Compressive Measurements Generated by Structurally Random Matrices: Asymptotic Normality and Quantization

Raziel Haimi-Cohen and Yenming Mark Lai

Abstract— Structurally random matrices (SRMs) are a practical alternative to fully random matrices (FRMs) when generating compressive sensing measurements because of their computational efficiency and their universality with respect to the sparsifying basis. In this work we derive the statistical distribution of compressive measurements generated by various types of SRMs, as a function of the signal properties. We show that under a wide range of conditions, that distribution is a mixture of asymptotically multi-variate normal components. We point out the implications for quantization and coding of the measurements and discuss design consideration for measurements transmission systems. Simulations on real-world video signals confirm the theoretical findings and show that the signal randomization of SRMs yields a dramatic improvement in quantization properties.

Keywords— Compressed Sensing; Quantization; Structurally random Matrices.

I. INTRODUCTION

Compressed sensing [1] is concerned with determining a signal \( x \in \mathbb{R}^n \) from a vector of measurements,

\[
y = \Phi x
\]

(1)

where \( \Phi \in \mathbb{R}^{m \times n}, m \ll n \), is a sensing matrix, and \( x \) is \( k \)-sparse representation in the column space of a sparsifier \( \Psi \),

\[
x = \Psi \zeta, \quad \|\zeta\|_0 \leq k
\]

(2)

where \( \Psi \) is an orthogonal or a tight frame matrix and \( \|\zeta\|_0 \) denotes the number of non-zero entries in \( \zeta \). If \( \Phi \Psi \) meets certain conditions, \( \zeta \) and hence \( x \) can be reconstructed from \( y \) by solving the constrained minimization problem

\[
\min \|\zeta\|_0 \quad \text{s.t. } y = \Phi \Psi \zeta
\]

(3)

Other results in the same vein extend the results to compressible signals (signals which can be approximated by sparse signals), or provide error bounds on the reconstructed solution when the measurements contain noise (In this case (3) may also be modified to account for the noise).

A. Sensing Matrix Design

Various design methods attempt to generate a sensing matrix \( \Phi \) that enables correct reconstruction of \( x \) from a small number of measurements in a computationally efficient way. Generally this goal is achieved only with very high probability (w.h.p.): either \( \Phi \) is a random matrix and w.h.p., the selected instance of \( \Phi \) enables correct and efficient reconstruction of every possible \( x, \zeta \) pair which satisfy (2); or the \( x, \zeta \) pair are random signals which satisfy (2) and \( \Phi \) is deterministic such that the pair \( x, \zeta \) can be reconstructed efficiently w.h.p. [2]. In this paper, we are interested in the first option.

A fully random matrix (FRM) is a matrix whose entries are independent, identically distributed (IID) Gaussian or Bernoulli random variables (RVs) [3][4]. If \( m \geq O(k \log(n/k)) \), then for any given \( \Psi \), w.h.p., \( \Phi \Psi \) is such that every \( x \) and \( \zeta \) which satisfy (2) can be reconstructed by solving (3). FRMs are universal, that is, the design of \( \Phi \) is independent of \( \Psi \), hence the choice of sparsifier can be deferred to the reconstruction stage, which is of significant practical importance. However, because of their completely unstructured nature, FRMs are computationally unwieldy in large scale applications since the random matrix needs to be both computed and stored.

Randomly sampled transforms (RST) address the computational complexity problem by imposing structural constraints on the randomness. Let

\[
\Phi = \sqrt{n/m} S W
\]

where \( W \in \mathbb{R}^{m \times n} \) is a square, orthonormal matrix having a fast transform, and \( S \in \mathbb{R}^{m \times m} \) is a random entries selection matrix, that is, a matrix whose rows are selected randomly, with uniform distribution, from the rows of \( I_n \), the \( n \times n \) identity matrix. \( \Phi x \) can then be computed efficiently by calculating the fast transform \( W x \) and selecting a random subset of the transform coefficients. RSTs guarantee a correct solution, w.h.p., if

\[
m \geq O\left( \mu^2(W, \Psi) k \log n \right)
\]

(4)

where \( \mu(W, \Psi) \), the mutual coherence of \( W \) and \( \Psi \), is:

\[
\mu(W, \Psi) = \sqrt{n} \max_{1 \leq i \leq n, 1 \leq j \leq n} \left| w_i \psi_j \right| / \left( \|w\|_2 \|\psi\|_2 \right)
\]

where \( w_i, \psi_j \) are the \( i \)th row and \( j \)th column of \( W, \Psi \), respectively [5]. Since \( 1 \leq \mu(W, \Psi) \leq \sqrt{n} \), we can choose \( m \ll n \) which satisfies (4) only if \( W \) is selected so that \( \mu(W, \Psi) \) is small. Therefore, RSTs are not universal.

The universality issue was addressed by the introduction of structurally random matrices (SRM) [6][7]:

\[
\Phi = \sqrt{n/m} S W R
\]

(5)

where \( S, W \) are as above and \( R \in \mathbb{R}^{m \times m} \), the randomizer, is a random square orthonormal matrix. Hence

\[
\Phi x = \sqrt{n/m} S W (R x) = \sqrt{n/m} S W (R \Psi) \zeta
\]

Therefore, a SRM with a given sparsifier \( \Psi \) behaves as the
RST $\sqrt{n/mSW}$ with the random sparsifier $R^Ψ$. If $R^Ψ$ and $W$ are mutually incoherent w.h.p., then SRMs are universal, and the known results for RSTs with incoherent sparsifiers (e.g. performance with compressible signals or noisy measurements) hold w.h.p.

Two types of randomization were proposed: **Local randomization** (LR), where each entry of $x$ is multiplied by $\pm 1$ with equal probability; and **global randomization** (GR), where the entries of $x$ are randomly shuffled. Both forms are computationally simple and were shown, for a large class of transforms $W$, to be universal, that is, the randomizers make the SRM incoherent with any given sparsifier w.h.p. [7].

Random convolution (RC) [8], is another method for generating universal and computationally efficient sensing matrices. We show in Sec. IV that this method is a frequency domain equivalent of SRM with LR. The concept of LR was also studied in [9] for the more general case where $\sqrt{n/mSW}$ in (5) is replaced by any matrix with the restricted isometry property (RIP).

**B. Quantization and coding of measurements**

As a data compression method, compressed sensing has some unique features. For example, the same measurements vector can be used by different recovery algorithms and different sparsifiers. Moreover, successful signal recovery is possible even if some measurements are lost. In addition, the balance of computational complexity between compression and reconstruction is sharply skewed towards the latter. While the right hand side of (1) is a simple linear operation, signal reconstruction requires solving a constrained minimization such as (3). These properties make compressed sensing attractive for applications such as video transmission over lossy channels[10]–[12], video transmission where the same signal may be decoded by different types of receivers [10][13] and video surveillance applications, where only a small part of the video stream needs to be reconstructed [14]–[16]. In all these applications the transmission of measurements requires a **coding scheme**, which entails source coding that is typically implemented by quantization followed by channel coding of the quantization codewords.

Conventional media coding standards are efficient over a wide range of input signals and operating conditions. One of the keys to this robustness is the usage of various signal-adaptive techniques in order to control the bit rate and improve performance. These techniques are applied before, during, and after quantization. For example, a linear prediction model may be estimated for the signal and the quantization may be performed on the prediction error, which reduces the bit rates needed to achieve specific quantization accuracies; the granularity of the quantizer may be varied according the signal content; and one out of several possible variable length coding schemes may be selected to achieve low rate lossless coding of the quantization codewords. The parameters of the linear prediction model, the quantizer granularity, and the lossless coding scheme need to be shared with the decoder, and hence they are encoded and sent as side information. Since the side information is critical for the decoding of the signal as a whole, it is typically encoded with higher accuracy and, in noisy channels, with better error protection, than the rest of the data. The bit rate overhead caused by sending the side information is usually small in comparison to the performance gain because the number of parameters which comprise the side information is relatively small.

The preferred coding scheme for compressive measurements depends on a variety of factors, but it is invariably based on assumptions about the probability distribution of the measurements, which is determined by the type of sensing matrix used. Furthermore, applying any of the signal-adaptive techniques described above requires having a parametric model where this distribution is specified by parameters estimated from the signal and transmitted to the decoder as side information.

The quantization of compressive measurements has recently received significant attention. Dai et al. [17][18] studied the effect of quantization on reconstruction accuracy with various quantizer designs — scalar (uniform and non-uniform), vector and entropy-coded, and provide asymptotic boundaries on the rate-distortion function when quantization is followed by Huffman coding. The efficacy of uniform vs. non-uniform scalar measurement quantization was compared specifically for video signals in [14][19]. Unlike all other quantizer designs we reviewed, the quantizer of [14] is signal-adaptive: its operation is controlled by the variance of the measurements in each frame, which is sent to the decoder as side information. A quantizer optimized for compressed sensing reconstruction is presented in [20]. Laska et al. studied the effect of saturation [21], the trade-off between number of measurements and quantization accuracy [22] and the extreme case of 1-bit quantizers [22]. Modifications to the reconstruction algorithms to address quantization effects were proposed in [21][24]. In all these papers, the sensing matrix was a FRM, and in many of them the input signal was random with a known distribution. Also, in most cases there was no attempt to reduce the bit rate of the quantizer codewords variable rate coding. Therefore, the results are of limited practical use for designing signal adaptive quantizers and channel coders for measurements generated by SRMs. Do et al. showed that under certain conditions, if $Φ$ is a SRM, the entries of $ΦΨ$ are asymptotically normally distributed [7], but since those entries are not independent it is difficult to draw conclusions about the actual measurements’ distribution from that work.

**C. Our contributions**

In this paper we study the distribution of measurements generated by SRMs and draw conclusions for the design of SRMs measurements coding schemes. We demonstrate our results by simulation on real-world signals and show that SRMs are superior to RSTs not only in being universal, but also by the fact that the measurements’ distributions are approximately normal and hence lend themselves to effective quantization.

The design of a coding scheme for compressive measurements depends on application specific considerations. Our goal here is not to propose a particular coding scheme, but to study the properties of the measurements’ distribution which need to underlie the design considerations. However, throughout the paper, we point out ways in which these
properties may be utilized in designing such a coding scheme.

This work differs from previous work on compressive measurement quantization in the following aspects: We study measurements generated by practical sensing matrices, such as SRMs and the closely related RC; we provide a rigorous characterization of the measurements distribution, including cross-correlation between measurements; and we propose methods for signal-adaptive quantization and channel-coding of measurements. For the latter, we assess the amount of side information needed by the decoder.

The paper is organized as follows: Sec. II introduces notations and concepts which are common to both LR and GR SRMs and provides examples of transforms suitable for use with SRMs. Sec. III introduces methods for SRMs measurement quantization and channel coding. Sec. IV and Sec. V characterize the distribution of measurements generated by SRMs with LR and GR, respectively. Sec. VI shows simulated SRMs measurements distributions and Sec. VII presents conclusions.

II. GENERAL PROPERTIES AND EXAMPLES OF SRMs

A. Notation

Let \( \Phi^{(n)} \triangleq \sqrt{n/m}S^{(n)}W^{(n)}R^{(n)} \in \mathbb{R}^{m \times n} \) be a SRM as defined by (5) and denote the jth row of \( W^{(n)} \) by \( w_j^{(n)} \triangleq [w_{j1}^{(n)}, \ldots, w_{jn}^{(n)}] \). Let \( \{x_k\}_{k=1,2,\ldots} \) be a deterministic bounded signal, \( |x_k| \leq x_{\text{max}} \) and let \( y^{(n)} = \Phi^{(n)}x^{(n)} \) be a vector of \( m \) compressive sensing measurements obtained from the \( n \)-dimensional signal vector \( x^{(n)} \triangleq [x_1, \ldots, x_n]^T \). Often the fast transform \( W^{(n)} \) exists only for specific orders (e.g. powers of 2). If the source signal is finite and of a dimension for which the transform does not exist, we assume that the signal is zero padded to the next available transform order.

