Detection of unknown signals in arbitrary noise

Glenn Ierley [*]
Department of Mathematical Sciences, Michigan Technological University, Houghton, Michigan 49931, USA and Scripps Institution of Oceanography, University of California San Diego, San Diego, California 92093, USA

Alex Kostinski [†]
Department of Physics, Michigan Technological University, Houghton, Michigan 49934, USA

(Received 14 April 2020; accepted 1 September 2020; published 29 September 2020)

We devise a simple method for detecting signals of unknown form buried in any noise, including heavy tailed. The method centers on 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 data points in a time series. Signals of any kind distort this uniformity. Such distortions are captured by rank-time cumulative distributions permitting all-purpose efficient detection, even for single time series and noise of infinite variance.

DOI: 10.1103/PhysRevE.102.032221

I. INTRODUCTION

Signal detection is ubiquitous in physics experiments, be it a passage of an elementary particle, arrival of a gravitational wave, or a radar echo. However, instrumental noise, clutter, unwanted fluctuations, and signal interference are also ever present. Thus, signal separation from noise is an essential part of any experiment. Although the literature on this problem is vast and contains a bewildering variety of solutions, these are mostly specialized, e.g., lock-in amplifiers perform remarkably in detecting low-level signals obscured by noise, but the signal form must be known. Here we address situations where signal form is not known nor is the noise of a conventional variety. We propose a rather general yet simple approach to signal-noise decomposition and apply it to the detection of a variety of signals.

To that end, consider n "time" series (i.e, any integer-indexed serially ordered data), each composed of N real-valued elements. Such n time series may represent n seismic or electroencephalogram (EEG) detectors, n hot wires in a turbulent flow, multichannel radar measurements, multicolor satellite measurements, n individual stock prices in a portfolio, etc., each acquiring N realizations (entries of a time series) of a noisy fluctuating process. Each series can be viewed as a signal (to be detected) buried in noise. However, despite its ubiquity, noise is difficult to define precisely yet generally (e.g., Refs. [1–3]) because, as the adage goes, one man's noise is another man's signal [4]. In particular, signal detection and processing literature has been dominated, for over a century, by the additive Gaussian white noise model and the least-squares approach of maximum likelihood [5–7]. Yet pronounced fluctuations associated with "black swan" events and heavy-tailed (algebraic) distributions have become commonplace [8,9], ranging from photonics to air pollution [10–13]. One example is Cauchy noise, whose pdf is

$$p(x)=\frac{1}{\pi\sigma}\left(1+\left[\frac{x-\mu}{\sigma^2}\right]^2\right)^{-1}, \tag{1}$$

possessing infinite variance. It belongs to a stable family of distributions, meaning that the sum of Cauchy-distributed random variables is also Cauchy-distributed, and with the same parameter [14]. Thus, the central limit theorem is violated and no benefit in measurement accuracy is obtained from averaging over n independent measurements, that is, no $1/\sqrt{n}$ improvement. Yet, Cauchy noise is ubiquitous, e.g., arising in the optics of Huygens' principle, as a quotient of two normally distributed variables [15], the lighthouse problem, [16], signal detection [17], synthetic aperture radar speckle, [18], etc.

The purpose of this work is to propose a method that largely bypasses the dependence on the probability distribution of noise. This is accomplished by working solely with ranked (sorted by magnitude) data[1] and introducing a mapping from raw data to rank-order space. A simple distribution-invariant characterization of independent identically distributed noise is then given. Rank-based statistics have been used before, e.g., in binary decision applications of mathematical statistics and estimation theory [20,21] and have even been applied to signal detection, e.g., the Wilkinson sum test or the Kendall τ rank correlation coefficient [22] but these were strictly one dimensional (1D), did not employ symmetries, and so are far narrower in scope than the proposed method to be explained next.

