Extraction of unknown signals in arbitrary noise

Glenn Ierley and Alex Kostinski

Phys. Rev. E **103**, 022130 — Published 18 February 2021

DOI: 10.1103/PhysRevE.103.022130

Extraction of Unknown Signals in Arbitrary Noise

Glenn Ierley*
Department of Mathematical Sciences, Michigan Technological University 1400 Townsend Drive, Houghton, Michigan 39921, USA and Scripps Institution of Oceanography, UC San Diego, 9500 Gilman Drive, La Jolla, California 92093-0225, USA

Alex Kostinski†
Department of Physics, Michigan Technological University 1400 Townsend Drive, Houghton, Michigan 49931, USA
(Dated: January 29, 2021)

We devise a general method to extract weak signals of unknown form, buried in noise of arbitrary distribution. Central to it is signal-noise decomposition in rank and time: only stationary white noise generates data with a jointly uniform rank-time probability distribution, $U(1,N)\times U(1,N)$, for N points in a data sequence. We show that rank, averaged across jointly indexed series of noisy data, tracks the underlying weak signal via a simple relation, for all noise distributions. We derive an exact analytic, distribution-independent, form for the discrete covariance matrix of cumulative distributions for independent and identically distributed noise and employ its eigenfunctions to extract unknown signals from single time series.

I. INTRODUCTION

Signal separation from noise is an essential part of any experiment, be it a passage of an elementary particle, arrival of a gravitational wave, or a radar echo. Instrumental noise, clutter, unwanted fluctuations are inevitable [1] and the literature on the topic is vast, crossing many fields, and containing a great variety of specialized solutions [2–4]. For example, lock-in amplifiers detect low-level signals obscured by noise, but the signal form must be known.

Approaches to extraction of signals of unknown form (non-parametric) typically rely on the assumption of additive normally distributed noise, e.g., [4] and signal processing literature has been dominated, for over a century, by the additive Gaussian white noise model, e.g., the least squares approach of maximum likelihood [5, 6]. But pronounced fluctuations associated with "black swan" events and heavy-tailed (power-law) distributions have become increasingly common in statistical physics [7, 8], e.g., ranging from photonics to air pollution [9–12]. To that end, here we address situations where neither signal form is known nor is the noise of a conventional variety.

Despite its ubiquity, noise is notoriously difficult to define (e.g., [1, 2, 13]) – as the old adage goes, one man's noise is another man's signal [14]. To that end, based on a simple signal-noise decomposition in a rank-time plane [15, 16], we propose a general all-purpose signal extraction method. Although there is a rich literature on rank-based approaches to non-parametric hypothesis testing [17, 18], there is a dearth of rank-based signal retrievals. In fact, to the best of our knowledge, rank-based unknown signal extraction has not been considered. For example, the comprehensive three-volume set [19] on statistical signal processing does not suggest a single ordinal method. Rank offers a broadly applicable framework with no constraints; noise can be white or colored, additive (e.g., "dark" noise in detectors) or signal-dependent such as fluctuations caused by an atmospheric turbulent propagation channel [3].

With that in mind, consider data from n serially-ordered channels (e.g., synchronously acquired time series), each comprised of N real-valued elements $x_k^{(j)}$ as depicted in Fig. 1. The superscript and subscript identify the time series and an observation within a given series, respectively. Such n time series might represent seismic or EEG detectors, elements of an antenna array, stock prices in a portfolio, weather stations around the globe, hot wires in a turbulent flow, etc., with each series containing N measurements of a noisy fluctuating process. Several copies of an unknown sought signal s are thus embedded in the several samples of noise. The ubiquitous case of $n = 1$ is also treated in the material below.

In applications, the unknown signal is typically calculated as the sample (arithmetic) mean

$$\widehat{s}_k = \frac{1}{n}(x_k^{(1)} + x_k^{(2)} + \ldots x_k^{(n)}) \qquad k = 1, \ldots N \qquad (1)$$

where the observations $x_k^{(j)}$ are of the signal buried in noise. The $\widehat{\ }$ denotes the *estimate* of the true mean s_k. It is one of the cornerstones of probability theory that the sample mean in (1) converges to the true (ensemble) mean as $n \to \infty$ (law of large numbers) [3, 20], except for fluctuations with infinite variance (e.g., Cauchy noise) [21]. Perhaps no calculation is more commonplace in data analysis than the sample mean yet, when first advanced, it was hotly debated [22], making an early appearance as the "Arithmeticall meane" in [23]. An alternative to the ubiquitous (1) is proposed below and it applies to any serially ordered data of *arbitrary* distribution.

* Emeritus, Scripps Institution of Oceanography. grierley@ucsd.edu
† kostinsk@mtu.edu

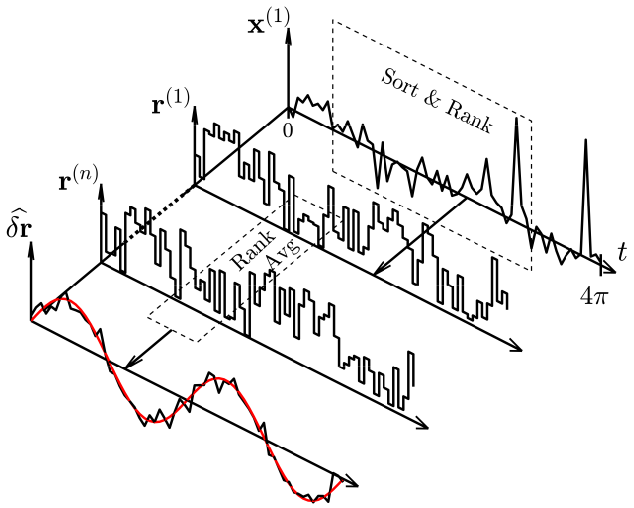

FIG. 1. **Mean rank tracks the weak signal via (2)**. The data value $x_k^{(j)}$ for each series $j = 1 \ldots n$ and time $k = 1 \ldots N$ is the sum of a random variable drawn from the Cauchy pdf $f(x) = (\pi[1+x^2])^{-1}$ and signal $s(t) = \sin(t)/2$ on $[0, 4\pi]$, with $N = 64$ and $n = 500$. Each data series is sorted to yield the rank vs. time series which are then averaged "cross-track". That result for mean rank is shown in black on the last trace. The red trace shows the ensemble limit for mean rank, which clearly reproduces the *form* of $s(t)$. The signal estimate itself, $\widehat{\mathbf{s}}$, is then derived from (2) with $\rho_1 = (2\pi)^{-1}$. The standard deviation of the error in that estimated signal, $\varepsilon = \widehat{\mathbf{s}} - \mathbf{s}$, is $\sigma_\varepsilon \approx 0.0573$. While the formal variance of Cauchy noise is infinite, the sample-based signal-to-noise ratio is $\sim 10^{-5}$.

In order to develop a robust, distribution-independent approach to signal extraction, we turn to ranking of data [15], i.e., values in each of the n time series are sorted by magnitude in ascending order, and N integer-valued ranks recorded in the corresponding N "time slots" [5]. To illustrate this process and the notation used in Fig. 1, consider the following example:

Mock Data

$\mathbf{x}$	-0.45	0.35	0.56	0.13	0.64	0.65	1.37
$\mathbf{t}$	1	2	3	4	5	6	7
$\mathbf{r}$	1	3	4	2	5	6	7
$\delta\mathbf{r}$	-3	-1	0	-2	1	2	3

where the top row $\mathbf{x}$ is a realization of raw data with signal $y(x) = x$ plus noise drawn from a uniform distribution on the interval $[-1/2, 1/2]$, the second row $\mathbf{t}$ is the time index, the third row $\mathbf{r}$ is the first row data reduced to rank, and the bottom row $\delta\mathbf{r}$ is the deviation from the expected mean rank for pure noise, $(N+1)/2 = 4$. The rank information in this single trace can also be stored in 2D rank-time grid as a (permutation) matrix P, whose rows and columns represent the time $\mathbf{t}$ and rank $\mathbf{r}$ vectors respectively. The N unit entries of P, indicating occupancy, are read from the vectors, e.g., $P_{4,2} = 1$ from the fourth entry and $P_{4,j} = 0$ for $j \neq 2$. For n traces, the sum of the n Ps, normalized by nN, yields the traditional (empirical) probability mass function (pmf), i.e., a normalized 2D histogram.

At first glance, ranking appears to hold little promise as, for example, all strong monotonic signals (low noise) such as increasing linear, logarithmic, and exponential functions yield the identical monotone rank distribution $1, 2, ..N$. Yet, Fig. 1 demonstrates that *deviation mean rank* recovers the form of the weak signal faithfully. How can this be? Paradoxically, even a tiny amount of noise, it turns out, restores the magnitude information, up to an additive constant.

An essential ingredient is the observation that only stationary white noise generates all permutations of N ranks among the N indexed slots with equal probability. This equipartition $U(1, N)$ is *perturbed by weak signals* so that the change in equal rank probabilities of pure noise is linearly proportional to (weak) signal amplitudes and recorded by the deviation mean rank as illustrated in the final trace of Fig. 1. Sampling variability, of course, is present in all traces, including the final (black) trace.

Mean rank, unlike the arithmetic mean, couples the "cross-track" and "along-track" directions as illustrated in Fig. 1, sorting within a trace and then averaging across traces, the last line of the above "mock data" illustrating such a sample trace numerically and sample traces seen graphically with $\mathbf{r}^{(1)}$ to $\mathbf{r}^{(n)}$ in the figure. As can be seen in the final trace, despite the loss of magnitude information, the (weak) signal-induced perturbation of the (pure noise) uniform rank probabilities suffices for the deviation mean rank, $\widehat{\delta\mathbf{r}}$, to track the underlying signal faithfully. In the weak signal limit one anticipates the signal-induced perturbation of equal rank probabilities to be linearly proportional to signal amplitude and it then remains to discover the constant of proportionality.

II. SUMMARY OF THE MAIN RESULTS

Fig. 1 demonstrates that for weak signals and large n, mean rank accurately tracks the signal form. The (signal-driven) perturbation from uniformity is defined by the rank deviation $\delta\widehat{r}_k = r_k - (N+1)/2$, where $(N+1)/2$ is the expected value for pure noise for all time slots. Then, the deviation mean rank provides an estimate of the exact signal via a remarkably simple relation:

$$\widehat{\mathbf{s}} \equiv \frac{1}{N\,\rho_1}\,\widehat{\delta\mathbf{r}} \approx \mathbf{s}, \qquad \rho_1 \equiv \int_{-\infty}^{\infty} f(x)^2\,dx\,. \tag{2}$$

where $\widehat{\mathbf{s}}$ is the rank-derived signal estimate and $f(x)$ is the probability density function (pdf) of the background noise, (1). Note that being a pdf, $f(x)$ has units so that inverse ρ_1 in (2) has dimensions of the actual measured signal amplitude. The pre-factor of $1/n$ in (1) is subsumed in the computation of mean rank and $\widehat{\mathbf{s}}$ is defined up to an additive constant. As explained in Section IIIC, the estimate in (2) is the leading order term in an asymptotic expansion for an ensemble, $n \to \infty$.