Let \( z^{(n)} \triangleq \sqrt{n/m}W^{(n)}R^{(n)}x^{(n)} \) be a random variable (RV) defined on the probability space of the randomizer \( R^{(n)} \). For \( 1 \leq j, h \leq n \), define:

\[
\mu_{jh} \triangleq E[z_j^{(n)}z_h^{(n)}] \quad \text{cov} \{z_j^{(n)}, z_h^{(n)}\} \triangleq E[z_j^{(n)}z_h^{(n)}] - \mu_{jh}\mu_{hn} \quad (6)
\]

\[
\sigma^2_{jh} \triangleq \text{var} \{z_j^{(n)}\} = \text{cov} \{z_j^{(n)}, z_h^{(n)}\} \quad (7)
\]

If \( V \) is a \( r \)-dimensional random vector, \( \text{cov} \{V\} \) is the \( r \times r \) matrix defined by

\[
\text{cov} \{V\}_{ij} \triangleq E[V_iV_j] - E[V_i]E[V_j], \quad 1 \leq i, j \leq r.
\]

By substitution, \( y^{(n)} = S^{(n)}z^{(n)} \) and since \( S^{(n)} \) is a random entry selection matrix, the compressive sensing measurements, are RVs given by \( y_k^{(n)} = z_{ck}^{(n)}, 1 \leq k \leq m \), where the measurements indices \( c_1(1), \ldots, c_m(m) \) are a random sample from \( \{1, \ldots, n\} \), with or without replacement, that is, \( c_n \triangleq [c_1(1), \ldots, c_m(m)]^T \) is uniformly distributed, either in \( C_n^{(n)} \triangleq \{1, \ldots, n\}^m \), for sampling with replacement or in \( C_n^{(n)} \triangleq \{g_1, \ldots, g_m\}^T \in C_n \mid g_1, \ldots, g_m \) are distinct \}, for sampling without replacement. Each of the measurements \( y_1^{(n)}, \ldots, y_m^{(n)} \) is a mixture, with equal probabilities, of the mixture components \( z_{c_1(k)}, \ldots, z_{c_m(k)} \), hence \( y_1^{(n)}, \ldots, y_m^{(n)} \) are identically distributed. When discussing the measurements’ distribution, \( y^{(n)} \) stands for any one of \( y_1^{(n)}, \ldots, y_m^{(n)} \). Thus,

\[
\mu_{m} \triangleq E\{y^{(n)}\} = n^{-1}\sum_{j=1}^{n}y_j^{(n)} \quad (8)
\]

\[
\sigma^2_{m} \triangleq \text{var} \{y^{(n)}\} = E\{(y^{(n)})^2\} - \mu^2_{m} \quad (9)
\]

\( \langle a \rangle_n \) denotes the value \( 0 < a' \leq n \) such that \( a = a' \mod(n) \).

When discussing asymptotic behavior as \( n \to \infty \) the notation is usually simplified, without loss of generality, by not indicating that \( n \) is restricted to orders at which the transform exists and that \( m \) is a function of \( n \). If \( n \) is fixed we usually omit the superscript \( (n) \) and subscript \( n \) for clarity.

For vectors \( u, v \in \mathbb{R}^n \), \( v * u \) and \( v \circ u \) denote convolution and pointwise multiplication, respectively, that is

\[
(v * u)_{i} \triangleq \sum_{j=1}^{n}v_i u_{i-j}, \quad 1 \leq j \leq n
\]

\[
(v \circ u)_{i} \triangleq v_i u_i, \quad 1 \leq j \leq n
\]

A bar over a vector or matrix symbol indicates the mean of the entries, e.g. \( \bar{x} \triangleq n^{-1}1^T x \), \( \bar{W} \triangleq n^{-2}1^TW1 \). A bar over a scalar has the usual meaning of complex conjugate.

Let \( r > 0 \) be a fixed integer. A sequence \( \{u^{(n)}\} \) of \( r \)-dimensional RVs is asymptotically multivariate normal (AMN) if for sufficiently large \( n \), \( \text{cov} \{u^{(n)}\} \) exists, is positive definite, and

\[
\left( \text{cov} \{u^{(n)}\} \right)^{-1/2} \left[ u^{(n)} - E\{u^{(n)}\} \right] \xrightarrow{n \to \infty} N(0, I_r). \]

B. Statistical Properties

If \( c_1(1), \ldots, c_m(m) \) are selected with replacement, then the measurements are independent. Even if they are selected with replacement, for large \( n \), if \( m \ll n \) the measurements are approximately independent, as stated by the Lemma 1 below:

**Lemma 1:** Let \( m = m(n) = o \left( \sqrt{n} \right) \) and let \( F_0^{(n)}, F_1^{(n)} \) be the multivariate distribution functions of \( y^{(n)} \) with and without replacement, respectively. Then for any sequence \( h^{(n)} \in \mathbb{R}^{m(n)}, \quad n = 1, \ldots, \) such that \( \liminf_{n \to \infty} F_0^{(n)}(h^{(n)}) > 0 \),

\[
F_1^{(n)}(h^{(n)})/F_0^{(n)}(h^{(n)}) \xrightarrow{n \to \infty} 1. \]

We characterize the distribution of the measurements in the following steps: First we derive expressions for the expectation and covariances of the mixture components \( z_{c_1(k)}, \ldots, z_{c_m(k)} \). The expressions for the variance of each of the mixture components, \( \sigma^2_{1}, \ldots, \sigma^2_{m} \), and the mean of each of the compressive measurements, \( \mu_{m} \), follow immediately from (7) and (8), respectively. Since \( R \) and \( W \) are orthonormal:

\[
E\{y^2\} = \frac{1}{n} \sum_{i=1}^{n} E\{z_i^2\} = \frac{1}{n} \sum_{i=1}^{n} E\{HRS_i^2\} = m^{-1}\|s\|^2.
\]

Therefore:

\[
\sigma^2 = m^{-1}\|s\|^2 - \mu^2. \quad (10)
\]
Then we provide an upper bound on the tail distribution of the mixture components \( z_1, \ldots, z_n \). Assume \( \sigma_j > 0 \) and let
\[
\tau_j \triangleq \sqrt{\frac{m}{n} \sigma_j \| Tw_j \|_2^2} \| T \xi_k \|_2
\]
where \( T : \mathbb{R}^n \to \mathbb{R}^a \) is a linear transformation. We show that
\[
\Pr\left( \| z_j - \mu_j \| > t \sigma_j \right) \leq 2 \exp\left( -t^2 \eta_n(t, \tau_j) \right), \quad t \geq 0
\]
\( \eta_n \) and \( T \) are determined by the type of randomization. \( \eta_n(t, \tau_j) \) is continuous, increasing in \( \tau_j \) and decreasing in \( t \), but the decrease in \( t \) is generally slow and \( \eta_n(t, \tau_j) \) is increasing. It is often desired to determine a range such that the probability of the tail of \( z_j \) outside is less than a given \( 0 < \delta < 1 \). If \( \sigma_j, \tau_j \) are known, this can be done by solving
\[
2 \exp\left( -t^2 \eta_n(t, \tau_j) \right) = \delta \quad \text{for } t. \quad \text{The following lemma shows how this can be done efficiently in a small number of iterations:}
\]
\[ \text{Lemma 2: For } t \geq 0 \text{ let } \eta(t) \text{ be a continuous, positive, and decreasing function such that } t \eta(t) \text{ is increasing. Let } 0 < \delta < 1. \text{Select } t_0 > 0 \text{ arbitrarily and define}
\]
\[
t_{n+1} \triangleq t_{n}^{1/(2n)} \left( \log \left( 2 / \delta \right) / \eta(t_{n}) \right)^{1/(2n)}, \quad n = 1, 2, \ldots
\]
Then \( t_n \to t \) such that,
\[
2 \exp\left( -t_{n}^2 \eta_{n}(t_{n}) \right) = \delta \quad \text{and}
\]
\[
\left| \log t_{n+1} - \log t_n \right| \leq 0.5 \left| \log t_n - - \log t_{n+1} \right| \quad \square
\]
Finally, we give conditions for asymptotic normality of the mixture components. For any fixed integer \( r > 0 \), let \( \{u^{(n)}\} \) be a sequence of \( r \)-dimensional RVs such that,
\[
\text{For sufficiently large } n: \quad u_k^{(n)} = z_{j_k}^{(n)}, \quad 1 \leq k \leq r,
\]
and \( 1 \leq j_k(1), \ldots, j_k(r) \leq n \) are distinct.
\[
\text{We show that } \{u^{(n)}\} \text{ is AMN if the following two conditions are met:}
\]
\[
\lim_{n \to \infty} \max_{1 \leq k \leq r} \| W_{j_k}^{(n)} \|_2 = 0 \quad \text{and } \quad \lim_{n \to \infty} \max_{1 \leq k \leq r} \left( m/n \right) \alpha^T \text{cov}(u^{(n)}) \alpha = 0
\]
In many practical cases, the convergence to normality is uniform, over all choices of distinct \( z_{j_k(i)}^{(n)}, z_{j_k(r)}^{(n)} \), as is stated formally below:
\[ \text{Lemma 3: Let } r > 0 \text{ be a fixed integer and let } J^{(n)} \subseteq \{1, \ldots, n\}, \quad n = 1, \ldots \text{ be sets of indices. Suppose that if, for sufficiently large } n, \quad \{u^{(n)}\} \text{ satisfies (14) and } \{ j_k(1), \ldots, j_k(r) \} \subseteq J^{(n)} \text{ then } \{u^{(n)}\} \text{ is AMN. Then the probability distribution of } \{u^{(n)}\} \text{ converges to } \mathcal{N}(0, I_r) \text{ uniformly over all sets of distinct indices } j_k(1), \ldots, j_k(r) \in J^{(n)} \quad \square}
\]
\[ C. \text{ Examples of Transforms Used by SRMs} \]
Common examples for the transform matrix \( W \) used in (5) are the Discrete Cosine Transform (DCT) and the Walsh-Hadamard Transform (WHT) matrices, each with rows \( w_j, j = 1, \ldots, n \) scaled so that \( W^T W = I_n \) (in our discussion of specific transforms, we always assume this normalization).

The Discrete Fourier Transform (DFT) may also be used as long as it is treated as a real (orthonormal) matrix, that is, the real and imaginary parts of each row of the original DFT are separated to form two unit-norm rows of real coefficients. Also, since the input is real, redundant rows should be omitted to make \( W \) a \( n \times n \) matrix. Other transform matrices may be formed as a Kronecker product of any orthonormal matrices [25]. By construction, WHT, DCT and DFT satisfy condition (15), which limits the concentration of energy within each row of \( W \). Also, \( W^{(m)} = W_n \otimes W_n \) satisfies this condition if either \( W_n^{(m)} \) or \( W_n^\text{DCT} \) satisfies it. Then \( W^{(m)} = W_n \otimes W_n \) will also satisfy this condition. Condition (15) is a weaker form of the conditions for a SRM with large \( n \) to be incoherent with any sparsifier w.h.p. [7]. Therefore transform matrices \( W^{(m)} \) for which condition (15) does not hold, such as wavelet transforms or the trivial \( W^{(m)} = I_n \), would not make good SRMs to begin with.