To that end, for each of the n time series, sort the data by magnitude in ascending order and record those N ranks in

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

[1]Rank is not always uniquely defined as ties are possible [19]. Throughout this work we break the ties by either adding tiny white noise or assigning fractional ties.

the corresponding N "time slots" or indices, e.g., Ref. [5]. The ranks are integer valued. Observe that because of invariance with respect to *reshuffling*, only when such data are generated by stationary white noise is each permutation of N ranks among the N indexed "time" slots equally likely, occurring with probability $1/N!$. Guided by this perfect duality (indistinguishability and independence) of rank (r) and time index (t) for this independent and identically distributed noise, we introduce a mapping to the rank-time (r, t) plane. Then, for independent and identically distributed noise, as $n \to \infty$ rank-time duality is expressed by the fact that the 2D discrete probability density (mass) function (pmf) is jointly uniform, i.e., $p(r, t) = U(1, N) \times U(1, N)$. This property of $p(r, t)$ uniquely characterizes independent and identically distributed noise in the rank-time plane, akin to the fundamental assumption of equilibrium statistical mechanics: All possible microstates (rank-time partitions) are equally likely.

This characterization of stationary white noise defines the *absence of a signal*, delivers a simple signal-noise decomposition, and allows one to understand signals broadly as departures from pure randomness. This approach can be used to test random number generators, e.g., Refs. [23,24], quantify instrumental errors, and so on, but here we focus on signal detection. Signals can be deterministic or random, steady or transient, and the same framework holds. We illustrate the method throughout on various heavy-tailed distributions, including Cauchy, even when only a single realization (one time series, $n = 1$) is available. To the best of our knowledge, there are no comparable approaches for detection of unknown signals.

A uniform pdf $p(x, y) = 1$ over a square is the continuous analog to $U(1, N) \times U(1, N)$, for N data points, and the associated cumulative distribution function (cdf) is simply $\mathcal{C}(x, y) = xy$. Given the cdf for a sample time series, the natural metric for signal detection is the deviation cumulative distribution function, implemented on the discrete lattice as

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

where $p_{i,j}$ is the probability mass function (pmf) and the symbol δ indicates deviation.[2]

Signals generally cause $\delta\mathcal{C}$ values to deviate from zero. The simplest scalar measure then is the arithmetic average over all matrix entries, $\overline{\delta\mathcal{C}}$, calculated as

$$\overline{\delta\mathcal{C}} \equiv \frac{1}{(N-1)^2} \sum_{j=1}^{N-1} \sum_{k=1}^{N-1} \delta\mathcal{C}_{j,k}. \tag{3}$$

This metric quantifies a trend in a particular single realization of N samples but, due to sample variability, $\overline{\delta\mathcal{C}}$ is a random variable; zero mean for the pure noise ensemble.

After some algebra on (2) and (3), one obtains for a single input data vector of N elements that

$$\overline{\delta\mathcal{C}} = \frac{1}{(N-1)^2} \left[\frac{1}{N} \sum_{k=1}^{N} k\, r_k - \frac{(N+1)^2}{4} \right], \tag{4}$$

where the kth raw entry maps to rank r_k. To relate (4) to a trend, note that for any monotonically increasing signal, $r_k = k$ and so (4) attains its maximum, namely

$$\max \overline{\delta\mathcal{C}} = \frac{1}{12} \frac{N+1}{N-1}. \tag{5}$$

The distribution for $\overline{\delta\mathcal{C}}$ is discrete for all N. The number of distinct values of (4) for pure noise, valid for $N \geqslant 4$, is[3]

$$a(N) = \binom{N+3}{3} + 1 = \frac{N^3 - N}{6} + 1. \tag{6}$$

The $\overline{\delta\mathcal{C}}$ distribution approaches the Gaussian one for large N with an asymptotic expansion for the standard deviation:

$$\sigma_{\overline{\delta\mathcal{C}}} \sim \frac{0.0826}{\sqrt{N}} \left[1 + \frac{2.9499}{N} + \ldots \right]. \tag{7}$$

In practice, (7) is quite adequate as the standard of measure against which to judge the statistical significance of $\overline{\delta\mathcal{C}}$.[4] Finally, for general n, the count of distinct values becomes $n\,a(N) - n + 1$ and (7) is multiplied by $1\sqrt{n}$.

