



Random subsets of structured deterministic frames have MANOVA spectra

Marina Haikin^a, Ram Zamir^a, and Matan Gavish^{b,1}

^aSchool of Electrical Engineering, Systems Department, Tel Aviv University, Tel Aviv 6997801, Israel; and ^bSchool of Computer Science and Engineering, The Hebrew University, Jerusalem 9190416, Israel

Edited by David L. Donoho, Stanford University, Stanford, CA, and approved May 2, 2017 (received for review January 4, 2017)

We draw a random subset of k rows from a frame with n rows (vectors) and m columns (dimensions), where k and m are proportional to n . For a variety of important deterministic equiangular tight frames (ETFs) and tight non-ETFs, we consider the distribution of singular values of the k -subset matrix. We observe that, for large n , they can be precisely described by a known probability distribution—Wachter’s MANOVA (multivariate ANOVA) spectral distribution, a phenomenon that was previously known only for two types of random frames. In terms of convergence to this limit, the k -subset matrix from all of these frames is shown to be empirically indistinguishable from the classical MANOVA (Jacobi) random matrix ensemble. Thus, empirically, the MANOVA ensemble offers a universal description of the spectra of randomly selected k subframes, even those taken from deterministic frames. The same universality phenomena is shown to hold for notable random frames as well. This description enables exact calculations of properties of solutions for systems of linear equations based on a random choice of k frame vectors of n possible vectors and has a variety of implications for erasure coding, compressed sensing, and sparse recovery. When the aspect ratio m/n is small, the MANOVA spectrum tends to the well-known Marčenko–Pastur distribution of the singular values of a Gaussian matrix, in agreement with previous work on highly redundant frames. Our results are empirical, but they are exhaustive, precise, and fully reproducible.

deterministic frames | MANOVA | analog source coding | equiangular tight frames | restricted isometry property

Consider a frame $\{\mathbf{x}_i\}_{i=1}^n \subset \mathbb{R}^m$ or \mathbb{C}^m , and stack the vectors as rows to obtain the n -by- m frame matrix X . Assume that $\|\mathbf{x}_i\|_2 = 1$ (deterministic frames) or $\lim_{n \rightarrow \infty} \|\mathbf{x}_i\| = 1$ almost surely (random frames). This paper studies properties of a random subframe $\{\mathbf{x}_i\}_{i \in K}$, where K is chosen uniformly at random from $[n] = \{1, \dots, n\}$ and $|K| = k \leq n$. We let X_K denote the k -by- m submatrix of X created by picking only the rows $\{\mathbf{x}_i\}_{i \in K}$; call this object a typical k submatrix of X . We consider a collection of well-known deterministic frames, listed in Table 1, which we denote by \mathcal{X} . Most of the frames in \mathcal{X} are equiangular tight frames (ETFs), and some are near-ETFs.

This paper suggests that, for a frame in \mathcal{X} , it is possible to calculate quantities of the form $\mathbb{E}_K \Psi(\lambda(G_K))$, where $\lambda(G_K) = (\lambda_1(G_K), \dots, \lambda_k(G_K))$ is the vector of eigenvalues of the k -by- k Gram matrix $G_K = X_K X_K'$ and Ψ is a functional of these eigenvalues. As discussed below, such quantities are of considerable interest in various applications where frames are used across a variety of domains, including compressed sensing, sparse recovery, and erasure coding.

We present a simple and explicit formula for calculating $\mathbb{E}_K \Psi(\lambda(G_K))$ for a given frame in \mathcal{X} and a given spectral functional Ψ . Specifically, for the case $k \leq m$,

$$\mathbb{E}_K \Psi(\lambda(G_K)) \approx \Psi \left(f_{\beta, \gamma}^{MANOVA} \right),$$

where $\beta = k/m$, $\gamma = m/n$, and $f_{\beta, \gamma}^{MANOVA}$ is the density of Wachter’s classical multivariate ANOVA limiting distribution

(1), which we denote here by $\text{MANOVA}(\beta, \gamma)$. The fluctuations about this approximate value are given exactly by

$$\mathbb{E}_K \left| \Psi(\lambda(G_K)) - \Psi \left(f_{\beta, \gamma}^{MANOVA} \right) \right|^2 = C n^{-b} \log^{-a}(n). \quad [1]$$

Although the constant C may depend on the frame, the exponents a and b are universal and depend only on Ψ and the aspect ratios β and γ . Evidently, the precision of the MANOVA-based approximation is good, is known, and improves as m and k both grow proportionally to n .

Eq. 1 is based on a far-reaching universality hypothesis. For all frames in \mathcal{X} as well as well-known random frames also listed in Table 1, we find that the spectrum of the typical k -submatrix ensemble is indistinguishable from that of the classical MANOVA (Jacobi) random matrix ensemble (2) of the same size. (Interestingly, it will be shown that, for deterministic ETFs, this indistinguishability holds in a stronger sense than for deterministic non-ETFs.) This universality is not asymptotic and concerns finite n -by- m frames. However, it does imply that the spectrum of the typical k -submatrix ensemble converges to a universal limiting distribution, which is none other than Wachter’s $\text{MANOVA}(\beta, \gamma)$ limiting distribution (1). It also implies that the universal exponents a and b in Eq. 1 are previously unknown, universal quantities corresponding to the classical MANOVA (Jacobi) random matrix ensemble.

Significance

A frame (overcomplete set of vectors) represents an analog coding scheme. Deterministic frame constructions offer useful codes for communication and signal processing tasks. When the coded signal only uses a random subset of the frame vectors (for example, in compressed sensing), the coding quality is determined by the typical covariances within subsets of frame vectors. We provide a method to calculate functions of these typical covariances, which predict specific performance measures of the corresponding coding scheme. Our method uses a universality property: for many well-known deterministic and random frames, typical covariances within subsets of frame vectors do not depend on the frame and are described by the MANOVA (multivariate ANOVA) ensemble, a classical object in statistics and random matrix theory.

Author contributions: R.Z. and M.G. designed research; M.H., R.Z., and M.G. performed research; M.G. contributed new reagents/analytic tools; M.H. analyzed data; and M.H., R.Z., and M.G. wrote the paper.

Conflict of interest statement: Matan Gavish is a former student of David L. Donoho, and they have published together, most recently in 2014.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The code and data supplement is available online at <https://publ.stanford.edu/qq138qm8653>.

¹To whom correspondence should be addressed. Email: gavish@cs.huji.ac.il.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1700203114/-DCSupplemental.

Table 1. Frames under study

Label	Name	\mathbb{R} or \mathbb{C}	Natural γ	Tight frame	Equiangular	Refs.
Deterministic frames						
DSS	Difference set spectrum	\mathbb{C}		Yes	Yes	36
GF	Grassmannian frame	\mathbb{C}	1/2	Yes	Yes	37, corollary 2.6b
RealPF	Real Paley's construction	\mathbb{R}	1/2	Yes	Yes	37, corollary 2.6a
ComplexPF	Complex Paley's construction	\mathbb{C}	1/2	Yes	Yes	38
Alltop	Quadratic Phase Chirp	\mathbb{C}	1/L	Yes	No	26, equation S4
SS	Spikes and Sines	\mathbb{C}	1/2	Yes	No	6
SH	Spikes and Hadamard	\mathbb{R}	1/2	Yes	No	6
Random frames						
HAAR	Unitary Haar frame	\mathbb{C}		Yes	No	3, 4
RealHAAR	Orthogonal Haar frame	\mathbb{R}		Yes	No	4
RandDFT	Random Fourier transform	\mathbb{C}		Yes	No	3
RandDCT	Random Cosine transform	\mathbb{R}		Yes	No	

This brief announcement tests Eq. 1 and the underlying universality hypothesis by conducting substantial computer experiments, in which a large number of random k submatrices are generated. We study a large variety of deterministic frames, both real and complex. In addition to the universal object (the MANOVA ensemble) itself, we study difference set spectrum (DSS) frames, Grassmannian frame (GF), real Paley (RealPF) frames, complex Paley (ComplexPF) frames, quadratic phase chirp (Alltop) frames, Spikes and Sines (SS) frames, and Spikes and Hadamard (SH) frames.

We report compelling empirical evidence, systematically documented and analyzed, which fully supports the universality hypothesis and Eq. 1. Our results are empirical, but they are exhaustive, precise, and reproducible, and they meet the best standards of empirical science.

For this purpose, we develop a natural framework for empirically testing such hypotheses regarding limiting distribution and convergence rates of random matrix ensembles. Before turning to deterministic frames, we validate our framework on well-known random frames, including real orthogonal Haar frames, complex unitary Haar frames, real random Cosine frames, and complex random Fourier frames. Interestingly, rigorous proofs that identify the MANOVA distribution as the limiting spectral distribution of typical k submatrices can be found in the literature for two of these random frames, namely the random Fourier frame (3) and the unitary Haar frame (4).

Motivation

Frames can be viewed as an analog counterpart for digital coding. They provide overcomplete representation of signals, adding redundancy and increasing immunity to noise. Indeed, they are used in many branches of science and engineering for stable signal representation as well as error and erasure correction.

