Learning-based design of random measurement matrix for compressed sensing with inter-column correlation using copula function

Mahdi Parchami, Hamidreza Amindavar, Wei-Ping Zhu

1Department of Electrical Engineering, Amirkabir University of Technology, 424 Hafez Ave., Tehran, Iran
2Department of Electrical and Computer Engineering, Concordia University, 1455 De Maisonneuve Blvd. W., Montreal, Canada
E-mail: mparchami@aut.ac.ir

Abstract: In this work, a novel learning-based approach for the design of a compressed sensing measurement matrix is proposed. In contrast with the state-of-the-art methods, the suggested approach takes into account the correlation within entries of each column of the measurement matrix, namely, the inter-column correlation (ICC). The new method makes use of a rather small number of training sparse signal vectors in a recursive scheme to obtain their corresponding measurement vectors. The latter is exploited to estimate the copula function of measurements which, in turn, is used to generate an arbitrary number of measurement vector ensembles. By using the latter, the autocorrelation of the measurement vectors is estimated precisely and then, the ICC of measurement matrix under design is obtained from the autocorrelation. Given the resulting ICC, the measurement matrix columns are to be generated independently, e.g. by employing a proper random Gaussian vector generator. Performance evaluations using both synthetic and real-world data confirm the superiority of the proposed approach to the less recent methods.

1 Introduction

Over the past two decades, an extensive amount of research has been devoted to the problem of capturing a compact (compressed) set of measurements from a potentially sparse signal, i.e. one with many zero coefficients in some domain. Properly designed, such a measurement set of a sparse signal preserves its essential content, whereas, at the same time, it reduces the costs in storage, processing, and transmission of the signal. This problem, often termed as compressed sensing (CS), can be represented as follows: \( y = \Phi x + v \) (1)

where \( y \in \mathbb{R}^n \) denotes the real vector of observations, \( \Phi \in \mathbb{R}^{n \times p} \) is a known (designed) measurement matrix with \( n < p, x \in \mathbb{R}^p \) is a so-called \( m \)-sparse signal (i.e. with \( m \) non-zero entries at most) to be estimated, and \( v \in \mathbb{R}^n \) represents the additive measurement noise [1]. Such a problem is raised in a wide range of applications including medical resonance imaging (MRI), spectroscopy, radar, Fourier optics, shortwave-infrared cameras, facial recognition, and network tomography [2, 3]. Moreover, sampling and image super-resolution reconstruction in the transform domain are important categories of information processing, which are associated with the topic of CS, e.g. [4, 5]. Furthermore, rather recently, electrical impedance tomography, which calculates the internal conductivity distribution within a body using electrical contact measurements, has benefited from reconstruction methods based on CS [6].

Considering the problem statement in (1), the most primary question is that how many measurements suffice to perfectly recover some sparse signal, \( x \), with high probability. The following corollary is an answer to this remarkable question [3].

**Corollary 1:** Let us draw each column of the measurement matrix, \( \Phi \), independently from the standard Gaussian distribution on \( \mathbb{R}^p \). If \( n \geq Cm \log(p/m) \), it is possible to reconstruct any \( m \)-sparse signal, \( x \in \mathbb{R}^p \), from the observations \( y \) in (1) with the probability exceeding \( 1 - e^{-Cn} \).

It is conventionally known that the optimal solution to the problem in (1) can be cast as the following minimisation problem

\[
\min \| x \|_0 \quad \text{subject to} \quad y = \Phi x.
\]

with \( \| . \|_0 \) being the \( \ell_0 \)-norm, i.e. the number of non-zero entries in its argument. Yet, in practice, solving this \( \ell_0 \)-minimisation problem is NP-hard and computationally infeasible for rather large signal dimensions [1]. As an alternative, the basis pursuit (BP) solution can be used [7], which replaces the \( \ell_0 \)-norm by an \( \ell_1 \)-norm, converting the CS reconstruction problem into a convex optimisation. There are also numerous variations of the latter method in the recent literature, e.g. [8, 9]. Moreover, in [10], the authors aim at directly minimising the \( \ell_0 \)-norm by a computationally fast method called smoothed \( \ell_0 \)-norm. On the other hand, greedy algorithms with the orthogonal matching pursuit (OMP) being the most popular, have been developed [11]. Moreover, in the statistical framework, Bayesian formulations for the estimation of sparse signals were suggested, wherein confidence measures for the estimated signal can be calculated [12]. A few major developments of the latter work have also been proposed, e.g. in [13], for the sparse component analysis in the presence of noise.

To date, the most fundamental directions in the development of the CS theory include the following: the creation of efficient yet practical sparse signal reconstruction algorithms [14–17], learning-based adoption of suitable structures/models within the sparse signal [18–20], taking into account the suppression of the additive noise [21–23], and design of the CS measurement matrix under different practical circumstances. In this work, we basically focus on the latter topic, i.e. the design of the measurement matrix for CS in a generic sense. To the best of our knowledge, compared to the tremendous amount of the existing research on CS topic, fewer methods exist that focus on the design of suitable measurement matrices under general assumptions. Traditionally, most suggested CS algorithms tend to employ the Gaussian and Bernoulli measurement matrices with independent and identically distributed (i.i.d.) entries [1]. Although this choice of the sensing matrix provides robust and desirable performance in many applications of...
CS to reduce the reconstruction error and the computational load involved, a number of methods to design the matrix \( \Phi \) have been suggested in the past. From a general point of view, the methods of CS matrix design can be categorised into deterministic and random measurement matrices. We present a brief extract of this literature in the following.