Both signal estimates, (1) for arithmetic mean and (2) for mean rank, converge to the true signal as $n \to \infty$, but the arithmetic mean requires finite variance and fails for heavy-tailed noise such as Pareto or Cauchy. The mean rank converges via (2) only in the weak signal limit. Convergence for stronger signals also holds but requires that (2) be used in an iterative scheme, see section III D.

The rates of convergence as given by the standard deviation in the mean rank and arithmetic mean signal estimates are, respectively

$$\frac{1}{\sqrt{12}\,\rho_1}\,\frac{1}{\sqrt{n}} \quad \text{and} \quad \sigma\,\frac{1}{\sqrt{n}}, \tag{3}$$

where σ is the standard deviation of noise. The mean rank expression arises from the zero-signal limit where mean rank is uniformly distributed and its fluctuations, (integer-valued counts) are described by the Poisson process. Thus, the convergence comparison is governed by the quantity $\sqrt{12}\,\rho_1\,\sigma$. (The square of this quantity, the Pitman asymptotic relative efficiency for rank spread test vs. the t test [24, 25], arose in mathematical statistics, in the context of hypothesis testing. A detailed comparison is supplied by the extensive table surveying the factors for Pitman efficiency for not only rank vs. arithmetic mean but also rank vs. median presented in Appendix B.2.)

A skeptic might object that $n = 500$ in Fig. 1 is rather confining. Indeed, while some applications such as ocean floats or earthquake sensors would allow large n [15], others such as EEG measurements do not. Probably the most common case in practice is that of a single time series ($n = 1$). Of course, the arithmetic mean, median estimates do no apply then. Remarkably, mean rank performs well even in this $n = 1$ case, as demonstrated in panels (a) and (b) of Fig. 2 where a weak sinc signal buried in overwhelming noise of infinite variance is retrieved with high fidelity from a single time series. So in what sense can we speak of "mean rank" when $n = 1$? Next, we highlight the key ideas.

The space of rank vectors consists of the $N!$ permutations of the integers $1, \ldots, N$ and one can compute the exact covariance matrix for this space which, up to a scale factor, assumes the "canonical" (indicated by *) form

$$\Sigma^*_{j,k} := N\,\min(j,k) - jk\,, \tag{4}$$

whose normalized eigenvectors of $\mathbf{\Sigma}^*\psi_m = \lambda_m\psi_m$ emerge as harmonics

$$\psi_m(x_k) = \sqrt{\frac{2}{N}}\begin{cases}\sin(m\,x_k) & m \quad \text{even}\\ \cos(m\,x_k) & m \quad \text{odd}\end{cases} \tag{5}$$

where $x_k = k\pi/N - \pi/2$ for $k = 1, \ldots (N-1)$ and

$$\lambda_m = \frac{N}{2\,(1-\cos(m\pi/N))}\,. \tag{6}$$

What is the relation of this result to iid noise? On recognizing that the mapping from stationary white noise to rank generates all permutations with equal probability of $1/N!$, it follows that the same covariance matrix also applies to the infinite ensemble. The harmonics in (5), thus linked to white noise, are related in the next section, via cumulative distribution fluctuations, to the so-called Brownian bridge (Weiner process pinned at the ends or returns of a zero-inertia Brownian particle)[26], but with the Brownian part rendered inessential and the eigenvalues for a Weiner process similar to (6) only for

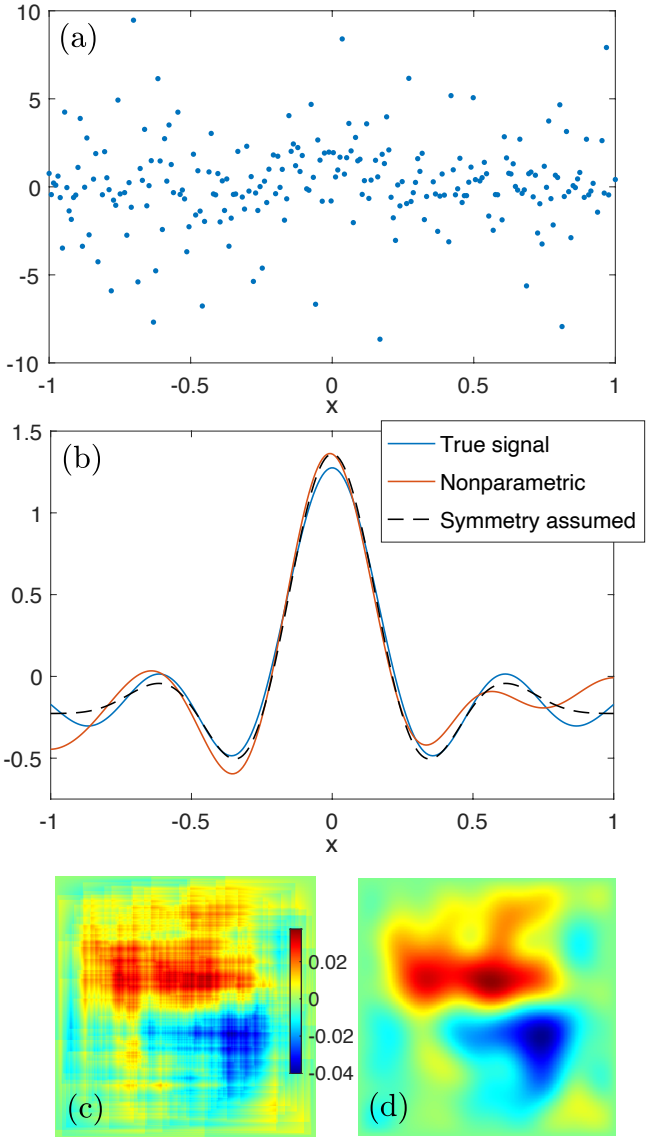

FIG. 2. **Nonparametric extraction for** $n = 1$. (a) A series consisting of a signal $s(x) = 1.44\,\mathrm{sinc}(4x)$ plus Cauchy noise (full vertical scale is $[-171, 888]$). (b) True (unknown) signal (blue), assumption-free estimate (red), and estimate assuming signal symmetry about the peak (dashed black). (c) Color-coded values of $\delta\mathcal{C}$ from (7) with $N = 256$. (d) A Fourier series expansion (8) of $\delta\mathcal{C}$ with $M = 10$. Note the similarity of large-scale patterns in (c) and (d). P is found from this $\delta\widetilde{\mathcal{C}}$ via (10), whose first moment yields the mean rank $\delta\mathbf{r}$ in (2). Then $\rho_1 = (2\pi)^{-1}$ yields the curves in (b).

small m, but not in the tails.

Towards signal extraction, consider raw data yielding a perfectly ordered succession of ranks $[1, 2, 3, ..N]$. Despite appearing "orderly", this rank partition is just as likely as any other to be generated by iid noise. But, given this data, we can ask: is the white noise model *as likely* as a signal plus noise model? This "likelihood" type of question is quantified here beyond the confines of parametric estimation and even for single trials ($n = 1$).

Guided by the equivalence of rank (r) and time (t) for independent and identically distributed (iid) noise, we map the raw data to the rank-time (r, t) plane. As $n \to \infty$, the resulting 2D discrete probability density (mass) function (pmf) is jointly uniform, $U(1, N) \times U(1, N)$. This characterization of stationary white noise defines the *absence of a signal*, delivers a simple signal-noise decomposition, and regards signals broadly as deviations from pure randomness. This symmetry-breaking property of signals then forms the basis for universal (distribution-independent) signal extraction.

To implement this program, we continue with a cumulative distribution function (cdf) of the iid noise, with a particularly simple form (xy in the continuous case) and whose fluctuations turn out to have universal correlation properties. With rare exceptions, e.g., the cdf of optical depth in [27] or improvements of statistical tests [28], physicists prefer pdfs to cdfs. Yet, the cdf has important advantages both, fundamental and practical. On the fundamental side, there is the general propensity for the cdf perspective to yield universal results, e.g., *distribution-independent* convergence theorems of Kolmogorov and Smirnov, as well as the celebrated Kolmogorov-Smirnov (KS) test itself [5, 29] and the three-term cdf decomposition theorem [29]. Perhaps most important, in our view, is the following remarkable yet seldom mentioned result. Let a random variable X have a continuous distribution for which the cdf is F_X. Then the random variable $Y = F_X(X)$ has a *standard uniform* distribution, for *any* F_X [30]. Dealing with data at the cdf level also has its perks. For instance, it is not necessary to bin the data to form a natural empirical cdf staircase whereas some form of binning is required for the pdf. Single-pass estimation of arbitrary quantiles is often desired, e.g., from time to time one might need the median value (or 95th percentile) of data seen thus far [5].

Returning to signal extraction, the ensemble limit of a jointly uniform pmf, $p(r, t) = U(1, N) \times U(1, N)$, and its associated discrete cumulative distribution function (simple xy for the continuous analog), $\mathcal{C}$ are both perturbed by signals. Here we isolate that perturbation by means of the 2D deviation cumulative distribution function $\delta\mathcal{C}$ defined as:

$$\delta\mathcal{C}_{k,l} = \sum_{i=1}^{k} \sum_{j=1}^{l} \left(p_{i,j} - \frac{1}{N^2} \right) \quad \{k, l\} = 1, \dots (N-1) \quad (7)$$

where $p_{i,j}$ is the pmf and the δ denotes deviation. The subtraction of $1/N^2$ removes the jointly uniform base state [31]. In practice, one obtains $p_{i,j}$s from ranked data via the normalized 2D histogram.

We show that $\delta\mathcal{C}$, generated by the iid noise, fluctuates in a distribution-independent manner just as rank vectors do. Moreover, the $\delta\mathcal{C}$ covariance matrix is the Kronecker tensor product of $\Sigma^*_{j,k}$ with itself. This, in turn, yields a convergent 2D (Fourier) eigenfunction expansion formed by the direct product of all $(N-1)^2$ pairs from (5) and corresponding products of the associated 1D eigenvalues.