Let $\lambda(G)$ denote the vector of nonzero eigenvalues of $G = X'X$, and let $\lambda_{max}(G)$ and $\lambda_{min}(G)$ denote its maximum and minimum, respectively. Frames were traditionally designed to achieve frame bounds $\lambda_{min}(G)$ as high as possible [$\lambda_{max}(G)$ as low as possible]. Alternatively, they were designed to minimize mutual coherence (5, 6), the maximal pairwise correlation between any two frame vectors.

In the passing decade, it has become apparent that neither frame bounds (a global criterion) nor coherence (a local pairwise criterion) are sufficient to explain various phenomena related to overcomplete representations and that one should also look at collective behavior of k frame vectors from the frame, $2 \leq k \leq n$. Although different applications focus on different properties of the submatrix G_K , most of these properties can be expressed as a function of $\lambda(G_K)$ and even just an average of a scalar function of the eigenvalues. Here are a few notable examples.

Restricted Isometry Property. Recovery of any $k/2$ -sparse signal $v \in \mathbb{R}^n$ from its linear measurement $F'v$ using ℓ_1 minimization is guaranteed if the spectral radius of $G_K - I$ [restricted isometry property (RIP)], namely

$$\Psi_{RIP}(\lambda(G_K)) = \max\{\lambda_{max}(G_K) - 1, 1 - \lambda_{min}(G_K)\}, \quad [2]$$

is uniformly bounded by some $\delta < 0.4531$ on all $K \subset [n]$ (7–9).

Statistical RIP. Numerous authors have studied a relaxation of the RIP condition suggested in ref. 10. Define

$$\Psi_{StRIP,\delta}(\lambda(G_K)) = \begin{cases} 1 & \Psi_{RIP}(\lambda(G_K)) \leq \delta \\ 0 & \text{otherwise} \end{cases} \quad [3]$$

Then, $\mathbb{E}_K \Psi_{StRIP,\delta}(\lambda(G_K))$ is the probability that the RIP condition with bound δ holds when X acts on a signal supported on a random set of k coordinates.

Analog Coding of a Source with Erasures. In ref. 11, two of us considered a typical erasure pattern of $n - k$ random samples known at the transmitter but not known at the receiver. The rate distortion function of the coding scheme suggested in ref. 11 is determined by $\mathbb{E}_K \log(\beta \Psi_{AC}(\lambda(G_K)))$, with

$$\Psi_{AC}(\lambda(G_K)) = \frac{\frac{1}{k} \text{tr}[(G_K)^{-1}]}{(\frac{1}{k} \text{tr}[G_K])^{-1}}, \quad [4]$$

[i.e., $\Psi_{AC}(\lambda(G_K))$ is the arithmetic-to-harmonic means ratio of the eigenvalues (the arithmetic mean is 1 because of the normalization of frames)]. This quantity is the signal amplification responsible for the excess rate of the suggested coding scheme. Note that β here is the inverse of β defined in ref. 11.

Shannon Transform. The quantity

$$\begin{aligned} \Psi_{Shannon}(\lambda(G_K)) &= \frac{1}{k} \log(\det(I + \alpha G_K)) \\ &= \frac{1}{k} \text{tr}(\log(I + \alpha G_K)), \end{aligned} \quad [5]$$

which was suggested in ref. 12, measures the capacity of a linear Gaussian erasure channel. Specifically, it assumes $y = XX'x + z$ (where x and y are the channel input and output) followed by $n - k$ random erasures. The quantity α in Eq. 5 is the signal-to-noise ratio $SNR = \alpha \geq 0$.

In this paper, we focus on typical case performance criteria [those that seek to optimize $\mathbb{E}_K \Psi(\lambda(G_K))$ over random choice of K] rather than worst case performance criteria [those that seek to optimize $\max_{K \subset [n]} \Psi(\lambda(G_K))$, such as RIP]. For the remainder of this paper, $K \subset [n]$ will denote a uniformly distributed random subset of size k . Importantly, k should be allowed to be large, even as large as m .

For a given Ψ , one would like to design frames that optimize $\mathbb{E}_K \Psi(\lambda(G_K))$. This optimization turns out to be a difficult task; in fact, it is not even known how to calculate $\mathbb{E}_K \Psi(\lambda(G_K))$ for a given frame X . Indeed, to calculate this quantity, one effectively has to average Ψ over the spectrum $\lambda(G_K)$ for all $\binom{n}{k}$ subsets $K \subset [n]$. It is of little surprise to the information theorist that the first frame designs, for which performance was formally bounded (and still not calculated exactly), consisted of random vectors (8, 13–17).

Random Frames

When the frame is random, namely when X is drawn from some ensemble of random matrices, each k submatrix X_K is also a random matrix. Given a specific Ψ , rather than seeking to bound $\mathbb{E}_K \Psi(\lambda(G_K))$ for specific n and m , it can be extremely rewarding to study the limit of $\Psi(\lambda(G_K))$ as the frame sizes n and m grow. The reason is that tools from random matrix theory become available, which allow exact asymptotic calculation of $\lambda(G_K)$ and $\Psi(\lambda(G_K))$, and also because their limiting values are usually very close to their corresponding values for finite n and m , even for low values of n .

Let us consider then a sequence of dimensions m_n with $m_n/n = \gamma_n \rightarrow \gamma$ and a sequence of random frame matrices $X^{(n)} \subset \mathbb{R}^{n \times m_n}$ or $\mathbb{C}^{n \times m_n}$. To characterize the collective behavior of k submatrices, we choose a sequence k_n with $k_n/m_n = \beta_n \rightarrow \beta$ and look at the spectrum $\lambda(G_{K_n})$ of the random matrix X_{K_n} as $n \rightarrow \infty$, where $K_n \subset [n]$ is a randomly chosen subset with $|K_n| = k_n$. Here and below, to avoid cumbersome notation, we omit the subscript n and write m , k , and K for m_n , k_n , and K_n .

A mainstay of random matrix theory is the celebrated convergence of the empirical spectral distribution of random matrices, drawn from a certain ensemble, to a limiting spectral distribution corresponding to that ensemble. Such convergence has indeed been established for three random frames:

1. Gaussian i.i.d. Frame. Let $X_{normal}^{(n)}$ have i.i.d. (independent and identically distributed) normal entries with mean zero and variance $1/m$. The empirical distribution of $\lambda(G_K)$ famously converges, almost surely in distribution, to the Marčenko–Pastur density (18) with parameter β :

$$f_{\beta}^{MP}(x) = \frac{\sqrt{(x - \lambda_-^{MP})(\lambda_+^{MP} - x)}}{2\beta\pi x} \cdot I_{(\lambda_-^{MP}, \lambda_+^{MP})}(x), \quad [6]$$

supported on $[\lambda_-^{MP}, \lambda_+^{MP}]$, where $\lambda_{\pm}^{MP} = (1 \pm \sqrt{\beta})^2$. Moreover, almost surely $\lambda_{max}(G_{normal}^{(n)}) \rightarrow \lambda_+$ and $\lambda_{min}(G_{normal}^{(n)}) \rightarrow \lambda_-$; in other words, the maximal and minimal empirical eigenvalues converge almost surely to the edges of the support of the limiting spectral distribution (19).

2. Random Fourier Frame. Consider the random Fourier frame, in which the m_n columns of $X_{Fourier}^{(n)}$ are drawn uniformly at random from the columns of the n -by- n discrete Fourier transform (DFT) matrix (normalized such that the absolute value of matrix entries is $1/\sqrt{m}$). Farrell (3) has proved that the empirical distribution of $\lambda(G_K)$ converges, almost surely in distribution, as $n \rightarrow \infty$ and m and k grow proportionally to n to the so-called MANOVA limiting distribution, which we now describe briefly.

The classical MANOVA(n, m, k, \mathcal{F}) ensemble,* with $\mathcal{F} \in \{\mathbb{R}, \mathbb{C}\}$, is the distribution of the random matrix

$$\frac{n}{m} (AA' + BB')^{-\frac{1}{2}} BB' (AA' + BB')^{-\frac{1}{2}}, \quad [7]$$

where $A_{k \times (n-m)}$, $B_{k \times m}$ are random standard Gaussian i.i.d. matrices with entries in \mathcal{F} . Wachter (1) discovered that, as

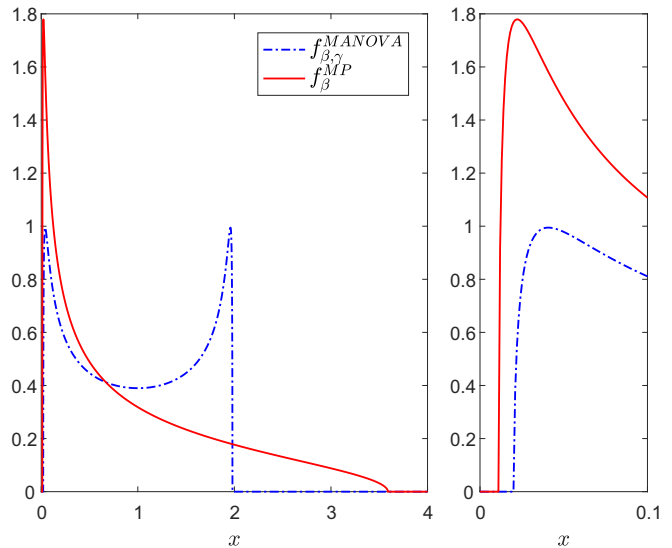


Fig. 1. Limiting MANOVA ($\beta = 0.8, \gamma = 0.5$) and Marčenko–Pastur ($\beta = 0.8$) density functions. (Left) Density on the interval $x \in [0, 4]$. (Right) Zoom in on the interval $x \in [0, 0.1]$.