Figure 1 illustrates application of another $\delta\mathcal{C}$ metric to detection. The time series in Fig. 1(a) contains four Gaussian pulses, $5\,e^{-4t^2}$ (typical shape in lasers or radar transmission, etc.), embedded in Cauchy noise. The detection is always blind as the shape is assumed unknown throughout. While the variance of Cauchy noise is formally infinite, the nominal sample-based signal-to-noise power (variance) ratio (SNR) is 4.4×10^{-4}. A sliding window of 41 samples ($\Delta t = 1/14$) is used to compute $p(r, t)$ and then the *deviation* cumulative distribution function $\delta\mathcal{C}$. Plotted in Fig. 1(b) is the second scalar detection metric derived from $\delta\mathcal{C}$: the rms value of a particular group component of the deviation cdf, as described further below. The four prominent spikes line up perfectly with the true pulse locations as demarcated by the red line. The black dashed line indicates our detection threshold, set at the 99.9% confidence limit for pure noise. Neither misses nor false alarms occur.

To illustrate the importance of representing 1D time series in the 2D rank-time plane, contrast the proposed approach with the widely used matched filter where a conjugate replica of the (known) waveform is used for detection [6]. The matched filter has a major advantage as the signal form is known but here fares poorly: A threshold that captures two pulses yields 4 false alarms, for three pulses 32 false alarms,

[2]To avoid cumbersome notation, we shall use $\mathcal{C}$ to denote either the ensemble cdf or the empirical one, the latter based on finite n including a single realization, depending on the context.

[3]The On-Line Encyclopedia of Integer Sequences gives this expression for the number of distinct values taken by the entropy for permutations $\pi(k)$ of $[1..N]$. It also gives the number of distinct, equally spaced, values taken by the sum of $k\,\pi(k)$, or $k\,r_k$ in our terms.

[4]For a thorough application to the key problem of linear regression see Appendix A.