III. MEASUREMENT QUANTIZATION AND CHANNEL CODING

A. Determining Variables for Quantization and Coding

Since the compressive measurements \( y_1, \ldots, y_m \) are, at least approximately, mutually independent, the advantage of vector quantization over scalar quantization is small compared to the cost of added complexity ([18] reaches a similar conclusion for FRM generated measurements). Therefore, we can apply a coding scheme in which each measurement is quantized by the same scalar quantizer and then lossless channel coding is applied to the quantization codewords to represent them as a bit sequence. The parameters of this coding scheme depend on the signal and need to be shared with the decoder. As we will show, under a wide range of conditions the mixture components \( z_1, \ldots, z_n \) are AMN, and the distribution of each of them is fully characterized by their mean and variance. In some important cases, \( z_1, \ldots, z_n \) are identically distributed and hence the coding scheme is parametrized by a single mean and variance. If this is not the case, the mixture of \( z_1, \ldots, z_n \) can be approximated by a mixture of a small number of Gaussians. The means, variances, and weights of the approximate mixture are sent as side information and used by both encoder and decoder to compute the parameters of the coding scheme.

While this method is simple, it is suboptimal if the distributions of \( z_1, \ldots, z_n \) are different. In this case, a lower bit rate can be achieved at the same distortion level by adapting the coding scheme for each of the measurements \( y_k, 1 \leq k \leq m \), to the distribution of \( z_{(k)} \), the corresponding mixture component. This requires sharing the coding scheme of each measurement with the decoder. As we will show, under a wide range of conditions the means and variances of \( z_1, \ldots, z_n \) may be represented by a parametric model. The model parameters may be sent as side information, and both encoder and decoder may derive the coding scheme for each measurement from the mean and variance computed by the model.

Quantizing mixture components, rather than measurements, eliminates the need for independence the measurements but
raises the question of independence among the mixture components. Fortunately, when the latter are AMN, independence can be achieved by decorrelation. One relatively simple way to achieve this is by linear prediction, as outlined in Appendix A. In this approach some mixture components are linearly predicted from other mixture components and the coding scheme is applied to the linear prediction residuals rather than to the original mixture components. If the measurements are highly correlated their variances are much larger than those of the residuals, hence can be coded more efficiently.

In Sec. IV and V below we show how the parameters of the distribution of the mixture components \( z_1, \ldots, z_n \) can be derived from the properties of the signal. In some CS applications the measurements are computed in the analog domain and the signal is not available. In such cases, the theoretical results inform us of the appropriate model for the measurements’ distribution, and the model parameters are estimated directly from the available measurements.

B. Scalar Quantization and Channel Coding

We turn our attention to the design of a coding scheme based on scalar quantization. The data is a sequence of \( m \) transmitted variables (TVs), which may be measurements, selected mixture components, or linear prediction residuals. The distributions of the TVs are given by a parametric model and they may or may not be identical. The coding schemes for each TV is adapted to its specific distribution and accordingly, they may or may not be identical.

A scalar quantizer \( Q(y) \triangleq \arg \min_{c \in \mathcal{C}} |y-c| \) maps a TV \( y \) into a finite codebook. The quantization region of a codeword \( c \in \mathcal{C} \) is \( Q^{-1}(c) \triangleq \{ u \in \mathbb{R} | Q(u) = c \} \). A codeword \( c \in \mathcal{C} \) approximates the values in its quantization region, \( Q^{-1}(c) \triangleq \{ u \in \mathbb{R} | Q(u) = c \} \) and the codeword distortion is \( E[|y-c|^2] \mid y \in Q(c) \). If \( Q^{-1}(c) \) is a finite interval \( c \) is unsaturated and the quantization error in \( Q^{-1}(c) \) is bounded; otherwise \( c \) is saturated and the quantization error in \( Q^{-1}(c) \) is unbounded. The range of the quantizer is the union of all bounded quantization regions. The accuracy of signal reconstruction from quantized measurements is severely degraded if even a small number of input codewords are saturated (the same would be true for measurements derived from saturated TVs, as such residuals). The reconstruction algorithm may be modified to prevent this degradation by a special handling of saturated measurements, but the results are only slightly better than those obtained when saturated measurements are simply discarded [21]. Therefore, the quantizer’s range should be wide enough to make saturation a rare event. On the other hand an excessively large quantizer’s range yields little performance gain and can add unnecessary complexity. Suppose that an upper bound for saturation frequency is specified as \( 1 > \delta > 0 \). If the same quantizer is used for all measurements, the quantizer range can be set empirically so that no more than \( m \delta \) measurements are saturated. However, if different quantizers are used, e.g. if the quantized entities are mixture components with different variances or linear prediction residuals, one may invoke the theoretical results provided here about the measurements’ distribution, such as asymptotic normality or the concentration result (12) and Lemma 2. The latter may be necessary because convergence to normal distribution may be slower at the tail. In addition, those results may be useful in order to assess the ranges that will be needed for a given class of input signals.

After quantization, channel coding is applied to the codewords to represent them as a bit sequence. The entropy
\[
H(Q(y)) \triangleq E[\log_2 P_r(Q(y))] \]
which is bounded from below by \( \log_2 |c| \) bit/measurement. This bound is attainable if \( \log_2 |c| \) is an integer. Otherwise, data rates arbitrarily close to the bound can be achieved by jointly coding sequences of several codewords.

We consider two types of scalar quantizers. First, we consider an optimal quantizer [27] which, based on knowledge of the measurements’ distribution, minimizes the distortion subject to a constraint on either the codebook size or the quantizer’s entropy. The optimization usually causes the codewords to have similar probabilities, which makes the bit rate of FLC close to that of VLC. Second, we consider a uniform quantizer where the quantization regions of all unsaturated codewords are of the same length. For a memoryless source, under a wide range of conditions, the entropy rate of a uniform quantizer is within a fraction of a bit from that of an optimal quantizer with the same distortion [27][28]. Therefore, there is a trade-off between complex optimal quantization followed by simple FLC versus simple uniform quantization followed by complex VLC.

The measurements may contain noise, e.g. noise propagated from a noisy source signal. With optimal quantization the quantization distortion differs from codeword to codeword. At high bit rates, some codeword distortions may be lower than the measurements’ noise floor, effectively wasting bits on representing the noise, while at low bit rates, the distortions of some unsaturated codewords may be large enough to induce severe reconstruction degradation as seen with saturated codewords. Therefore, uniform quantization with VLC appears to be a more robust approach.

Since the coding scheme is signal-adaptive, its settings must be shared with the decoder as side information. Whether we use an optimal quantizer with FLC or a uniform quantizer with VLC, this may require transmitting \( |c| \) parameters to specify the values, or the probabilities, of each codeword, respectively. That may be acceptable if the codebook is small, but for a large codebook this amount of side information may be too much. Instead, the measurements’ distribution may be modeled by a parametric model, with the side information consisting only of estimates of the model parameters. Both encoder and decoder can then derive the coding scheme from the parametric distributions. Some of the estimated parameters, e.g. the signal mean, may, in fact, be compressive measurements and may be used as such. The designation of values as side information indicates that those values are used
to specify the coding scheme and therefore are treated differently: they are guaranteed to be transmitted regardless of the random measurements selection, and they are often transmitted with higher accuracy or more error protection than the rest of the measurements.

IV. LOCAL RANDOMIZATION

A locally randomized SRM (LR-SRM) is SRM whose randomizer is defined by $R_k \triangleq \delta_{jk} b_j$, where $b_1, \ldots, b_n$ are IID Rademacher RVs, each getting the values $\pm 1$ with equal probability.

A. Measurements’ Distribution with LR-SRM

Theorem 1 (distribution of LR-SRM mixture components): With local randomization, for $1 \leq j, h \leq n$:

(a) The distribution of each mixture components $z_j$ is symmetric and

$$E \{ z_j \} = 0$$

$$\text{cov} \{ z_j, z_k \} = -\frac{n}{m} \sum_{t=1}^{m} w_{kt} x_t^2 = -\frac{n}{m} (w_j \circ w_h)(x \circ x)$$

(b) If $\sigma_j > 0$ then the bound (12) holds with

$$\xi(u) = \frac{u}{4} (1 + u) \log (1 + u)$$

and $\tau_j$, defined by (11) with $T = I_n$, is

$$\tau_j \triangleq \inf_{1 \leq a \leq n} \left\{ \| x \|_2^2 \right\} - \frac{1}{\sqrt{\sum_{i=1}^{n} \| w_i \|_2^2}}$$

(c) For any fixed integer $r = 0$ a sequence $\{ u^{(n)} \}$ defined by (14) is AMN if conditions (15) and (16) are satisfied $\square$.

By (17) and (10), $\mu_j = 0$ and $\sigma_j^2 = m^{-1} \| x \|_2^2$. Equation (18) shows that if $|x_k|, k = 1, \ldots, n$ are constant then $z_j, 1 \leq j \leq n$ are uncorrelated, but otherwise they can be highly correlated. Since the rows of $W$ form a basis, $w_j \circ w_h$ can be represented by a linear combination of $w_1, \ldots, w_n$. For the WHT, DCT and DFT, the number of non-zero terms is very small hence (18) can be simplified by substituting

$$w_j \circ w_h = n^{-1/2} \sum_{t=1}^{p} \gamma_{t}(j,h)w_{l}(j,h)$$

where $p$ is the number of non-zero terms and $\gamma_{t}(j,h), k = 1, \ldots, p$, are weights and indices, respectively. For the WHT, $p = 1$, $\gamma_{1}(j,h) = 1$ and $l(1,h)$ is derived by bitwise modulo-2 addition of the binary representations of $j$ and $h$. For the DCT and DFT, $p = 2$, $\gamma_{2}(j,h) = \pm 2^{-1/2}$, and $w_{l}(j,h)$ correspond to frequencies which are sums or differences of the frequencies of $w_j$ and $w_h$. Therefore, by (18) the covariance can be computed from $W(x \circ x)$, the transform of $x \circ x$, as:

$$\text{cov}(z_j, z_k) = \left(n^{1/2}/m\right) \sum_{t=1}^{p} \gamma_{t}(j,h) w_{l}(j,h) (x \circ x)$$

Furthermore, it can also be shown that if (20), (21) hold for transforms $W', W''$ with $p = p', p''$ in the right hand side, respectively, then (20), (21) hold for the rows of $W' \otimes W''$ with $p = p'p''$.

If $x$ is a typical media signal, $W(x \circ x)$ is often highly compressible and hence can be approximated by a model with a small number of parameters. For example, the approximation can be found by first saving a few dominant entries of $W(x \circ x)$ and setting the rest to zero. For the WHT, DFT and DCT, $l(1,1), \ldots, l(1,n) = (1, \ldots, n)$ for $1 \leq j \leq n$. Therefore in the WHT case, by (21), each saved entry of $W(x \circ x)$ determines one entry in each of the $n$ rows of $W(x \circ x)$. Therefore, the signal covariance may be estimated by (21), using an approximation of $W(x \circ x)$.

The probability of the tail of the distribution of the mixture component $z_j$ is limited by (12), which is particularly simple here, because $\eta(t, \tau_j)$ does not vary with $n$ and $\tau_j$ is a merely a scaling factor of $t$. $\xi(0)$ of (19) is a slowly decreasing function, with $\xi(0) = 0.5$ (Fig. 1). The values of $\| W_{t} \|_\infty, 1 \leq j \leq n$ are generally similar to each other and in the order of $n^{-1/2}$. Therefore, by (11), the mixture components in which $\tau_j$ is very small, and therefore the bound (12) is weak, are the components with very small variance. This happens if $W_j \circ x \approx 0_n$, that is if $x$ has significant energy only in regions where $W_j$ has no significant energy. Otherwise, in the range of interest for $t$, $\tau_j \gg t$ even for moderate $n$, hence the bound on the tail distribution in (12) is very close to $2 \exp(-t^2/2)$.