The main and seemingly contradictory idea now is to apply (2) even at $n = 1$ as one estimates in Fig. 2(d) a small subset of the 2D expansion coefficients for the ensemble limit ($n \to \infty$) of $\delta\mathcal{C}$ based on the single realization of $\delta\mathcal{C}$ in 2(c). As indicated in Fig. 7 of Appendix B.4, a spectral signature guides the choice of the truncation M for the expansion:

$$\delta\widetilde{\mathcal{C}}_{k,j} = \sum_{n=1}^{M} \sum_{m=1}^{M} c_{n,m}\, \psi_n(y_k)\, \psi_m(x_j)\,. \quad (8)$$

where

$$c_{n,m} = \sum_{k=1}^{N-1} \sum_{j=1}^{N-1} \psi_n(y_k)\, \psi_m(x_j)\, \delta\mathcal{C}_{k,j}$$

From these estimated coefficients one obtains the smooth $\delta\widetilde{\mathcal{C}}$ and hence $P(r, t)$ from

$$P(x, y) = \frac{\partial^2 \delta\widetilde{\mathcal{C}}(x, y)}{\partial x \partial y} \quad (9)$$

which, in discrete form, means using the harmonics

$$\frac{d\psi_m(x'_k)}{dx} = \frac{\pi}{N} \sqrt{\frac{2}{N}} \begin{cases} m\, \cos(m\, x'_k) & m \quad \text{even} \\ -m\, \sin(m\, x'_k) & m \quad \text{odd} \end{cases}$$

where $x'_k = k\pi/N - \pi/2\,(1 - 1/N)$ for $k = 0, \dots (N-1)$. Then

$$P_{j,k} \approx \sum_{n=1}^{M} \sum_{m=1}^{M} c_{n,m}\, \frac{d\psi_n(y'_k)}{dy}\, \frac{d\psi_m(x'_j)}{dx}\,. \quad (10)$$

and $\delta\widehat{\mathbf{r}}/N = \mathbf{P}[1 \dots N]_0^T$ (subscript 0 indicates zero mean) from which we obtain the signal estimate plotted in (2)b. The dashed curve is a better match yet signal parity is used as prior information. An iterative approach, quantifying the "weak signal" range of validity and extending (2) to weaker noise (summarized in the caption of Fig. 3), again yields retrievals unmatched by other methods.

III. UNDERLYING THEORETICAL DEVELOPMENT

We begin with the derivation of (4) for the fluctuation of rank vectors (discrete case) and then proceed with the parallel derivation for cdf fluctuations (continuous case) by treating cumulative distributions as order statistics [32]. This culminates with the perturbation expansion to arrive at (2) and higher order terms.

A. Discrete Covariance Matrix: rank

Consider iid noise-generated fluctuations of the sample vector with N entries. All possible permutations of N ranks in N slots occur with the same probability $1/N!$. Thus, an ergodic behavior is ensured as our "system" samples all available "microstates" without bias or preference and the ensemble limit holds ($n \to \infty$).

Elements of the i-th permutation vector are denoted by $r_k^{(i)}$ for $k = 1, 2, \ldots N$ and $i = 1, 2, \ldots N_p$ where N_p is the size of the permutation space, which depends upon the allowed set of entries $X = \{\mathbf{x}\}$. Towards calculating the covariance matrix, we then define zero-mean partial sums as

$$s_k^{(i)} = \sum_{j=1}^{k} \left(r_k^{(i)} - \sum_{j=1}^{N} r_j^{(i)}/N \right) \qquad k = 1, 2, \ldots (N-1)\,. \tag{11}$$

and calculate the covariance matrix from

$$\Sigma_{k,l} = \frac{1}{N_p} \sum_{i=1}^{N_p} s_k^{(i)} s_l^{(i)} \qquad (k,l) = 1, 2, \ldots (N-1)\,. \tag{12}$$

The discrete covariance matrix "standard" is given by (4) and the simplest set of permutations yielding it is $N\, I_N$, where I_N is the identity matrix of order N. A broader context for this choice is the set of $N!$ permutation matrices P_N, regarding the $r_k^{(i)}$ as either the rows or columns. Each yields the identity case covariance matrix. Any particular permutation matrix can be regarded as the rank-time realization [15]. Using the identity matrix, we have $r_k^{(i)} = \delta_{i,k}$ and the zero mean partial sums are given by

$$s_k^{(i)} = NH(k - i + 1) - k \qquad k = 1, \ldots (N-1)$$

where H denotes the Heaviside function with the convention $H(0) = 0$. Elements in the covariance matrix reduce to evaluation of the product

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^{N} & \left(NH(j-i+1) - j\right) \left(NH(k-i+1) - k\right) \\ &= N \sum_{i=1}^{N} H(j-i+1)\, H(k-i+1) \\ &\quad - \sum_{i=1}^{N} \left[jH(k-i+1) + kH(j-i+1) \right] + j\,k \\ &= N\, \min(j,k) - j\,k \quad (j,k) = 1, \ldots (N-1) \end{aligned}$$

For *all* general permutation classes in (12) then $\Sigma_{j,k} = \alpha(N)\, \Sigma^*_{j,k}$ and it suffices to simply compute $\Sigma_{1,1}$ in (12) as a function of N to fix $\alpha(N) = \Sigma_{1,1}/(N-1)$. A variety of examples of $\alpha(N)$ is presented in Appendix C. That this result emerges, wholly independent of details of the x_i save through the vestigial factor of $\alpha(N)$ is paralleled in the order statistics argument below when the perfect differential drops all reference to the underlying noise distribution $f(x)$.

Application of this covariance matrix to $\delta\mathcal{C}$ rests upon the assumption of equal *a priori* occupancy of the corresponding finite state space for an infinite ensemble of white noise input vectors. Individual realizations of $\delta\mathcal{C}$ derive from an $N \times N$ permutation matrix P:

$$N\delta\mathcal{C}_{k,l} = \sum_{i=1}^{k} \sum_{j=1}^{l} (P_{i,j} - 1/N) \qquad \{k,l\} = 1, 2, \ldots (N-1)$$

Based on either rows or columns alone, this is equivalent to (4) but for I_N rather than NI_N, e.g, 1D $\alpha(N)/N^2$. The double sum taken here then leads to a covariance matrix which is $(N-1)^2 \times (N-1)^2$ given by the Kronecker tensor product of $\alpha(N)\, \Sigma^*$ with itself. The singular value decomposition of the cdf ensemble for $\delta\mathcal{C}$ is then given by $\mathbf{U}\, \boldsymbol{\sigma}\, \mathbf{V}^T$ where the columns of $\mathbf{U}$ are the reshaped $(N-1)^2$ 2D eigenvectors and $\boldsymbol{\sigma}$ is a diagonal matrix with entries that derive from corresponding products of (6), subject to the scale factor for $\delta\mathcal{C}$, and the standard factor of $1/\sqrt{N-1}$ for the conversion from eigenvalues of the covariance matrix to singular values[33]:

$$\lambda_{n,m} = \frac{1}{2N\sqrt{N-1}} \sqrt{\frac{1}{\left(1 - \cos(m\pi/N)\right)\left(1 - \cos(n\pi/N)\right)}}\,, \tag{13}$$

where $(m,n) = 1 \ldots (N-1)$ with the values sorted in decreasing order, commencing with $(m = 1, n = 1)$, and columns of $\mathbf{U}$ and rows of $\mathbf{V}$ shuffled correspondingly.

B. Continuous Covariance Matrix: order statistics

We shall now broaden the above derivation to the continuous case. There are two major reasons for doing so. First, the approach is needed to derive the main result, equation (2). Second, we establish universality of fluctuations of the empirical cdf via order statistics [32, 34] and link it with the universality of the pinned Wiener process (so-called Brownian bridge [26, 28]). Our derivation does *not* invoke the central limit theorem, thereby disentangling Brownian and bridge parts, an important result in itself.

To that end, we adopt the perspective of *order statistics* on the empirical cdf, e.g., [29, 34]. Let $\mathcal{X} = (X_1, X_1 \ldots X_n)$ be a random vector (sample) of size n and $(x_1, x_2, \ldots x_n)$ be a realization of the random vector $\mathcal{X}$. For the iid case, let $f(x)$ and $F(x)$ be their pdf and cdf, respectively. When the observations are arranged by ascending magnitude, $X_{(j)}$ denotes the jth order statistic, that is, the jth smallest of the continuous iid random variables $X_1, X_2, \ldots, X_n$, [29, 32]. The empirical cdf (data staircase) can now be defined via the Heaviside

(step) function $H(x)$ as a random process

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} H(x - X_i)$$

whose particular realization is obtained (measured) for $X_i = x_i$. So the step heights of the increasing staircase $F_n(x)$ are multiples of $1/n$. The key idea here is that by proving the invariance of the iid-generated covariance matrix we also show the universality (distribution-independence) of $F_n(x)$ fluctuations.

Returning now to the order statistics, for the density function of $X_{(J)}$ of an iid random process, one has

$$f_{X(j)}(x) = \frac{n!}{(n-j)!\,(j-1)!}\,[F(x)]^{j-1}\,[1-F(x)]^{n-j}\,f(x)\,.$$

e.g., see p. 276 of [3] or Eq. (6.2) in [32].

The expectation value of $F(X_{(J)})$ is then

$$E[F(X_{(j)})] = \frac{n!}{(n-j)!\,(j-1)!}\int_a^b F(x) \times [F(x)]^{j-1}\,[1-F(x)]^{n-j}\,f(x)\,dx \tag{14}$$

The substitution $s = F(x)$ and $ds = f(x)\,dx$ yields

$$\int_0^1 s^j\,[1-s]^{n-j}\,ds = \frac{(n-j)!\,j!}{(n+1)!}\,.$$

Hence $E[F(X_{(j)})] = j/(n+1) \quad j = 1, 2, \ldots n$.

Similarly from Eq. (6.6) in [32], the joint density of $X_{(j)}$ and $X_{(k)}$

$$f_{X(j),X(k)}(x_j, x_k) = \frac{n!}{(j-1)!\,(k-j-1)!\,(n-k)!} \times [F(x_j)]^{j-1}\,[F(x_k)-F(x_j)]^{k-j-1} \times [1-F(x_k)]^{n-k}\,f(x_j)\,f(x_k)\,. \tag{15}$$

and, with the same variable change, $E[X_{(j)}\,X_{(k)}]$

$$E[X_{(j)}\,X_{(k)}] = \frac{n!}{(j-1)!\,(k-j-1)!\,(n-k)!} \times \int_0^1 dt\,(1-t)^{n-k}\,t\int_0^t ds\; s^j\,(t-s)^{k-j-1} \tag{16}$$

This integral is evaluated in terms of the standard beta function [29] and one arrives at the sought result for the order statistics covariance matrix ($j < k$)

$$E[X_{(j)}\,X_{(k)}] - E[F(X_{(j)})]E[F(X_{(k)})] = \frac{j\,(n+1-k)}{(n+1)^2\,(n+2)}\,. \tag{17}$$

Note j and k are interchanged for $j \geq k$. Equation (17) is the same as the earlier result (4) to within a scaling factor (see Appendix C).

Although we are not aware of the universality question being posed in terms of the covariance matrix, a related result exists in the theory of stochastic processes. Namely eigenvalues for the Brownian bridge (pinned Wiener process) are similar to (5) but rely on the normal approximation. Thus, it is particularly noteworthy that the result (17) is distribution-independent and yet neither this derivation nor the discrete one above invokes the central limit theorem.

