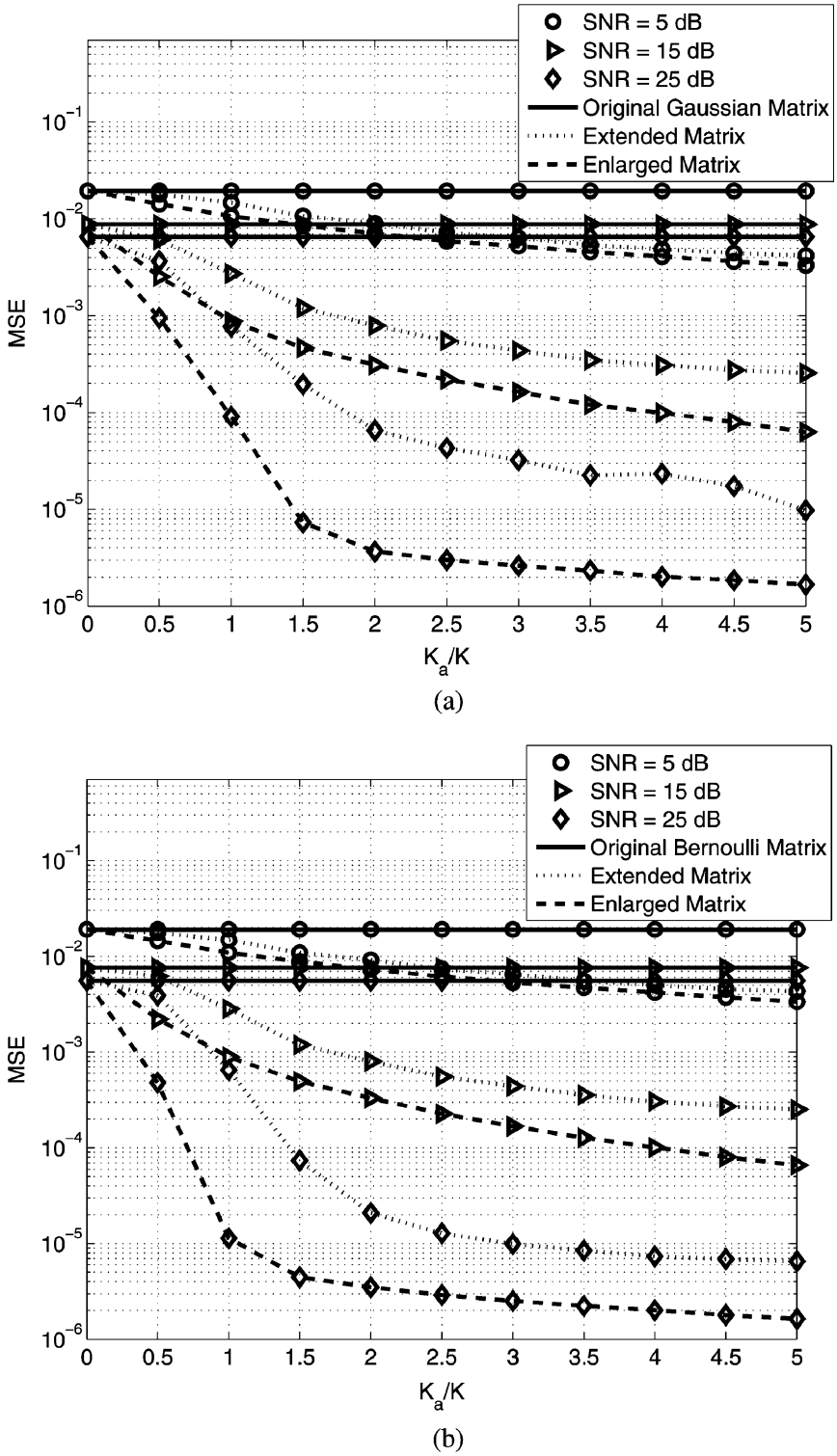


Fig. 5. Recovery of the time-sparse signal based on the empirical risk minimization method: MSEs versus K_a/K. (a) Measurement matrix with Gaussian distributed entries; (b) Measurement matrix with Bernoulli distributed entries.

and solved using the iterative bound optimization procedure [4]. Here $\chi \triangleq \{\boldsymbol{x}: \ ||\boldsymbol{\Psi}^T\boldsymbol{x}||^2 \leq NB^2\}$. This procedure uses the threshold $\sqrt{2\log 2\log N/\lambda\epsilon}$ where λ is the largest eigenvalue of the matrix $\boldsymbol{\Phi}^T\boldsymbol{\Phi}$. In our simulations, this threshold is set to 0.035 for the case of the extended measurement matrix and 0.05 for the cases of the original and enlarged measurement matrices. These threshold values are optimized as recommended in [4]. The stopping criterion for the iterative bound optimization procedure is $||\hat{\boldsymbol{x}}^{(i+1)} - \hat{\boldsymbol{x}}^{(i)}||_{l_\infty} \leq \theta$ where $||\cdot||_{l_\infty}$ is the l_∞ norm and $\hat{\boldsymbol{x}}^{(i)}$ denotes the value of $\hat{\boldsymbol{x}}$ obtained in the ith iteration. The value $\theta = 0.001$ is selected.