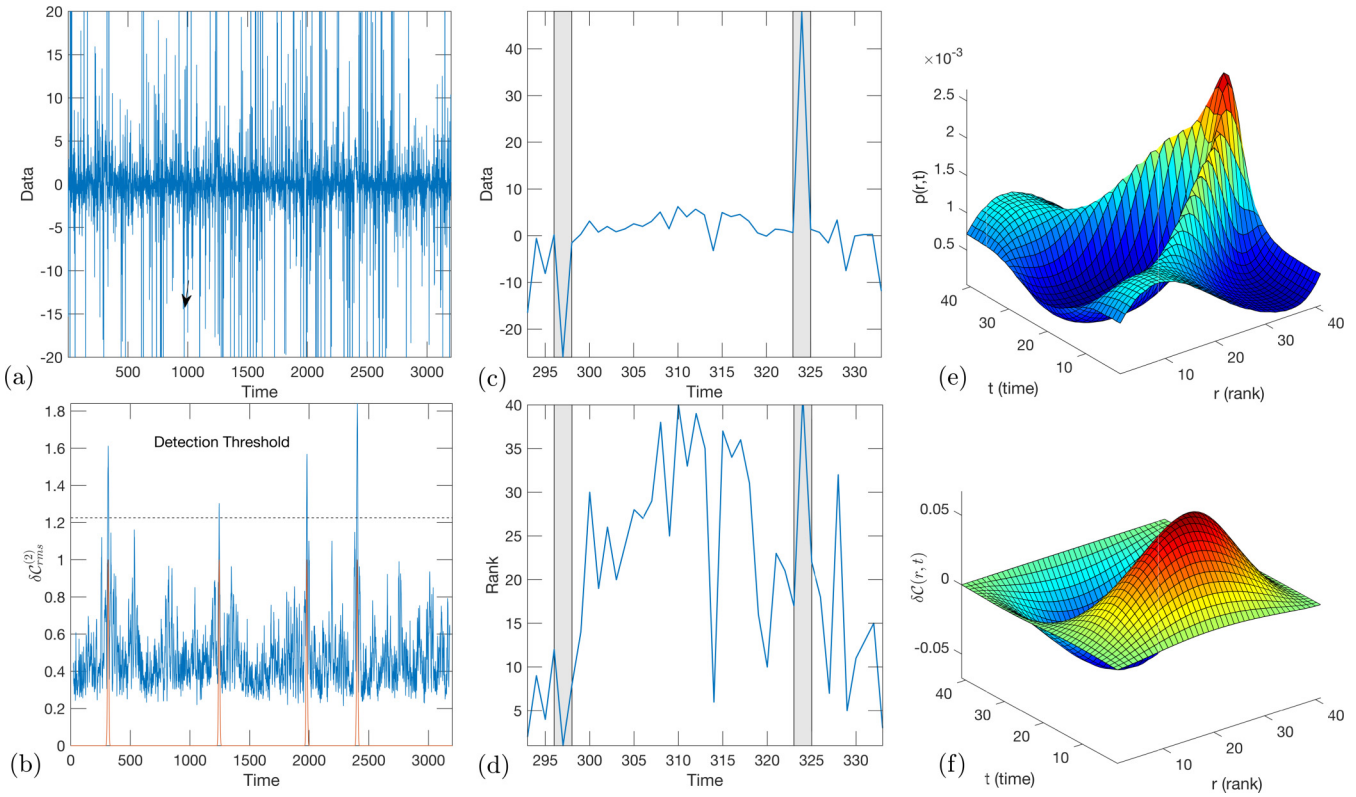

FIG. 1. Proposed method illustrated on a Gaussian pulse $S(t) = 5\ \exp(-4t^2)$ buried in Cauchy noise. (a) Time series of Cauchy noise with four (unknown) Gaussian pulses to detect. Vertical scale is clipped to show detail. (b) The detection metric (see text) is shown, where the four spikes line up with the true pulse locations (red). The detection threshold is marked by the black dashed line. (c) Raw data for the sliding window of $N = 41$ centered about the first detected pulse, plotted full scale. (d) The raw data of panel (b) converted to rank from 1 to 41 so, e.g., the sharp negative spike highlighted in gray at left maps to 1 while the large positive outlier in gray at right maps to 41. This collapse of outliers permits the underlying signal to be detected. Ratio of variance or signal-to-noise power ratio is technically zero for the Cauchy noise but the nominal, sample-based value is $\approx 4 \times 10^{-4}$; (e) $p(r, t)$, rank-order probability mass function (pmf) of signal plus noise ($n = 10^6$), with the Gaussian "ridge" seen in the maximum of p vs. t. (f) The $\delta\mathcal{C}$ of signal plus noise (see text), nearly exact with $n = 10^6$.

and for all four pulses an overwhelming 69 false alarms. One can improve the matched filter with a 2D rank-time representation, using a dot product with the ensemble average $\delta\mathcal{C}$ (normalized and reshaped as a unit vector) of Fig. 1(e). While this then does match our detector in Fig. 1(b), it requires prior knowledge of the signal form.

For weak signal detection in additive Cauchy noise, the form of the locally optimal detector is given in Ref. [17] and denoted there as $\Lambda(\mathbf{x})$. The detector is nonlinear in the data $\mathbf{x}$ and, as with matched filters, assumes the form of the signal is known. The detector further relies on the specific assumption of Cauchy noise with a known location parameter. Despite the special purpose design and the *a priori information*, for the time series in Fig. 1(a), a threshold for $\Lambda(\mathbf{x})$ that just captures all four pulses still gives two false alarms, thereby underperforming the proposed general method.