To link (17) explicitly to the Wiener process, we employ the Gnedenko-Koroliuk map from the cdf difference onto a pinned random walk [35]. Letting $\{Y_k\}$ be another set of n iid random variables with the same cdf $F(x)$, one orders the $2n$-tuple composed of $(X_1, X_2 \ldots X_n)$ and $(Y_1, Y_2 \ldots Y_n)$ in ascending order, and puts either 1 or -1 in the j-th entry, to replace X_j or Y_j respectively. One thereby obtains a map onto a discrete random-walk of $2n$ steps that returns to the origin. In the Gaussian limit this yields the Brownian bridge but for finite N, the exact covariance matrix generated by all $(2N)!/(N!)^2$ permutations of this Gnedenko-Koroliuk map is a constant times (4). On the assumption that all permutations are equally likely, this then is the exact ensemble limit for the finite random walk.

In the continuous limit, equations (4) and (17) become the function $K(x, y) = \min(x, y) - x\,y$ which is recognized as the correlation function for the Brownian bridge [26, 28]. Principal components from an orthogonal function expansion of the Karhunen-Loeve type [20, 36] are then eigensolution pairs

$$f(x) = \sin(k\pi x) \qquad g(y) = \frac{1}{k^2\pi^2}\,\sin(k\pi y), \quad k = 1, 2, \ldots\,.$$

arising from the integral equation

$$g(y) = \int_0^1 K(x, y)\,f(x)\,dx \qquad K(x, y) = \min(x, y) - x\,y$$

for the Brownian bridge [26]. The iid assumption is essential for these results, e.g., our computations (to be reported) show that a sufficiently strong negative correlation, by forcing fine scale wrinkles in δC fluctuations, completely removes the principal mode (half-wave).

C. Mean rank: proof of (2) for weak signals

We prove (2) for $N = 3$ to distill the bare essentials and the general case is treated in (B2). Following the above machinery for order statistics, we introduce the pdf, f_k and cdf F_k. The three conditional pdfs for s_k to map to each of ranks $r_{1,2,3}$ are then

$$\begin{aligned} f_1(s_k \to r_1 | X = x) &= (1 - F_i)\,(1 - F_j)\,f_k \\ f_2(s_k \to r_2 | X = x) &= [F_i\,(1 - F_j) + (1 - F_i)\,F_j]\;f_k \\ f_3(s_k \to r_3 | X = x) &= F_i\,F_j\,f_k \end{aligned} \tag{18}$$

These expressions for the mutually exclusive events are exact and, as a check, sum to f_k.

Next, let the iid noise rank uniformity be perturbed by a weak signal as $f_k = f(x - \epsilon s_k)$ where $\epsilon \ll 1$ and similarly for F_k. (For $\epsilon = 0$, $P_1 = P_2 = P_3 = 1/3$.) From its Taylor series, the $\mathcal{O}(\epsilon)$ perturbation of pdf for expected mean rank is

$$\frac{d}{dx}\Big|_{\epsilon=0} (r_1 \cdot f_1 + r_2 \cdot f_2 + r_3 \cdot f_3) = \\ -s_k\, f'(x) - 2s_k F(x)\, f'(x) - (s_i + s_j) f(x)^2\,.$$

To find the mean rank perturbation at t_k, we integrate on $[-\infty, \infty]$ [37], appeal to vanishing boundary conditions on $f(x)$, and treat the second term by parts to obtain

$$2s_k \int_{-\infty}^{\infty} f(x)^2\, dx - (s_i + s_j) \int_{-\infty}^{\infty} f(x)^2\, dx \\ = 3\left[(s_k - (s_i + s_j + s_k)/3\right] \int_{-\infty}^{\infty} f(x)^2\, dx\,.$$

This can be rewritten as $\delta\widehat{r}_k = 3\rho_1\, \widehat{s}_k$, a simple rearrangement of (2) for $N = 3$, and one sees that the constant offset of $\widehat{s}_k$ amounts to a subtraction of the signal mean. A generalization to all orders of the perturbed integral of (18) for arbitrary N is given in (B2) of the Appendix.

D. Mean rank nonlinearity and iterative error reduction

Although weak signals are of most interest and (2) works well in that limit, one needs the means to evaluate the "weakness" of an *unknown* signal. To that end, the validity of the linear approximation can be assessed by the next order perturbation in signal amplitude. From (B2) in the Appendix, a cubic truncation assumes the form

$$\langle r_k \rangle = \rho_1\, \epsilon \left[N\, s_k - \sum_i s_i \right] \qquad (19)$$
$$- \frac{\rho_3\, \epsilon^3}{3!} \left[N s_k^3 - 3 s_k^2 \sum_i s_i + 3 s_k \sum_i s_i^2 - \sum_i s_i^3 \right] + \mathcal{O}(\epsilon^5)$$

where

$$\rho_3 \equiv \int_{-\infty}^{\infty} (f'(x))^2\, dx\,.$$

As earlier noted, since $f(x)$ is a pdf, the units of ρ_1 are those inverse signal amplitude. Those of ρ_3, are the inverse cube of the amplitude. Correction terms to (19) up to $\mathcal{O}(\epsilon^7)$ are illustrated in Fig. 4 of Appendix B.1. Equations (2) and (19) apply "as is" to correlated noise and also to non-stationary noise pdfs by computing $\int f(x)^2\, dx$ as a function of a relevant parameter and then averaging over the range of that parameter.

Whereas the signal plotted in Fig. (2) derives from the $M \times M$ approximation of $\delta\mathcal{C}$ given in (8), we now approximate instead with only $M \times 1$ lowest harmonics in rank, $\psi_1(x_k)$. This is because signals in the time domain drive predominantly the first harmonic, with the excitation of higher harmonics a function of moments of the noise pdf, $f(x)$. On the assumption that all of the amplitude $c_{k,1}$ derives from signal, removing that signal from the time series will simultaneously zero out both $c_{k,1}$ and the *signal-induced* contribution to higher harmonics, $c_{k,(2,3,\ldots)}$. In addition, the amplitude of that first rank harmonic is a nonlinear function of forcing hence even after obtaining the $M \times 1$ approximation of $\delta\mathcal{C}$, solving for P and using (2), for all but very weak signals, the M harmonics $c_{k,1}, k = 1..M$ will retain significant signal. It is hence necessary to iterate, accumulating successive contributions to the signal.

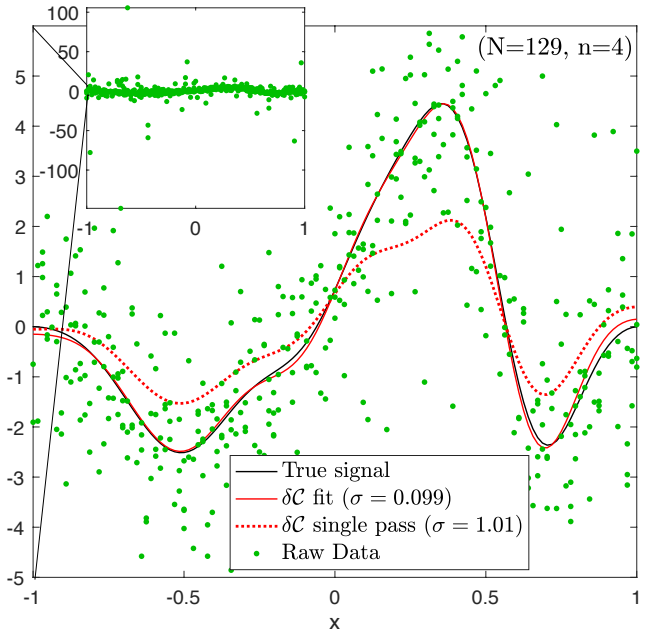

FIG. 3. **Iterative reduction of rank nonlinearity yields excellent signal retrieval.** The signal (blue solid line) is deliberately generic: a Fourier synthesis with randomly chosen coefficients. Green dots (data) include added Cauchy (infinite variance) noise. The dotted red line shows the first pass signal estimate from (20), leaving $\sigma_{\boldsymbol{\varepsilon}}^2 = 0.529$. The amplitude error is due to cubic (and higher) terms in (19). Iteration (solid red) annuls the M mode coefficients ($\sigma_{\boldsymbol{\varepsilon}}^2 = 10^{-6}$), hence eliminating the cubic error. To the best of our knowledge, such performance with merely $n = 4$ is unrivaled by other methods.

The modified (scalar) form of (2) becomes

$$s_k \approx \frac{1}{N\,\rho_1} \left[\sqrt{\frac{2}{N}}\, \frac{\pi \cos(\pi/2N)}{1 - \cos(\pi/N)} \sum_{i=1}^{M} \varepsilon_i \frac{d\psi_i(y'_k)}{dx} \right] \qquad (20)$$

where

$$\varepsilon_i = -\sum_{j=1}^{N-1} \sum_{k=1}^{N-1} \psi_i(y_j)\, \psi_1(x_k)\, \delta\mathcal{C}_{j,k} \quad i = 1, \ldots M \qquad (21)$$

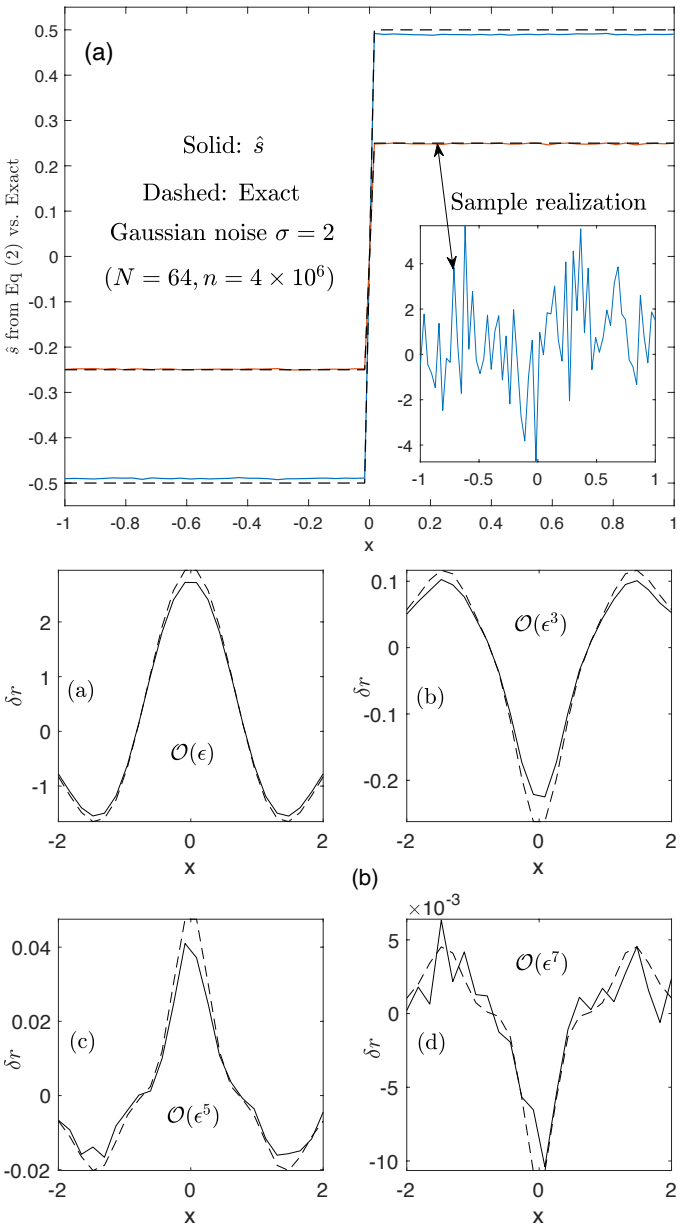

FIG. 4. **Tests of expansions (19) and (B2)**. Panel (a): The test signals are step functions $y(x) = H(x)$ and $y(x) = H(x)/2$, captured well by the leading order estimate (2), but each with a small gap caused by rank nonlinearity. The lesser amplitude deficit, $\approx 10^{-3}$ for $H(x)/2$, is barely discernible, indicating validity of the "weak signal" (small perturbation) regime for this $\sigma = 2$ Gaussian noise (see text). Inset: a sample time series gives a visual sense of the weak signal relative to noise. Panel (b) quartet: The signal, $\mathrm{sinc}(x)$ on $x = [-2, 2]$ is buried in additive Cauchy noise [38] ($x_0 = 0, \gamma = 1$) with ($N = 24, n = 2 \times 10^7$). Solid line is the residual Monte Carlo δr, progressively decomposed. Dashed line is the analytic correction from (B2) removed at each order in ϵ. Along-track sorting confers a global character to rank, exemplified by the $\mathcal{O}(\epsilon^3)$ term in (19) with nonlocal quadratic and cubic sums, while the linear sum over s_i reflects rank invariance to a constant offset.