Fig. 5(a) and (b) shows the MSEs for all three measurement matrices tested versus the ratio K_a/K for the Gaussian and Bernoulli cases, respectively. The results are shown for three

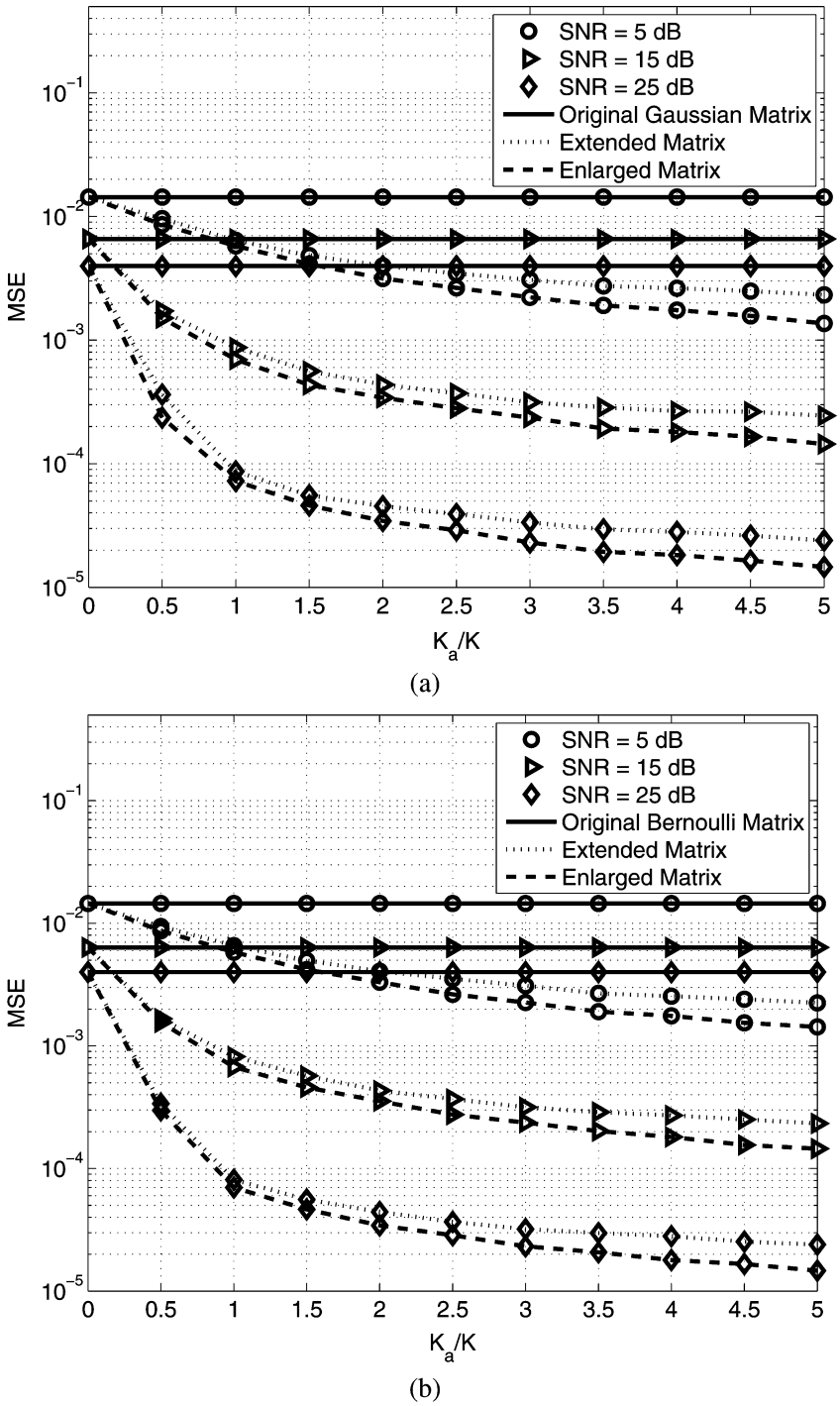


Fig. 6. Recovery of the sparse OFDM signal based on the l_1-norm minimization algorithm: MSEs versus K_a/K. (a) Measurement matrix with Gaussian distributed entries; (b) Measurement matrix with Bernoulli distributed entries.

different SNR values of 5, 15, and 25 dB. The same conclusions as in the first example can be drawn in this example. Therefore, the proposed segmented AIC indeed leads to significantly improved signal recovery performance without increasing the number of BMIs.