$k/m \rightarrow \beta \leq 1$ and $m/n \rightarrow \gamma$, the empirical spectral distribution of the MANOVA(n, m, k, \mathbb{R}) ensemble converges, almost surely in distribution, to the so-called MANOVA(β, γ) limiting spectral distribution,[†] with density that is given by

$$f_{\beta, \gamma}^{MANOVA}(x) = \frac{\sqrt{(x - r_-)(r_+ - x)}}{2\beta\pi x(1 - \gamma x)} \cdot I_{(r_-, r_+)}(x) + \left(1 + \frac{1}{\beta} - \frac{1}{\beta\gamma}\right)^+ \cdot \delta\left(x - \frac{1}{\gamma}\right), \quad [8]$$

where $(x)^+ = \max(0, x)$. The limiting MANOVA distribution is compactly supported on $[r_-, r_+]$ with

$$r_{\pm} = \left(\sqrt{\beta(1 - \gamma)} \pm \sqrt{1 - \beta\gamma}\right)^2. \quad [9]$$

The same holds for the MANOVA(n, m, k, \mathbb{C}) ensemble.

Note that the support of the MANOVA(β, γ) distribution is smaller than that of the corresponding Marčenko–Pastur law for the same aspect ratios. Fig. 1 shows these two densities for $\beta = 0.8$ and $\gamma = 0.5$. Nevertheless, as the MANOVA dimension ratio becomes small, its distribution tends to the Marčenko–Pastur distribution (Eq. 6) [i.e., $f_{\beta, \gamma}^{MANOVA}(x) \rightarrow f_{\beta}^{MP}(x)$ as $\gamma \rightarrow 0$]. Thus, a highly redundant random Fourier frame behaves like a Gaussian i.i.d. frame.

3. Unitary Haar Frame. Let $X_{haar}^{(n)}$ consist of the first m columns of a Haar-distributed n -by- n unitary matrix normalized by $\sqrt{n/m}$ (the Haar distribution being the uniform distribution over the group of n -by- n unitary matrices). Edelman and Sutton (4) proved that the empirical spectral distribution of $\lambda(G_K)$ also converges, almost surely in distribution, to the MANOVA limiting spectral distribution (refs. 1 and 3, closing remarks).

The maximal and minimal eigenvalues of a matrix from the MANOVA(n, m, k, \mathcal{F}) ensemble ($\mathcal{F} \in \{\mathbb{R}, \mathbb{C}\}$) are known to

*Also known as the beta-Jacobi ensemble with $\beta = 1$ (orthogonal) for $\mathcal{F} = \mathbb{R}$ and $\beta = 2$ (unitary) for $\mathcal{F} = \mathbb{C}$.

[†]The literature uses the term MANOVA to refer to both the random matrix ensemble, which we denote here by MANOVA(n, m, k, \mathcal{F}), and the limiting spectral distribution, which we denote here by MANOVA(β, γ).

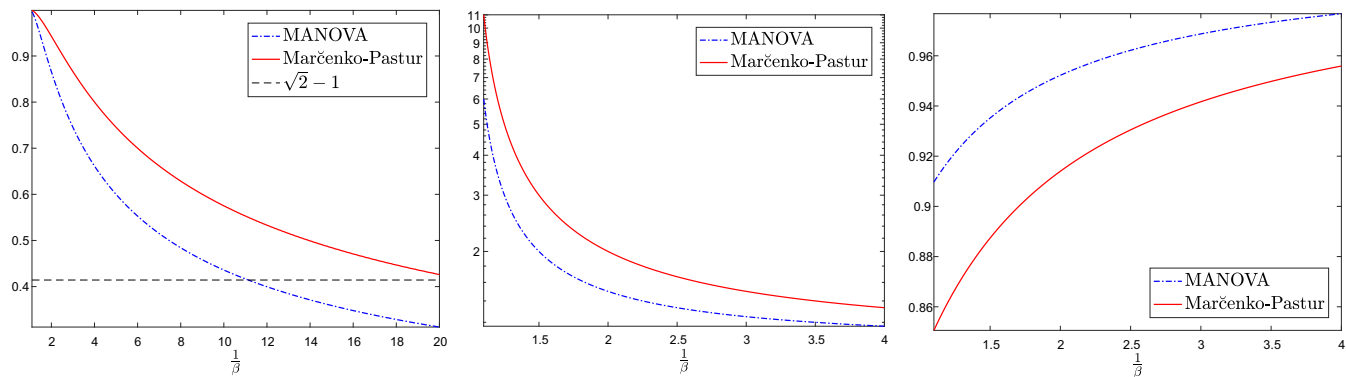


Fig. 2. Comparison of limiting values of $\mathbb{E}_K \Psi(\lambda(G_K))$ for the three functions Ψ discussed in *Motivation* between the Marčenko–Pastur limiting distribution and the MANOVA distribution. (Left) Ψ_{RIP} (lower is better). (Center) Ψ_{AC} (lower is better). (Right) $\Psi_{Shannon}$ (higher is better).

converge almost surely to r_+ and r_- , respectively (20). Although we are not aware of any parallel results for the random Fourier and Haar frames, the empirical evidence in this paper shows that it must be the case.

These random matrix phenomena have practical significance for evaluations of functions of the form $\Psi(\lambda(G_K))$, such as those mentioned above. The functions Ψ_{AC} and $\Psi_{Shannon}$, for example, are called linear spectral statistics in ref. 21, namely functions of $\lambda(G_K)$ that may be written as an integral of a scalar function against the empirical measure of $\lambda(G_K)$. Convergence of the empirical distribution of $\lambda(G_K^{(n)})$ to the limiting MANOVA distribution with density $f_{\beta,\gamma}^{MANOVA}$ implies

$$\lim_{n \rightarrow \infty} \Psi_{AC}(\lambda(G_K^{(n)})) = \int \frac{1}{x} f_{\beta,\gamma}^{MANOVA}(x) dx \quad [10]$$

$$\lim_{n \rightarrow \infty} \Psi_{Shannon}(\lambda(G_K^{(n)})) = \int \log(1 + \alpha x) f_{\beta,\gamma}^{MANOVA}(x) dx$$

for both the random Fourier and Haar frames; the integrals on the right-hand side may be evaluated explicitly. Similarly, convergence of $\lambda_{max}(G_K)$ and $\lambda_{min}(G_K)$ to r_+ and r_- implies, for example, that

$$\lim_{n \rightarrow \infty} \Psi_{RIP}(\lambda(G_K^{(n)})) = \max(r_+ - 1, 1 - r_-). \quad [11]$$

To show why such calculations are significant, we note that Eqs. 10 and 11 immediately allow us to compare the Gaussian i.i.d. frame with the random Fourier and Haar frames in terms of their limiting value of functions of interest. Fig. 2 compares the limiting value of Ψ_{RIP} , Ψ_{AC} , and $\Psi_{Shannon}$ over varying values of $\beta = \lim_{n \rightarrow \infty} k/m$. The plots clearly show that frames with a typical k submatrix that exhibits a MANOVA spectrum are superior to frames with a typical k submatrix that exhibits a Marčenko–Pastur spectrum across the performance measures.

Deterministic Frames: Universality Hypothesis

Deterministic frames, namely frames with design that involves no randomness, have so far eluded this kind of asymptotically exact analysis. Although there are results regarding RIP (22, 23) and statistical RIP (10, 24, 25), for example, of deterministic frame designs, they are mostly focused on highly redundant frames ($\gamma \rightarrow 0$) and the wide submatrix ($\beta \rightarrow 0$) case, where the spectrum tends to the Marčenko–Pastur distribution. Furthermore, nothing analogous, say, to the precise comparisons of Fig. 2 exists in the literature to the best of our knowledge. Specifically, no results

analogous to Eqs. 10 and 11 are known for deterministic frames, let alone the associated convergence rates, if any.

To subject deterministic frames to an asymptotic analysis, we shift our focus from a single frame X to a family of deterministic frames $\{X^{(n)}\}$ created by a common construction. The frame matrix $X^{(n)}$ is n -by- m . Each frame family determines allowable subsequences (n, m) ; to simplify notation, we leave the subsequence implicit and index the frame sequence simply by n . The frame family also determines the aspect ratio limit $\gamma = \lim_{n \rightarrow \infty} m/n$. In what follows, we also fix a sequence k with $\beta = \lim_{n \rightarrow \infty} k/m$ and let $K \subset [n]$ denote a uniformly distributed random subset.

Frames Under Study. The different frames that we studied are listed in Table 1 in a manner inspired by ref. 26. In addition to our deterministic frames of interest (the set \mathcal{X}), Table 1 also contains two examples of random frames (real and complex variants for each) for validation and convergence analysis purposes.