Whereas random measurement matrices ensure sparse recovery via \( \ell_p \)-minimisation with an optimal bound on the number of required measurements, deterministic matrices can be more suitable for applications subject to physical or other constraints and are less computationally expensive [1]. In [19], a framework to construct deterministic measurement matrices and to optimise them along with a sparsifying dictionary was proposed. Moreover, in [24], the concept of block \( p \)-ary codes is exploited to design deterministic matrices, which are able to perform almost as well as random matrices. More recently in [25], a subsampled structure of the form \( P_\Psi \Psi \) was considered for \( \Phi \) with \( \Psi \) and \( P_\Delta \) respectively, being the basis matrix and the subsampling operator. Therein, data-driven optimisation procedures were sought to obtain the index set, \( \Omega \), of the projector \( P_\Delta \).

Measurement matrices with random entries have the advantage of being robust in terms of their performance, e.g. great reconstruction power, in most CS applications. Moreover, they universally satisfy the fundamental theoretical requirements of a generic CS matrix such as the restricted isometry property and the mutual incoherence [26]. Nevertheless, regarding the design of fast random sensing matrices, limited research studies exist within the literature. In chapter 9 of [1], a thorough review of the topic of sparse signal recovery with random matrices has been presented. Therein, generic random matrices whose entries are drawn independently according to a sub-Gaussian distribution have been studied with Gaussian and Rademacher distributions as special cases. Even though the incoherence property imposes that a universally suitable measurement matrix should have uncorrelated columns, there are still a few fundamental remarks that can delineate the employment of traditional random matrices (see Section III.A of [11] for a discussion of this). One such remark is the fact that each independent column of a random \( \Phi \) may have some correlation within its own entries. This correlation, namely, the inter-column correlation (ICC), has been taken into account in [27] to design the independent columns of \( \Phi \). In the latter work, based on a variant of the structure for \( \Phi \) suggested in [25], namely, \( P_\nu \Phi \) with \( \nu \) as an unknown parameter controlling the ICC, a mixture of training sparse signals are used to estimate \( \nu \). Not using the resulting \( P_\nu \Phi \), an estimation for the ICC is performed which is, in turn, exploited in a Gaussian random vector generator to produce the columns of the ultimate random \( \Phi \). Looking into the prospective directions in CS, there are more futuristic concepts to come into use in measurement matrix design, such as the learning-based method aiming at maximising the entropy of measurement vectors in [28].

In this study, towards developing a more generic structure for a fully random measurement matrix, we take into account the aforementioned notion of ICC. In this regard, we make use of the concept of copula functions which has been successfully used in [29] to develop a CS framework with a multivariate Gaussian copula. The notion of ICC, apart from providing a more robust structure for the measurement matrix, performs as a degree-of-freedom to adapt the measurement matrix to input observations, i.e. the sparse signal of interest. To this end, in contrast with [27], rather than limiting the proposed solution to a particular structure in the estimation of the ICC as well as only using the limited number of training signals, by making use of the concept of copula functions, we first generate an arbitrary number of measurement vectors, \( y \), and then use these pseudo-measurements to estimate the present correlation across the elements of the vector \( y \). This correlation is subsequently exploited to calculate the ICC by means of a straightforward yet accurate relationship. The i.i.d. columns of the measurement matrix under design are then generated using a random vector generator with the estimated ICC.

The remainder of this paper is organised as follows. In Section 2, the proposed algorithm for the design of \( \Phi \) is outlined, which is based on the concept of copula from the statistics literature. In Section 3, the proposed approach to design the random measurement matrix is investigated and compared against the robust choices for the random \( \Phi \). Therein, the advantage of the proposed method is validated by means of objective error measures as well as subjective assessment. Section 4 is devoted to concluding remarks.

### 2 Proposed algorithm for measurement matrix design

In this section, we present in detail our approach for designing the random measurement matrix with ICC by leveraging the concept of copula. We present our methodology in the real field, \( \mathbb{R} \), although it can be generalised to the complex field. We make no restricting assumptions on the structure/nature of the underlying sparse signal nor the measurement matrix in general.

Denoting the \( k \)-th column of the measurement matrix \( \Phi = [\phi_1 \ldots \phi_p] \) by \( \phi_k = [\phi_{k1} \ldots \phi_{kn}]^\top \) for \( 1 \leq k \leq p \), with \( T \) denoting the matrix transpose, we have for the elements of the ICC matrix of interest: \( \rho_{ij}^\nu \equiv \text{E}\{\phi_{ik}\phi_{jk}\} \) for \( 1 \leq i<j \leq n \), or in the matrix form: \( \mathbf{R}_\nu = \text{E}\{\phi_i\phi_j^\top\} \equiv \text{E}\{\nu\} \) denoting the statistical expectation. Now, in accordance with [1, 3], assuming i.i.d. and zero-mean columns for the measurement matrix under study, the subscript \( k \) can be dropped and we have that \( \rho_{ij}^\nu \equiv r_{ij} \) or equivalently, \( \mathbf{R}_\nu \equiv \mathbf{R} \). Our goal within the rest of this section is to provide an appropriate approach for the estimation of the ICC to generate the random \( \Phi \). To this end, considering the problem statement in (1), by denoting \( y = [y_1 \ldots y_n]^\top \), \( x = [x_1, x_2, \ldots, x_n]^\top \), and \( v = [v_1, v_2, \ldots, v_n]^\top \), it is deduced that

\[
y_i = \sum_{k=1}^p \phi_{ik} x_k + v_i, \quad 1 \leq i \leq n
\]

Using (3) to calculate the autocorrelation of \( y \) gives

\[
\text{E}\{y_i y_j\} = \sum_{k=1}^p \sum_{l=1}^p \text{E}\{\phi_{ik}\phi_{jl}\} + \text{E}\{v_i v_j\}
\]

\[
= \sum_{k=1}^p \sum_{l=1}^p \text{E}\{\phi_{ik}\phi_{jl}\} x_k x_l + \text{E}\{v_i v_j\}, \quad 1 \leq i,j \leq n
\]

where we used the fact that the additive noise, \( v_i \), is zero-mean and independent from the summation in (3) and also that \( \Phi \) and \( x \) are independent. Assuming the independence of columns of \( \Phi \) and having zero-mean elements at each column, it follows that \( \text{E}\{\phi_{ik}\phi_{jk}\} = \text{E}\{\phi_{ik}\}\text{E}\{\phi_{jk}\} = 0 \) unless \( k = k' \). Applying the latter identity to (4) results in the following:

\[
\text{E}\{y_i y_j\} = \sum_{k=1}^p \text{E}\{\phi_{ik}\phi_{jk}\} \text{E}\{x_k \} + \text{E}\{v_i v_j\}, \quad 1 \leq i,j \leq n
\]