Condition (16) can be written as:

$$\forall x \in \mathbb{R}^r, x \neq 0: \liminf_{n \to \infty} \left\{ \sum_{t=1}^{n} \alpha_t W_{(t,n)} \circ x \right\}_\infty > 0$$

which, roughly speaking, requires that any linear combination of the relevant rows should have some of its energy in areas where the signal has energy.

The univariate distribution of the mixture components $z_1, \ldots, z_n$ is particularly simple if all the entries of $W$ have the same magnitude, as with the WHT. In this case $z_1, \ldots, z_n$ are identically distributed with $\sigma_j = m^{-1} \| x \|_2^2$, $\tau_j = n^{-1/2} \| x \|_\infty$, $1 \leq j \leq n$. In addition (15) is satisfied by design, and (16)
becomes, for \( r = 1 \), a condition on the signal norm:

\[
\liminf_{n \to \infty} n^{-1} \left\| x^{(r)} \right\| > 0.
\]  

(22)

If this is true, then \( \left\| x^{(r)} \right\| \) grows at least as \( \sqrt{n} \) while \( \left\| x^{(r)} \right\| \leq \chi_{\max} \). Therefore, unless the signal is compressible, which make \( \chi_{\max} \) large without increasing \( \left\| x \right\| \) by much, the right hand side of (12) decays reasonably quickly in the practical range of interest of \( I \).

Suppose that the entries of \( W \) do not have the same magnitude, and for some \( n \), \( z_1, \ldots, z_n \) are approximately normal. Then the tail distribution is determined by the largest among the variance of the mixture components \( \sigma^2_0, \ldots, \sigma^2_n \).

This is bounded because, by (18),

\[
\sigma^2_0 \leq (n/m) \left\| w \right\|^2 \left\| \chi \right\|^2 = \left( n \left\| \chi \right\|^2 \right) \sigma^2_0
\]

In both the DFT and DCT, \( \left\| w \right\|_2 \leq \sqrt{2/n} \), hence \( \sigma^2_0 \leq 2\sigma^2_0 \).

Furthermore, in the DFT case, let \( (w_j, w_k) \) be a pair of rows which correspond to the real and imaginary parts, respectively, of the \( h \)th complex DFT coefficient, with DFT coefficient numbering starting at 1. Calculation using (18) yields

\[
\sigma^2_j = \sigma^2_0 + q_h, \quad \sigma^2_j = \sigma^2_0 - q_h
\]

where

\[
q_h = \frac{1}{m} \sum_{i=1}^{m} x_i \cos \frac{4\pi(h_1-1)(k-1)}{n}
\]

Let \( g \) be the greatest common divisor of \( n \) and \( 2(h-1) \) and let \( d \equiv n/g \). \( q_h = \sigma^2_0 \) or \( q_h = -\sigma^2_0 \) only if \( x_i = 0 \) when \( \langle k \rangle_1 \neq 0 \) or \( \langle k \rangle_2 \neq (1+d)/2 \), respectively. In both cases, \( \{x_i\} \) is a train of spikes at intervals of \( d \) with a specific phase. In order for \( q_h \) or \( -q_h \) to be close to \( \sigma^2_j \), the signal has to approximate such a spike train, that is, \( x_i \) must be near zero except when \( \langle k \rangle_d \neq 1 \) or \( \langle k \rangle_d \neq (1+d/2) \), respectively. While such signals are certainly possible, in many practical applications they are uncommon, in which case \( |q_h| \ll \sigma^2_0 \) and therefore, \( \sigma^2_0, \sigma^2_j \approx \sigma^2_0 \).

### B. Measurements’ Distribution with Random Convolution

Let \( F^{(n)} \) be the \( n \)th-order complex DFT matrix, given by

\[
f^{(n)}_{ij} \triangleq n^{-1/2} \exp[-2\pi i (j-1)(k-1)/n].
\]

Note that indexing starts at 1 and \( F^{(n)} F^{(n)*} = I \). Let \( R \) be a random diagonal matrix with diagonal elements \( b_1, \ldots, b_n \) such that \( FRF^* \) is real. The RC sensing matrix is the real random [8]:

\[
\Phi = \sqrt{n/m} SF^* FRF^*
\]

where \( S \in \mathbb{R}^{n \times m} \) is a random selection matrix as in (5). The name “random convolution” is because, by the properties of DFT,

\[
z \triangleq F^* FRF \Phi = F^* (b \circ (Fx)) = n^{1/2} (F^* b) \ast x
\]

where \( b \triangleq [b_1, \ldots, b_n]^T \). The measurement vector is obtained by multiplying \( Fx \), the DFT representation of the signal, by \( SF^* R \). If \( F \) had been real, \( SF^* R \) could have been a LR-SRM and the replacement of \( x \) by \( Fx \) would make no difference because of the universality of SRMs. The distribution of \( z \) could then be determined by Theorem 1, using \( Fx \) instead of \( x \). However, since \( F \) is complex but \( FRF^* \) must be real, \( b_1, \ldots, b_n \) cannot be IID Rademacher. Instead, to define the distribution \( b_1, \ldots, b_n \), we let

\[
\chi_n(k) \triangleq \begin{cases} \frac{1}{2} & \text{for } 2k = 2n \text{ or } \frac{1}{2} & \text{otherwise} \end{cases}
\]

and let \( b_i \triangleq \exp(i \beta_k), 1 \leq k \leq n \) where \( \{\beta_k\} \) are independent, \( \beta_k \sim \mathcal{U}(0, \pi) \), and

\[
\beta_k \sim \mathcal{U}(0, 2\pi), \quad \chi_n(k) = 1 \text{ or } k \leq (n/2 + 1) \quad \beta_k = 2\pi - \beta_{n-2k}, \quad (n/2 + 1) < k \leq n
\]

where \( \mathcal{U}(A) \) denotes uniform distribution on \( A \). Therefore, \( b_1, \ldots, b_n \) are unit-magnitude, uniformly distributed RVs, and RCs can be seen as a frequency domain variant of LR-SRM.

**Theorem 2 (distribution of RC mixture components):** With RC, for \( 1 \leq j, h \leq n \), we have the three following results:

(a) The distribution of \( z_j \) is zero mean, symmetric, and

\[
\text{cov}(z_j, z_h) = m^{-1} \rho_j(j - h)
\]

where \( \rho_j(l) \) is the circular autocorrelation of \( x \), that is

\[
\rho_j(l) \triangleq \sum_{i=1}^{n} x_i x_{(i+j) \mod n}, \quad |l| < n.
\]

(b) If \( x \neq 0 \) then (12) holds with \( \zeta(t, r_j) = \xi(t/r_j) \), where \( \xi(t) \) is defined by (19) and \( r_j \) is given by

\[
r_j \triangleq \|x\| / \max_{h \neq k} \left( 2 - \chi_h(k) \right) \|Fx_k\| \]

(c) a sequence \( \{u^{(n)}\} \) as defined by (14) is AMN if condition (16) is satisfied and

\[
v_{\max} \triangleq \sup_n \left\| F^{(n)} x^{(n)} \right\| < \infty
\]

As in the LR-SRM case, \( \mu_j = 0 \) and \( \sigma_j^2 = m^{-1} \left\| Fx \right\|^2 \). The covariance of \( z \) can be efficiently computed by noting that the autocorrelation of the signal is the inverse DFT of

\[
\left[ \left\| (Fx)_0 \right\|, \ldots, \left\| (Fx)_n \right\| \right]^T
\]

and \( Fx \) is computed during the measurements computation. The covariance matrices \( \Sigma_i \) are symmetric Toeplitz. Therefore checking condition (16) for \( r > 1 \) is particularly simple.

Formally, \( r_j \) cannot be derived from (11) because the operator \( T \) is defined as real, while \( w_j \) here is a row of the complex matrix \( F \). However, we can get a similar representation by defining \( T_w \triangleq F, \quad T_x \triangleq (2I - G)F \), where \( G \) is a diagonal matrix with \( \chi_1(1), \ldots, \chi_n(n) \) as diagonal entries. It is easy to verify that \( r_j \triangleq m^{-1} \sigma_j \left\| F^{(n)} x^{(n)} \right\| \left\| F^{(n)} x^{(n)} \right\|^{-1} \). The
difference between $T_w$ and $T_s$ is a consequence of the need to represent an inherently complex method in the real domain.

With standard LR-SRM, the analysis of the univariate case was especially simple when the entries of the transform matrix had the same magnitude. With RC the transform matrix is $F^*$, in which all entries have the same magnitude, and we see the same behavior, where $\sigma_k = \ldots = \sigma_n^2$ and $\tau_i = \ldots = \tau_n$.

In the following, because of the similarities, when referring to LR-SRM we mean to include Random Convolution as well.

C. Quantization and Coding of LR-SRM Measurements

We first consider coding LR measurements without linear prediction. In this case, only univariate distribution information needs to be shared with the decoder. If all the entries of the transform matrix have the same magnitude, as in the WHT case and the RC case, then the mixture component variances are the same and the conditions for asymptotic normality are satisfied with well behaved signals. Therefore $\sigma_y$ is the only parameter that needs to be shared as side information and all measurements can be coded by the same coding scheme and optimized for $N(0, \sigma_y^2)$. This can be a good approximation for other transforms with entries of unequal magnitude, if $\sigma_1, \ldots, \sigma_n$ are similar. If there are significant differences among $\sigma_1, \ldots, \sigma_n$, the distribution of the measurements, which is a mixture of $z_1, \ldots, z_n$, can be approximated by a mixture of a smaller number of Gaussians, with the weights and variances of the approximation being shared as side information. The coding scheme is then optimized for this approximate mixture distribution.

For the DCT and DFT, $\sigma_1, \ldots, \sigma_n$ may be approximated by some $\hat{\sigma}_1, \ldots, \hat{\sigma}_n$ using (21) and a parametric estimate of $W(x \cdot x)$. In this case it is possible to encode each measurement $y_k, 1 \leq k \leq m$ with a different coding scheme matched to $z_{(k)}$, by assuming a distribution of $N(0, \hat{\sigma}_{(k)}^2)$.

The above approaches, however, ignore the correlation among the mixture components, as shown by (18) and (24). Linear prediction can significantly reduce the data rate of the quantized measurements as outlined in Appendix A. The key to that approach is the finding of an approximation for $\text{cov}(z_j, z_k), 1 \leq j, k \leq q$, which can be shared with the decoder as side information. For LR-SRM with the WHT, DFT and DCT this can be done using (20), (21), as explained in Sec. IV.A above. For RC, this can be done even more effectively because the covariance matrix is a Toeplitz matrix, hence the number of distinct matrix entries that need to be approximated is $n$ rather than $n(n+1)/2$. Furthermore, these distinct $n$ entries are the circular autocorrelation sequence of $x$, which, for a typical media signal, can be represented by various alternative forms which are suitable for transmission and quantization [29]. If the conditions for asymptotic normality of the mixture components hold and $n$ is large enough, then the residuals are distributed normally and their estimated variance is given by (31), hence a coding scheme can be optimally adapted for each residual.