Functionals Under Study. We studied the functionals Ψ_{STRIP} from Eq. 3, Ψ_{AC} from Eq. 4, and $\Psi_{Shannon}$ from Eq. 5. In addition, we studied the maximal and minimal eigenvalues of G_K and its condition number:

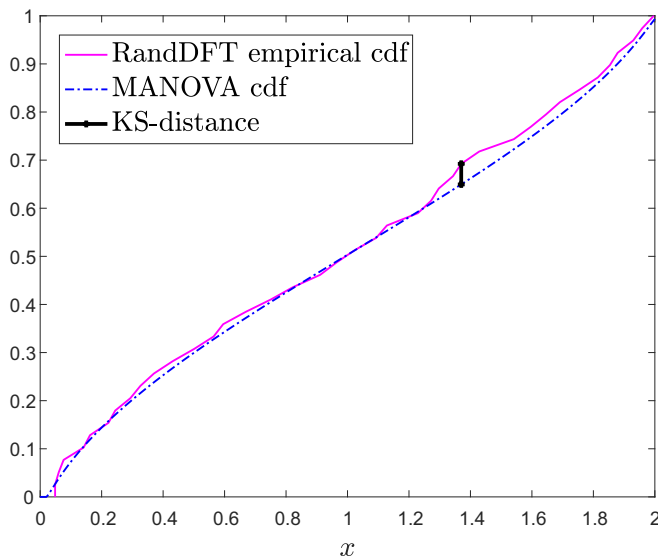


Fig. 3. KS distance of random DFT subframe: $\beta = 0.8$, $\gamma = 0.5$, and $n = 100$.

$$\begin{aligned}\Psi_{max}(\lambda(G_K)) &= \lambda_{max}(G_K) \\ \Psi_{min}(\lambda(G_K)) &= \lambda_{min}(G_K) \\ \Psi_{cond}(\lambda(G_K)) &= \frac{\lambda_{max}(G_K)}{\lambda_{min}(G_K)}.\end{aligned}$$

Measuring the Rate of Convergence. To quantify the rate of convergence of the entire spectrum of the k -by- m matrix X_K , which is a k submatrix of an n -by- m frame matrix X , to a limiting distribution, we let $F[X_K]$ denote the empirical cumulative distribution function (CDF) of $\lambda(G_K)$ and $F_{\beta,\gamma}^{MANOVA}(x) = \int_{r_-}^x f_{\beta,\gamma}^{MANOVA}(z) dz$ denote the CDF of the MANOVA(β, γ) limiting distribution. The quantity

$$\Delta_{KS}(X_K) = \left\| F[X_K] - F_{\beta_n, \gamma_n}^{MANOVA} \right\|_{KS},$$

where $\|\cdot\|_{KS}$ is the Kolmogorov–Smirnov (KS) distance between CDFs, measures the distance to the hypothesized limit. Here, $\beta_n = k/m$ and $\gamma_n = m/n$ are the actual aspect ratios for the matrix X_K at hand. As a baseline, we use $\Delta_{KS}(Y_{n,m,k,\mathcal{F}})$, where $Y_{n,m,k,\mathcal{F}}$ is a matrix from the MANOVA(n, m, k, \mathcal{F}) ensemble, with $\mathcal{F} = \mathbb{R}$ if X_K is real and $\mathcal{F} = \mathbb{C}$ if X_K is complex. Fig. 3 illustrates the KS distance between an empirical CDF and the limiting MANOVA CDF.

Similarly, to quantify the rate of convergence of a functional Ψ , the quantity

$$\Delta_{\Psi}(X_K; n, m, k) = |\Psi(\lambda(G_K)) - \Psi(f_{\beta_n, \gamma_n}^{MANOVA})|$$

is the distance between the measured value of Ψ on a given k submatrix X_K and its hypothesized limiting value. For a baseline, we can use $\Delta_{\Psi}(Y_{n,m,k,\mathcal{F}})$, with $\mathcal{F} = \mathbb{R}$ if X_K is real and $\mathcal{F} = \mathbb{C}$ if X_K is complex. For linear spectral functionals, like Ψ_{AC} and $\Psi_{Shannon}$, which may be written as $\Psi(\lambda(G_K)) = \int \psi dF[X_K]$ for some kernel ψ , we have $\Psi(f_{\beta,\gamma}^{MANOVA}) = \int \psi dF_{\beta,\gamma}^{MANOVA}$. For Ψ_{RIP} that depends on $\lambda_{max}(G_K)$ and $\lambda_{min}(G_K)$, we have $\Psi_{RIP}(f_{\beta,\gamma}^{MANOVA}) = \max\{r_+ - 1, 1 - r_-\}$.

Universality Hypothesis. The contributions of this paper are based on the following assertions on the typical k -submatrix ensemble X_K corresponding to a frame family $X^{(n)}$. This family may be random or deterministic, real or complex.

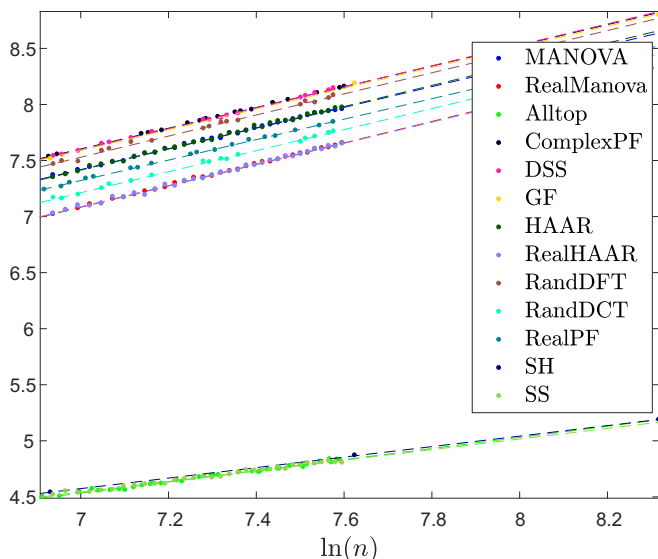


Fig. 4. Test 1 for $\gamma = 0.5$ and $\beta = 0.8$. The plot shows $-1/2 \ln \text{Var}_K(\Delta_{KS}(X_K^{(n)}))$ over $\ln(n)$.

Table 2. Results of test 1 for $\gamma = 0.5$ and $\beta = 0.8$

Frame	R^2	\hat{b}	$SE(\hat{b})$	P value $b = b_{MANOVA}$
MANOVA	0.99828	0.92505	0.00690	1
DSS	0.99858	0.93652	0.00911	0.32089
GF	0.99921	0.92474	0.02608	0.99082
ComplexPF	0.99950	0.92454	0.00535	0.95390
Alltop	0.98906	0.49660	0.00883	9.4651e-47
SS	0.98767	0.47354	0.00950	5.8136e-45
HAAR	0.99736	0.94421	0.00873	0.09019
RandDFT	0.99544	0.94127	0.01644	0.36788
RealMANOVA	0.99873	0.95610	0.00613	1
RealPF	0.99871	0.91244	0.00821	9.7174e-05
SH	0.99989	0.46822	0.00492	6.3109e-35
RealHAAR	0.99596	0.94456	0.01081	0.35675
RandDCT	0.99773	0.93859	0.01156	0.18737

H1. Existence of a Limiting Spectral Distribution. The empirical spectral distribution of $X_K^{(n)}$, namely the distribution of $\lambda(G_K^{(n)})$, converges, as $n \rightarrow \infty$, to a compactly supported limiting distribution; furthermore, $\lambda_{max}(G_K^{(n)})$ and $\lambda_{min}(G_K^{(n)})$ converge to the edges of that compact support.

H2. Universality of the Limiting Spectral Distribution. The limiting spectral distribution of $X_K^{(n)}$ is the MANOVA(β, γ) distribution (1) with density that is shown in Eq. 8. Also, $\lambda_{max}(G_K^{(n)}) \rightarrow r_+$ and $\lambda_{min}(G_K^{(n)}) \rightarrow r_-$, where r_{\pm} is given by Eq. 9.

H3. Exact Power Law Rate of Convergence for the Entire Spectrum. The spectrum of $X_K^{(n)}$ converges to the limiting MANOVA(β, γ) distribution

$$\left(\mathbb{E}_{K_n} \left(\Delta_{KS}(X_K^{(n)}) \right) \right)^2 \searrow 0,$$

and in fact, its fluctuations are given by the law

$$\text{Var}_K(\Delta_{KS}(X_K^{(n)})) = Cn^{-2b} \quad [12]$$

for some constants C, b , which may depend on the frame family.

H4. Universality of the Rate of Convergence for the Entire Spectrum of ETFs. For an ETF family, the exponent b in Eq. 12 is universal and does not depend on the frame. Furthermore, Eq. 12 also holds, with the same universal exponent, replacing $G_K^{(n)}$ with a same-sized matrix from the MANOVA(n, m, k, \mathcal{F}) distribution defined in Eq. 7 with $\mathcal{F} = \mathbb{R}$ if $X^{(n)}$ is a real frame family and $\mathcal{F} = \mathbb{C}$ if $X^{(n)}$ is a complex frame family. In other words, the universal exponent b for ETFs is a property of the MANOVA (Jacobi) random matrix ensemble.

H5. Exact Power Law Rate of Convergence for Functionals. For a “nice” functional Ψ , the value of $\Psi(\lambda(G_K^{(n)}))$ converges to $\Psi(f_{\beta,\gamma}^{MANOVA})$ according to the law

$$\mathbb{E}_K(\Delta_{\Psi}(X_K^{(n)}))^2 = Cn^{-b} \log^{-a}(n) \quad [13]$$

for some constants C, b, a .