As an iterative scheme, $s_k \to s_k^{(n)}$ and $\varepsilon_i \to \varepsilon_i^{(n)}$ with $\varepsilon_i^{(1)} = c_{k,1}$, that is, the modal coefficients of the initial $\delta\mathcal{C}$. One then subtracts the resulting mean rank prediction of $s_k^{(1)}$ from the raw data and recomputes $\delta\mathcal{C}$. In practice, after four or so iterations, convergence in $\{\varepsilon_i^{(n)}\}$ as measured by

$$\sigma_{\boldsymbol{\varepsilon}}^2 = \frac{1}{M-1}\sum_{i-1}^{M}(\varepsilon_i - \bar{\varepsilon})^2 \tag{22}$$

saturates at about 10^{-4} [39].

In contrast to Fig. (2), for Fig. (3) the iteration is vital; the first pass error is reduced by more than an order of magnitude. Iteration is appropriate when a suitably weighted value of σ_{cubic}, the standard deviation of the bracketed portion of the cubic term in (19), is large in comparison to the noise contribution noted below in (23). Further details are discussed in Appendix B.3.

A second trial with this signal was run using: (i) *correlated* Cauchy noise given by $z_i = (x_i + x_{i+1})/2$, and (ii) an incorrect value for ρ_1, namely that for Gaussian noise. Iteration converged to a solution with $\sigma = 0.18$ quite similar to that plotted. This success illustrates two key points; first, noise correlation does not alter that there remains a proportionality between mean rank and signal and second, when iteration is applicable, that errors in ρ_1 are self-compensating.

The expected value of the noise standard deviation for this $M \times 1$ approach can be approximated as

$$\sigma_{noise} \sim \frac{1}{\sqrt{12}\rho_1}\frac{1}{\sqrt{n}}\left[\sqrt{\frac{M}{N}} - \frac{0.1410}{\sqrt{N}} + \mathcal{O}\left(\frac{1}{N}\right)\right] \tag{23}$$

where M is the order of the Fourier expansion. This takes the place of (3). The condition for successful extraction of a weak signal for small n is that the needed number of harmonics M for the signal is a small fraction of N. To the best of our knowledge, for this mixed regime of moderate n and M with arbitrary noise, there is no method that can rival the accuracy exhibited here.

IV. CONCLUDING REMARKS

The main theme of this paper is that departures from iid white noise are identified as "signals". To gauge such departures we began by noting that the ensemble limit of the rank-time probability distribution for any N point sequence is uniform and, in consequence, the deviation cumulative distribution function $\delta\mathcal{C}$ vanishes. We then establish the needed benchmarks for expected sample variability with finite data sets, replacing our previously purely computational singular value results in [15] by far deeper analytical results, including an exact distribution-independent form for the discrete covariance matrix of cumulative distributions for iid noise, whether of discrete

(4) or continuous (17) origin, and the exact accompanying eigenvalues and eigenvectors for noise-induced fluctuations about the equilibrium state.

We also prove that mean rank from an average across jointly indexed series of noisy data, tracks an arbitrary underlying weak signal via the simple, leading order, relation (2) for all stationary noise distributions. An asymptotic expansion of this relation in (19) and (B2) supplies precise meaning to signal "weakness". When nonlinear terms become significant, (2) becomes an iterative scheme.

Figures 1–3 and 6 illustrate the successful extraction of (presumedly) deterministic signals, both weak and strong, embedded in arbitrary iid noise, where other methods are of little or no avail. Not only do we extract unknown signals for small n, as in Fig. 3, but even single trials, as in Fig. 2, constituting perhaps the greatest advance of this paper. The reason for the success with $n = 1$ is that the mean rank utilizes both cross-track information (N) and along-track (n), thus feasible when the latter is unavailable. A devil's advocate might argue that, in lieu of cross-track coupling, one can also form a local running mean (or median) in the along-track direction. But that window makes no use of the universal, global, characterization of expected fluctuations for stationary white noise.

Taking a more general view, noise other than iid *itself* constitutes a "signal", reflected in a structured $\delta\mathcal{C}$ improbable under iid sampling variability. Perhaps the most common departure is colored, i.e., correlated, noise which, though stationary, has a spectrum that is not flat. That "signal" is not expressed in mean rank, which is still zero (in the limit), but in the mean second moment of rank. Positive correlation has in addition a systematic effect mainly on the eigenvalues of the covariance matrix, whereas negative correlation is more dramatically seen in a sharp phase transition with a vanishing of the lowest mode of the covariance matrix. Similarly, uncorrelated but nonstationary noise, where the variance is a function of time, also leaves mean rank zero, with the signal again in the second moment of rank. In a variety of problems, noise of either character – colored or nonstationary – is present but there is also a deterministic signal derived from mean rank that is of principal interest. We have noted for such cases that, in application of (2), for colored noise ρ_1 is unaffected and, for nonstationary noise, ρ_1 is modified by averaging over time.

Still farther from these generalized forms of noise, chaos all the more departs from norms in (13) for sample variability. Nonetheless, for discrete examples such as the logistic and tent maps, chaos can maintain mixing of rank with sufficient uniformity that one can extract lower frequency signals buried within chaos based on the mean rank relation, proceeding as in Section III.D, illustrating the wider applicability of (2). The proposed method can also be used to test random number generators, e.g., [40, 41], quantify instrumental errors, detect over-fitting (when the δC residual suggests anti-correlation), etc.

ACKNOWLEDGMENTS

We thank two anonymous reviewers for helpful comments. This work was supported by the NSF grant AGS-1639868.

Appendix A: Brief overview

Below we address technical but important issues such as a more detailed characterization of the expansion for mean rank and its limitations, the relation of mean rank to arithmetic mean and median, error analysis that accounts for nonlinearity and simultaneously gives guidance on when iteration is needed, and extension of the idea of mean rank to the second moment, employing symmetries in rank-time plane. We also give a brief illustration of other instances of the discrete covariance matrix as well as brief consideration of the kurtosis deficit that is present for finite N in the pdfs for Fourier modes of $\delta\mathcal{C}$.

Appendix B: On Mean Rank

1. Nonlinearity

Proof of the relation between mean rank and signal is outlined in the main text for $N = 3$ and here we describe it for a general n and in more detail. We begin with:

$$P_{n,k}(\epsilon\mathbf{s}) = \int_a^b dx\, f(x-\epsilon s_n) \sum_{j=1}^{\binom{K-1}{k-1}} \prod_{n=1}^{k-1} F(x-\epsilon s_{\tau_{j,n}}) \prod_{m=1}^{K-k} \left(1-F(x-\epsilon s_{\tilde{\tau}_{j,m}})\right) \tag{B1}$$

where $f(x)$ and $F(x)$ are probability density and cumulative distributions, respectively on the interval $[a, b]$. Here $\binom{K-1}{k-1}$ is the binomial coefficient, τ is a matrix whose rows contain all possible choices of $k-1$ elements from the set $\{1, 2, \ldots, K\}_{\underline{n}}$ and

$$\{\tilde{\tau}_{j,m}\} \equiv \{1, 2, \ldots, K\}_{\underline{n}} \backslash \{\tau_{j,n}\}\,.$$

where $\backslash$ denotes the relative complement (set exclusion).

In contrast to the earlier, strictly formal, appearance of (B1) as (B1) in [15] (with a slightly different notation), it is used here to obtain an expansion for mean rank. Using (B1), the ensemble mean rank is

$$\langle r_n\rangle = \sum_{k=1}^{N} k\, P_{n,k}(\epsilon\mathbf{s}) = \frac{N+1}{2} - \sum_{j=1}^{\infty} \frac{(-1)^j\, \epsilon^{2j-1}\rho_{2j-1}}{(2j-1)!} \sum_{i=0}^{2j-1} \binom{2j-1}{i} (-1)^i\, s_n^{2j-1-i} \sum_{k=1}^{N} s_k^i \tag{B2}$$