Considering that \( \text{E}\{\phi_{ik}\phi_{jk}\} \) is independent of \( k \) and equal to the ICC term \( r_{ij} \), it is inferred that

\[
r_{ij} = \frac{\text{E}\{y_i y_j\} - \text{E}\{v_i v_j\}}{\sum_{k=1}^p \text{E}\{x_k \}}, \quad 1 \leq i,j \leq n
\]

To obtain the ICC, i.e. \( r_{ij} \), using the above, the correlation of the additive noise has to be known or estimated in advance. Also, it is observed that the ICC of the measurement matrix scales with the autocorrelation of the measurements, \( \text{E}\{y_i y_j\} \), whereas the latter is unknown. Directly estimating this term would require an enough number of training samples of the measurement \( y_i \), which is not available prior to designing the measurement matrix. To overcome this complication, we employ the concept of copula functions to
properly generate a sufficient number of samples of the measurement vector, \( y \), and then, use the generated samples in an averaging scheme to estimate the correlation \( E\{yy^\top\} \) in (6).

Algorithm 1 in the sequel outlines the suggested algorithm to determine the a priori measurements whereas Algorithm 2 details the overall framework proposed for the design of the measurement matrix. In the following subsections, we will elaborate on the different steps of these algorithms.

**Algorithm 1:** Suggested approach for obtaining the a priori measurements

*Input:* the training sparse signal samples, \( \{x_i^j\}_{i=1}^\Lambda \), with \( \Lambda \) denoting the sample index and the number of training signal vectors, respectively, the initial measurement matrix with i.i.d. Gaussian entries, \( \Phi^0 \), and the number of desired iterations, \( J \).

*Output:* the desired a priori measurements, \( \{y_i^j\}_{i=1}^\Lambda \).

- Set \( y_i^{0\beta} = \Phi^0 x_i^j \), for \( \lambda = 1, 2, \ldots, \Lambda \).
- Repeat the subsequent steps for \( j = 1, 2, \ldots, J \).
- Set \( E\{yy^\top\} \approx (1/\Lambda) \sum_{\lambda=1}^\Lambda y_{\lambda}^{(\lambda-1)} y_{\lambda}^{(\lambda-1)\top} \).
- Use (6) to obtain the ICC as
  \[
  R_p^\lambda = E\{yy^\top\} - E[yy^\top] \sum_{\lambda=1}^\Lambda E[xx^\top] \tag{7}
  \]
- Employ a Gaussian random vector generator to generate the updated measurement matrix, \( \Phi^\lambda \), with the ICC matrix as in step 2.
- Use \( \Phi^\lambda \) to sense the measurements as \( y_i^{\lambda\beta} = \Phi^\lambda x_i^j \), for \( \lambda = 1, 2, \ldots, \Lambda \).
- \( y_i^{\lambda\beta} \) for \( \lambda = 1, 2, \ldots, \Lambda \) is the set of desired a priori measurements.

2.1 Determination of the a priori measurements

According to (1), without an assumption for the CS measurement matrix, there will be no prior knowledge of training samples for \( y \) obtained by employing the available training samples of the sparse signal of interest, \( x \). To overcome this limitation, in an iterative fashion, we start by assuming an initial measurement matrix, \( \Phi^0 \), e.g. the Gaussian matrix with i.i.d. entries herein, and use it to sense the training sparse signal, \( \{x_i^j\}_{i=1}^\Lambda \), to obtain the corresponding set of training, or namely, a priori measurements, \( \{y_i^j\}_{i=1}^\Lambda \), at the first iteration. Next, we use the latter to provide an estimate of the autocorrelation of measurements as \( E\{yy^\top\} \approx (1/\Lambda) \sum_{\lambda=1}^\Lambda y_{\lambda}^{(\lambda-1)} y_{\lambda}^{(\lambda-1)\top} \). This correlation is then plugged into (6) to obtain a preliminary estimate of the ICC of the measurement matrix, \( \Phi \), which is exploited by a proper Gaussian random vector generator to generate the i.i.d. columns of the new measurement matrix \( \Phi^\lambda \). Next, the updated \( \Phi \) is used to sense \( \{x_i^j\}_{i=1}^\Lambda \) again to obtain the updated measurement vectors, which are then used in the averaging scheme to estimate the updated autocorrelation, \( E[yy^\top] \). The latter is again used in (6), resulting in the updated ICC that is used for generating the next measurement matrix, \( \Phi^\lambda \), to sense \( \{x_i^j\}_{i=1}^\Lambda \). Again, at the end of a few iterations, the ultimate set of the measurement vectors, \( \{y_i^j\}_{i=1}^\Lambda \), is to be used in the estimation of the copula function of \( y \), as will be discussed in the following subsection.

**Algorithm 2:** Proposed approach for the design of random measurement matrix

*Input:* the training sparse signal vectors, \( \{x_i^j\}_{i=1}^\Lambda \), and the initial measurement matrix with i.i.d. Gaussian entries, \( \Phi^0 \).

*Output:* the i.i.d. columns of the measurement matrix, \( \Phi \), with ICC.