H6. Universality of the Rate of Convergence for Functionals. Although the constant C in Eq. 13 may depend on the frame, the exponents a, b are universal. Eq. 13 also holds, with the same universal exponents, replacing $G_K^{(n)}$ with a same-sized matrix from the MANOVA(n, m, k, \mathcal{F}) ensemble defined in Eq. 7, with

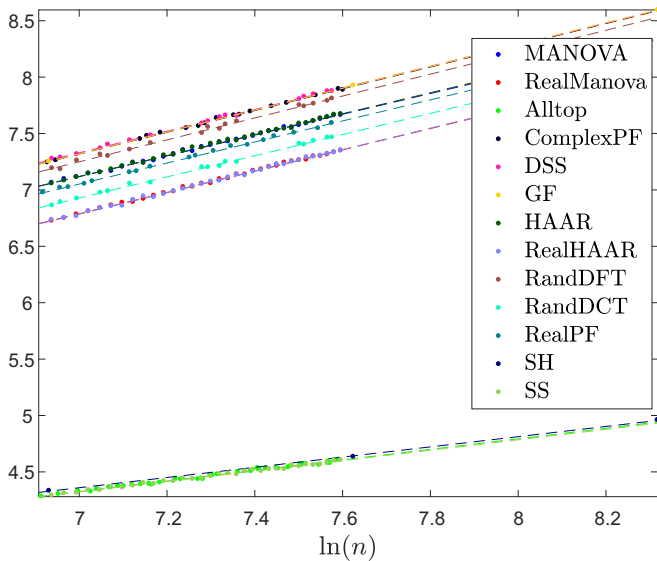


Fig. 5. Test 1 for $\gamma = 0.5$ and $\beta = 0.6$. The plot shows $-1/2 \ln \text{Var}_K(\Delta_{KS}(X_K^{(n)}))$ over $\ln(n)$.

$\mathcal{F} = \mathbb{R}$ if $X^{(n)}$ is a real frame family and $\mathcal{F} = \mathbb{C}$ if $X^{(n)}$ is a complex frame family. In other words, the universal exponents a, b are a property of the MANOVA (Jacobi) random matrix ensemble.

Nonstandard Aspect Ratio $\beta > 1$. Although the classical MANOVA ensemble and limiting density are not defined for $\beta > 1$, in our case, it is certainly possible to sample $k > m$ vectors from the n possible frame vectors, resulting in a situation with $\beta > 1$. In this situation, the hypotheses above require slight modifications. Specifically, the limiting spectral distribution of $X_K^{(n)}$ for $\beta > 1$ is

$$\left(1 - \frac{1}{\beta}\right) \delta(x) + f_{\beta, \gamma}^{\text{MANOVA}}(x), \quad [14]$$

where $f_{\beta, \gamma}^{\text{MANOVA}}(x)$ is the function (no longer a density) defined in Eq. 8. The rate of convergence of the distribution of non-zero eigenvalues to the limiting density $1/\beta f_{1/\beta, \beta\gamma}^{\text{MANOVA}}(1/\beta x) = \beta f_{\beta, \gamma}^{\text{MANOVA}}(x)$ is compared with the baseline $\beta \cdot Y_{n, k, m, \mathcal{F}}$, where $Y_{n, k, m, \mathcal{F}}$ is a matrix from the MANOVA(n, k, m, \mathcal{F}) ensemble (i.e., with reversed order of k and m).

Methods

The software that we developed has been permanently deposited in the data and code supplement (<https://purl.stanford.edu/qg138qm8653>). Because many of the deterministic frames under study are only defined for $\gamma = 0.5$, we primarily studied the aspect ratios ($\gamma = 0.5, \beta$) with $\beta \in \{0.3, 0.5, 0.6, 0.7, 0.8, 0.9\}$. In addition, we inspected all frames under study that are defined for the aspect ratios ($\gamma = 0.25, \beta = 0.6$) and ($\gamma = 0.25, \beta = 0.8$) (all random frames as well as DSS and Alltop). We also studied nonstandard aspect ratios $\beta > 1$ as described in SI Appendix (<https://purl.stanford.edu/qg138qm8653>). For deterministic frames, n took allowed values in the ranges (240, 2,000), ($2^5, 2^{12}$) for Grassmannian and SH frames, and (600, 4,000) for DSS frame with $\gamma = 0.25$. For random frames and MANOVA ensemble, we used dense grid of values in the range (240, 2,000). Hypothesis testing as discussed below was based on a subset of these values, where $n \geq 1,000$. For each of the frame families under study and each value of β and γ under study, we selected a sequence (n, m, k). The values n and m were selected so that m/n will be as close as possible to γ ; however, because of different aspect ratio constraints by the different frames, occasionally, we had m/n close but not equal to γ . We then determined k , such that k/m will be as close as possible to β . For each n , we generated a single n -by- m frame matrix $X^{(n)}$. We then produced T independent samples from the uniform distribution on k_n subsets, $K[1], \dots, K[T] \subset [n]$, and generated their corresponding k submatrices $X_{K[i]}^{(n)}$ ($1 \leq i \leq T$). Importantly, all of these are submatrices of the same frame

matrix $X^{(n)}$. We calculated $\overline{\Delta_{KS}^{\text{Var}}}(X_K^{(n)}) = \overline{\Delta_{KS}^2}(X_K^{(n)}) - \overline{\Delta_{KS}}^2(X_K^{(n)})$, the empirical variance of $\Delta_{KS}(X_{K[i]}^{(n)})$, and $\overline{\Delta_{KS}^2}(X_K^{(n)})$, the average value of $\Delta_{KS}^2(X_{K[i]}^{(n)})$ on $1 \leq i \leq T$ as a Monte Carlo approximation to the left-hand side of Eq. 12, variance and MSE (mean square error), respectively. For each of the functionals under study, we also calculated $\overline{\Delta_{\Psi}^2}(X_{K[i]}^{(n)})$, the average value of $\Delta_{\Psi}^2(X_{K[i]}^{(n)})$ on $1 \leq i \leq T$, as a Monte Carlo approximation to the left-hand side of Eq. 13.

Separately, for each triplet (n, m, k) and $\mathcal{F} \in \{\mathbb{R}, \mathbb{C}\}$, we have performed T independent draws from the MANOVA(n, m, k, \mathcal{F}) ensembles [7] and calculated analogous quantities $\overline{\Delta_{KS}^{\text{Var}}}(Y_{n, m, k, \mathcal{F}})$, $\overline{\Delta_{KS}^2}(Y_{n, m, k, \mathcal{F}})$, and $\overline{\Delta_{\Psi}^2}(Y_{n, m, k, \mathcal{F}})$.

Test 1: Testing H1–H4. For each of the frames under study and each value of (β, γ) , we computed the KS distance for $T = 10^4$ submatrices and performed simple linear regression of $-1/2 \log(\overline{\Delta_{KS}^{\text{Var}}}(X_K^{(n)}))$ on $\log(n)$ with an intercept. We obtained the estimated linear coefficient \hat{b} as an estimate for the exponent b and its SE $\sigma(\hat{b})$. Similarly, we regressed $-1/2 \log(\overline{\Delta_{KS}^{\text{Var}}}(Y_{n, m, k, \mathcal{F}}))$ on $\log(n)$ to obtain \hat{b}_{MANOVA} and $\sigma(\hat{b}_{\text{MANOVA}})$. We performed Student's t test to test the null hypotheses $b = b_{\text{MANOVA}}$ using the test statistic

$$t = \frac{\hat{b} - b_{\text{MANOVA}}}{\sqrt{\sigma(\hat{b})^2 + \sigma(\hat{b}_{\text{MANOVA}})^2}}$$

Under the null hypothesis, the test statistic is distributed $t_{(N + N_{\text{MANOVA}} - 4)}$, where N and N_{MANOVA} are the numbers of different values of n for which we have collected the data for a frame and the MANOVA ensemble, respectively. We report the R^2 of the linear fit, the slope coefficient \hat{b} and its SE, and the P value of the above t test. We next regressed $-\log(\overline{\Delta_{KS}^2})$ on $\log(n)$. Because $\overline{\Delta_{KS}^2} = (\overline{\Delta_{KS}})^2 + \overline{\Delta_{KS}^{\text{Var}}}$, a linear fit verifies that $(\overline{\Delta_{KS}})^2 \searrow 0$.