C. Simulation Example 3: OFDM Signal With l_1 Norm Minimization-Based Recovery

In our third example, we consider an orthogonal frequency-division-multiplexed (OFDM) signal with three nonzero subcarriers out of 128 available frequency bins. Nonzero subcarriers are modulated with quadrature phase-shift keying (QPSK) symbols. The $N = 128$ basis functions $\{\psi_n(t)\}_{n=1}^{N}$ are

$$\psi_n(t) = \cos\left((n-1)\frac{2\pi}{T}t\right) \quad t \in [0,T], \quad n = 1,\ldots,N. \quad (77)$$

The number of chips per symbol duration in the sampling waveform is set to $N_c = 256$. It is because we need to ensure that the rows of the 128×256 sparsity matrix $\mathbf{\Psi}$, which is calcu-

TABLE I
PERCENTAGE THAT THE POSITIONS OF THE NONZERO SIGNAL VALUES ARE CORRECTLY IDENTIFIED

			K_a	0	9	18	27	36	45
Noiseless case	AIC with K_e branches	Percentage		0	22.45	90.5	100	100	100
		MSE		N/A	0	0	0	0	0
	Segmented AIC	Percentage		0	10.2	58.35	83.10	89.7	90.5
		MSE		N/A	0	0	0	0	0
SNR=25 dB	AIC with K_e branches	Percentage		0	11	71.45	98	100	100
		MSE		N/A	5.48×10^{-4}	4.91×10^{-4}	2.78×10^{-4}	1.53×10^{-4}	1.04×10^{-4}
	Segmented AIC	Percentage		0	4.85	34.30	57.20	70.5	73.65
		MSE		N/A	6.06×10^{-4}	5.88×10^{-4}	5.78×10^{-4}	4.82×10^{-4}	3.98×10^{-4}

lated according to (5), are approximately orthogonal. Then the SNR can be defined as $\|\boldsymbol{\Phi}'\boldsymbol{x}\|_{l_2}^2/\|\boldsymbol{w}\|_{l_2}^2$. Moreover, one can approximate $\|\boldsymbol{\Phi}'\boldsymbol{x}\|_{l_2}^2 = \|\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{x}\|_{l_2}^2$ by $\psi_{\text{ave}}(K'/N)\|\boldsymbol{x}\|_{l_2}^2$, where ψ_{ave} is the average norm of the N rows of the sparsity matrix $\boldsymbol{\Psi}$ and $K' = K$ for the sampling scheme based on the original measurement matrix, while $K' = K_e$ for the other two schemes. The number of original samples K is set to 16. Considering that the nonzero subcarriers are modulated with unit norm QPSK symbols, the approximate SNR in dBs can be calculated as $10\log_{10}(3\psi_{\text{ave}}/N\sigma^2)$. The l_1 norm minimization-based recovery method is used for signal recovery based on the compressed samples obtained using the three sampling schemes tested.

Fig. 6(a) and (b) shows the MSEs for all three measurement matrices tested versus the ratio K_a/K for the Gaussian and Bernoulli cases, respectively. Comparing the results in Fig. 6 and Fig. 4, one can deduce the universality of the Gaussian and Bernoulli measurement matrices, which means that we are able to recover the signal using the measurements collected with these measurement matrices regardless of the sparsity basis. As in the previous two simulation examples, the proposed segmented CS scheme significantly outperforms the original sampling scheme with the same number of BMIs, while slightly deteriorates in performance compared to the sampling scheme with enlarged number of BMIs.

It is also worth mentioning that the MSE of the recovered signal depends on the ratio between the sparsity and the extended number of samples (see for example (63)). Moreover, the RIP sets a bound for the number of samples required for successful recovery given the sparsity level of the signal. Thus, if the signal is not sparse enough and the number of collected samples is low, a recovery algorithm can fail to recover the signal from such small number of samples. If the number of samples is sufficient to ensure successful recovery, but the ratio between the sparsity and the number of samples is high, the MSE can be still high. By using the technique proposed in this paper for extending the number of samples, this situation can be improved as we show in our next example.

D. Simulation Example 4: the Number of BMIs in the Conventional AIC Is Insufficient for Successful Recovery

Our last simulation example considers the case when the number of original compressed samples, that is, the number of BMIs in the conventional AIC, is insufficient for successful recovery. The time-sparse signal described in our first simulation example is assumed. The number of basis functions is $N = 128$, however, the number of nonzero projections, i.e., the signal sparsity level is $S = 5$. The number of BMIs in the conventional AIC is $K = 9$ and the number of segments in the proposed segmented AIC is $M = 8$. Since generally speaking four times as many samples are needed as the sparsity level of the signal to guarantee exact recovery in noiseless case [1], the number of samples that can be collected by the conventional AIC with $K = 9$ BMIs is insufficient for exact recovery even in the noise free case. Thus, the conventional AIC is not applicable and only the sampling schemes based on the extended and enlarged measurement matrices are compared to each other in terms of the percentage that the positions of the nonzero values of the time-sparse signal are correctly identified. The number of samples can be increased to at most $K^2 = 81$ if the proposed segmented AIC is used. The MSEs averaged over all cases of successful recovery are also reported. Two different cases of (a) no measurement noise and (b) $\text{SNR} = 25$ dB are considered. The simulation results are gathered in Table I.

The results in Table I show that although the AIC with $K = 9$ (the column $K_a = 0$ in the table) BMIs cannot successfully recover the positions of nonzero entries of the signal, the segmented AIC is able to find those positions and the success rate increases as K_a increases. The success rate of the AIC with larger number of BMIs is higher as expected. For both schemes the lower success rates can be observed in the noisy case as compared to the noiseless case.