- Use Algorithm 1 to obtain the a priori measurements, \( y_i^j \), from \( x_i^j \) and \( \Phi^0 \), for \( \lambda = 1, 2, \ldots, \Lambda \).
- Estimate the marginal cumulative distribution function's (CDF's) and probability distribution function's (PDF's) of \( y \) as well as the copula density function of \( y \) by using \( \{y_i^j\}_{i=1}^\Lambda \), and adopt them to obtain the empirical joint PDF of \( y \), as explained in Section 2.2.
- Exploit the empirical joint PDF of \( y \) to generate the pseudo-measurement vectors, \( \{y_i^{\lambda\beta}\}_{\lambda=1}^\Lambda \), with \( \lambda \) and \( \Lambda \) denoting the sample index and the number of desired iterations, respectively, as discussed in Section 2.2.
- Use \( \{y_i^{\lambda\beta}\}_{\lambda=1}^\Lambda \) in a smoothing scheme to estimate the ICC matrix, \( R = [r_{ij}] \), as explained in Section 2.3.
- Employ a Gaussian random vector generator to independently generate \( \mu \) columns of the measurement matrix, \( \Phi \), with the correlation matrix equal to \( R \) in the above.

2.2 Making use of copula to generate pseudo measurement vectors

As the next step of the proposed algorithm, it is required to generate a sufficient number of samples for the pseudo-measurement vectors, \( \{y_i^{\lambda\beta}\}_{\lambda=1}^\Lambda \), having the set of a priori measurements, \( \{y_i^j\}_{i=1}^\Lambda \). The most straightforward technique to fulfill this purpose is to estimate the parametric joint distribution of the measurement vector, \( y \), using \( \{y_i^{\lambda\beta}\}_{\lambda=1}^\Lambda \). However, due to the inherent complexity in doing so, given the dimension of \( y \) in practice, we leverage the concept of copula and make use of the empirical copula to estimate the joint distribution. The latter is used to generate an arbitrary number of samples for \( y \).

In the probability theory and statistics, a copula is defined as a multivariate probability distribution, \( c(y_1, y_2, \ldots, y_n) \), for which the marginal distribution of each random variable, \( y_p \), is uniform [30]. According to Sklar’s theorem, any multivariate joint distribution can be expressed in terms of univariate marginal distributions and a copula function, as the following [31]:

\[
  f(y) = c(F_1(y_1), F_2(y_2), \ldots, F_n(y_n)) \prod_{i=1}^n f_i(y_i), \tag{8}
\]

where \( f(y) \) is the joint distribution of \( y \), \( F_i(y_i) \) and \( f_i(y_i) \) are, respectively, the marginal CDF and PDF of \( y_i \), and \( c(\cdot) \) is the so-called copula density, which is the joint distribution of \( F_i(y_i) \)'s. There are a few famous parametric copula families in the literature, which include parameters controlling the strength of dependence, e.g. the Archimedean copulas [30]. Although the estimation of copula parameters has been thoroughly studied in the literature, e.g. in [32], in our application, due to the high dimensionality of \( y \), direct estimation of parametric copulas was found to be a tedious task. Thus, to avoid further complexity, the empirical (non-parametric) copula is preferred for the purpose of this work.

To empirically estimate the copula-based joint PDF in (8), both the marginal PDF's and CDF's and the copula density need to be replaced by their empirical estimates, based on a given sample of \( y \), which is \( \{y_i^j\}_{i=1}^\Lambda \). To proceed further, having sorted the \( \Lambda \) entries along each dimension of \( \{y_i^{\lambda\beta}\}_{\lambda=1}^\Lambda \), from the smallest to the largest to form the vector set \( \{\hat{y}_i^{\lambda\beta}\}_{\lambda=1}^\Lambda \), we shall have for the empirical marginal CDF's the following:

\[
\begin{align*}
  0, \quad & y_i < \hat{y}_i^{\lambda\beta} \\
  1, \quad & y_i \geq \hat{y}_i^{\lambda\beta}
\end{align*}
\]

\[
\hat{F}_i(y_i) = \begin{cases} 
0, & y_i < \hat{y}_i^{\lambda\beta} \\
1, & y_i \geq \hat{y}_i^{\lambda\beta}
\end{cases} \tag{9}
\]
for \( \lambda = 1, 2, \ldots, \Lambda - 1 \), where \( y_i^{(k)} \) is the \( k \)th entry of \( y_i^{(k)} \) for \( i = 1, 2, \ldots, n \). Subsequently, for the empirical marginal PDFs, it follows that

\[
0, \quad y_i^{(k)} < y_{i-1}^{(k)} \\
\int_{y_i^{(k)}}^{y_{i+1}^{(k)}} f_i(y) \, dy, \quad y_{i-1}^{(k)} \leq y_i^{(k)} \leq y_{i+1}^{(k)} \\
0, \quad y_{\Lambda}^{(k)} \leq y_i^{(k)}
\]

where we used the fact that \( f_i(y) \) is the derivative (forward difference) of \( f(y) \). To derive an estimate for the copula density function, \( c(\cdot) \), in (8), we resort to the empirical copula proposed in [33]. In this sense, by taking \( u_i = F_i(y_i) \) for \( i = 1, 2, \ldots, n \), and also defining the positive integer \( K \) as a precision parameter, with \( \zeta = 1/K \), the following expression can be deduced for the empirical copula density function [33]

\[
\hat{c}(u) = \frac{N_{Ku}}{\Lambda^2} \quad u \in I_K
\]

where \( u = [u_1, u_2, \ldots, u_{\Lambda}]^T \), \( I_K \) is the \( K \times K \) block diagonal matrices, \( I_{\lambda, \lambda} \) with the subintervals of \([0, 1]\) as \( I_{\lambda, \lambda} \equiv [\{j \in \{1, \ldots, K\} \} \text{ for } i = 1, 2, \ldots, n, \text{ and } N_{Ku} \) equal to the cardinality of the set \( \{u_i | u_i \in I_K, \lambda = 1, 2, \ldots, \Lambda\} \) over \( \lambda \) with \( u_i \) calculated as \( \hat{F}_i(y_i^{(\lambda)}) \).