Distribution	$p(x)$	Mean Rank $\int p(x)^2\,dx$	Arithmetic Mean $\sigma(\mu_1)$	Median $\sigma(\mu_2)$
Gaussian	$\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$	$\frac{1}{2\sigma\sqrt{\pi}}$	σ	$\sqrt{\pi/2}\,\sigma$
Uniform	$\frac{1}{b-a}$	$\frac{1}{b-a}$	$\frac{b-a}{\sqrt{12}}$	$\frac{b-a}{2}$
Rayleigh	$\frac{x}{\sigma^2}\exp(-x^2/(2\sigma^2))$	$\frac{\sqrt{\pi}}{4\sigma}$	$\sqrt{\frac{4-\pi}{2}}\,\sigma$	$\frac{\sigma}{\sqrt{2\ln 2}}$
Exponential	$\lambda\exp(-\lambda x)$	$\frac{\lambda}{2}$	$\frac{1}{\lambda}$	$1/\lambda$
Chi-square	$\frac{1}{2^{k/2}\Gamma(k/2)}x^{k/2-1}\exp(-x/2)$	$\frac{\Gamma(k/2-1/2)}{4\sqrt{\pi}\,\Gamma(k/2)}$	$\sqrt{2k}$	$\sqrt{2k}$
Erlang	$\frac{\lambda^k}{\Gamma(k)}x^{k-1}\exp(-\lambda x)$	$\frac{\lambda\Gamma(k-1/2)}{2\sqrt{\pi}\,\Gamma(k)}$	$\frac{\sqrt{k}}{\lambda}$	no closed form
Beta	$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$	$\frac{\Gamma(2\alpha-1)\,\Gamma(2\beta-1)}{B(\alpha,\beta)^2\,\Gamma(2\alpha+2\beta-2)}\quad(\alpha,\beta)\geq 1$	$\sqrt{\frac{\alpha\beta}{\alpha+\beta+1}}\frac{1}{\alpha+\beta}$	$\frac{B(\alpha,\beta)}{2\,s^{\alpha-1}(1-s)^{\beta-1}}\quad s=I^{-1}_{1/2}(\alpha,\beta)$
Log-Normal	$\frac{1}{x\,\sigma\sqrt{2\pi}}\exp\left(-\frac{(\log(x)-\mu)^2}{2\,\sigma^2}\right)$	$\frac{\exp(\sigma^2/4-\mu)}{2\,\sigma\sqrt{\pi}}$	$(\exp(2\sigma^2)-\exp(\sigma^2))^{1/2}\exp(\mu)$	$\sqrt{\pi/2}\,\sigma\exp(\mu)$
Birnbaum-Saunders	$\frac{\sqrt{2}}{4\gamma\sqrt{\pi}x}\left(\sqrt{\frac{x}{\beta}}+\sqrt{\frac{\beta}{x}}\right)\exp\left(-\frac{1}{2\gamma^2}(\sqrt{x/\beta}-\sqrt{\beta/x})^2\right)$	$\frac{\exp\left(\frac{2}{\gamma^2}\right)}{4\pi\beta\gamma^2}\left(2\,K_0\left(\frac{2}{\gamma^2}\right)+(2+\gamma^2)\,K_1\left(\frac{2}{\gamma^2}\right)\right)$	$\beta\gamma\sqrt{1+5\gamma^2/4}$	$\sqrt{2\pi}\,\beta\gamma/2$
Student	$\frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\,\pi}\,\Gamma(\nu/2)}(1+x^2/\nu)^{-(\nu+1)/2}$	$\frac{2^{1-\nu}}{\nu^{3/2}}\frac{\Gamma((\nu+1)/2)\,\Gamma(\nu+1/2)}{\Gamma(\nu/2)^3}$	$\frac{\nu}{\nu-2}\quad(\nu>2)$	$\frac{\sqrt{\nu\pi}\,\Gamma(\nu/2)}{2\Gamma((\nu+1)/2)}$
Cauchy	$\frac{1}{\pi\gamma\left(1+\left(\frac{x-x_0}{\gamma}\right)^2\right)}$	$\frac{1}{2\,\pi\,\gamma}$	∞	$\pi\gamma/2$
Lévy	$\sqrt{\frac{c}{2\pi}}\frac{\exp(-c/(2(x-\mu)))}{(x-\mu)^{3/2}}$	$\frac{1}{2\,c\,\pi}$	∞	$\frac{\sqrt{\pi}c\exp(s)}{4s^{3/2}}\quad s=\left(\mathrm{erfc}^{-1}(1/2)\right)^2$

FIG. 5. **Convergence Table: Mean Rank Wins.** The $1/\sqrt{n}$ rate of decrease in sampling error is a broadly applicable result. While perhaps most familiar from characterizing accuracy of a sample mean, it is equally relevant here for addressing random error in signal extraction. For each of the twelve pdfs we list ρ_1, the relevant factor in (3) for mean rank error, the numerator $\sigma(\mu_1)$ for arithmetic mean error, and the numerator $\sigma(\mu_2)$ for median error. The color shading shows which of the three averages converges fastest, with blue denoting an unconditional advantage while green indicates conditional restrictions on parameters in the pdf. We note that the earliest invocation of this for mean rank was Pitman's characterization of the efficiency of the rank spread test in [24].

where

$$\rho_{2j-1}=\int_a^b dx\left(\frac{d^{j-1}f}{dx^{j-1}}\right)^2\ .$$

Recall that, being a pdf, $f(x)$ has units so that ρ_1 has units of the inverse signal amplitude, ρ_3 units of inverse cube of the amplitude, etc. The simplest application of this result is truncation at $\mathcal{O}(\epsilon^3)$, as in (19), by means of which we can precisely defines what is meant in this work by a "weak signal". For example, in Fig. 4a, we chose (as suggested by a referee), two Heaviside functions, $H(x)$ as for $H(x)/2$, differing by a factor of two in the step height. This example is ideally suited to separate magnitude and shape recovery. As Fig. 4a shows, the signals are well approximated by (2) and so the leading order weak signal approximation appears to hold. To check this more precisely we evaluate the cubic correction in (19).

Reading off values from Fig. 4a, we set $\epsilon=1$ and let s_k absorb both form and amplitude information. Removing the mean leaves odd step functions, hence the sums with odd powers of s_i vanish by antisymmetry. This leaves only the single sum $\sum s_i^2$ in (19) which, as the summand is constant, reduces to $N\,s_k^2$ so that the second bracketed term is $4Ns_k^3$. Dividing through by $N\rho_1$, yields a refined mean rank signal estimate, with a correction term of $-4\rho_3 s_k^3/(3!\rho_1)$. While for a general function, the form also influences the correction term, the main operative factors that define a weak signal are thus (i) the cube of the signal amplitude, scaled by, (ii) a proportionality factor ρ_3/ρ_1 reflecting a subtle characterization of the background noise. For the Gaussian pdf with $\sigma=2$ in particular, $\rho_3=(32\sqrt{\pi})^{-1}$ and using $s_k=1/4$ for the lesser step function $H(x)/2$ yields a correction of $-1/768\approx -0.00130$ for $x>0$, and the opposite signed correction for $x<0$. This can be checked by averaging the exact signal less the numerical mean rank signal estimator. On accounting for the sign change at $x=0$, the numerical average of the data over $[-1,1]$ gives -0.000125, in excellent agreement with the analytic cubic correction. The slight difference is due mainly to noise in the numerical mean rank, as the succeeding quintic correction is only 6.1×10^{-6}. For the larger step function, a prediction immediately follows of $8/768=1/96=-0.0104$, which also compares well to the corresponding numerical mean of -0.0101. One sees that judging a weak signal by the traditional signal-to-

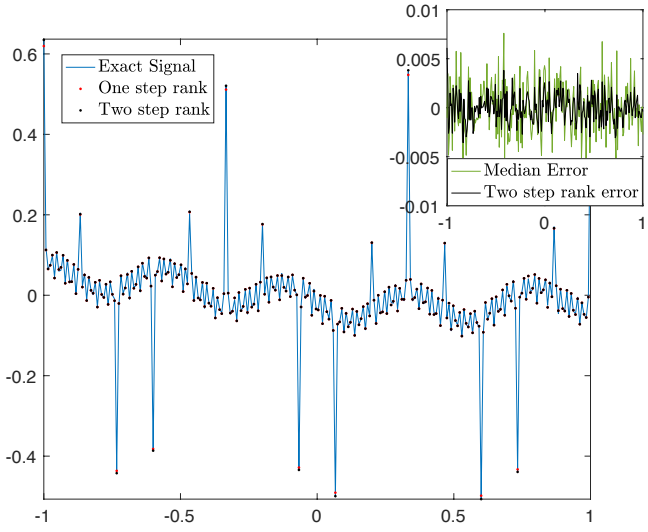

FIG. 6. **A fractal signal retrieval from log-normal noise.** The signal is the Weierstrass function with ($a = 0.99, b = 6$), of Hausdorff dimension 1.9944, and sampled at $N = 256$ points on the unit interval. The noise is log-normal with pdf parameters ($\mu_n = 0, \sigma_n = 4/3$). The mean rank is an average over $n = 4.27 \times 10^6$ time series. Magnitude of the noise notwithstanding, influence of the cubic nonlinearity can be seen in comparing the one (red) and two step (black) computations of mean rank at the highest peaks. The former has $\sigma = 0.00301$ while the latter with $\sigma = 0.00253$ is in excellent accord with a pure noise error estimate of 0.00258. The median error in the inset panel is seen to be about twice as large as the pure noise rank error, borne out both empirically with $\sigma = 0.00481$ and on the basis of the standard error prediction from the Pitman table of $\sigma = 0.00493$.

noise ratio has little relation to the terms in which it is here defined.

The infinite series in (B2) holds for many common C^∞ pdfs whose support is $\mathbb{R}$. Convergence of that series depends on both the signal s_k and the noise, through ρ_{2j-1}. For example, it converges for almost all weak signals for Gaussian and logistic pdfs but is always asymptotic for log-normal noise owing to a controlling factor for ρ_{2j-1} that grows like $\exp((2j-1)^2\sigma^2/4)$. In Fig. 4b, with Cauchy noise as an example, the terms in (B2) up to $\mathcal{O}(\epsilon^7)$ are shown. A second calculation (not shown) gave similar agreement up to and including $\mathcal{O}(\epsilon^{11})$ for a signal in Gaussian noise.

For pdfs with finite or semi-infinite support, (B2) as written can break down at an order that depends on properties of each particular $f(x)$. In all cases the first term in the expansion holds. But while e.g. a uniform distribution on $[0, 1]$ formally yields $\rho_3 = 0$, there is nonetheless a cubic correction to rank whose *form* in terms of s is still that given in (19). However, the expression for the coefficient in place of ρ_3 is not yet known.

2. Relation to other averages

Color shading in the table of Fig. 5 is used to denote the smallest factor controlling convergence as $n^{-1/2}$. Blue denotes unconditional dominance, e.g. for Gaussian noise, the arithmetic mean is best for all σ while for noise with a uniform distribution, mean rank and arithmetic mean are tied. Green indicates dominance over a broad, but not unlimited, range of parameter(s) in the pdf. Specifically: for a chi-square distribution mean rank is best out to $k = 33$. For a beta distribution, mean rank and arithmetic mean divide the (α, β) plane between them, with a wedge symmetrically disposed about $\alpha = \beta$ where the arithmetic mean converges faster, and the complementary region is mean rank. There is in addition a technical point: the expression given in the table for mean rank is limited to $(\alpha, \beta) \geq 1$. Outside of that region, e.g. along the line $\alpha = \beta$ on $[0, 1]$, the governing coefficient remains well defined through some form of regularization yet to be explained. For the Student distribution, one could also shade the arithmetic mean in green. Here is a triple exchange. For ν on $[0, 1.81]$ (including Cauchy noise) the median is best. Then for $[1.81, 61.1]$, mean rank. And finally, recovering the limiting case of Gaussian noise, on $[61.1, \infty]$, it is arithmetic mean.

The random errors in the table are due to noise. As shown above in (19) for mean rank, there are in addition systematic nonlinear corrections, commencing with cubic terms. When the leading (linear) mean rank signal estimate is dominated by random error, this precludes trying to make a cubic correction. But an example here, and also in the main paper, illustrate when and how that nonlinear error can be eliminated. For median and arithmetic mean, no such consideration arises.