V. GLOBAL RANDOMIZATION

A globally randomized SRM (GR-SRM) is a SRM whose randomizer is defined by $R_x \hat{=} \delta_{j, \pi(k)}$, where $\pi: \mathbb{Z}_n \to \mathbb{Z}_n$ is a uniformly distributed random permutation of $\{1, \ldots, n\}$. Therefore,

$$z_j = \sqrt{n/m} \sum_{k=1}^n w_{j, \pi(k)} x_k, \quad 1 \leq j \leq n. \quad (25)$$

A. Measurements’ Distribution with GR-SRM

With GR, any two mixture components, $z_j, z_h, 1 \leq j, h \leq n$, have the same distribution if the entries of $w_j$ are a permutation of the entries of $w_h$. Therefore, the number of distinct mixture components and their respective weights are determined solely by the transform $W$. For the WHT there are only two distinct mixture components: $\{z_1\}$ and $\{z_2, \ldots, z_n\}$, corresponding, respectively, to the constant $w_1$, and to $\{w_2, \ldots, w_n\}$, each of which contains an equal number of $+n^{-1/2}$ and $-n^{-1/2}$ entries. For the DCT and DFT, the number of distinct mixture components is the number of divisors of $n$. In particular, if $n$ is a power of 2 there are $1+\log_2 n$ distinct mixture components. If $W = W' \otimes W''$ the number of mixture components for $W$ does not exceed the product of the number of mixture components for $W'$ and $W''$. Thus, in these common examples, the number of distinct mixture components in the measurements’ distribution is much smaller than $n$.

Theorem 3 (distribution of GR-SRM mixture components): With global randomization, for $1 \leq j, h \leq n$:

(a) $E_{\{z_j\}} = n^{1/2} m^{-1/2} \hat{w}_j \hat{x}_j$, 

(b) $\text{cov}(z_j, z_h) = \frac{n}{m(n-1)} (\hat{\sigma}_j^2 - n \hat{w}_j \hat{w}_h) \left( \|\hat{x}_j\|^2 - n \hat{x}_j^2 \right)$

(c) $\tau_j = (n-1)^{-1/2} \frac{\|w_j\|}{\|w_j\|_2} \frac{\|w_j\|}{\|w_j\|_2} \left( \frac{\|w_j\|}{\|w_j\|_2} \right)^2 \lesssim \frac{n}{\sqrt{n-1}}$, 

and (12) holds with $\eta_n(t, \tau_j) \hat{=} \tau_j^2 / (8n + 4t \tau_j)$.

(c) For any integer $r > 0$ a sequence $\{u^{(r)}\}$ as defined by (4) is AMN if conditions (15) and (16) are satisfied □.

Corollary: With global randomization, for $1 \leq k \leq m$:

$$E_{\{y\}} = n^{1/2} m^{-1/2} \hat{w} \hat{x},$$

$$\sigma_y^2 \hat{=} \text{var} \{y\} = m^{-1} \left[ \|\hat{x}\|^2 - n^{-1} \left( \hat{w} \hat{x} \right)^2 \right].$$

Unlike the LR case, Theorem 3 cleanly separates the signal and the transform by representing all significant parameters as products of functions of only $x$ or only $W$. In addition, for a given $W$, the means and covariance structure of $z_1, \ldots, z_n$ are fully defined by $\hat{w}_1, \ldots, \hat{w}_n$, up to scaling factors determined.
by \(\|\mathbf{x}\|_2\) and \(\bar{\mathbf{x}}\). In particular \(\sigma_j = \sigma_0\) if \(\mathbf{w}_j = \bar{\mathbf{w}}_h\).

Suppose that there is only one row \(\mathbf{w}_j, 1 \leq j \leq n\) such that \(\mathbf{w}_j = 0\), say, \(j = 1\). This is true for the WHT, DCT and DFT, and if it satisfied by \(W', W^{*}\) it is also satisfied by \(W' \otimes W'\). For such transform matrices the means and covariances for \(1 \leq j, h \leq n\) are given by

\[
Ez_j = \delta_j nm^{-1/2} \bar{\mathbf{x}},
\]

\[
\text{cov}(z_j, z_h) = \begin{cases} 
\left(\|\mathbf{x}\|_2^2 - n\bar{\mathbf{x}}^2\right)n/(m(n-1)) & j = h > 1 \\
0 & \text{otherwise}
\end{cases}
\]

Therefore, for any fixed \(r > 0\), and any sequence \(\{\mathbf{u}^{(r)}\}\) as defined by (14), condition (16) is satisfied if

\[
\liminf_{n \rightarrow \infty} n^{-1} \left\|x^{(r)} - \bar{\mathbf{x}}\right\|_2^2 > 0
\]

(29)

\[
\liminf_{n \rightarrow \infty} \min_{1 \leq i \leq m} f_i(k) \geq 2
\]

(30)

The signal \(x^{(r)} - \bar{\mathbf{x}}\) is the original signal minus its mean, and (29) requires that for sufficiently large \(n\), the RMS of this signal will not fall below some \(\varepsilon > 0\). Therefore, if (15) and (29)-(30) are satisfied, then \(\{\mathbf{u}^{(r)}\}\) is AMN. In addition, since \(z^{(r)}_j, \ldots, z^{(r)}_m\) are uncorrelated, they are approximately independent for large \(n\).

\(T\) in part (b) of the theorem is the operator which removes the mean from the signal, \(\mathbf{T} = \mathbf{x} - \bar{\mathbf{x}}\). Let

\[
a_x \triangleq n^{-1/2} \left\|\mathbf{T}\right\|_2/\left\|\mathbf{x}\right\|_2 \leq 1, \quad b_j \triangleq n^{-1/2} \left\|\mathbf{T} \mathbf{w}_j^*\right\|_2/\left\|\mathbf{T}\mathbf{w}_j^*\right\|_2 \leq 1.
\]

Then \(\tau_j = a_x b_j (n/(n-1))^{1/2}\). For moderately large \(n\), \(n/(n-1) \approx 1\) and, in the range of interest for \(t\), \(4\tau_j \leq 4m/\sqrt{n-1} \ll 8n\).

Therefore, \(\eta(t, \tau_j) \approx a_x b_j^2/8\) and the upper bound in the right hand side of (12) is approximately \(e^{-(a_x b_j^2)^{1/2}}\). Note that this expression cannot be smaller than \(e^{-\tau_j^2/2}\), while in the LR case, the bound in the right hand side of (12) could be arbitrarily close to \(e^{-\tau_j^2/2}\).

\(a_x\) and \(b_j\) are the ratios between the root mean square (RMS) and the peak of \(\mathbf{T}\) and \(\mathbf{T}\mathbf{w}_j\), and thus characterize the signal and the transform, respectively. \(a_x\) can be very small only if \(\mathbf{T}\) is compressible or sparse. For \(j > 1\), \(b_j = 1/\sqrt{2}\) in the WHT and \(b_j = 1/\sqrt{2}\) in the DCT and DFT. Let \(\mathbf{w}_j^*, \mathbf{w}_j^*\) be rows of the transform matrices \(W, W^{*}\) of orders \(n', n''\), respectively, such that \(\mathbf{w}_j = \bar{\mathbf{w}}_j^* = 0\). Let \(n \triangleq n'n''\) and let \(\mathbf{w}_j = \mathbf{w}_j^* \otimes \mathbf{w}_j^*\) be the row in \(W \triangleq W' \otimes W^{*}\) which corresponds to \(\mathbf{w}_j, \mathbf{w}_j^*\). If \(b_j, b_j', b_j''\) are the constants for \(\mathbf{w}_j, \mathbf{w}_j^*, \mathbf{w}_j^*\), respectively, then

\[
b_j = n^{-1/2} \left\|\mathbf{w}_j^*\right\|_2 = \left(n^{-1/2} \left\|\mathbf{w}_j^*\right\|_2\right)^{(n-1)/2} \left(n^{-1/2} \left\|\mathbf{w}_j^*\right\|_2\right)^{(n-1)/2} = b_j' b_j''
\]

Therefore, in the examples we considered, \(b_j\) is a characteristic constant of the transform type and it is not very small for well-behaved transforms.

---

**B. Quantization and Coding of GR-SRM Measurements**

The only signal parameters which are involved in determining the mean and covariance of the measurements are \(\bar{\mathbf{x}}\) and \(\|\mathbf{x}\|_2\), therefore we assume that they are shared as side information. We also assume that \(\mathbf{w}_j = 0\) for \(1 \leq j \leq n\). Then \(z_j = n^{1/2} \bar{\mathbf{x}}\). Since \(\bar{\mathbf{x}}\) is received as side information we assume that \(S\) never selects \(z_j\).

By our assumptions, \(z_1, \ldots, z_n\) are uncorrelated and \(z_2, \ldots, z_n\) have the same means and covariances. Therefore, no redundancy can be removed by linear prediction. If condition (15) holds for the transform matrix and condition (29) holds for the signal, then for sufficiently large \(n\), \(z_2, \ldots, z_n\) can be treated as identically distributed normal RVs and a coding scheme, optimized for \(N(0, \sigma_j^2)\), may be used for all of them.

---

**VI. NORMALITY TESTING**

We checked the normality of \(z_j, j > 1\) for LR-SRM, GR-SRM and RST, each with the WHT, DCT and DFT. In the LR cases, the signal mean \(\bar{\mathbf{x}}\) was subtracted from the signal prior to multiplying the signal by the sensing matrix. The test material was comprised of 160 video signals, each consisting of 12 frames of \(88 \times 72\) 8-bit pixels. The 76032 pixels in each signal were zero padded to \(n = 2^{17}\). In each test case, each signal vector was multiplied by 8 matrices of a specified type, which differed only in the random number generator seeds used to create the random matrices \(S\) and \(R\). \(m\) was set to 76032, thus after removing the \(z\) entries we had \(8 \times 76031 = 608248\) measurements per signal. The mean and variance of those measurements were estimated and the measurements were normalized to zero mean and unit variance. If this set of measurements is a sample of a standard normal distribution, its quantile-quantile (Q-Q) plot [31] should be linear with a slope of 1. The Q-Q plots of the
normalized measurements were computed for each of the signals and the 160 Q-Q plots were overlaid in one graph.

The graphs for each test case are shown in Fig. 2. In all cases of local and global randomization, the curves appear to be very close to linear, which indicates that the sample distribution is quite close to normal. Moderate deviations from the straight line appear above 3.5 standard deviations and may be explained by the scarcity of the data points in those regions. On the other hand, for RSTs the measurements distribution is far from normal and the shape of the Q-Q plots indicates that these distributions have much heavier tails than normal distributions with the same variance. The wide spread among the Q-Q plots in RSTs shows that the shape of the measurements’ distribution is highly dependent on the input signal.

According to the analysis of the LR case (Sec. IV), asymptotic normality is guaranteed for measurements generated by WHT, but signal dependent deviations from normality may appear in measurements generated by DCT or DFT. Yet, these deviations are hardly noticeable in Fig. 2. Fig. 3 shows an enlarged section (the range of [1, 1.5]) of the same graphs. While the graphs of the WHT case retain its clean, linear shape, the graph of the DCT case is a wide stripe with irregular boundaries, indicating that its underlying individual Q-Q plots deviated slightly from the linear shape expected for a normal distribution, and that these deviations are signal dependent, making each Q-Q plot different. The Q-Q graph for the DFT case is as clean as the one of the WHT, which confirms our expectation that the constraint (23) will cause $\sigma_1^2, \ldots, \sigma_n^2$ to be close to $\sigma_j^2$.

VII. DISCUSSION

SRMs were introduced in order to make transform based sensing matrices universal. Our work shows that the signal randomization of SRMs makes either the measurements or the mixture components they are derived from, suitable for quantization by making their distributions approximately normal, under a wide variety of conditions. In contrast, our simulations show that it would be quite difficult to efficiently quantize measurements generated by RSTs because of the wide spread of their distributions and because of the complex dependence of the distribution on the input signal.

Random Convolution [8] is another method which was proposed independently for the same purpose after SRMs were introduced [6]. We showed that it can be viewed a variant of LR-SRM, with similar statistical properties of measurements.