VII. CONCLUSION

A new segmented CS method for AIC has been proposed. According to this method, an analog signal measured by K parallel BMIs, each characterized by a specific random sampling waveform, is first segmented in time into M segments so that a $K \times M$ matrix of subsamples is obtained. Then the subsamples collected on different segments and different BMIs are reused so that a larger number of samples (at most K^2) than the number of BMIs is collected. Such samples are correlated to each other over at most one segment and the technique is shown to be equivalent to extending the measurement matrix consisting of the BMI sampling waveforms by adding new

rows without actually increasing the number of BMIs. Such extended measurement matrix satisfies the RIP with overwhelming probability if the original measurement matrix of BMI sampling waveforms satisfies it. Due to the inherent structure of the proposed segmented CS method, the complexity of the sampling device is slightly increased, while the signal recovery performance is shown to be significantly improved. Specifically, we have proved that the performance of the signal recovery based on the empirical risk minimization improves when the segmented AIC is used for sampling instead of the conventional AIC with the same number of BMIs. Remarkably, if the number of BMIs is insufficient in the conventional AIC to guarantee successful recovery, the proposed segmented AIC supplies the recovery algorithm with additional samples so that successful recovery becomes possible. At the same time, the complexity increase is only due to the $M \leq K$ times higher sampling rate and the necessity to solve a larger size optimization problem at the recovery stage, while the number of BMIs remains the same at the sampling stage. The validity, effectiveness, and superiority of the proposed segmented AIC over the conventional AIC is also justified based on our simulation results.

APPENDIX

Proof of Theorem 1: The total number of possible permutations of $\boldsymbol{z}$ is $K!$. Let $\mathcal{A}$ be the set of permutations $\pi_s, s = 1,\dots,|\mathcal{A}|$ that satisfy the following condition

$$\begin{aligned} &\pi_s(k) \neq \pi_t(k), \quad s \neq t, \\ &\forall s,t \in \{1,\dots,|\mathcal{A}|\}, \quad \forall k \in \{1,\dots,K\}. \end{aligned} \tag{78}$$

It is easy to see that the number of distinct permutations satisfying the condition (78) is K, so $|\mathcal{A}| = K$. It is also straightforward to see that the choice of such K distinct permutations is not unique. As a specific choice, let the elements of $\mathcal{A}$, i.e., the permutations $\pi_s,\ s = 1,\dots,K$, be

$$\pi_s(k) = ((s+k-2) \bmod K) + 1, \quad s,k = 1,\dots,K \tag{79}$$

with π_1 being the identity permutation, i.e., the permutations that does not change $\boldsymbol{z}$.

Consider now the matrix $\boldsymbol{Z}$ which consists of M identical columns $\boldsymbol{z}$. The ith set of column permutations of matrix $\boldsymbol{Z}$ is $\mathcal{P}^{(i)} = \{\pi_1^{(i)},\dots,\pi_M^{(i)}\}$ and the corresponding permuted matrix is $\boldsymbol{Z}^{\mathcal{P}^{(i)}}$. Let $\{\pi_1^{(i)},\dots,\pi_M^{(i)}\}$ be any combination of the K permutations in (79). Then there are K^M possible choices for $\mathcal{P}^{(i)}$. However, not all of these possible choices are permissible by the conditions of the theorem.

Indeed, let the set $\mathcal{P}^{(1)}$ be a combination of permutations from $\mathcal{A}$ that satisfies (22). There are $I-1$ other sets $\mathcal{P}^{(i)},\ i = 2,\dots,I$ which satisfy both (22) and (23). Gathering all such sets in one set, we obtain the set $\mathcal{P} = \{\mathcal{P}^{(1)},\dots,\mathcal{P}^{(I)}\}$. Now let $\mathcal{P}^{(I+1)} = [\pi_1^{(I+1)},\dots,\pi_M^{(I+1)}]$ be one more set of permutations where $\exists \pi_m^{(I+1)},\ m = 1,\dots,M$ such that $\pi_m^{(I+1)} \notin \mathcal{A}$. An arbitrary kth row of $\boldsymbol{Z}^{\mathcal{P}^{(I+1)}}$ is $([\boldsymbol{Z}^{\mathcal{P}^{(I+1)}}]_{k,1},\dots,[\boldsymbol{Z}^{\mathcal{P}^{(I+1)}}]_{k,M})$ where $[\boldsymbol{Z}^{\mathcal{P}^{(I+1)}}]_{k,1},\dots,[\boldsymbol{Z}^{\mathcal{P}^{(I+1)}}]_{k,M} \in \{1,\dots,K\}$. This exact same row can be found as the first row of one of the permuted matrices $\boldsymbol{Z}^{\mathcal{P}^{(i)}},\ \mathcal{P}^{(i)} \in \mathcal{P}$. Specifically, this is the permuted matrix $\boldsymbol{Z}^{\mathcal{P}^{(i)}}$ that is obtained by applying the permutations $\mathcal{P}^{(i)} = \{\pi_{[\boldsymbol{Z}^{\mathcal{P}^{(I+1)}}]_{k,1}},\dots,\pi_{[\boldsymbol{Z}^{\mathcal{P}^{(I+1)}}]_{k,M}}\}$. The permutations $\mathcal{P}^{(i)}$ either has to belong to $\mathcal{P}$ or being crossed out from $\mathcal{P}$ because of conflicting with another element $\mathcal{P}^{(l)} \in \mathcal{P}$, $l \neq i$. In both cases, $\mathcal{P}^{(I+1)}$ can not be added to $\mathcal{P}$ because it will contradict the conditions (22) and (23).