This empirical copula density estimate at \( \lambda \) follows that

\[
\int_{y_i^{(k)}}^{y_{i+1}^{(k)}} f_i(y) \, dy, \quad y_{i-1}^{(k)} \leq y_i^{(k)} \leq y_{i+1}^{(k)}
\]

In (11), it is observed that the copula density estimate at \( u \) located in a specific subcube, i.e. \( I_K \), is proportional to the number of measurements, \( u_i \), located in \( I_K \) divided over the entire number of measurements, \( \Lambda \). This actually justifies the expression for \( \hat{c}(u) \) as the joint density of the vector \( u \), which is in fact a piecewise constant function in each subcube of the unit n-cube.

At this stage, having estimated the marginal CDF’s and PDF’s as well as the copula density through (9)–(11), the joint PDF of the unit -cube.

\[
\int_{y_i^{(k)}}^{y_{i+1}^{(k)}} f_i(y) \, dy, \quad y_{i-1}^{(k)} \leq y_i^{(k)} \leq y_{i+1}^{(k)}
\]

In (11), it is observed that the copula density estimate at \( u \) located in a specific subcube, i.e. \( I_K \), is proportional to the number of measurements, \( u_i \), located in \( I_K \) divided over the entire number of measurements, \( \Lambda \). This actually justifies the expression for \( \hat{c}(u) \) as the joint density of the vector \( u \), which is in fact a piecewise constant function in each subcube of the unit n-cube.

2.3 Smoothing scheme for the estimation of ICC

In this part, we use a straightforward smoothing scheme on the set of pseudo-random measurement vectors, \( \{y_i^{(k)}\}_{k=1}^{\Lambda} \), to obtain the autocorrelation of measurements, \( E(y_i y_j) \), and just the ICC, \( \rho_{ij} \), given by (6). For doing so, given a sufficient number of samples, \( \Lambda \), we approximate the expectation, \( E(\cdot) \), by an averaging, resulting in the estimation, \( \hat{E}(y_i y_j) = (1/\Lambda) \sum_{k=1}^{\Lambda} y_i^{(k)} y_j^{(k)} \), with \( y_i^{(k)} \) denoting the \( k \)th element of \( y_i^{(k)} \), \( 1 \leq i, j \leq \Lambda \). In the case of complex field, however, to further reduce the estimation error, one can take advantage of the fact that the correlation matrix, \( E(\Re y_i y_j) \), with \( \Re \) denoting the matrix Hermitian, is a Hermitian symmetric matrix, and therefore, it follows

\[
\hat{E}(y_i y_j) = \frac{\hat{E}(y_i y_j) + \hat{E}(y_j y_i)}{2}.
\]

where \( \hat{y}_j \) denotes the complex conjugate of \( y_j \) and \( \hat{E}(y_i y_j) \) denotes the estimation of \( E(y_i y_j) \) by the aforementioned averaging scheme for the upper-diagonal elements, i.e. for \( j > i \). Obviously, the above strategy is useful only for the off-diagonal elements of the correlation matrix. Finally, exploiting the estimated autocorrelation from (12) in (6) provides the desired ICC to be used as the correlation matrix in a proper Gaussian random vector generator to generate the columns of the measurement matrix, i.e. \( q_{ij} \) for \( 1 \leq k \leq p \).

It is worth mentioning that since each column of the proposed measurement matrix is generated independently, there will be no mutual coherence between each pair of columns. Therefore, the proposed ICC-based measurement matrix satisfies the mutual incoherence condition within the CS theory.

3 Experimental results

In this section, we present a summary of the experiments demonstrating the advantage of the proposed approach. To implement the suggested approach in Section 2, we took the number of training sparse signals, \( \Lambda \), in the range of \([100, 200]\), the precision \( K = 20 \), and we subsequently generated \( L = 10^4 \) ensembles of the pseudo-measurement vectors, \( y_i^{(k)} \). Also, it was experimentally observed that having only \( J = 2 \) iterations suffices to obtain the set of \( K \) measurements in Algorithm 1, with higher \( J \) values providing no sensible improvement. In all conducted experiments, the set of training and test signals were chosen to be different but with similar signal parameters including \( d \) and \( m \) for the simulated signals. Finally, to independently generate the columns of the measurement matrix, \( \Phi \), we used the \texttt{mvnrnd} function of MATLAB with the parameters MU and SigmAam respectively, set to \( \{0\}_{\times 1} \) and the ICC matrix, \( \Omega \), as obtained in (6).

To thoroughly investigate the advantage of the proposed measurement matrix, \( \Phi \), in Section 2 w.r.t. other measurement matrices, we evaluate their performance by using them in several major CS methods, namely, the OMP [11], Compressive sampling matching pursuit (CoSaMP) [35], Bayesian compressed sensing (BCS) [12], and BP [7] methods. We consider the traditional \( \Phi \) with i.i.d. Gaussian elements, the deterministic measurement matrix suggested in [25], a Gaussian \( \Phi \) with fixed ICC, as well as the suggested ICC-based \( \Phi \) in [27]. The measurement matrix with deterministic elements in [25] is a subsampled structure of the form \( F \Psi \Psi^T \) with \( \Psi \) and \( P_{\Omega} \), respectively, being the basis matrix and subsampling operator. Therein, a data-driven optimisation procedure is sought to obtain the index set, \( \Omega \), of projector \( P_{\Omega} \). In our implementation of the aforementioned method, we take the corresponding objective function to be \( f_{\text{avg}} \) which suggests an averaging over training signals prior to optimising for the index set \( \Omega \) used in \( P_{\Omega} \). As for the \( \Phi \) with a fixed ICC, we use an exponentially decaying correlation structure, namely, \( e^{-\alpha \tau} \), with \( \alpha \) and \( \tau \), respectively, being a fixed adjustable constant and the lag between each two entries of a column. Also, as with the suggested \( \Phi \) in [27] with random elements, some structures are assumed for the ICC within the columns of \( \Phi \), which is controllable through a parameter that is optimised in a training-based framework.