Our analysis shows that while the univariate statistical properties of the measurements are similar for global and local randomization, there is a fundamental difference between the multivariate properties of the two: the mixture components $z_1, \ldots, z_n$ generated by the GR-SRMs are uncorrelated, if $\mathbf{W}_j = 0, 1 < j \leq n$, as is the case in all our example transforms (DCT, DFT, WHT), while mixture components generated by LR-SRMs may be highly correlated. This can have a significant impact on the design of the coding scheme, as pointed out in Sec. III. A possible reason for this difference is that while both LR and GR generate an uncorrelated random vector $\mathbf{x} = \mathbf{A} \mathbf{x}$, GR does a more thorough randomization job: with GR the sequence $\left\{ x_k^2 \right\}$ is also uncorrelated, whereas with LR $\bar{x}_k^2 = \bar{x}_k^2, 1 \leq k \leq n$ (the same is true for higher even-order moments).

Within the LR family, there are differences between transform matrices with equal magnitude entries, such as the WHT and RC, and those with different magnitude entries, such as the DCT and DFT (in its real version). With the former, $z_1, \ldots, z_n$ have identical variances, while with the latter the variances are different, hence the measurements distribution is not asymptotically normal. As our normality testing simulation showed, the deviations from normality vary depending on the type of transform and may be small enough to be ignored.

To show the relevance of these findings to the actual design of a quantizer and channel coder, we proposed several methods to improve the quantization and coding of compressive measurements by using signal-specific parameters. These parameters are computed by the encoder and conveyed to the decoder as side information, enabling the latter to modify the dequantizer and channel decoder so that they match the quantizer and channel encoder.

The applicability of these methods and the amount of side information which needs to be shared depend on the flavor of SRM used and the properties of the input signal.

The differences in measurements distribution may lead to design preference for a particular type of SRM. However, the designer’s choice may be restricted by the application, especially if the measurements are generated in the analog domain. For example, LR may be impractical in a one pixel camera [30] because it requires negative optical amplification.

APPENDIX A: DECORRELATION BY LINEAR PREDICTION

Suppose we have a parametric model approximation to $\text{cov}(z_j, z_j), 1 \leq i, j \leq n$, and the parameters of the model are shared with the decoder as side information. Let $r > 0$ be a fixed integer and for simplicity assume $m$ is an integer multiple of $r$. Let $z_{(1)}, \ldots, z_{(m)}$ be grouped in $m/r$ sets $\{ z_{1p}, \ldots, z_{rp} \}, 1 \leq p \leq m/r$ such that the $r$ members of each set are significantly correlated. For $1 \leq q \leq r$, $1 \leq p \leq m/r$, let $\sigma_{q,p}$ be the linear prediction coefficients which minimize $\text{var}(u_{qp})$, where the residual $u_{qp}$ is defined by

$$u_{qp} = \left( z_{qp} - E\{ z_{qp} \} \right) - \sum_{q=1}^{q-1} \sigma_{q,p} \left( z_{qp} - E\{ z_{qp} \} \right)$$

As is well known, $\sigma_{q,p}$ are the solution of a set of $q-1$ linear equations whose coefficients are $\text{cov}(z_{1p}, z_{qp}), 1 \leq k, l \leq q \cdot \{ u_{1p}, \ldots, u_{qp} \}$ are zero mean, uncorrelated and
var[u_{pq}] = var[z_{pq}] - \sum_{k=p+1}^{q-1} a_{k,p,q} \text{cov}(z_{pq}, z_{kp}). \tag{31}

If \{z_{p1}, \ldots, z_{pq}\} are multivariate normal, then \{u_{p1}, \ldots, u_{pq}\} are multivariate normal and independent. In our case the linear prediction coefficients and var[u_{pq}] are computed using the parametric approximation of the covariance. Let Q_{pq} be a quantizer optimized for the residual u_{pq}. Quantization is performed recursively for q = 1, \ldots, r by
\[ \hat{u}_{pq} = Q_p (z_{pq} - E[z_{pq}]) - \sum_{k=p+1}^{q-1} a_{k,p,q} (z_{pq} - E[z_{kp}]), \]

\[ \hat{z}_{pq} = \hat{u}_{pq} + E[z_{pq}] + \sum_{k=p+1}^{q-1} a_{k,p,q} (\hat{z}_{pq} - E[z_{kp}]). \tag{32} \]

The code words \{\hat{u}_{pq}, 1 \leq q \leq r, 1 \leq p \leq n/r\} are channel-coded and transmitted as side information. The decoder uses the approximate covariance model to group \{z_{11}, \ldots, z_{1n}\} into r-sized sets and to estimate the prediction coefficients in the same way sets as the encoder. Then, the decoder uses (32) to compute the dequantized measurements \{\hat{z}_{11}, \ldots, \hat{z}_{1n}\}.

\section*{APPENDIX B: PROOFS}

\textbf{Proof of Lemma 1:} Since \( C_n = \log(n) \),
\[ |C_n| = n(n-1) \ldots (n-m+1), \quad m = m(n) = o(\sqrt{n}) \]
0 \geq \log \left[ \frac{C_n}{C_i} \right] = \sum_{k=0}^{m-1} \log \left( 1 - \frac{k}{n} \right) \geq \sum_{k=0}^{m-1} \frac{k}{n} = \frac{m(m-1)}{2n} \rightarrow 0
\]

hence \( \frac{C_n}{C_i} \rightarrow 1 \). For any \( g \in C_n \) let \( G_{k,n} \) be the multivariate distribution function of \( \{z_{11}, \ldots, z_{1n}\} \). Then
\[ F_i(n) = |C_i|^{-1} \sum_{g \in C_i} G_{k,n}(h^{(o)}), \quad i = 0, 1 \]

Therefore,
\[ F_i(n) = |C_i|^{-1} \left[ \sum_{g \in C_i} G_{k,n}(h^{(o)}) - \sum_{g \in C_i - C_i} G_{k,n}(h^{(o)}) \right] = \frac{F_i(n)}{C_i} \left[ \frac{1}{C_i} \sum_{g \in C_i} G_{k,n}(h^{(o)}) - \frac{1}{C_i} \sum_{g \in C_i - C_i} G_{k,n}(h^{(o)}) \right] \]

\[ \frac{F_i(n)}{F_i(n)} = 1 \left[ \frac{1}{C_i} \sum_{g \in C_i} G_{k,n}(h^{(o)}) - \frac{1}{C_i} \sum_{g \in C_i - C_i} G_{k,n}(h^{(o)}) \right] \]

The second term on the right hand side vanishes as \( n \rightarrow \infty \)
because \( \liminf_{n \rightarrow \infty} F_i(n) > 0 \)
and
\[ 0 \leq |C_i|^{-1} \sum_{g \in C_i - C_i} G_{k,n}(h^{(o)}) \leq \frac{1}{C_i} \left[ |C_i| - |C_i| \right] \rightarrow 0. \]

Therefore
\[ \lim_{n \rightarrow \infty} \frac{F_i(n)}{F_i(n)} = \lim_{n \rightarrow \infty} \frac{|C_i|}{C_i} = 1. \square \]

\textbf{Proof of Lemma 2:} \( t_i \) exists because \( t_i^2 \eta(t) \) is continuous and increasing from zero to infinity for \( t \in [0, \infty) \). Define
\[ L(t) = \sqrt{\log(2/\delta)} \eta(t) = t_i \frac{\eta(t_i)}{\eta(t)} = t_i \left( \frac{t_i \eta(t_i)}{\eta(t)} \right) \]

\[ U(t) = \log(2/\delta) / \eta(t) = t_i \left( \eta(t_i) \right) / \eta(t), \]

then \( L(t)/U(t) = t_i/L(t) \) and because of the monotonicity properties of \( \eta(t) \) and \( \eta(t) \):
\[ t_i \leq t \Rightarrow U(t) \leq t_i \leq L(t), \quad t_i \leq t \Rightarrow U(t) \geq t_i \geq L(t) \leq t \]

By the definition, \( t_n = \sqrt{L(t_n)/U(t_n)} \), hence \( \log t_n \) is the midpoint of the interval between \( \log U(t_n) \) and \( \log L(t_n) \), \( t_i \) is inside this interval while \( t_n \) is outside it. Therefore,
\[ \log \left( \frac{t_n}{t_i} \right) \leq 0.5 \log \left( \frac{L(t_n)}{U(t_n)} \right) = 0.5 \log \left( \frac{t_n}{U(t_n)} \right) \]

This proves (13) and therefore, \( t_n \rightarrow t_i \square \).

\textbf{Proof of Lemma 3:} We need to show that for any \( \epsilon > 0 \) and \( v \in \mathbb{R}^r \) there is \( n_\epsilon(\nu, v) \) such that if \( n \geq n_\epsilon(\nu, v) \) and \( \{u^{(n)}\} \) satisfies (14) with \( \{j_p(1), \ldots, j_p(r)\} \subseteq J^{(n)}, \) then
\[ F(v) = F_{u^{(n)}}(v) \leq \epsilon, \quad \text{where } F(v), \ F_{u^{(n)}}(v) \text{ denote the } \]

probability distributions of \( \mathcal{N}(0, I) \) and of \( \{u^{(n)}\}, \) respectively. Suppose this was not true. Then we could create a sequence \( \{u^{(\nu)}\} \) which, for sufficiently large \( n, \) satisfies (14) and for which \( \{j_p(1), \ldots, j_p(r)\} \subseteq J^{(n)}, \) but
\[ \limsup_n \left[ F(v) - F_{u^{(n)}}(v) \right] \geq \epsilon, \quad \text{in contradiction to the lemma’s assumption that } \{u^{(n)}\} \text{ is AMN.} \square \]

The proofs of theorems 1-3 rely on the following theorems and on Lemma 4 below:

\textbf{Bennett’s Concentration Inequality} [32]: Let \( X \mid k = 1, \ldots, n \) be independent RVs such that
\[ \Pr\left[ X_k \leq D \right] = 1. \]

\[ \Pr\left[ \sum_{k=1}^n (X_k - E(X_k)) \geq \epsilon \right] \leq \exp \left[ -\frac{\epsilon^2}{D^2} \left( eD \sqrt{\epsilon} \right)^{-1} \right] \]

where
\[ \theta(u) = (1 + u) \log(1 + u) - u \]

and consequently
\[ \Pr\left[ \sum_{k=1}^n (X_k - E(X_k)) \geq \epsilon \right] \leq 2 \exp \left[ -\frac{\epsilon^2}{D^2} \left( eD \sqrt{\epsilon} \right)^{-1} \right] \square. \tag{35} \]

\textbf{Lyapunov Central Limit Theorem for Triangular Arrays} [33]: Let \( X_{nk}, 1 \leq k \leq p_n, n = 1, 2, \ldots \) be RVs, \( E(X_{nk}) = \mu_{nk}, \)
\[ \text{var}(X_{nk}) = \sigma_{nk}^2 \ll \infty, \text{ such that for each } n, \]
\( X_{nk}, \ldots, X_{nq_n} \) are independent. Let \( s_n^2 = \sum_{k=1}^{p_n} (X_{nk} - \mu_{nk})^2 \]
\[ \lim_{n \rightarrow \infty} s_n^{-2} \sum_{k=1}^{p_n} (X_{nk} - \mu_{nk})^2 = 0 \]

for some \( \delta > 0, \) then:
\[ s_n^{-1} \sum_{k=1}^{p_n} (X_{nk} - \mu_{nk}) \rightarrow \mathcal{N}(0, 1) \square. \tag{37} \]

\textbf{Chatterjee Concentration Inequality} [34]: Let \( a_{ij} \in [0, 1], 1 \leq i, j \leq n \) and let \( S = \sum_{k=1}^n a_{k,j}(x_k), \) where \( \pi \) is a...
uniformly distributed, random permutation of \{1,\ldots,n\}. For \( t \geq 0 \)

\[
\Pr\left[|S - E\{S\}| \geq rt\right] \leq 2\exp\left[-t^2/(4E\{S\} + 2r^2)\right].
\]

(38)

**Combinatorial Central Limit Theorem** [35]: For \( n=1,2,\ldots \)

let \( \{a_n(i)\}, \{b_n(i)\}, 1 \leq i \leq n \) be constant sequences with means \( \bar{a}_n, \bar{b}_n \), respectively. Let \( \pi_n \) be a uniformly distributed permutation of \{1,\ldots,n\}. Define \( S_n \equiv \sum_{i=1}^{n} a_n(i)b_n(\pi_n(i)) \).

\[
\lim_{n \to \infty} \frac{\max_{1 \leq i \leq n} (a_n(i) - \bar{a}_n)^2}{\sum_{i=1}^{n} (a_n(i) - \bar{a}_n)^2} = 0
\]

(39)

then \( S_n \) is asymptotically normally distributed.