Therefore, the set $\mathcal{P}$ can be built using only the permutations from the set $\mathcal{A}$, i.e., the K permutations in (79). Rearranging the rows of $\boldsymbol{Z}^{\mathcal{P}^{(i)}}$ in a certain way, one can force the elements in the first column of $\boldsymbol{Z}^{\mathcal{P}^{(i)}}$ to appear in the original increasing order, i.e., enforce the first column to be equivalent to the vector of indexes $\boldsymbol{z}$. It can be done by applying to each permutation in the set $\mathcal{P}^{(i)}$ the inverse permutation $(\pi_1^{(i)})^{-1}$, which itself is one of the permutations in (79). Therefore, the set $\mathcal{P}^{(i)} = \{\pi_1^{(i)},\dots,\pi_M^{(i)}\}$ can be replaced by the equivalent set $\{(\pi_1^{(i)})^{-1}\pi_1^{(i)},\dots,(\pi_1^{(i)})^{-1}\pi_M^{(i)}\} = \{\pi_1,\dots,(\pi_1^{(i)})^{-1}\pi_M^{(i)}\}$ where $(\pi_1^{(i)})^{-1}\pi_j^{(i)} \in \mathcal{A}$. Hence, we can consider only the permutations of the form $\mathcal{P}^{(i)} = \{\pi_1,\dots,\pi_j^{(i)},\dots,\pi_M^{(i)}\}$. Since the condition (22) requires that $\pi_2^{(i)}$ should be different from π_1, the only available options for the permutations on the second column of $\boldsymbol{Z}$ are the $K-1$ permutations $\pi_2,\dots,\pi_K$ in (79). Therefore, I at most equals $K-1$. Note that I can be smaller than $K-1$ if for some $i \in \{1,\dots,K-1\}$, $K/\gcd(i,K) < M$ (also see Example 1 after Theorem 1). Thus, in general $I \leq K-1$.

Proof of Lemma 1: Let all the rows of $(\boldsymbol{\Phi}_e)_{\mathcal{I}}$ be partitioned into two sets of sizes (cardinality) as close as possible to each other, where all elements in each set are guaranteed to be statistically independent. In particular, note that the elements of the new K_a rows of $\boldsymbol{\Phi}_e$ are chosen either from the first $K_a + M - 1$ rows of $\boldsymbol{\Phi}$ if $K_a + M - 1 < K$ or from the whole matrix $\boldsymbol{\Phi}$. Therefore, if $K_a + M - 1 < K$, the last $K - K_a - M + 1$ rows of $\boldsymbol{\Phi}$ play no role whatsoever in the process of extending the measurement matrix and they are independent on the rows of $\boldsymbol{\Phi}_1$ in (28). These rows are called unused rows. Thus, one can freely add any number of such unused rows to the set of rows in $\boldsymbol{\Phi}_1$ without disrupting its status of being formed by independent Gaussian variables. Since $\min\{K, K_a + M - 1\} \leq \lceil (K + K_a)/2 \rceil$, there exist at least $\lfloor (K + K_a)/2 \rfloor - K_a$ unused rows which can be added to the set of rows in $\boldsymbol{\Phi}_1$. Such process describes how the rows of $(\boldsymbol{\Phi}_e)_{\mathcal{I}}$ are split into the desired sets $(\boldsymbol{\Phi}_e)^1_{\mathcal{I}}$ and $(\boldsymbol{\Phi}_e)^2_{\mathcal{I}}$ of statistically independent elements. As a result, the first matrix $(\boldsymbol{\Phi}_e)^1_{\mathcal{I}}$ includes the first $\lceil (K + K_a)/2 \rceil$ rows of $(\boldsymbol{\Phi}_e)_{\mathcal{I}}$, while the rest of the rows are included in $(\boldsymbol{\Phi}_e)^2_{\mathcal{I}}$.

Since the elements of the matrices $(\boldsymbol{\Phi}_e)^1_{\mathcal{I}}$ and $(\boldsymbol{\Phi}_e)^2_{\mathcal{I}}$ are i.i.d. Gaussian, these matrices will satisfy (7) with probabilities equal or larger than $1 - 2(12/\delta_S)^S e^{-C_0 \lceil K_e/2 \rceil}$ and $1 - 2(12/\delta_S)^S e^{-C_0 \lfloor K_e/2 \rfloor}$, respectively. Therefore, both matrices $(\boldsymbol{\Phi}_e)^1_{\mathcal{I}}$ and $(\boldsymbol{\Phi}_e)^2_{\mathcal{I}}$ satisfy (7) simultaneously with the common probability

$$\begin{aligned} \Pr\{(\boldsymbol{\Phi}_e)^i_{\mathcal{I}} \text{ satisfies (7)}\} \geq 1 - 2(12/\delta_S)^S e^{-C_0 \lfloor K_e/2 \rfloor}, \\ i = 1,2. \end{aligned} \tag{80}$$