3. Error analysis: noise + nonlinearity

Errors owing to (19) and the standard error (3) arise from opposing tendencies with respect to signal amplitude. Characterizing a general signal as $\epsilon\,\mathbf{s}$ where $\mathbf{s}$ itself is normalized in some convenient fashion and the scale varied with ϵ, the random noise errors diminish with increasing ϵ while the cubic deviations increase. The specific form of the error $\mathcal{E}$ (in units of signal, $1/\rho_1$) is given by

$$\mathcal{E}(\epsilon, n) = \frac{1}{\sqrt{12}\,\rho_1}\sqrt{\frac{1}{\epsilon^2\,n} + \frac{\rho_3^2\,\epsilon^4}{3N^2}\sigma_{cubic}^2} \quad , \qquad \text{(B3)}$$

where σ_{cubic} is the standard deviation of the bracketed portion of the cubic term in (19). Note that σ_{cubic}^2 is a rational function of N which, to leading order, is $\mathcal{O}(N^2)$. This cancels against the N^2 dependence of the denominator. For fixed ϵ, the standard deviation $\mathcal{E}(\epsilon, n)$ at first decreases sensibly with increasing n but ultimately saturates at a level controlled by the cubic term.

Further demonstration of mean rank-based signal extraction is seen in Fig. 6, where a small amplitude fractal, nowhere differentiable, Weierstrass function is faithfully recovered, demonstrating that there is no implicit smoothness, or other, constraint on the form of the signal that can be extracted by mean rank. Log-normal noise was used with $(\mu_n = 0, \sigma_n = 4/3)$. Applications may include fractional Brownian motion, etc.

The plot shows extraction of the original signal $\mathbf{s}$ itself. The imposed signal was $\epsilon\mathbf{s}$ with $\epsilon = 0.164$. A companion value of $n = 4.27 \times 10^6$ was used. For this noise

$$\rho_1 = \frac{\exp \sigma_n^2/4}{2\sigma_n\sqrt{\pi}} = \frac{3\exp 4/9}{8\sqrt{\pi}} = 0.3299$$
$$\rho_3 = \frac{\exp(9\sigma_n^2/4)\,(\sigma_n^2+2)}{8\sigma_n^3\,\sqrt{\pi}} = \frac{51\exp(4)}{256\sqrt{\pi}} = 6.1366$$

and from (3) the predicted standard error for pure noise (incorporating a factor of $1/\epsilon$) is 0.00258. With $\sigma_{cubic} = 7.9267$, the estimated cubic error contribution is identical. (The value of n was chosen purposely to bring this about.)

The predicted combined error in (B3) is hence a factor of $\sqrt{2}$ larger, or $\sigma = 0.00365$. The actual error after one step was $\sigma = 0.00301$, i.e. less than expected. However, $\rho_5 = 2.13 \times 10^4$, leading to a quintic term in (19) opposite in sign, and twice as large as, the cubic. This is indicative of the asymptotic character of the expansion for log-normal noise and so one should not be surprised that a more precise estimate for nonlinear error is problematic. In any event, with just one further iteration, the standard deviation is reduced to $\sigma = 0.00253$, almost exactly the predicted value noted above for pure noise with a mean rank estimator.

For comparison in the inset, we show the error for the signal estimate based on the median. The indicated tabular value of the factor for the median standard error in Fig. 5 predicts 0.00493, in close conformity with the actual numerical value of 0.00481. And the pointwise error of the iterated rank solution is indeed visually seen to be about one-half that for the median-based extraction.

That all the nonlinearity has been accounted for is suggested on noting that the correlation of the one-step residual with the original signal is -0.49 (the negative sign pairs with that of the cubic correction) while after two steps that residual correlation is 0.05. Interestingly, the correlation of the final mean rank residual and median residual is 0.84, suggesting an intriguing and unexpected relation between global (along track) rank and strictly local (cross track) median methods.

We note that another approach to signal extraction can be suggested here which has, for unclear reasons, never achieved much prominence. Remarkable efforts have gone into the derivation of countless special purpose algorithms for parameter estimation. In particular we note the more than 200 page compendium [42]. As an example, for the Pareto distribution, defined by

$$f(x;\lambda,\alpha) = \frac{\alpha\,\lambda^\alpha}{x^{\alpha+1}}\,,$$

a particular parameter estimation algorithm appears as Ex. 2 of [42]. Namely, denoting $X = (X_1, X_2, \ldots, X_n)$ where $X_i \sim f(x;\lambda,\alpha)$ we let $X_{(1)} \equiv \min_i X_i$, setting

$$S_n = \sum_{i=1}^{n} \ln(X_1/X_{(1)}) \quad \text{and} \quad T_n = X_{(1)}\left(1 - \frac{S_n}{n(n-1)}\right).$$

Then $T_n \to \lambda$. In comparison, the median approaches $\sqrt[\alpha]{2}\,\lambda$. Though presented as an exercise in parameter estimation, this same algorithm is equally applicable to signal estimation. For example with $(\lambda = 3/2, \alpha = 2)$ the standard deviation as a signal estimator scales as

$$\frac{0.7474}{n} + \frac{0.7009}{n^2} + \mathcal{O}(n^{-3})\,, \tag{B4}$$

standing in sharp contrast to the convergence of the mean rank for signal extraction, which scales as

$$\frac{5\sqrt{3}}{16}\,\frac{1}{n^{1/2}} + \mathcal{O}(n^{-1})\,.$$

(The leading coefficient for median is $3/2\sqrt{2}$.) Clearly (B4) is preferable, with a rate of convergence beyond the typical $n^{-1/2}$. At the same time, it works for precisely one noise pdf, Pareto. We can thus distinguish very general methods of arithmetic mean, median, and mean rank from special purpose, and sometimes very potent, means of signal extraction for large n. But mean rank stands out among *all* these competitors because of its effective extension to $n = 1$.

4. Rank and Symmetries

Here we link the rank expression (2) developed here with earlier results on symmetries of δC [16]. These links open up new vistas and, in particular, motivate another new result for the mean rank as sketched below.

First we note for the discussion on an $M\times 1$ approximation of $\delta\mathcal{C}$ that one might like some guidance on choosing M. For this purpose, the restricted set of singular values

$$\lambda_{j,1} = \frac{1}{2N\sqrt{N-1}}\sqrt{\frac{1}{(1-\cos(\pi/N))\,(1-\cos(j\pi/N))}} \tag{B5}$$

is helpful. In Fig. 7, we plot the log of

$$\mathcal{S}_k = \frac{\sum_{j=N-k}^{N-1} c_{j,1}^2}{\sum_{j=N-k}^{N-1}\lambda_{j,1}^2} \qquad k = 1, \ldots, (N-1)\,.$$

Note the inverted order of summation: from smallest to largest. In the ensemble limit over many realizations of the $\{c_{j,1}\}$ with pure noise, $\mathcal{S}_j = 1$ for all j. The distribution of $\mathcal{S}_j$ is a function of j, approximately Gaussian in the middle, but exponential at the ends. This is taken into account with the dashed lines, which show the 5% and 95% confidence limits. The very clear spike at the

Scheme	$\alpha(N)$	$\{\mathbf{x}\}$	N_p
$N\,I_N$	1	$x_1 = 1 \quad x_{2,3,\ldots N} = 0$	N
Rank	$\frac{N+1}{12}$	$x_k = k \quad k = 1..N$	$N!$
Gnedenko-Koroliuk	$\frac{1}{2N-1}$	$x_{1,2,\ldots N} = 1 \quad x_{(N+1),\ldots 2N} = -1$	$\frac{(2N)!}{(N!)^2}$
Generalized Gnedenko	$\frac{4N(N-k)}{(2N-k)^2(2N-k-1)}$	$x_{1,2,\ldots N} = 1 \quad x_{(N+1),\ldots 2N-k} = -1$	$\frac{(2N-k)!}{N!(N-k)!}$
Coin toss	$\frac{1}{N}$	$x_1 = 1, x_2 = -1$	2^N
N-sided dice	$\frac{N^2-1}{12N}$	$x_k = k \quad k = 1..N$	N^N
Order statistics	$\frac{1}{N^2\,(N+1)}$	$\mathbb{R}$	∞

TABLE I. **Comparison of $\alpha(N)$ for various permutation classes.** Third column shows the elements in the set. The final column, N_p, shows the number of permutations. The first four examples are sampling without replacement. By design to map the statistics of the K-S test to a discrete random walk, the x_i for the Gnedenko-Koroliuk case sum to zero, i.e. the "Brownian bridge" is automatically enforced. The succeeding generalized case simply requires a fixed subtraction of $-k/(2N-k)$. The next two cases are sampling with replacement; a fair two-sided coin and fair N-sided die. Allowed permutations for the two-sided coin range from all heads to all tails, with the Genedenko-Koroliuk and generalized Gnendenko families both embedded within. For completeness, the final case recasts (17) also within this same framework of a universal covariance matrix. (For convenience we make the variable change $n+1 \to N$.) Explicit combinatorial arguments for $\alpha(N)$ in these cases differ greatly in details and none resembles the reasoning in III.B for order statistics. But the common theme is that each final covariance matrix differs only by a simple factor from Σ^* contingent – in the discrete cases – on the key assumption that all permutations in each class are equally likely. Thus, the *ratio* of any two elements of the covariance matrix is universal.

right hand side gives the guidance needed to choose a sensible value of M for both Figs. 2 and 3 in the main text: the base of the very sharp rise.

As presented in the main body of the paper, a suitable basis for signal extraction is (20) from which one can now see the analytic basis of Fig. 6(a,b) in [15]. In particular the mode $\psi_n(y_j)\,\psi_m(x_k)$ is excited by forcing of the form $d\psi_n(y'_j)/dx$ provided m is odd, i.e. even in rank (horizontal) and either even or odd in time. This explains the contribution from group symmetries D_2, D_4 and $C_1^{(y)}$. What then of the other two point groups?

In a striking parallel to (19), we conjecture (with extensive numerical evidence) that

$$\delta\widehat{\boldsymbol{\sigma}} \equiv \frac{1}{N^2\,\nu_1}\,\delta\widehat{\mathbf{r}\circ\mathbf{r}} \;\approx\; \delta\boldsymbol{\sigma} \tag{B6}$$

where $\delta\widehat{\boldsymbol{\sigma}}$ is the rank-based estimate for a weak signal in a form of non-stationary standard deviation (e.g., caused by multiplicative noise [3]) and $\delta\widehat{\mathbf{r}\circ\mathbf{r}}$ is the deviation of mean *square* rank from uniformity (with $\circ$ denoting the Hadamard product). As with $\mathbf{s}$ previously, so too here $\delta\boldsymbol{\sigma}$ represents variation relative to the mean. The constant ν_1 is given by an integral whose dependence on $f(x)$ is still to be determined. Note that contributions to (B6) derive from the $C_1^{(x)}$ and R_2 group components of $\delta\mathcal{C}$, cleanly divided from the $D_{2,4}$ and $C_1^{(y)}$ components that drive the nonstationary mean. Not surprisingly, higher rank moments similarly pair with nonstationary skewness, kurtosis, etc. of the raw data.

Appendix C: $\alpha(N)$ for Covariance Matrices

See Table I for six discrete examples – sampling with and without replacement – that give a covariance matrix differing only in the prefactor $\alpha(N)$ from the universal form Σ^* given by 4 of the main text.