**Lemma 4**: If \( \{u_n^{(1)}\} \) as defined by (14) and condition (16) is satisfied, then for sufficiently large \( n \) there is \( \varepsilon > 0 \) such that

\[
\forall \alpha \in \mathbb{R}^t : (m/n)\alpha^T \text{cov}(u^{(1)}(\alpha)) \geq \varepsilon \|\alpha\|^2.
\]

**Proof**: Since \( W^{(n)}(R^n) \) is orthonormal, for any \( \mathbf{b} \in \mathbb{R}^t \),

\[
\|\mathbf{b}^TW^{(n)}(R^n)x^{(n)}\|^2 \leq \max_{1 \leq n \leq N} \|\mathbf{b}\|^2.
\]

Suppose that \( j_1(1), \ldots, j_n(r) \) are distinct. Let \( \mathbf{b}_1, \mathbf{b}_2 \in \mathbb{R}^t \) and \( i=1,2 \) define \( \mathbf{b}_1, \mathbf{b}_2 \in \mathbb{R}^t \) by

\[
\left(\mathbf{b}_j\right)_h = \begin{cases} (\mathbf{b}_1)_h & \text{if } h = j(k) \text{ for some } 1 \leq k \leq r \varepsilon \|\alpha\|^2. \\
0 & \text{otherwise} \end{cases}
\]

Then

\[
\|\mathbf{b}_1\|_2 = \|\mathbf{b}_2\|_2, \text{ and}
\]

\[
\|\mathbf{b}_1^T \text{cov}(u^{(n)}(\alpha)) \mathbf{b}_1\| = \|\mathbf{b}_1^T \text{cov}(z^{(n)}(\alpha)) \mathbf{b}_1\| \leq E\left[\|\mathbf{b}_1^TW^{(n)}(R^n)x^{(n)}\|^2\right] + E\left[\|\mathbf{b}_1^TW^{(n)}(R^n)x^{(n)}\|^2\right] - 2\max_{1 \leq n \leq N} \|\mathbf{b}_1\|_2 \leq 2\max_{1 \leq n \leq N} \|\mathbf{b}_1\|_2.
\]

Condition (36) of Lyapunov CLT for triangular arrays is satisfied because for any \( \delta > 0 \), using (15):

\[
\left(\frac{1}{n}\sum_{i=1}^{n} E\left[|X_{n,i} - E|X_{n,i}|\right]\right) \leq s_n^{-2}\max_{1 \leq i \leq n} \|\mathbf{w}_i\|^2.
\]

Then \( X_{n,i} \), \( X_{n,m} \) are independent, zero mean RVs and:

\[
E\left[X_{n,i}^2\right] = \sum_{j=1}^{i} \sum_{k=1}^{j} \text{cov}(z^{(n)}(\alpha)) \mathbf{w}_i \mathbf{w}_k, 1 \leq i \leq n
\]

Then \( X_{n,i} \), \( X_{n,m} \) are independent, zero mean RVs and:

\[
\sum_{i=1}^{n} E\left[X_{n,i}^2\right] = (m/n)\sum_{i=1}^{n} \text{cov}(z^{(n)}(\alpha)) \mathbf{w}_i \mathbf{w}_k
\]

Then:

\[
\sum_{i=1}^{n} E\left[X_{n,i}^2\right] = (m/n)\sum_{i=1}^{n} \text{cov}(z^{(n)}(\alpha)) \mathbf{w}_i \mathbf{w}_k
\]

Then:

\[
\sum_{i=1}^{n} E\left[X_{n,i}^2\right] = (m/n)\sum_{i=1}^{n} \text{cov}(z^{(n)}(\alpha)) \mathbf{w}_i \mathbf{w}_k
\]

Then:

\[
\sum_{i=1}^{n} E\left[X_{n,i}^2\right] = (m/n)\sum_{i=1}^{n} \text{cov}(z^{(n)}(\alpha)) \mathbf{w}_i \mathbf{w}_k
\]

Then:

\[
\sum_{i=1}^{n} E\left[X_{n,i}^2\right] = (m/n)\sum_{i=1}^{n} \text{cov}(z^{(n)}(\alpha)) \mathbf{w}_i \mathbf{w}_k
\]

Then:

\[
\sum_{i=1}^{n} E\left[X_{n,i}^2\right] = (m/n)\sum_{i=1}^{n} \text{cov}(z^{(n)}(\alpha)) \mathbf{w}_i \mathbf{w}_k
\]

Then:

\[
\sum_{i=1}^{n} E\left[X_{n,i}^2\right] = (m/n)\sum_{i=1}^{n} \text{cov}(z^{(n)}(\alpha)) \mathbf{w}_i \mathbf{w}_k
\]

Therefore

\[
\mathbf{e} \cdot \mathbf{u}^{(n)}(\alpha) = \mathbf{e} \cdot \mathbf{u}^{(n)}(\alpha)
\]

is true for all \( \mathbf{e} \in \mathbb{R}^t \), \( \mathbf{e} \neq 0 \), then by Cramér-Wold Theorem [33].

**Proof of Theorem 2**: (a) Let \( \mathbf{v} = [v_1, \ldots, v_n]^T = F\mathbf{x} \). For \( 1 \leq j, k \leq n \) let

\[
\alpha_j \triangleq \arg(\alpha_k) + 2\pi n^{-1}(j-1)(k-1)
\]

\[
X_{\alpha} \triangleq m^{-\frac{1}{2}} (2 - \chi_{\alpha}(k)) |v_j| \cos(\beta_k + \phi_{\alpha})
\]

Note that \( \arg(\alpha_k) \) and \( \alpha_j \) are multiples of \( \pi \) if \( \chi_{\alpha}(k) = 1 \).

\[
z_j = m^{-\frac{1}{2}} \text{Re} \left\{ \sum_{k=1}^{n} \bar{r}_j \bar{r}_k b_k w_k \right\} = \sum_{k=1}^{n} \text{Re} \left\{ X_{\alpha} \right\}
\]

The distribution of \( z_j \), \( 1 \leq j \leq n \) is symmetric and \( E(z_j) = 0 \) because the distribution of each \( X_{\alpha} \) is symmetric. For any \( 1 \leq j, h \leq n \), \( 1 \leq k, l \leq n/2+1 \)
\[ E \left\{ \cos(\beta_i + \varphi_{\beta}) \cos(\beta_i + \varphi_{\beta}) \right\} = \]
\[ \frac{\delta_{\mu}}{2} E \left\{ \cos(\varphi_{\beta} - \varphi_{\beta}) + \cos(2\beta_i + \varphi_{\beta} + \varphi_{\beta}) \right\} = \]
\[ \frac{\delta_{\mu}}{2} (1 + \chi(k)) \cos(\varphi_{\beta} - \varphi_{\beta}) \]
hence
\[ E \left\{ X_{X_{\alpha}} \right\} = m^4 \delta_{\mu} (2 - \chi(k)) \mathbb{E} \left[ \cos(\varphi_{\beta} - \varphi_{\beta}) \right]. \]
Let \( \tilde{v}_n \triangleq v_n \), \( 1 \leq k \leq n \), \( \tilde{v}_n \triangleq [\tilde{v}_1, \ldots, \tilde{v}_n]^T \). By the DFT properties, \( \rho_n(j) = n^{2}(F^{t} \tilde{v})_{j+1} \). Therefore,
\[
\begin{align*}
\text{cov} (z_j, z_k) &= E \left\{ z_j z_k \right\} = \sum_{k \in \mathbb{Z}^2/\{0\}} E \left\{ X_{X_{\alpha}} \right\} = \\
&= m^4 \sum_{k \in \mathbb{Z}^2/\{0\}} (2 - \chi(k)) \tilde{v}_n \cos(\varphi_{\beta} - \varphi_{\beta}) = \\
&= m^4 \max_{\delta_{\mu}} \Re \left\{ \left( \tilde{v}_n \cos(\varphi_{\beta} - \varphi_{\beta}) \right) \right\} = m^4 \rho_\mu (j - h) \\
\end{align*}
\]
(b) \( X_{X_{\alpha}} \) are independent and
\[
\sum_{k \in \mathbb{Z}^2/\{0\}} E \left\{ X_{X_{\alpha}} \right\} = m^4 \max_{\delta_{\mu}} \left( 2 - \chi(k) \right) \mathbb{E} \left[ (F \tilde{v}_n)_{j+1} \right] \geq |X_{X_{\alpha}}|. \]
Since \( x \neq 0 \), \( \sigma_j, D_j \) are positive. By Bennett concentration inequality, since \( \tau_j = \sigma_j / D_j \):
\[
\Pr \left\{ \left| z_j \right| > \sigma_j \right\} \leq 2 \exp \left[ - \tau_j^2 / (\tau / r_j) \right] = 2 \exp \left[ - \tau^2 / (\tau / r_j) \right]. \]
(c) Let \( C_n \triangleq \text{cov}(u^{(n)}) \). By Lemma 4 that there is \( \epsilon > 0 \) such that for sufficiently large \( n \), \( C_n \) is positive definite and for \( \alpha \in \mathbb{R}^r \), \( \alpha^T C_n \alpha \geq \epsilon^2 (n/m) \| \alpha \|^2 \). For such \( n \) and any \( \alpha \in \mathbb{R}^r \) define \( \tilde{\alpha}_n \triangleq C_n^{1/2} \alpha \), \( [v_{(1)}^n, \ldots, v_{(n)}^n]^T \triangleq F^{(n)} x^{(n)} \). For \( 1 \leq k \leq n/2 + 1 \) define
\[
\begin{align*}
\tilde{w}_n^{(k)} &\triangleq \left[ f_{(1),k,j}, \ldots, f_{(n),k,j} \right]^T \\
\varphi_{\beta_k} &\triangleq \arg \left( v_k \tilde{\alpha}^T \tilde{w}_n^{(k)} \right) \\
X_{X_{\alpha}} \triangleq \left( n/m \right)^{1/2} (2 - \chi(k)) \Re (v_k \tilde{\alpha}^T \tilde{w}_n^{(k)}) = \\
&= \left( n/m \right)^{1/2} (2 - \chi(k)) \left| v_k \tilde{\alpha}^T \tilde{w}_n^{(k)} \right| \cos(\beta_k + \varphi_{\beta_k}) \\
\end{align*}
\]
(26) follows from
\[
E \left\{ z_j \right\} = \left( n/m \right)^{1/2} \sum_{k \leq n/2} x_k E \left\{ w_{j,k} \right\} = n^{1/2} m^{1/2} \tilde{w}_n. \]
(27) is proved by computing \( E \{ z_{j,h} \} \):
\[
E \left\{ z_{j,h} \right\} = \frac{n}{m} \sum_{k \leq n/2} x_k x_{j,h} E \left\{ w_{j,k} \right\} = \\
= \frac{n}{m} \sum_{k \leq n/2} x_k x_{j,h} E \left\{ w_{j,k} \right\} + \frac{n}{m} \sum_{k \leq n/2} x_k x_{j,h} E \left\{ w_{j,h} \right\} \]
Using the results above and the identity
\[
\sum_{k \leq n/2} x_k x_{j,k} = \left( \sum_{k \leq n/2} x_k \right)^2 - \sum_{k \leq n/2} x_k^2 = n^2 \tilde{w} - \|s\|^2 \]
we get
The conditions of Chatterjee concentration inequality are satisfied for $S_j$. Since $E(S_j) = \sum_{k=4}^{2a} E(a_{\alpha(k)}) = n/2$, let $z_{j} - E\{z_{j}\} = \left(2\sigma_j / \tau_j\right)\left(S_j - E\{S_j\}\right)$

$$\Pr\{z_{j} - E\{z_{j}\} \geq \sigma_j\} = \Pr\{S_j - E\{S_j\} \geq \tau_j / 2\} \leq 2\exp\left[-(\tau_j / 2)^2/(2n + \tau_j)^2\right] = 2\exp\left[-\pi^2 \eta_j(t, \tau_j)\right].$$