One can discern the Gaussian-like ridge in $p(r, t)$ of Fig. 1(e), a signal-induced distortion of the uniform pmf of pure independent and identically distributed noise. Note that parity with respect to both arguments (r, t) of $\mathcal{C}$ flips from that of $p(r, t)$ (e.g., even to odd) as indicated by the mixed derivative relation for the continuous analog between cdf and pdf:

$$p(x, y) = \frac{\partial^2 \mathcal{C}(x, y)}{\partial x \partial y}. \tag{8}$$

From Fig. 1(f) we see that the area integral of $\delta\mathcal{C}$ vanishes because $\delta\mathcal{C}$ is even in rank and odd in time. Any 1D $f(x)$ can be decomposed into odd and even component as $1/2[f(x) + f(-x)] + 1/2[f(x) - f(-x)]$ but how does one extend the notion of parity to 2D? A natural extension to consider is each argument at a time, i.e.,

$$f(x, y) = f_0 + f_1 + f_2 + f_3, \tag{9}$$

where:

$$\begin{aligned} f_0 &= [f(x, y) + f(-x, y) + f(x, -y) + f(-x, -y)]/4 \\ f_1 &= [f(x, y) + f(-x, y) - f(x, -y) - f(-x, -y)]/4 \\ f_2 &= [f(x, y) - f(-x, y) + f(x, -y) - f(-x, -y)]/4 \\ f_3 &= [f(x, y) - f(-x, y) - f(x, -y) + f(-x, -y)]/4, \end{aligned}$$

as given in, e.g., Ref. [25], motivated by the discussion in Ref. [26]. However, this seemingly straightforward odd or even decomposition does not quite suit our purposes here.

II. SYMMETRIES OF NOISE VIA $\delta\mathcal{C}(r,t)$

Even in a featureless white noise, transient patterns and spurious trends arise by chance. This *sampling variability* in the r,t plane must be characterized in order to detect weak signals buried in noise. In contrast to signals, the hallmark of pure noise is rank-time duality. For the ensemble limit this duality is seen for example in that the pdf $p(r,t)$ is invariant with respect to rotation by $\pi/2$ in the (r,t) plane. For fluctuations about that equilibrium, duality means that the probability of any perturbation $\delta p(r,t)$ is the same as for $\delta p(t,r)$, suggesting the relevance of symmetries.

We now proceed to extend the notion of parity to two dimensions based on symmetries and the associated 2D point groups. The f_1, f_2, f_3 terms in (9) already achieve this association, corresponding to C_1^x, C_1^y, R_2 symmetries, respectively, that is, reflections (about x and about y axes) and a rotation. But the 2D point group characterization is completed by two dihedral elements: D_4, representing the eight symmetries (rotations and reflections) of the $N\times N$ square [27], and D_2, the reduced four symmetries of the rectangle, and these are merged in the single remaining term f_0. The proper expression of rank-time duality requires these be distinguished. To that end, we split f_0 as:

$$f^{(D_4)} = [f(x,y)+f(-x,y)+f(x,-y)+f(-x,-y) + f(y,x)+f(-y,x)+f(y,-x)+f(-y,-x)]/8, \tag{10}$$

$$f^{(D_2)} = [f(x,y)+f(-x,y)+f(x,-y)+f(-x,-y) - f(y,x)-f(-y,x)-f(y,-x)-f(-y,-x)]/8, \tag{11}$$

with the first term capturing the key rank-time invariance of pure noise. Expanding $\delta\mathcal{C}$ in place of the general function f above, one thus obtains five symmetry components for the deviation cdf. This decomposition is orthogonal, i.e., for continuous functions, the integral of any dissimilar pair over the symmetric domain vanishes.[5] This five-term decomposition (D_4 and D_2, reflections $C_1^{(x)}$ and $C_1^{(y)}$, and rotation R_2) forms the basis for improved detection as the signal-induced broken rank-time symmetry affects the five group components differently.