For the evaluation of the methods under study, we consider both synthetic and real-world signals. In case of synthetic signals, we distinguish two different scenarios, one with a randomly generated binary signal and the other with a random zero-mean Gaussian signal. In the former case, we generate a binary signal of length \( p \), consisting of \( m \) ones (sparsity degree) randomly distributed along with the signal array with the rest of its elements set to zero. This signal is to be recovered using \( n \) linear measurements as given by (1). Since perfect reconstruction is possible in this case, the evaluations can be performed in terms of the perfect reconstruction probability of the sparse signal.

We first investigate the performance of the proposed measurement matrix in Section 2 with different numbers of the training sparse signal vectors, \( \Lambda \). Shown in Fig. 1 is the

\[\text{IET Signal Process., 2020, Vol. 14 Iss. 6, pp. 385-395} \]

© The Institution of Engineering and Technology 2020
reconstruction error obtained by exploiting the proposed \( \Phi \) in the OMP method versus \( \Lambda \).

In this experiment, we considered reconstructing randomly generated Gaussian sparse signals contaminated by white noise in an signal-to-noise ratio (SNR) value of 10 dB; and we computed the normalised mean squared error (MSE) between the true and reconstructed signal. It is seen that by increasing the number of training signals from 20 to 100, the normalised MSE reduces dramatically but it remains almost fixed with higher \( \Lambda \) values for the parameter setting under study. This shows the number of required training samples for precisely estimating the underlying copula function, and therefore, the joint distribution of the measurement vector, which is considered in performing the foregoing experiments.

Next, we study the effect of the number of iterations used in Algorithm 1, namely, \( J \), on the overall performance of the proposed measurement matrix in Section 2. To this end, using the previous setup, we obtained the reconstruction errors versus \( J \), as shown in Fig. 2.

It is inferred that using only \( J = 2 \) iterations suffices for the aforementioned algorithm to provide the desired performance and that increasing \( J \) to higher values only results in trivial improvements. This fact led us to use no more than two iterations for all the other experiments and enabled us to avoid extra computational burden. Unless otherwise stated, we used the parameter settings \( J = 2 \) and \( \Lambda = 120 \).

Next, in Fig. 3, the reconstruction probability of a randomly generated binary signal is illustrated versus the sparsity level for the different measurement matrices. As viewed, independent of the four underlying CS approaches, the proposed \( \Phi \) is capable of providing larger reconstruction probabilities for a reasonable range of the sparsity level, \( m \). Furthermore, with increasing \( m \), it is observed that the reconstruction probability obtained by the suggested \( \Phi \) decays less sharply as compared to the other methods. These demonstrate the robust advantage in taking into account the ICC in designing the measurement matrix with different reconstruction methods.

Moreover, in the case of non-binary (continuous) signals, we reconstruct randomly generated Gaussian samples in white noise at different SNR values. Next, we calculated the normalised MSE between the true and reconstructed signals versus SNR, as illustrated in Fig. 4. It is observed that using the proposed ICC-based measurement matrix helps further in reducing the reconstruction error w.r.t. the other methods. This advantage is more pronounced at higher SNR values, since none of the measurement matrices are basically designed to deal with adverse noisy conditions. Such a concept is to be considered as a topic of future research. It should be noted that, however, to fairly compare the proposed approach with the other methods in Fig. 3, we did not take into account the noise correlation matrix, \( E\{vv^T\} \), in implementing Algorithm 1; whereas integrating a proper estimate of \( E\{vv^T\} \) into the latter, as performed before and shown in Fig. 4 (the plot named as proposed \( \Phi \) with known noise), leads to superior performance.

Often in the relevant literature, the performance of CS methods against the number of observations is of high interest. For this reason, in Fig. 5, we have demonstrated the error performance of the proposed measurement matrix along with others with the number of measurements, \( n \), changing from 60 to 140. As viewed in this experiment, the performance advantage of the proposed measurement matrix is apparent w.r.t. the less recent methods in consistency with \( n \).

The majority of the superiority of the proposed \( \Phi \) in Section 2 is due to making use of the learning-based process in Algorithm 2 to update the present ICC in accordance with the given training sparse signals, as against the \( \Phi \) containing a fixed ICC. Moreover, as opposed to the deterministic learning-based \( \Phi \) suggested in [25], the randomness of the proposed \( \Phi \) allows for a better...
reconstruction accuracy. Pursuing the performance evaluation for the case of binary signals, we here investigate the number of required measurements, $n$, to identify a sparse signal with a fixed reconstruction probability. Fig. 6 displays the relationship between $n$ and the sparsity level, $m$, to achieve a recovery probability of 95% with $p = 250$ using the OMP method with different measurement matrices presuming the noise-free condition. A trend of almost linearly growing $n$ with increasing $m$ is observed in this experiment. The advantage of using the proposed measurement matrix is evident by the smaller number of measurements it requires.

Subsequently, to investigate the relative performance of the proposed measurement matrix using a real-world data set, we implement the four reconstruction methods, namely, OMP, CoSaMP, BCS, and BP, using different measurement matrices on a dataset consisting of MRI images of the knee available in ‘mridata’ as discussed in [36]. For this purpose, we represent the MRI images in the discrete cosine transform (DCT) domain where the corresponding coefficient vectors of the images are sparse, and therefore, CS measurements can be realised efficiently. We also downsampled the original image files from 512 × 512 to smaller sizes to examine the results for different image resolutions. To avoid excessive complexity, we exploited the block-based CS, as

---

**Fig. 3** Reconstruction probability versus sparsity level for a random binary signal with noise-free measurements for $n = 100$ and $p = 250$
(a) OMP, (b) CoSaMP, (c) BCS, (d) BP

**Fig. 4** Normalised reconstruction MSE versus SNR for a random Gaussian signal with $n = 100$, $p = 250$, and $m = 30$
(a) OMP, (b) CoSaMP, (c) BCS, (d) BP
explained in Section 3.4 of [37]. In this strategy, an image is divided into $B \times B$ non-overlapping blocks and then acquired using the appropriately-sized measurement matrix. In this sense, we implemented the CS method on $16 \times 16$ image blocks with overlapping of 50% to reduce artefacts, resulting in the entire length of $p = 256$ for the sparse signal under reconstruction.