Test 2: Testing H5–H6. For each of the frames under study, each of the functionals Ψ under study, and each value of (β, γ) , we computed the empirical value of the functionals on $T = 10^3$ submatrices. We first performed linear regression of $-\log(\overline{\Delta_{\Psi}^2}(Y_{n, m, k, \mathcal{F}}))$ on $\log(n)$ and $\log(\log(n))$ with an intercept for $\mathcal{F} \in \{\mathbb{R}, \mathbb{C}\}$. Let a_0 denote the fitted coefficient for $\log(n)$, and let b_0 denote the fitted coefficient for $\log(\log(n))$. This step was based on triplets (n, m, k), yielding accurate aspect ratios in the range $240 \leq n \leq 2,000$. We then performed simple linear regression of $-\log(\overline{\Delta_{\Psi}^2}(X_K^{(n)}; n, m, k))$ on $\log(n) + (a_0/b_0) \cdot \log(\log(n))$. The estimated linear regression coefficient \hat{b} is the estimate for the exponent b in Eq. 13, and $\sigma(\hat{b})$ is its SE. We used $\hat{b} \cdot (a_0/b_0)$ as an estimate for the exponent a in Eq. 13. We proceeded as above to test the null hypothesis $b = b_0$. We report

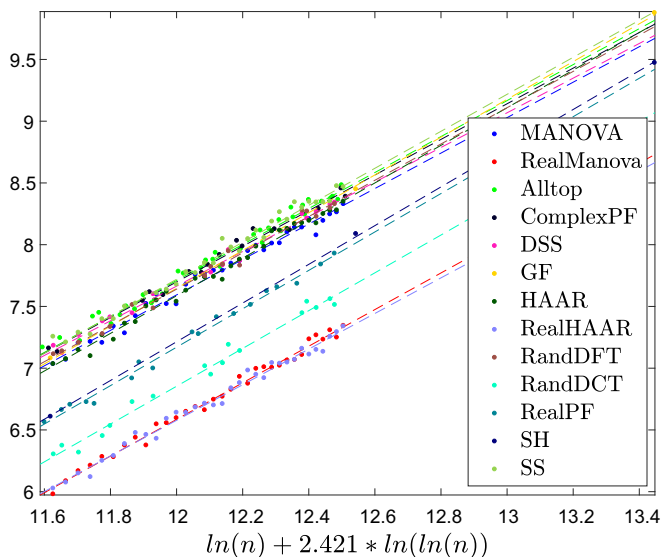


Fig. 6. Test 2 for Ψ_{AC} , $\gamma = 0.5$, and $\beta = 0.8$. The plot shows $-\ln \mathbb{E}_K(\Delta_{\Psi}(X_K^{(n)})^2)$.

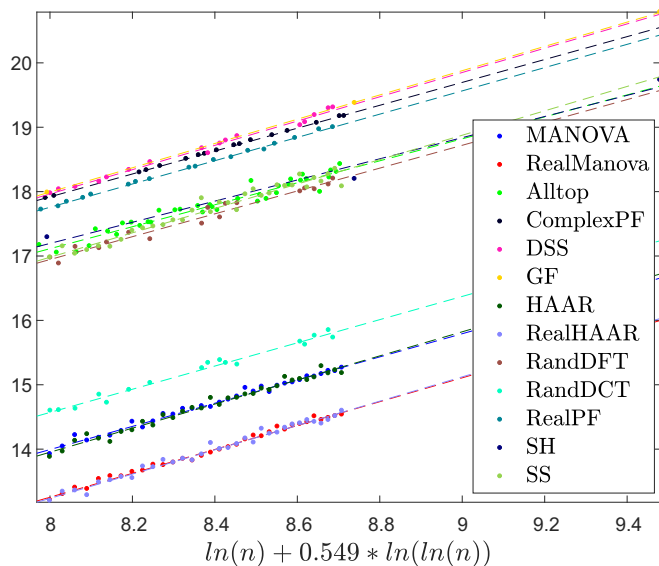


Fig. 7. Test 2 for $\Psi_{Shannon}$, $\gamma=0.5$, and $\beta=0.8$. The plot shows $-\ln \mathbb{E}_K(\Delta_\Psi(X_K^{(n)}))^2$.

the R^2 of the linear fit, the slope coefficient \hat{b} and its SE, and the P value of the test above.

Computing. To allow the number of Monte Carlo samples to be as large as $T=10^4$ and n to be as large as 2,000, we used a large Matlab cluster running on Amazon Web Services. We used 32 logical core machines with 240 GB RAM (random access memory) each, which were running several hundred hours in total. The code that we executed has been deposited (<https://purl.stanford.edu/qg138qm8653>); it may easily be executed for smaller values of T and n on smaller machines.

Results

The raw results obtained in our experiments as well as the analysis results of each experiment have been deposited with their generating code (<https://purl.stanford.edu/qg138qm8653>).

For space considerations, the full documentation of our results is deferred to *SI Appendix* (<https://purl.stanford.edu/qg138qm8653>). To offer a few examples, Fig. 4 and Table 2 show the linear fit to $\overline{\Delta}_{KS}^{Var}$ for $(\gamma=0.5, \beta=0.8)$. Fig. 5 shows the linear fit to $\overline{\Delta}_{KS}^{Var}$ for a different value of β , namely $(\gamma=0.5, \beta=0.6)$. Fig. 6 shows the linear fit to $\overline{\Delta}_{AC}$ for $(\gamma=0.5, \beta=0.8)$. Fig. 7 and Table 3 show the linear fit to $\Delta_\Psi^{Shannon}$ for $(\gamma=0.5, \beta=0.8)$. Similar figures and tables for the other values (γ, β) , in particular, $(\beta=0.3, \gamma=0.5)$, $(\beta=0.5, \gamma=0.5)$, $(\beta=0.7, \gamma=0.5)$, $(\beta=0.9, \gamma=0.5)$, $(\beta=0.6, \gamma=0.25)$, and $(\beta=0.8, \gamma=0.25)$, are deferred to *SI Appendix*. Note that, in all coefficient tables, both those shown here and those deferred to *SI Appendix*, the upper boxes show complex frames (with t test comparison with the complex MANOVA ensemble of the same size denoted “MANOVA”), and the bottom boxes show real frames (with t test comparison with the real MANOVA ensemble of the same size denoted “RealMANOVA”). In each box, the top rows are deterministic frames, and the bottom rows are random frames. Furthermore, note that, in plots for test 2, the horizontal axis is slightly different for real and complex frames, because the preliminary step described above was performed separately for real and complex frames. In the interest of space, we plot all frames over the horizontal axis calculated for complex frames.

Validation on Random Frames. Although our primary interest was in deterministic frames, we included in the frames under study random frames. For the complex Haar frame and ran-

dom Fourier frame, convergence of the empirical CDF of the spectrum to the limiting $\text{MANOVA}(\beta, \gamma)$ distribution has been proved in refs. 3 and 4. To our surprise, not only was our framework validated on the four random frames under study, in the sense of asymptotic empirical spectral distribution, but also, all universality hypotheses H1–H6 were accepted (not rejected at the 0.001 significance level, with very few exceptions).

Test Results on Deterministic Frames. A tabular summary of our results, per hypothesis and per frame under study, is included for convenience in *SI Appendix*. Universality hypotheses H1–H3 were accepted on all deterministic frames. For H1 and H2, convergence of the empirical spectral distribution to the $\text{MANOVA}(\beta, \gamma)$ limit has been observed in all cases. For H3, the linear fit in all cases was excellent with $R^2 > 0.99$ without exception, confirming the power law in Eq. 12 and the polynomial decrease of $\overline{\Delta}_{KS}^2$ with n . Universality hypothesis H4 was accepted (not rejected) for deterministic ETFs at the 0.001 significance level, with few exceptions (Table 2; full results and a summary table are in *SI Appendix*); it was rejected for deterministic non-ETFs. For $\gamma=0.25$, hypothesis H4 has also been accepted for the Alltop frame (*SI Appendix*). Universality hypothesis H5 was accepted for all deterministic frames, with excellent linear fits ($R^2 > 0.97$ without exception), confirming the power law in Eq. 13. Universality hypothesis H6 was accepted (not rejected) at the 0.001 significance level (and even 0.05 with few exceptions) for all deterministic frames. For the reader’s convenience, Table 4 summarizes the universal exponents for convergence of the entire spectrum (H4) and the universal exponents for convergence of the functionals under study (H6) for $(\beta, \gamma) = (0.8, 0.5)$. The framework developed in this paper readily allows tabulation of these universal exponents for any value of (β, γ) . We have observed that the universal exponents are slightly sensitive to the random seed. However, exact evaluation of this variability requires very significant computational resources and is beyond our scope. Similarly, some sensitivity of the P values to random seed has been observed.

Reproducibility Advisory. All of the figures and tables in this paper, including those in *SI Appendix*, are fully reproducible from our raw results and code deposited in the data and code supplement (<https://purl.stanford.edu/qg138qm8653>).