For example, the Gaussian pulse in Fig. 1 is even in time and so then is the signal-induced deviation from the uniform $p(r,t)$ of noise. From (8), the resulting $\delta\mathcal{C}$ is odd in time, as seen in Fig. 1(f). The latter is observed to be even in rank. This pairing of even in rank (x) and odd in time (y) indicates use of the symmetry-matched even-odd $C_1^{(y)}$ component of $\delta\mathcal{C}$ for such detection. The strongest response of $C_1^{(y)}$ is to a parabolic signal in time. Similarly, $\delta\mathcal{C}$ of even-even D_4 symmetry for pure noise is most readily perturbed by a signal linear in time (or any monotonic transformation thereof), see Appendix C for details. For signals of unknown parity, or none at all, one monitors all five components looking for significant departures from pure noise.

As this work is devoted to (threshold) detection, one needs a simple scalar metric based on $\delta\mathcal{C}$. Furthermore, to set confidence limits for detection the sampling variability of pure noise must be characterized for each of the five components. The mean value of $\delta\mathcal{C}$ introduced at (3) is available only for D_4 as the remaining group averages vanish identically. Another natural detection metric (statistic) is that already used in Fig. 1, the rms value. For $\delta\mathcal{C}$ itself this is calculated as

$$\delta\mathcal{C}_{\text{rms}} \equiv \sqrt{\overline{\delta\mathcal{C}^2}} \equiv \frac{1}{(N-1)}\left[\sum_{j=1}^{N-1}\sum_{k=1}^{N-1}(\delta\mathcal{C}_{j,k})^2\right]^{1/2} \tag{12}$$

and similarly for the five group components of $\delta\mathcal{C}$.

The resulting five rms values are evaluated to yield the cdfs plotted in Fig. 2. Results for $(N=320, n=100)$ (blue curves) are from a Monte Carlo run of 10^5 trials. The $\delta\mathcal{C}^{(2)}_{\text{rms}}$, for reflection group $C_1^{(y)}$, is used for detection of Gaussian signals in the bottom trace in Fig. 1(a) (y is time).

The $N=10, n=1$ results (red curves) are exact as we take advantage of the fact that, once reduced to rank, there are precisely $N!$ permutations of the integers from 1 to N, each equally likely. The group cdfs for each rank permutation are readily computed by direct enumeration of all $10! = 3\,628\,800$ microstates. This is equivalent to the true ensemble limit as $n\to\infty$. The distributions in Fig. 2 for $(N=10, n=1)$ stand apart mainly due to $n=1$. Increasing just to $n=2$ at fixed N notably reduces the gap in D_2 and $C_1^{(x,y)}$ in Fig. 2(a) due to a cancellation that is absent when $n=1$.[6]

Figure 2(b) is a detailed exploration of the rms distribution for $C_1^{(y)}$ in order to further interpret the detection of the Gaussian pulses of Fig. 1 where the $C_1^{(y)}$ was used. Plotting vs. x^2 clarifies the asymptotic form, and a fit for slope and intercept yield the dashed line, thereby extending the reliable range for meaningful estimates. All four pulse detections are seen to lie well beyond the threshold.

III. DETECTION EXAMPLES FOR A SUITE OF SIGNALS

To showcase this single unified all-purpose detection scheme for any signal of *unknown form* buried in heavy-tailed noise, four disparate examples are given in Fig. 3. To the best of our knowledge, no method of such breadth is available elsewhere. The four signals are the chirp, the sinc function: $\sin(x)/x$, evaluated on $x=[-4,4]$, the Weierstrass function (interpolating between regular and algebraic tails; see Ref. [9], p. 13): $f(x)=\sum_{n=0}^{\infty} a^n\cos(b^n\pi x)$, $a=0.99$ $b=6$, and the Devil's staircase [28].