Let $K_1' \triangleq \lceil K_e/2 \rceil$ and $K_2' \triangleq \lfloor K_e/2 \rfloor$. Consider the event when both $(\mathbf{\Phi}_e)_{\mathcal{T}}^1$ and $(\mathbf{\Phi}_e)_{\mathcal{T}}^2$ satisfy (7). Then the following inequality holds for any vector $\boldsymbol{c} \in \mathbb{R}^S$:

$$\sum_{i=1}^{2} \frac{K_i'}{N}(1-\delta_S)\|\boldsymbol{c}\|_{l_2}^2 \leq \sum_{i=1}^{2} \|(\mathbf{\Phi}_e)_{\mathcal{T}}^i \boldsymbol{c}\|_{l_2}^2 \leq \sum_{i=1}^{2} \frac{K_i'}{N}(1+\delta_S)\|\boldsymbol{c}\|_{l_2}^2 \tag{81}$$

or, equivalently,

$$\frac{K_e}{N}(1-\delta_S)\|\boldsymbol{c}\|_{l_2}^2 \leq \|(\mathbf{\Phi}_e)_{\mathcal{T}}\boldsymbol{c}\|_{l_2}^2 \leq \frac{K_e}{N}(1+\delta_S)\|\boldsymbol{c}\|_{l_2}^2. \tag{82}$$

Therefore, if both matrices $(\mathbf{\Phi}_e)_{\mathcal{T}}^1$ and $(\mathbf{\Phi}_e)_{\mathcal{T}}^2$ satisfy (7), then the matrix $(\mathbf{\Phi}_e)_{\mathcal{T}}$ also satisfies (7). Moreover, the probability that $(\mathbf{\Phi}_e)_{\mathcal{T}}$ does not satisfy (7) can be found as

$$\begin{aligned}
&\Pr\{(\mathbf{\Phi}_e)_{\mathcal{T}} \text{ does not satisfy (7)}\} \\
&\quad \leq \Pr\{(\mathbf{\Phi}_e)_{\mathcal{T}}^1 \text{ or } (\mathbf{\Phi}_e)_{\mathcal{T}}^2 \text{ does not satisfy (7)}\} \\
&\quad \overset{(a)}{\leq} \sum_{i=1}^{2} \Pr\{(\mathbf{\Phi}_e)_{\mathcal{T}}^i \text{ does not satisfy (7)}\} \\
&\quad \overset{(b)}{\leq} 4(12/\delta_S)^S e^{-C_0\lfloor K_e/2 \rfloor}
\end{aligned} \tag{83}$$

where the inequality (a) follows from the union bounding and the inequality (b) follows from (80). Thus, the inequality (30) holds.

Proof of Theorem 2: According to (30), the matrix $(\mathbf{\Phi}_e)_{\mathcal{T}}$ does not satisfy (7) with probability less than or equal to $4(12/\delta_S)^S e^{-C_0\lfloor K_e/2 \rfloor}$ for any subset $\mathcal{T} \subset \{1,\ldots,N\}$ of cardinality S. Since there are $\binom{N}{S} \leq (Ne/S)^S$ different subsets $\mathcal{T}$ of cardinality S, $\mathbf{\Phi}_e$ does not satisfy the RIP with probability

$$\begin{aligned}
&\Pr\{\mathbf{\Phi}_e \text{ does not satisfy RIP}\} \\
&\quad \leq 4\binom{N}{S}(12/\delta_S)^S e^{-C_0\lfloor K_e/2 \rfloor} \\
&\quad \leq 4(Ne/S)^S(12/\delta_S)^S e^{-C_0\lfloor K_e/2 \rfloor} \\
&\quad = 4e^{-(C_0\lfloor K_e/2 \rfloor - S[\log(Ne/S)+\log(12/\delta_S)])} \\
&\quad \leq 4e^{-\{C_0\lfloor K_e/2 \rfloor - C_3[\log(Ne/S)+\log(12/\delta_S)]\lfloor K_e/2 \rfloor/\log(N/S)\}} \\
&\quad = 4e^{-\{C_0 - C_3[1+(1+\log(12/\delta_S))/\log(N/S)]\}\lfloor K_e/2 \rfloor}.
\end{aligned} \tag{84}$$

Setting $C_4 = C_0 - C_3[1 + (1 + \log(12/\delta_S))/\log(N/S)]$ and choosing C_3 small enough that guarantees that C_4 is positive, we obtain (31).

Proof of Lemma 2: The method of the proof is the same as the one used to prove Lemma 1 and is based on splitting the rows of $\mathbf{\Phi}_e$ into a number of sets with independent entries. Here, the splitting is carried out as shown in (33).