Discussion

The Hypotheses. Our universality hypotheses may be surprising in several aspects. First, the frames examined were designed

Table 3. Results of test 2 for $\Psi_{Shannon}$, $\gamma=0.5$, and $\beta=0.8$

Frame	R^2	\hat{b}	$SE(\hat{b})$	P value $b = b_{MANOVA}$
MANOVA	0.98721	1.79936	0.03678	1
DSS	0.99110	1.88674	0.04615	0.14551
GF	0.99997	1.88548	0.01073	0.03161
ComplexPF	0.99977	1.77783	0.00701	0.56808
Alltop	0.93841	1.70618	0.07388	0.26297
SS	0.95539	1.89501	0.07355	0.24922
HAAR	0.97971	1.87082	0.04836	0.24400
RandDFT	0.96928	1.77454	0.08157	0.78270
RealMANOVA	0.99202	2.05451	0.03309	1
RealPF	0.99834	2.00345	0.02045	0.19576
SH	0.97850	1.81297	0.26874	0.37904
RealHAAR	0.98287	2.09078	0.04958	0.54503
RandDCT	0.98364	1.99663	0.06648	0.43977

Table 4. Summary of universal exponents for convergence: $\gamma = 0.5$, $\beta = 0.8$, and $(\Psi_S = \Psi_{Shannon})$

Frame	b_{spectrum}	$b_{\Psi_{\text{RIP}}}$	$a_{\Psi_{\text{RIP}}}$	$b_{\Psi_{\text{AC}}}$	$a_{\Psi_{\text{AC}}}$	b_{Ψ_S}	a_{Ψ_S}	$b_{\Psi_{\text{max}}}$	$a_{\Psi_{\text{max}}}$	$b_{\Psi_{\text{min}}}$	$a_{\Psi_{\text{min}}}$	$b_{\Psi_{\text{cond}}}$	$a_{\Psi_{\text{cond}}}$
MANOVA	0.93	1.15	2.21	1.44	3.48	1.80	0.99	1.13	2.48	1.00	3.09	1.87	-4.55
DSS	0.94	1.14	2.18	1.40	3.40	1.89	1.04	1.10	2.41	1.00	3.11	1.87	-4.56
GF	0.92	1.17	2.23	1.53	3.70	1.89	1.03	1.13	2.48	1.04	3.22	1.95	-4.76
ComplexPF	0.92	1.13	2.17	1.44	3.49	1.78	0.98	1.10	2.41	1.00	3.12	1.87	-4.56
Alltop	0.50	1.14	2.18	1.46	3.53	1.71	0.94	1.11	2.42	1.01	3.13	1.86	-4.54
SS	0.47	1.11	2.13	1.50	3.63	1.90	1.04	1.08	2.36	0.98	3.06	1.83	-4.47
HAAR	0.94	1.10	2.11	1.52	3.69	1.87	1.03	1.09	2.37	1.01	3.13	1.88	-4.59
RandDFT	0.94	1.21	2.32	1.47	3.56	1.77	0.97	1.11	2.42	1.03	3.18	1.93	-4.70
RealMANOVA	0.96	0.87	3.58	1.26	5.21	1.27	5.26	0.90	3.73	0.87	3.58	0.77	3.17
RealPF	0.91	0.92	3.82	1.32	5.46	1.24	5.12	0.94	3.88	0.94	3.88	0.81	3.36
SH	0.47	0.93	3.82	1.34	5.53	1.14	4.71	0.93	3.82	0.93	3.82	0.85	3.51
RealHAAR	0.94	0.86	3.54	1.23	5.07	1.29	5.35	0.89	3.68	0.90	3.73	0.79	3.28
RandDCT	0.94	0.99	4.08	1.30	5.38	1.24	5.10	0.94	3.89	0.95	3.93	0.82	3.40

to minimize frame bounds and worse case pairwise correlations. Still, it seems that they perform well when the performance criterion is based on spectrum of the typical selection of k frame vectors. Second, under the universality hypotheses, all of these deterministic frames perform exactly as well as random frame designs, such as the random Fourier frame. Inasmuch as frames are continuous codes, we find deterministic codes matching the performance of random codes. Third, the hypotheses suggest an extremely broad universality property: many different ensembles of random matrices asymptotically exhibit the limiting MANOVA spectrum.

All of the deterministic frames under study satisfy the universality hypotheses (with hypothesis H4 satisfied only for ETFs). This finding should not give the impression that any deterministic frame satisfies these hypotheses! First, the empirical measures of an arbitrary sequence of frames rarely converge (thus violating hypothesis H1). Second, even if they converge, a too simplistic frame design often leads to concentration of the lower edge of the empirical spectrum near zero, resulting in a non-MANOVA spectrum and poor performance. For example, if the frame is sparse, say, consisting of some m columns of the n -by- n identity matrix, then a fraction $(n - m)/n$ of the singular values of a typical submatrix is exactly zero.

The frames under study are all ETFs or near-ETFs, all with favorable frame properties. To make this point, we have included in *SI Appendix* (<https://purl.stanford.edu/qg138qm8653>) study of a low-pass frame, in which the Fourier frequencies included in the frame are the lowest ones. This construction is in contrast with the clever choice of frequencies leading to the DSS frame. Indeed, the low-pass frame does not have appealing frame properties. It is quite obvious from the results in *SI Appendix* as well as the results regarding the closely related random Vandermonde ensemble (27) that such frames do not satisfy any of the universality hypotheses H2–H6.

We note that convergence rates of the form Eqs. 12 and 13 are known for other classical random matrix ensembles (28–31).

We further note that hypotheses H1–H4 do not imply hypotheses H5 and H6. Even if the empirical CDF converges in the KS metric to the limiting MANOVA(β, γ) distribution, functionals that are not continuous in the KS metric do not necessarily converge, and moreover, no uniform rate of convergence is a priori implied.

Our Contributions. This paper presents a simple method for approximate computation (with known and good approximation error) of spectral functionals of k -submatrix ensemble for a variety of random and deterministic frames using Eq. 1. Our results make it possible to tabulate these approximate values, creating a useful resource for scientists. As an example, we include Table 5, a lookup table for the value of the functional Ψ_{AC} on the DSS deterministic frame family, listing by values of n and k the asymptotic (approximate) value calculated analytically from the limiting $f_{\beta, \gamma}^{\text{MANOVA}}$ distribution, and the standard approximation error.

To this end, we developed a systematic empirical framework, which allows validation of Eq. 1 and discovery of the exponents there. Our work is fully reproducible, and our framework is available (along with the rest of our results and code) in the code and data supplement (<https://purl.stanford.edu/qg138qm8653>). In addition, our results provide overwhelming empirical evidence for a number of phenomena.

- The typical k -submatrix ensemble of deterministic frames is an object of interest. Although there is absolutely no randomness involved in the submatrix X_K of a deterministic frame (other than the choice of subset K), the typical k submatrix seems to be an ensemble in its own right, with properties so far attributed only to random matrix ensembles—including a universal, compactly supported limiting spectral distribution and convergence of the maximal (minimal) singular value to the upper (lower) edges of the limiting distribution.
- MANOVA(β, γ) as a universal limiting spectral distribution. Wachter's MANOVA(β, γ) distribution is the limiting spectral distribution of $\lambda(G_K)$ as $k/m \rightarrow \beta$ and $m/n \rightarrow \gamma$ for the typical k -submatrix ensemble of deterministic frames (including difference set, Grassmannian, real Paley, complex Paley, quadratic chirp, SS, and SH). The same is true for real random frames—random cosine transform and random Haar.
- Convergence of the edge spectrum. For all of the deterministic frames above as well as the random frames (random cosine, random Fourier, complex Haar, and real Haar), the maximal and minimal eigenvalues of the k -typical submatrix ensemble converge to the support edges of the

Table 5. Root mean square error $\Psi(f_{\beta, \gamma}^{\text{MANOVA}}) \pm \sqrt{\Delta_{\Psi}(X_K^{(n)})^2}$ for Ψ_{AC} and DSS frame, $m = (n - 1)/2$ and $k = \beta \cdot m$

n	1,031	1,151	1,291	1,451	1,571	1,811	1,951
$\beta = 0.8$	3 ± 0.0281	3 ± 0.0253	3 ± 0.0227	3 ± 0.0204	3 ± 0.0189	3 ± 0.0166	3 ± 0.0155
$\beta = 0.6$	1.75 ± 0.0073	1.75 ± 0.0065	1.75 ± 0.0058	1.75 ± 0.0051	1.75 ± 0.0048	1.75 ± 0.0041	1.75 ± 0.0038

- MANOVA(β, γ) limiting distribution. The convergence follows a universal power law rate.
- iv) A definite power law rate of convergence for the entire spectrum of the MANOVA(n, m, k, \mathcal{F}) ensemble to its MANOVA(β, γ) limit, with different exponents in the real and complex cases.
 - v) Universality of the power law exponents for the entire spectrum. The complex deterministic ETFs (difference set, Grassmannian, and complex Paley) share the power law exponents with the MANOVA(n, m, k, \mathbb{C}) ensemble. The same is true for the complex random frames (random Fourier and complex Haar). The complex tight nonequiangular Alltop frame, which can be constructed for various aspect ratios, also shares the power law exponents with the MANOVA(n, m, k, \mathbb{C}) ensemble for $\gamma < 0.5$. The real deterministic ETF (real Paley) shares the exponent with the MANOVA(n, m, k, \mathbb{R}). The same is true for real random frames (random cosine and real Haar). All non-ETFs under study, with $\gamma = 0.5$, share different power law exponents (slower convergence).
 - vi) A definite power law rate of convergence for functionals, including Ψ_{StRIP} , Ψ_{AC} , and $\Psi_{Shannon}$.
 - vii) Universality of the power law exponents for functionals. For practically all frames under study, both random and deterministic, the power law exponents for functionals agree with those of the MANOVA(n, m, k, \mathbb{R}) (real frames) and MANOVA(n, m, k, \mathbb{C}) (complex frames).

Intercepts. Our results showed a surprising categorization of the deterministic and random frames under study according to the constant C in Eq. 12 or equivalently, according to the intercept (vertical shift) in the linear regression on $\log(n)$. Figs. 4 and 5 clearly show that the regression lines, while having identical slopes (as predicated by hypothesis H3), are grouped according to their intercepts into the following seven categories: complex MANOVA ensemble and complex Haar (Manova and HAAR); real MANOVA ensemble and real Haar (RealManova and RealHAAR); complex ETFs (DSS, GF, and ComplexPF); non-ETFs (SS, SH, and Alltop); real ETF (RealPF); complex random Fourier (RandDFT); and real random Fourier (RandDCT).