To embed weak (unknown) signals in a various types of noise we use a convenient one-parameter (ν) noise distribution

[5] The discrete equivalent is that the double sum over the corresponding Hadamard product vanishes.

[6] The linearity of the mean value implies that ensemble average of $\delta\mathcal{C}$ (over n) and matrix average (over N) commute; noise cancellation yields $1/\sqrt{n}$ regardless of the order of operations. By contrast, for rms values the ensemble average of the mean square for pure noise does not converge to zero, only the mean square of the ensemble average.

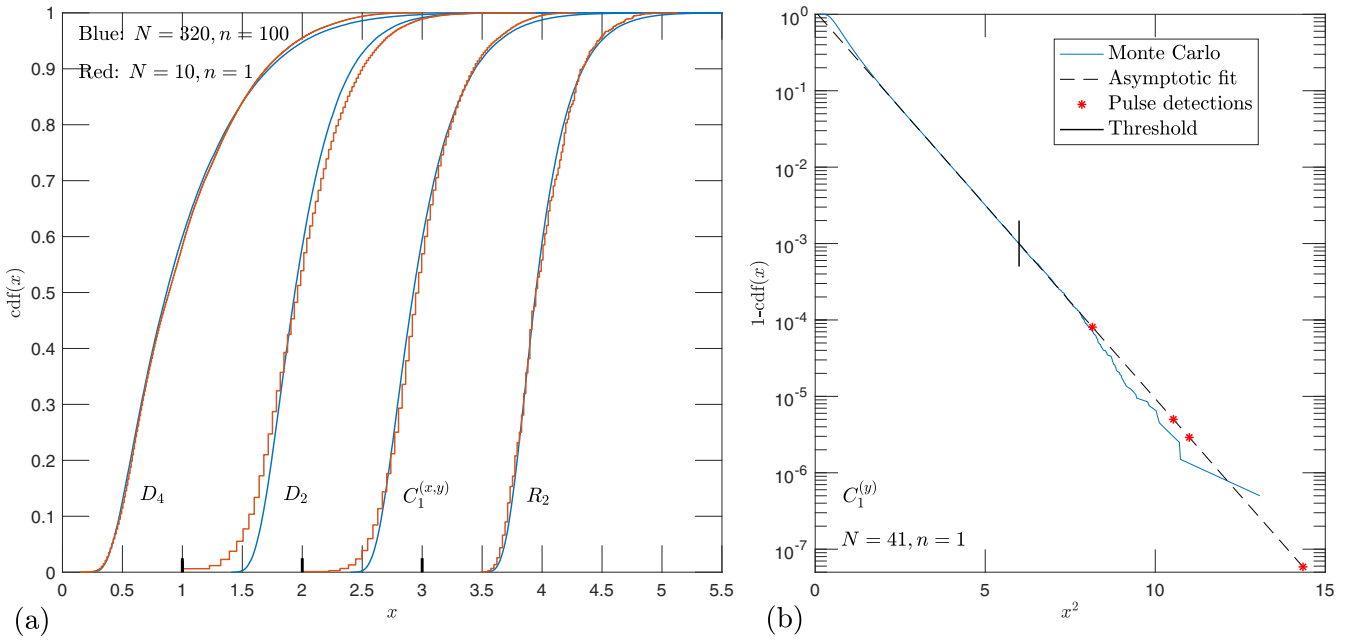

FIG. 2. (a) The cdfs of rms values (scaled by their means) of the five-term orthogonal expansion of $\delta\mathcal{C}$ for pure noise: dihedrals D_4 and D_2, reflections $C_1^{(x)}$ and $C_1^{(y)}$, and rotation R_2. $C_1^{(x)}$ and $C_1^{(y)}$ share a cdf so only four curves are plotted and offset by 1 at the indicated bold tick marks for clarity. The staircases for the coarse discretization of $N = 10$ and a single trial ($n = 1$) are surprisingly close to the smooth curves for ($N = 320, n = 100$). The D_4 cdf is notably broader and, in the case of $N = 10$, smoother than the other four. (b) As in Fig. 1(b), the rms value for the $C_1^{(y)}$ component of $\delta\mathcal{C}$ for ($N = 41, n = 1$) is shown. $C_1^{(y)}$ works well because the signal is even in time when the sliding window is centered at the peak. There is a noise-induced spike at about $t = 500$ in the plotted test statistic (Fig. 1) which lies near the indicated detection threshold, while the weakest pulse near $t = 1250$ lies just above that threshold. Seemingly, this suggests that, in a long run, false negatives and positives are commonplace. Not so, as panel (b) shows, by mapping the test statistic to the complement probability. The $\log[1 - \mathrm{cdf}(x)]$ approaches a simple asymptote (dashed line), indicating that the cdf itself approaches unity as $\exp(-0.0759\,x^2)$ for $x^2 > 2.5$. The threshold in Fig. 1(b), marked by the vertical bar, maps to a probability of 10^{-3} while the weakest of the four pulses (leftmost red star) is more than 10-fold less likely. Based on the asymptotic expansion, the strongest pulse detection is seen to have a probability $<10^{-7}$, well beyond the range of a reasonably accessible Monte Carlo run.

(Student) [5]:

$$p(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\,\nu}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \tag{13}$$

spanning the gamut from the normal distribution in the limit of $\nu \to \infty$ to the Cauchy distribution ($\nu = 1$).