Let $(\mathbf{\Phi}_e)_{\mathcal{T}}^i$, $i = 1,\ldots,n_p - 1$ be the matrix containing the $(i-1)K + 1$st to the iKth rows of $(\mathbf{\Phi}_e)_{\mathcal{T}}$. The last $K_e - (n_p - 1)K$ rows of $(\mathbf{\Phi}_e)_{\mathcal{T}}$ form the matrix $(\mathbf{\Phi}_e)_{\mathcal{T}}^{n_p}$. Since the matrices $(\mathbf{\Phi}_e)_{\mathcal{T}}^i$, $i = 1,\ldots,n_p - 1$ consist of independent entries, they satisfy (7) each with probability of at least $1 - 2(12/\delta_S)^S e^{-C_0 K}$. For the same reason, the matrix $(\mathbf{\Phi}_e)_{\mathcal{T}}^{n_p}$ satisfies (7) with probability greater than or equal to $1 - 2(12/\delta_S)^S e^{-C_0 K_{n_p}}$. In the event that all the matrices $(\mathbf{\Phi}_e)_{\mathcal{T}}^i$, $i = 1,\ldots,n_p$ satisfy (7) simultaneously for $\boldsymbol{c} \in \mathbb{R}^S$ we have

$$\begin{aligned}
\sum_{i=1}^{n_p} \frac{K_i}{N}(1-\delta_S)\|\boldsymbol{c}\|_{l_2}^2 &\leq \sum_{i=1}^{n_p} \|(\mathbf{\Phi}_e)_{\mathcal{T}}^i \boldsymbol{c}\|_{l_2}^2 \\
&\leq \sum_{i=1}^{n_p} \frac{K_i}{N}(1+\delta_S)\|\boldsymbol{c}\|_{l_2}^2 \\
\Rightarrow \quad \frac{K_e}{N}(1-\delta_S)\|\boldsymbol{c}\|_{l_2}^2 &\leq \|(\mathbf{\Phi}_e)_{\mathcal{T}}\boldsymbol{c}\|_{l_2}^2 \\
&\leq \frac{K_e}{N}(1+\delta_S)\|\boldsymbol{c}\|_{l_2}^2.
\end{aligned} \tag{85}$$

Therefore, using the union bound and (85), we can conclude that

$$\begin{aligned}
&\Pr\{(\mathbf{\Phi}_e)_{\mathcal{T}} \text{ does not satisfy (7)}\} \\
&\quad \leq \sum_{i=1}^{n_p} \Pr\{(\mathbf{\Phi}_e)_{\mathcal{T}}^i \text{ does not satisfy (7)}\} \\
&\quad \leq 2(n_p - 1)(12/\delta_S)^S(e^{-C_0 K}) + 2(12/\delta_S)^S\left(e^{-C_0 K_{n_p}}\right)
\end{aligned} \tag{86}$$

which proves the lemma.

Proof of Theorem 3: According to Lemma 2, for any subset $\mathcal{T} \subset \{1,\ldots,N\}$ of cardinality S, the probability that $(\mathbf{\Phi}_e)_{\mathcal{T}}$ does not satisfy (7) is less than or equal to $2(n_p - 1)(12/\delta_S)^S(e^{-C_0 K}) + 2(12/\delta_S)^S(e^{-C_0 K_{n_p}})$. Using the fact that there are $\binom{N}{S} \leq (Ne/S)^S$ different subsets $\mathcal{T}$, the probability that the extended measurement matrix $\mathbf{\Phi}_e$ does not satisfy the RIP can be computed as

$$\begin{aligned}
&\Pr\{\mathbf{\Phi}_e \text{ does not satisfy the RIP}\} \\
&\quad \leq 2(n_p - 1)\binom{N}{S}(12/\delta_S)^S e^{-C_0 K} \\
&\qquad + 2\binom{N}{S}(12/\delta_S)^S e^{-C_0 K_{n_p}} \\
&\quad \leq 2(n_p - 1)(Ne/S)^S(12/\delta_S)^S e^{-C_0 K} \\
&\qquad + 2(Ne/S)^S(12/\delta_S)^S e^{-C_0 K_{n_p}} \\
&\quad = 2(n_p - 1) \\
&\qquad \times e^{-(C_0 K - S[\log(Ne/S)+\log(12/\delta_S)])} \\
&\qquad + 2e^{-(C_0 K_{n_p} - S[\log(Ne/S)+\log(12/\delta_S)])} \\
&\quad \leq 2(n_p - 1) \\
&\qquad \times e^{-\{C_0 K - \frac{C_3 K_{n_p}}{K}[\log(Ne/S)+\log(12/\delta_S)]K/\log(N/S)\}} \\
&\qquad + 2e^{-\{C_0 K_{n_p} - C_3 K_{n_p}[\log(Ne/S)+\log(12/\delta_S)]K_{n_p}/\log(N/S)\}} \\
&\quad = 2(n_p - 1)e^{-\{C_0 - \frac{C_3 K_{n_p}}{K}[1+(1+\log(12/\delta_S))/\log(N/S)]\}K} \\
&\qquad + 2e^{-\{C_0 - C_3[1+(1+\log(12/\delta_S))/\log(N/S)]\}K_{n_p}}.
\end{aligned} \tag{87}$$

Denoting the constant terms as $C_4 = C_0 - C_3[1 + (1 + \log(12/\delta_S))/\log(N/S)]$ and $C_4' = C_0 - (C_3 K_{n_p}/K)[1 + (1+\log(12/\delta_S))/\log(N/S)]$ and choosing C_3 small enough in order to guarantee that C_4 and C_4' are positive, we obtain (35).

REFERENCES

[1] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[2] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[3] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[4] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4036–4048, Sep. 2006.

[5] M. Wakin, J. N. Laska, M. F. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "An architecture for compressive imaging," in *Proc. IEEE ICIP*, Atlanta, GA, Oct. 2006, pp. 1273–1276.