Interestingly, intercepts of all complex frames are larger (meaning that the linear coefficient C in Eq. 12 is smaller) than those of all real frames. Also, the less randomness exists in the frame, the higher the intercept: intercepts of deterministic ETFs are higher than those of random Fourier and random cosine, which are in turn, higher than those of Haar frames and the MANOVA ensembles.

Related Work. Farrell (3) has conjectured that the phenomenon of convergence of the spectrum of typical k submatrices to the limiting MANOVA distribution is indeed much broader and extends beyond the partial Fourier frame that he considered. A related empirical study was conducted by Monajemi et al. (26). In

it, the authors considered the so-called sparsity-undersampling phase transition in compressed sensing. This asymptotic quantity poses a performance criterion for frames that interacts with the typical k submatrix X_K in a manner possibly more complicated than the spectrum $\lambda(G_K)$. The authors investigated various deterministic frames, most of which are studied in this paper, and brought empirical evidence that the phase transition for each of these deterministic frames is identical to the phase transition of Gaussian frames. Gurevich and Hadani (24) proposed certain deterministic frame construction and effectively proved that the empirical spectral distribution of their typical k submatrix converges to a semicircle, assuming $k = m^{1-\epsilon}$, a scaling relation different from the one considered here. The work in refs. 32 and 33 also considered deterministic frame designs, chirp sensing codes, and binary linear codes, with a random sampling. In their design, the aspect ratios are large (e.g., in ref. 32, $m \sim k^2$ and $n \sim m^2$), and therefore, the spectrum converges to the Marčenko–Pastur distribution. Tropp (34) provided bounds for $\lambda_{max}(G_K)$ and $\lambda_{min}(G_K)$ when X is a general dictionary. Collins (35) has shown that the spectrum of a matrix model deriving from random projections has the same eigenvalue distribution of the MANOVA ensemble in finite n . Wachter (1) used a connection between the MANOVA ensemble and submatrices of Haar matrices to derive the asymptotic spectral distribution MANOVA(β, γ).

Conclusions

We have observed a surprising universality property for the k -submatrix ensemble corresponding to various well-known deterministic frames as well as well-known random frames. The MANOVA ensemble and the MANOVA limiting distribution emerge as key objects in the study of frames, both random and deterministic, in the context of sparse signals and erasure channels. We hope that our findings will invite rigorous mathematical study of these fascinating phenomena.

In any frame where our universality hypotheses hold (including all of the frames under study here), Fig. 2 correctly describes the limiting values of f_{RIP} , f_{AC} , and $f_{Shannon}$ and shows that codes based on deterministic frames (involving no randomness and allowing fast implementations) are better across performance measures than i.i.d. random codes.

The empirical framework that we proposed in this paper may be easily applied to new frame families $X^{(n)}$ and new functionals Ψ , extending our results further and mapping the frontiers of the universality property. In any frame family and for any functional where our universality hypotheses hold, we have proposed a simple, effective method for calculating quantities of the form $\mathbb{E}_K \Psi(\lambda(G_K))$ to known approximation, which improves polynomially with n .

ACKNOWLEDGMENTS. We thank the anonymous referees for their helpful comments. This work was partially supported by Israeli Science Foundation Grant 1523/16.

1. Wachter KW (1980) The limiting empirical measure of multiple discriminant ratios. *Ann Math Stat* 8:937–957.
2. Forrester PJ (2010) *Log-Gases and Random Matrices (LMS-34)* (Princeton Univ Press, Princeton).
3. Farrell B (2011) Limiting empirical singular value distribution of restrictions of discrete Fourier transform matrices. *J Fourier Anal Appl* 17:733–753.
4. Edelman A, Sutton BD (2008) The beta-Jacobi matrix model, the CS decomposition, and generalized singular value problems. *Found Comput Math* 8:259–285.
5. Donoho DL, Elad M, Temlyakov VN (2006) Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans Inf Theory* 52:6–18.
6. Elad M (2010) *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing* (Springer, New York).
7. Candes EJ (2008) The restricted isometry property and its implications for compressed sensing. *Compt Rendus Math* 346:589–592.
8. Candes EJ, Tao T (2006) Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans Inf Theory* 52:5406–5425.
9. Foucart S, Lai MJ (2009) Sparsest solutions of underdetermined linear systems via ℓ_q -minimization for $0 < q \leq 1$. *Appl Comput Harmon Anal* 26:395–407.
10. Calderbank R, Howard S, Jafarpour S (2010) Construction of a large class of deterministic sensing matrices that satisfy a statistical isometry property. *IEEE J Sel Top Signal Process* 4:358–374.
11. Haikin M, Zamir R (2016) Analog coding of a source with erasures. *Proceedings of the IEEE International Symposium on Information Theory Proceedings (ISIT)* (IEEE, New York), pp 2074–2078.
12. Tulino AM, Verdú S (2004) *Random Matrix Theory and Wireless Communications* (Now Publishers Inc., Delft, The Netherlands), Vol 1.
13. Haviv I, Regev O (2016) The restricted isometry property of subsampled Fourier matrices. *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, ed Krauthgamer R (SIAM, Philadelphia), pp 288–297.

14. Rudelson M, Vershynin R (2008) On sparse reconstruction from Fourier and Gaussian measurements. *Commun Pure Appl Math* 61:1025–1045.
15. Nelson J, Price E, Wootters M (2014) New constructions of RIP matrices with fast multiplication and fewer rows. *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, ed Chekuri C (SIAM, Philadelphia), pp 1515–1528.
16. Pfander GE, Rauhut H, Tropp JA (2013) The restricted isometry property for time-frequency structured random matrices. *Probab Theory Relat Fields* 156:707–737.
17. Cheraghchi M, Guruswami V, Velingker A (2013) Restricted isometry of Fourier matrices and list decodability of random linear codes. *SIAM J Comput* 42:1888–1914.
18. Marčenko VA, Pastur LA (1967) Distribution of eigenvalues for some sets of random matrices. *Math USSR-Sbornik* 1:457–483.
19. Bai Z, Silverstein JW (2010) *Spectral Analysis of Large Dimensional Random Matrices* (Springer, New York), Vol 20.
20. Johnstone IM (2008) Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy–Widom limits and rates of convergence. *Ann Stat* 36:2638.
21. Yao J, Bai Z, Zheng S (2015) *Large Sample Covariance Matrices and High-Dimensional Data Analysis (No. 39)* (Cambridge Univ Press, Cambridge, UK).
22. Bandeira AS, Fickus M, Mixon DG, Wong P (2013) The road to deterministic matrices with the restricted isometry property. *J Fourier Anal Appl* 19:1123–1149.
23. Fickus M, Jasper J, Mixon DG, Peterson J (2015) Group-theoretic constructions of erasure-robust frames. *Linear Algebra Appl* 479:131–154.
24. Gurevich S, Hadani R (2008) The statistical restricted isometry property and the Wigner semicircle distribution of incoherent dictionaries. arXiv:0812.2602.
25. Mazumdar A, Barg A (2011) General constructions of deterministic (s) rip matrices for compressive sampling. *Proceedings of the IEEE International Symposium on Information Theory Proceedings (ISIT)* (IEEE, New York), pp 678–682.
26. Monajemi H, Jafarpour S, Gavish M, Donoho DL (2013) Deterministic matrices matching the compressed sensing phase transitions of Gaussian random matrices. *Proc Natl Acad Sci USA* 110:1181–1186.
27. Debbah M (2008) Asymptotic behaviour of random Vandermonde matrices with entries on the unit circle. arXiv:0802.3570.
28. Götze F, Tikhomirov A (2011) On the rate of convergence to the Marchenko–Pastur distribution. arXiv:1110.1284.
29. Götze F, Tikhomirov A (2016) Optimal bounds for convergence of expected spectral distributions to the semi-circular law. *Probab Theory Relat Fields* 165:163–233.
30. Chatterjee S, Bose A (2004) A new method for bounding rates of convergence of empirical spectral distributions. *J Theor Probab* 17:1003–1019.
31. Meckes ES, Meckes MW (2016) Rates of convergence for empirical spectral measures: A soft approach. arXiv:1601.03720.
32. Applebaum L, Howard SD, Searle S, Calderbank R (2009) Chirp sensing codes: Deterministic compressed sensing measurements for fast recovery. *Appl Comput Harmon Anal* 26:283–290.
33. Babadi B, Tarokh V (2011) Spectral distribution of random matrices from binary linear block codes. *IEEE Trans Inf Theory* 57:3955–3962.
34. Tropp JA (2008) On the conditioning of random subdictionaries. *Appl Comput Harmon Anal* 25:1–24.
35. Collins B (2005) Product of random projections, Jacobi ensembles and universality problems arising from free probability. *Probab Theor Relat Field* 133:315–344.
36. Xia P, Zhou S, Giannakis GB (2005) Achieving the Welch bound with difference sets. *IEEE Trans Inf Theory* 51:1900–1907.
37. Strohmer T, Heath RW (2003) Grassmannian frames with applications to coding and communication. *Appl Comput Harmon Anal* 14:257–275.
38. Paley RE (1933) On orthogonal matrices. *J Math Phys* 12:311–320.