In Fig. 7, the normalised MSE in the reconstruction of the MRI images versus the image resolution is shown for a fixed compression ratio, i.e. $p/n = 4$. It can be viewed that the proposed approach consistently provides less reconstruction errors for different image sizes. In a similar fashion, the normalised MSE versus the compression ratio for $256 \times 256$ images is demonstrated in Fig. 8. Under the constraint of $p = 256$, the compression ratio, in fact, controls the number of measurements, $n$. Therefore, this experiment also investigates the impact of the number of measurements on the performance. It is observed in Fig. 8 that by increasing the compression ratio, or equivalently, reducing the number of measurements taken, the reconstruction errors monotonically increase. Comparatively, the smaller MSE values observed consistently confirm the advantage of using the proposed measurement matrix design.

To conduct further analysis of the real-world data, we here take advantage of an image quality model, the peak SNR (PSNR) which is defined as [38]

$$PSNR = 10 \log \left( \frac{2^b - 1}{\text{MSE}} \right).$$

Fig. 5 Normalised reconstruction MSE versus the number of measurements for a noisy random Gaussian signal at SNR = 20 dB with $p = 250$ and $m = 30$

(a) OMP, (b) CoSaMP, (c) BCS, (d) BP

Fig. 6 Number of measurements, $n$, required to recover an $m$-sparse binary signal with dimension $p = 256$ at least 95% of the time

Next, to provide an estimate for the processing costs involved in implementing the measurement matrices, we employ them in the OMP method to reconstruct binary signals with specified parameters and measure the execution times in MATLAB. The corresponding results for the case of 1000 trials of signal reconstruction are shown in Fig. 10. Please note that this chart displays the execution time to complete 1000 trials, which includes the time to generate 1000 sparse signals in addition to the time required by 1000 invocations of the OMP algorithm. For the most computationally intensive method, each trial takes an average of 0.33 s. It should be noted that the smaller computational effort
required by the method in [25] is mostly due to the deterministic nature of the suggested measurement matrix therein as opposed to the random structure in others. We believe that the relatively higher complexity imposed by the correlation-based measurement matrices is an acceptable compromise, considering their superior reconstruction performance.

Finally, to visually demonstrate the effectiveness of the proposed measurement matrix as compared with other methods, we show the recovered 512 × 512 MRI images of the knee from ‘mridata’ using the OMP algorithm with different measurement matrices under study in Figs. 11–16. For this experiment, we used the parameter setting \( n = 100 \) and \( m = 50 \) within the block-based implementation of the CS for the images. The advantage in the reconstruction by using the proposed measurement matrix compared with the other CS matrices is observable in this figure by means of the higher quality of the recovered images. It should be
mentioned that the set of training images was taken from the same database with the test and training images both being the figure of a knee MRI image captured at different shots.

4 Concluding remarks

A novel approach for the design of random CS measurement matrices, $\Phi$, was proposed. The proposed $\Phi$ considers correlations within each column of $\Phi$, namely, the ICC, in a generalised
framework. The ICC is estimated using training sparse signals by means of a statistical tool, i.e. the copula function. The copula, employed in this work as an empirical estimate of the joint distribution of the measurement vector, is exploited to estimate the autocorrelation of the measurement vector, and subsequently, the ICC. We considered a random Gaussian measurement matrix as the designed $\Phi$ with the obtained ICC along each of its independent columns and compared its reconstruction performance with some of the state-of-the-art measurement matrices using different CS methods. Our results consistently confirmed the performance advantage of the suggested $\Phi$. Moreover, by integrating a proper measurement noise estimator into the proposed approach, even further improvements can be achieved in the case of highly noisy measurements. Prospective research topics in this direction can include:

- Investigating application-oriented structures for the ICC in $\Phi$ as opposed to blindly estimating it.
- Making use of the specific features of the sparse signal of interest in designing the measurement matrix.
- Looking into the portion of correlation between the entries of the measurement vector due to the correlation present within the non-zero samples of the sparse signal.
- A few practical extensions including the development and integration of a blind estimator for the measurement noise, using imperfectly-known training signals, and application-specific constraints on the measurement matrix.

5 References

[1] Foucart, S., Rauhut, H.: 'A mathematical introduction to compressive sensing' (Springer, New York, 2013)
[2] Candès, E., Wakin, M.B.: 'An introduction to compressive sampling', IEEE Signal Process. Mag., 2008, 25, (2), pp. 21–30
[3] Eldar, Y.C., Kutyniok, G.: 'Compressed sensing: theory and applications' (Cambridge University Press, United Kingdom, 2012)
[4] Wei, D., Li, Y.: 'Reconstruction of multidimensional bandlimited signals from multichannel samples in linear canonical transform domain', IET Signal Process., 2014, 8, pp. 647–657
[5] Wei, D., Li, Y.: 'Generalized sampling expansions with multiple sampling rates for lowpass and bandpass signals in the fractional Fourier transform domain', IEEE Trans. Signal Process., 2016, 64, (18), pp. 4861–4874
[6] Nasehi Tehrani, J., Jin, C., McEwan, A., et al.: 'A comparison between compressed sensing algorithms in electrical impedance tomography', 2010 Annual Int. Conf. of the IEEE Engineering in Medicine and Biology, Buenos Aires, Argentina, 2010, pp. 3109–3112
[7] Donoho, D.L.: 'Compressed sensing', IEEE Trans. Inf. Theory, 2006, 52, (4), pp. 1289–1306
[8] Mostaghim, A., Jacques, L., Cambareri, V., et al.: 'Consistent basis pursuit for signal and matrix estimates in quantized compressed sensing', IEEE Signal Process. Lett., 2016, 23, (1), pp. 25–29
[9] Goldstein, T., Studer, C.: 'Phasemax: convex phase retrieval via basis pursuit', IEEE Trans. Inf. Theory, 2018, 64, (4), pp. 2675–2689
[10] Mohimani, H., Babaei-Zadeh, M., Jutten, C.: 'A fast approach for overcomplete sparse decomposition based on smoothed $\ell^0$ norm', IEEE Trans. Signal Process., 2009, 57, (1), pp. 289–301
[11] Tropp, J.A., Gilbert, A.C.: 'Signal recovery from random measurements via orthogonal matching pursuit', IEEE Trans. Inf. Theory, 2007, 53, (12), pp. 4655–4666
[12] Ji, S., Xue, Y., Carin, L.: 'Bayesian compressive sensing', IEEE Trans. Signal Process., 2008, 56, (6), pp. 2346–2356
[13] Zayyani, H., Babaei-Zadeh, M., Jutten, C.: 'An iterative Bayesian algorithm for sparse component analysis in presence of noise', IEEE Trans. Signal Process., 2009, 57, (11), pp. 4378–4390
[14] Needell, D., Vershynin, R.: ‘Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit’, Found. Comput. Math., 2009, 9, (3), pp. 317–334