[6] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Joint sourcechannel communication for distributed estimation in sensor networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3629–3653, Oct. 2007.

[7] Z. Yu, S. Hoyos, and B. M. Sadler, "Mixed-signal parallel compressed sensing and reception for cognitive radio," in *Proc. IEEE ICASSP*, Las Vegas, NV, Apr. 2008, pp. 3861–3864.

[8] G. Taubock and F. Hlawatsch, "A compressed sensing technique for OFDM channel estimation in mobile environments: Exploiting channel sparsity for reducing pilots," in *Proc. IEEE ICASSP*, Las Vegas, NV, Apr. 2008, pp. 2885–2888.

[9] Y. M. Lu and M. N. Do, "A theory for sampling signals from a union of subspaces," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2334–2345, Jun. 2008.

[10] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.

[11] J. N. Laska, S. Kirolos, M. F. Duarte, T. S. Ragheb, R. G. Baraniuk, and Y. Massoud, "Theory and implementation of an analog-to-information converter using random demodulation," in *Proc. IEEE ISCAS*, New Orleans, LA, May 2007, pp. 1959–1962.

[12] O. Taheri and S. A. Vorobyov, "Segmented compressed sampling for analog-to-information conversion," in *Proc. IEEE CAMSAP*, Aruba, Dutch Antilles, Dec. 2010, pp. 113–116.

[13] W. Badjwa, J. D. Haupt, G. M. Raz, S. J. Wright, and R. D. Nowak, "Toeplitz-structured compressed sensing matrices," in *Proc. IEEE SSP*, Madison, WI, Aug. 2007, pp. 294–298.

[14] R. Baraniuk and P. Steeghs, "Compressive radar imaging," presented at the IEEE Radar Conf., Waltham, MA, Apr. 2007.

[15] O. Taheri and S. A. Vorobyov, "Empirical risk minimization-based analysis of segmented compressed sampling," presented at the 44th Annu. Asilomar Conf. Signals, Systems, Computers, Pacific Grove, CA, Nov. 2010.

[16] P. Vandewalle, J. Kovacevic, and M. Vetterli, "Reproducible research in signal processing," *IEEE Signal Process. Mag.*, vol. 26, no. 3, pp. 37–47, May 2009.

[17] D. L. Donoho and J. Tanner, "Counting faces of randomly projected polytopes when the projection radically lowers dimension," *J. Amer. Math. Soc.*, vol. 22, no. 1, pp. 1–53, Jan. 2009.

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

[19] R. G. Baraniuk, M. Davenport, R. De Vore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Construct. Approx.*, vol. 28, no. 3, pp. 253–263, Dec. 2008.

[20] D. Donoho, "For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest solution," *Commun. Pure Appl. Math.*, vol. 59, pp. 797–829, Jun. 2006.

[21] E. J. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, pp. 1207–1223, Aug. 2006.

[22] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.

[23] D. Angelosante and G. B. Giannakis, "RLS-weighted LASSO for adaptive estimation of sparse signals," in *Proc. IEEE ICASSP*, Taipei, Taiwan, R.O.C., Apr. 2009, pp. 3245–3248.

[24] C. Craig, "On the Tchebycheff inequality of Bernstein," *Ann. Math. Stat.*, vol. 4, no. 2, pp. 94–102, May 1933.

[25] Y. Zhang, "Theory of compressive sensing via l_1-minimization: A non-RIP analysis and extensions," Rice Univ. CAAM, Houston, TX, Tech. Rep. TR08-11, 2008 [Online]. Available: http://www.caam.rice.edu/~yzhang/reports/tr0811.pdf

Omid Taheri (S'10) received the B.Sc. and M.Sc. degrees in electrical engineering from Isfahan University of Technology, Isfahan, Iran, in 2005 and 2007, respectively.

He is currently working towards the Ph.D. degree in electrical engineering at the University of Alberta, Edmonton, AB, Canada. His research interests are in digital signal processing with emphasis on compressive sampling, analog-to-information conversion, and sparse channel estimation.

Sergiy A. Vorobyov (M'02–SM'05) received the M.Sc. and Ph.D. degrees in systems and control from Kharkiv National University of Radio Electronics, Ukraine, in 1994 and 1997, respectively.

Since 2006, he has been with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada, where he became an Associate Professor in 2010. Since his graduation, he also occupied various research and faculty positions in Kharkiv National University of Radio Electronics, Ukraine; Institute of Physical and Chemical Research (RIKEN), Japan; McMaster University, Canada; Duisburg-Essen University and Darmstadt University, both in Germany; and the Joint Research Institute, Heriot-Watt University, and Edinburgh Univerity, all in the United Kingdom. His research interests include statistical and array signal processing, applications of linear algebra, optimization, and game theory methods in signal processing and communications, estimation, detection, and sampling theories, and cognitive systems.

Dr. Vorobyov is a recipient of the 2004 IEEE Signal Processing Society Best Paper Award, the 2007 Alberta Ingenuity New Faculty Award, and other research awards. He was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2006 to 2010 and for the IEEE SIGNAL PROCESSING LETTERS from 2007 to 2009. He is a member of Sensor Array and Multi-Channel Signal Processing and Signal Processing for Communications and Networking Technical Committees of the IEEE Signal Processing Society.