(c) Let $C_\alpha \triangleq \text{cov}(u^{(\alpha)})$. By lemma 4 there is $\varepsilon > 0$ such that for sufficiently large $n$, $C_\alpha$ is positive definite and for any $\alpha \in \mathbb{R}^r$, $\alpha^T C_\alpha \alpha \geq \varepsilon (n/m) \|\alpha\|^2$. For any $\alpha \in \mathbb{R}^r$, $\alpha \neq 0$, and $1 \leq k \leq n$ define:

$$\tilde{\alpha}_k \triangleq [\tilde{\alpha}_{k(1)}, \ldots, \tilde{\alpha}_{k(n)}]^T \triangleq C_\alpha^{-1/2} \tilde{\alpha}_k,$$

$$\tilde{\alpha}_k^{(n)} \triangleq [\tilde{\alpha}_{k(1)}, \ldots, \tilde{\alpha}_{k(j)}]^T, \quad \tilde{\alpha}_k^{(n)} \triangleq [\tilde{\alpha}_{k(1)}, \ldots, \tilde{\alpha}_{k(n)}]^T,$$

$$a_{\alpha(k)} \triangleq \sqrt{n/m} (x_k - \bar{x}^{(\alpha)}), \quad b_{\alpha(k)} \triangleq \tilde{a}_{\alpha(k)} \tilde{w}_k^{(n)}.$$  

We show that the conditions of the combinatorial CLT hold for $\{a_{\alpha(k)}, \{b_{\alpha(k)}\}, 1 \leq k \leq n, n = 1, 2, \ldots$. Note that $\tilde{\alpha}_n = \tilde{\beta}_n = 0$.

$$n^{-1} \sum_{k=1}^{n} \left(a_{\alpha(k)} - \tilde{\alpha}_n\right)^2 \sum_{k=1}^{n} \left(b_{\alpha(k)} - \tilde{\beta}_n\right)^2 =$$

$$m^{-1} \sum_{k=1}^{n} \left(x_k - \bar{x}^{(\alpha)}\right)^2 \sum_{k=1}^{n} \left(\tilde{\alpha}_k^{(n)} \tilde{w}_k^{(n)} - \tilde{\alpha}_n^{(n)} \tilde{w}_k^{(n)}\right)^2 =$$

$$m^{-1} \left(\|\bar{x}^{(\alpha)}\|^2 - n\left(\bar{x}^{(\alpha)}\right)^2\right) \sum_{k=1}^{n} \left(\tilde{a}_{\alpha(k)} \tilde{\alpha}_k^{(n)} - \tilde{a}_{\alpha(k)} \tilde{w}_k^{(n)}\right)^2 =$$

$$(n-1)m^{-1} \tilde{\alpha}_n^T C_{\tilde{\alpha}_n^{(n)}} \tilde{\alpha}_n \geq (n-1)m^{-1} \|\tilde{\alpha}_n^{(n)}\|_E E_{\tilde{\alpha}_n^{(n)}}$$

The resulting CLT then follows from the combinatorial CLT.

Acknowledgment

We are grateful to Hong Jiang, Jin Cao, Tin Ho and Alexei Ashikhmin of Bell-Laboratories for helpful discussions and suggestions.

References

[1] M. A. Davenport, F. Duarte, Marco, Y. C. Eldar, and G. Kutyniok, “Introduction to Compressed Sensing,” in Compressed Sensing, Y.C. Eldar and G. Kutyniok, Eds., Cambridge, UK: Cambridge University Press, 2012.

[2] R. Calderbank, S. Howard, and S. Jafarpour, “Construction of a Large Class of Deterministic Sensing Matrices That Satisfy a Statistical Isometry Property,” IEEE J. Sel. Top. Signal Process., vol. 4, no. 2, pp. 358–374, Apr. 2010.

[3] E. J. Candès and T. Tao, “Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?,” IEEE Transactions on Information Theory, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[4] E. J. Candès and T. Tao, “Decoding by Linear Programming,” IEEE Trans. Inf. Theory, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[5] E. Candès and J. Romberg, “Sparsity and incoherence in compressive sampling,” Inverse Probl., vol. 23, no. 3, pp. 969–985, Jun. 2007.

Compressive Measurements Generated by Structurally Random Matrices: Asymptotic Normality and Quantization
[6] T. T. Do, T. D. Tran, and L. Gan, “Fast compressive sampling with structurally random matrices,” in ICASSP 2008 IEEE International Conference Acoustic Speech & Signal Processing, 2008, pp. 3369–3372.

[7] T. T. Do, L. Gan, N. H. Nguyen, and T. D. Tran, “Fast and Efficient Compressive Sensing Using Structurally Random Matrices,” *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 139–154, Jan. 2012.

[8] J. Romberg, “Compressive Sensing by Random Convolution,” *SIAM J. Imaging Sci.*, vol. 2, no. 4, pp. 1098–1128, Jan. 2009.

[9] F. Krahmer and R. Ward, “New and Improved Johnson–Lindenstrauss Embeddings via the Restricted Isometry Property,” *SIAM J. Math. Anal.*, vol. 43, no. 3, pp. 1269–1281, Jan. 2011.

[10] C. Li, H. Jiang, P. Wilford, and Y. Zhang, “Video coding using compressive sensing for wireless communications,” *2011 IEEE Wireless Communications and Networking Conference*, pp. 2077–2082, Mar. 2011.

[11] C. Li, H. Jiang, P. Wilford, Y. Zhang, and M. Scheutzow, “A new compressive video sensing framework for mobile broadcast,” *IEEE Transactions on Broadcasting*, vol. 59, no. 1, pp. 197–205, Mar. 2013.

[12] S. Pudlewski and T. Melodia, “On the performance of compressive video streaming for wireless multimedia sensor networks,” in *2010 IEEE International Conference on Communications*, 2010, pp. 1-5.

[13] H. Jiang, C. Li, R. Haimi-Cohen, P. A. Wilford, and Y. Zhang, “Scalable video coding using compressive sensing,” *Bell Labs Technical Journal*, vol. 16, no. 4, pp. 149–169, Mar. 2012.

[14] D. Venkatraman and A. Makur, “A compressive sensing approach to object-based surveillance video coding,” in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 3513–3516.

[15] A. Mahalanobis and R. Muise, “Object specific image reconstruction using a compressive sensing architecture for application in surveillance systems,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 3, pp. 1167–1180, Jul. 2009.

[16] H. Jiang, S. Zhao, Z. Shen, W. Deng, P. Wilford, and R. Haimi-Cohen, “Video Analysis using Compressive Sensing with Low Latency,” *Bell Labs Tech. J.*, vol. 18, no. 4, 2014.

[17] W. Dai, H. V. Pham, and O. Milenkovic, “Distortion-rate functions for quantized compressive sensing,” in *2009 IEEE Information Theory Workshop on Networking and Information Theory*, 2009, pp. 171–175.

[18] W. Dai, R. V. Pham, and O. Milenkovic, “A comparative study of quantized compressive sensing schemes,” in *2009 IEEE International Symposium on Information Theory*, 2009, pp. 11–15.

[19] Y. Baig, E. M.-K. Lai, and J. P. Lewis, “Quantization effects on compressed sensing video,” in *2010 17th International Conference on Telecommunications*, 2010, pp. 935–940.

[20] J. Z. Sun and V. K. Goyal, “Optimal quantization of random measurements in compressed sensing,” *2009 IEEE International Symposium on Information Theory*, vol. 3, no. 1, pp. 6–10, Jun. 2009.

[21] J. N. Laska, P. T. Boufounos, M. a. Davenport, and R. G. Baraniuk, “Democracy in action: Quantization, saturation, and compressive sensing,” *Appl. Comput. Harmon. Anal.*, vol. 31, no. 3, pp. 429–443, Nov. 2011.

[22] J. N. Laska and R. G. Baraniuk, “Regime change: Bit-depth versus measurement-rate in compressive sensing,” *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3496–3505, Jul. 2012.

[23] J. N. Laska and R. G. Baraniuk, “Trust, But Verify: Fast and Accurate Signal Recovery From 1-Bit Compressive Measurements,” *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5289–5301, Nov. 2011.

[24] L. Jacques, D. K. Hammond, and J. M. Fadili, “Dequantizing compressed sensing: When constraints combine,” *IEEE Transaction on Information Theory*, vol. 57, no. 1, pp. 559–571, 2011.A. Zymnis, S. Boyd, and E. Candès, “Compressed sensing with quantized measurements,” *IEEE Signal Processing Letters*, vol. 17, no. 2, pp. 149–152, 2010.

[25] M. F. Duarte and R. G. Baraniuk, “Kronecker compressive sensing,” *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 494–504, Feb. 2012.

[26] J. Rissanen and G. G. Langdon, “Arithmetic coding,” *IBM J. Res. Dev.*, vol. 23, no. 2, pp. 149–162, 1979.M. F. Durate et al., “Single-Pixel Imaging via Compressive Sampling,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, 2008.

[27] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression* Kluwer Academic Publishers, 1992, sec. 9.9, pp. 295-302.

[28] J. Ziv, “On universal quantization,” *IEEE Transactions on Information Theory*, vol. 31, no. 3, pp. 344–347, May 1985.
[29] A. M. Kondoz, *Digital Speech*, Sec. 3.3.2, 4.2. Chichester, 1994.

[30] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, “Single-Pixel Imaging via Compressive Sampling,” *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 83–91, 2008.

[31] H.C. Thode, *Testing for Normality* New York: Marcel Dekker Inc., 2002, pp. 21-23.

[32] G. Bennett, “Probability inequalities for the sum of independent random variables,” J. American Stat. Assoc., vol. 57(297), pp. 33-45, 1962

[33] P. Billingsley, *Probability and Measure*, 3rd ed. New York: Wiley, 1995.

[34] S. Chatterjee, “Stein’s method for concentration inequalities,” Probability Theory and Related Fields, vol. 138, no. 1–2, pp. 305–321, Apr. 2007.

[35] W. Hoeffding, “A Combinatorial Central Limit Theorem,” *The Annals of Mathematical Statistics*, vol. 22, no. 4, pp. 558–566, 1951.