The key defining property of all these discrete examples is that each of the vector spaces is invariant to the interchange of any possible pair of entries (j,k) applied to all members in the space.

Appendix D: Kurtosis corrections for finite N

Departure of modal pdfs from Gaussian form for finite N is important when devising KS-type tests of pure noise. The derivation of the covariance matrix supplies the elements of $\mathbf{U}$ and singular values $\boldsymbol{\sigma}$ (whose squares are pure noise variances), but about the pdf for normalized Fourier amplitude of mode k (the k^{th} row of $\mathbf{V}$) one can infer only that it is zero-mean, symmetric, and has unit variance. A suitable representation for the 1D problem

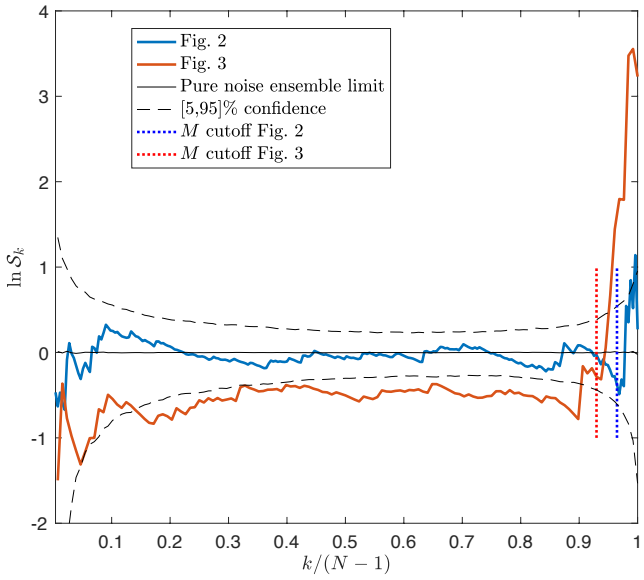

FIG. 7. **Running test of significance:** A running log of the ratio of the cumulative sum of the squares of the expansion coefficients $c_{j,1}$ to the cumulative sum of the square of the corresponding singular values derived from the covariance matrix of $\delta\mathcal{C}$ for white noise. The running ensemble limit is unity. As seen here, the sharp rise at the end indicates the departure from iid noise, albeit only marginally so for Fig. 2 (main text). The spike gives guidance on a suitable choice of M in signal extraction at moderate n (here $n = 1$ and 4 respectively). Black dashed lines show the 5% and 95% confidence limits for pure noise. In passing one notes the noise for Fig. 2 is as "middle of the road" as possible while Fig. 3 (main text) enjoys a decidedly low noise background. This contrast, *which has nothing to do with respective signal amplitudes*, though clear here, can in no way be apprehend from a direct visual comparison of the two plots of the raw data.

is then

$$f(x) \sim \exp\left(\sum_{j=0}^{J} c_j(N) x^{2j}\right) . \tag{D1}$$

For $N \to \infty$ we recover the Gaussian result with $c_0 = -1/2\log(2\pi) = -0.9189$, $c_1 = -1/2$, and remaining $c_j = 0$. Departure from these values is shown in Table II for $N = 32$.

However, for odd N, an idiosyncrasy arises in applying (D1) for the median Fourier mode $(N+1)/2$, whose entries are:

$$\sqrt{\frac{2}{N+1}}\,[1, 0, -1, 0, 1, \ldots] . \tag{D2}$$

For this single mode a distinct expansion applies, as given by the coefficients $c_j^{(2)}$ in Table II. Note that $\lim_{N\to\infty} c_0(N) \to \log(1/\sqrt{2\pi}), c_1 \to -1/2$, and all higher coefficients vanish, i.e., one recovers the standard zero mean unit variance normal distribution. The 2D extension for $\delta\mathcal{C}$ is straightforward.

[1] A. Van der Ziel, *Noise in Measurements* (Wiley, 1976).
[2] P. R. Bevington and D. K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences* (McGraw-Hill, New York, 1992).
[3] R. Frieden, *Probability, Statistical Optics, and Data Testing: A Problem Solving Approach*, Vol. 10 (Springer, 3rd edition, 2001).
[4] R. N. McDonough and A. D. Whalen, *Detection of Signals in Noise* (Academic Press, 1995).
[5] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing* (Cambridge University Press, New York, 2007).
[6] L. L. Scharf, *Statistical signal processing*, Vol. 98 (Addison-Wesley Reading, MA, 1991).
[7] P. L. Krapivsky, S. Redner, and E. Ben-Naim, *A kinetic view of statistical physics* (Cambridge University Press, 2010).
[8] J. Klafter and I. M. Sokolov, *First steps in random walks: from tools to applications* (Oxford University Press, 2011).
[9] M. Manceau, K. Y. Spasibko, G. Leuchs, R. Filip, and M. V. Chekhova, Physical review letters **123**, 123606 (2019).
[10] G. Williams, B. Schäfer, and C. Beck, Physical Review Research **2**, 013019 (2020).
[11] R. C. Dalang, T. Humeau, *et al.*, The Annals of Probability **45**, 4389 (2017).
[12] T. Srokowski, The European Physical Journal B **85**, 65 (2012).
[13] D. K. C. MacDonald, *Noise and fluctuations: an introduction* (Courier Corporation, 2006).
[14] R. Wittje, Physics Today **73**, 42 (2020).
[15] G. Ierley and A. Kostinski, Physical Review X **9**, 031039 (2019).
[16] G. Ierley and A. Kostinski, Physical Review E (2020).
[17] M. Hollander, D. A. Wolfe, and E. Chicken, *Nonparametric Statistical Methods*, Vol. 751 (John Wiley & Sons, 2013).
[18] E. L. Lehmann and H. J. D'Abrera, *Nonparametrics:*

j	$c_j^{(1)} \quad (N=32)$	$c_j^{(2)} \quad (N=33)$
0	-0.93029000	-0.94023252
1	-0.47720819	-0.45901106
2	$-0.37949305 \times 10^{-2}$	$-0.56800443 \times 10^{-2}$
3	$-0.15309200 \times 10^{-4}$	$-0.229661486 \times 10^{-3}$
4	$\mathcal{O}(10^{-8})$	$\mathcal{O}(10^{-7})$

TABLE II. Sample expansion coefficients for (D1). For a Gaussian distribution $J = 1$, $c_0 = -1/2\log(2\pi) = -0.9189$, and $c_1 = -1/2$. Entries here show the departure from this. With even N the first column, $c_j^{(1)}$, applies for all k. The negative definite coefficients show faster decay than a Gaussian, consistent with a kurtosis of $\kappa = 2.9152$ vs. $\kappa = 3$ for a normal distribution. As noted in the text, the second column, $c_j^{(2)}$ applies to a single mode, $k = (N+1)/2$ when N is odd. Here the kurtosis deficit is yet more marked at $\kappa = 2.8382$. For the listed values of $c_j^{(1)}$, $\int_{-\infty}^{\infty} f(x)\,dx = 1.000003$ and $\sigma = 1.00006$ compared to exact values of unity for each. Comparable accuracy obtains for $c_J^{(2)}$.

Statistical Methods Based on Ranks. (Holden-Day, San Francisco, 1975).

[19] S. M. Kay, *Fundamentals of Statistical Signal Processing Volume I: Estimation Theory; Volume II: Detection Theory; Volume III: Practical Algorithm Development.* (Prentice-Hall, 1998).

[20] A. Papoulis, *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill, 1984).

[21] The arithmetic mean is not suitable also when the data traces bear differing units, e.g, spaceborne radars, lidars and radiometers observing the same storm.

[22] S. M. Stigler, *The seven pillars of statistical wisdom* (Harvard University Press, 2016).

[23] H. Gellibrand, *A discourse mathematical on the variation of the magneticall needle, Together with Its Admirable Diminution Lately Discovered* (1635).

[24] E. G. J. Pitman, *Notes on Non-Parametric Statistical Inference*, Institute of Statistics Mimeo Series #27 p.56 (North Carolina State University, Raleigh NC, 1949).

[25] P. J. Bickel and K. A. Doksum, *Mathematical statistics: Basic Ideas and Selected Topics* (Simon and Schuster, NY, 1977) pp. 352–353.

[26] I. Gikhman, A. Skorokhod, and M. Yadrenko, *Probability Theory and Mathematical Statistics* (Vyshcha Shkola, in Russian, Kiev, 1988).

[27] C.-Z. Gao, C.-B. Zhang, C.-X. Yu, X.-F. Xu, S.-C. Wang, C. Yang, Z.-Y. Hong, Z.-F. Fan, and P. Wang, Physical Review E **102**, 022111 (2020).

[28] R. Chicheportiche and J.-P. Bouchaud, Physical Review E **86**, 041115 (2012).

[29] A. Renyi, *Probability Theory* (Dover, Mineola, New York, 2007).

[30] Y. Dodge and D. Commenges, *The Oxford dictionary of statistical terms* (Oxford University Press on Demand, 2006).

[31] For notational economy, we use $\mathcal{C}$ to denote the ensemble cdf as well as the empirical one.

[32] S. Ross, *A First Course in Probability* (Prentice-Hall, Upper Saddle River, 2009).

[33] The pdf for Fourier amplitudes of mode k (the k^{th} row of $\mathbf{V}$) tends to a zero-mean unit variance normal distribution only in the limit as $N \to \infty$, as documented in Appendix D.

[34] V. Voinov and M. Nikulin, *Unbiased Estimators and Their Applications* (Nauka Moscow, in Russian, 1989).

[35] W. Feller, *An Introduction to Probability Theory and its Applications V.2* (Wiley, 1966) p. 38.

[36] M. B. Priestley, *Spectral analysis and time series* (Academic press, 1981).

[37] Any pdf defined on a compact domain can be simply padded with zeroes.

[38] S. R. K. Vadali, P. Ray, S. Mula, and P. K. Varshney, IEEE Transactions on Communications **65**, 1061 (2017).

[39] While one might try to seek a zero of the vector-valued function $\boldsymbol{\varepsilon}$, owing to the discreteness of rank, no signal will exist that can exactly zero each component. However, the value of $\sigma_{\boldsymbol{\varepsilon}}$ can be minimized with the Nelder-Mead algorithm, which allows for the discrete structure induced by rank. This was done for Fig. 3 with a mild improvement to $\sigma_{\boldsymbol{\varepsilon}} = 0.001$. But in practice, simple iteration suffices.

[40] M. Herrero-Collantes and J. C. Garcia-Escartin, Reviews of Modern Physics **89**, 015004 (2017).

[41] C. Gabriel, C. Wittmann, D. Sych, R. Dong, W. Mauerer, U. L. Andersen, C. Marquardt, and G. Leuchs, Nature Photonics **4**, 711 (2010).

[42] V. Voinov and M. Nikulin, *Unbiased Estimators and Their Applications: Volume 1: Univariate Case*, Vol. 263 (Springer Science & Business Media, 2012).