[15] Donoho, D.L., Tsaig, Y., Drori, I., et al.: ‘Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit’, IEEE Trans. Inf. Theory, 2012, 58, (2), pp. 1094–1121

[16] Babacan, S.D., Molina, R., Katsaggelos, A.K.: ‘Bayesian compressive sensing using Laplace priors’, IEEE Trans. Image Process., 2010, 19, (1), pp. 53–63

[17] Baron, D., Sarvotham, S., Baraniuk, R.G.: ‘Bayesian compressive sensing via belief propagation’, IEEE Trans. Signal Process., 2010, 58, (1), pp. 269–280

[18] Duarte, M.F., Eldar, Y.C.: ‘Structured compressed sensing: from theory to applications’, IEEE Trans. Signal Process., 2011, 59, (9), pp. 4053–4085

[19] Duarte-Carvajalino, J.M., Sapiro, G.: ‘Learning to sense sparse signals: simultaneous sensing matrix and sparsifying dictionary optimization’, IEEE Trans. Image Process., 2006, 15, (7), pp. 1395–1408

[20] Baraniuk, R.G., Cevher, V., Duarte, M.F., et al.: ‘Model-based compressive sensing’, IEEE Trans. Inf. Theory, 2010, 56, (4), pp. 1982–2001

[21] Tsaig, Y., Donoho, D.L.: ‘Extensions of compressed sensing’, Signal Process., 2006, 86, (3), pp. 549–571

[22] Cai, T.T., Wang, L.: ‘Orthogonal matching pursuit for sparse signal recovery with noise’, IEEE Trans. Inf. Theory, 2011, 57, (7), pp. 4680–4688

[23] Malloy, M.L., Nowak, R.D.: ‘Near-optimal adaptive compressed sensing’, IEEE Trans. Inf. Theory, 2014, 60, (7), pp. 4001–4012

[24] Amini, A., Montazerhodjat, V., Marvasti, F.: ‘Matrices with small coherence using p-ary block codes’, IEEE Trans. Signal Process., 2012, 60, (1), pp. 172–181

[25] Baldassarre, L., Li, Y.H., Scarlett, J., et al.: ‘Learning-based compressive subsampling’, IEEE J. Sel. Top. Signal Process., 2016, 10, (4), pp. 809–822

[26] Candès, E.J., Eldar, Y.C., Needell, D., et al.: ‘Compressed sensing with coherent and redundant dictionaries’, Appl. Comput. Harmon. Anal., 2011, 31, pp. 59–73

[27] Parchami, M., Aminifar, H., Zhu, W.P.: ‘Learning-based design of measurement matrix with inter-column correlation for compressive sensing’. Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), Calgary, Canada, 2018

[28] Parpasarathly, G., Abhilash, G.: ‘Entropy-based learning of sensing matrices’, IET Signal Process., 2019, 13, pp. 650–660

[29] Sadeghigol, Z., Zayyani, H., Abin, H., et al.: ‘Multivariate copula spatial dependency in one bit compressed sensing’, arXiv:171109217, 2017

[30] Nelsen, R.B.: ‘An introduction to copulas (Springer series in statistics)’ (Springer-Verlag, Berlin, Heidelberg, 2006)

[31] Ushchendorf, L.R.: ‘On the distributional transform, Sklar's theorem, and the empirical copula process’, J. Stat. Plan. Inference, 2009, 139, (11), pp. 3921–3927

[32] Hofert, M., Mächler, M., McNeil, A.J.: ‘Likelihood inference for Archimedean copulas in high dimensions under known margins’, J. Multivariate Anal., 2012, 110, pp. 133–150

[33] Strelen, J.C., Nasuji, F.: ‘Analysis and generation of random vectors with copulas’. Proc. 39th Conf. on Winter Simulation: 40 Years! The Best is Yet to Come, Washington, D.C., US, 2007, pp. 488–496

[34] Strelen, J.C.: ‘Tools for dependent simulation input with copulas’. ICST, Paris, France, 2010

[35] Needell, D., Tropp, J.A.: ‘CoSaMP: iterative signal recovery from incomplete and inaccurate samples’, Appl. Comput. Harmon. Anal., 2009, 26, (3), pp. 301–321

[36] Lustig, M., Vasanawala, S.: ‘MRI datasets for compressed sensing’. Available at http://www.mridata.org/

[37] Fowler, J.E., Mun, S., Tramel, E.W.: ‘Block-based compressed sensing of images and video’ (Foundations and Trends in Signal Processing, United States, 2012)

[38] Do, T.T., Gan, L., Nguyen, N.H., et al.: ‘Fast and efficient compressive sensing using structurally random matrices’, IEEE Trans. Signal Process., 2012, 60, (1), pp. 139–154