Detection performance for the four examples of Fig. 3 depends on signal amplitude and the number n of available traces. In Figs. 3(a)–3(c), $n = 1$ while for Fig. 3(d) we illustrate a case with $n = 16$. For ease of comparison, signal amplitudes were selected to yield similar performance "DET" curves for Cauchy noise (blue).[7] Typically the role of noise in detection is accounted for through its variance in the denominator of the SNR. However, as earlier noted, variance is formally infinite for Cauchy noise and the finite samples here yield implausibly small values, with single realizations yielding SNR values in the range 10^{-4}–10^{-9}. This ambiguity of SNR in relation to rank-based detection is due to rank dispensing with all information on absolute magnitude. But, at least for $\nu = 4$ with $\sigma^2 = 2$, we can note the conventional and well-defined four values of SNR given in the caption. These span an order of magnitude, with small values for Figs. 3(a) and 3(c), and larger for Figs. 3(b) and 3(d), depending mostly on spectral content. Simply put, for a fixed signal variance, lower-frequency signals are easier to detect.

Figure 4(d) displays the Weierstrass function, whose spiky character is itself already sufficiently noiselike that, in spite of plotting $n = 16$ successive time traces in the second inset, no hint of any signal is discernible by eye. But note the dots overlain on the signal in the first inset. These are the (scaled) mean rank averaged over n. And just as the Gaussian pulse emerged in Fig. 1(d) for $n = 40$, so, too, here observe both the general pairing of the 10 most prominent peaks in the signal with outliers in rank (marked by the circled points), as well as the similarity in overall weak downward drift. Detection for this signal rests on the resulting anomalous value of the rms D_4 component of $\delta\mathcal{C}$.

For detection of four Gaussian pulses in Fig. 1, a perfect detection threshold was contrasted with performance of the locally optimal detector, $\Lambda(\mathbf{x})$ of Ref. [17]. Because $\Lambda(\mathbf{x})$ assumes the signal embedded in Cauchy noise is of *known* form, its optimal window size follows. For $\delta\mathcal{C}$, that window will generally be wider to benefit from the added statistics of arbitrary noise. Furthermore, as the signal in this work is assumed *unknown*, one cannot prescribe a suitable window *a priori* but rather must discover it during the detection process based on $\delta\mathcal{C}$. Nevertheless, to facilitate a simple qualitative comparison in Fig. 1, we used a fixed window size given by

[7] See Appendix B for a simple analytic model for these DET curves.

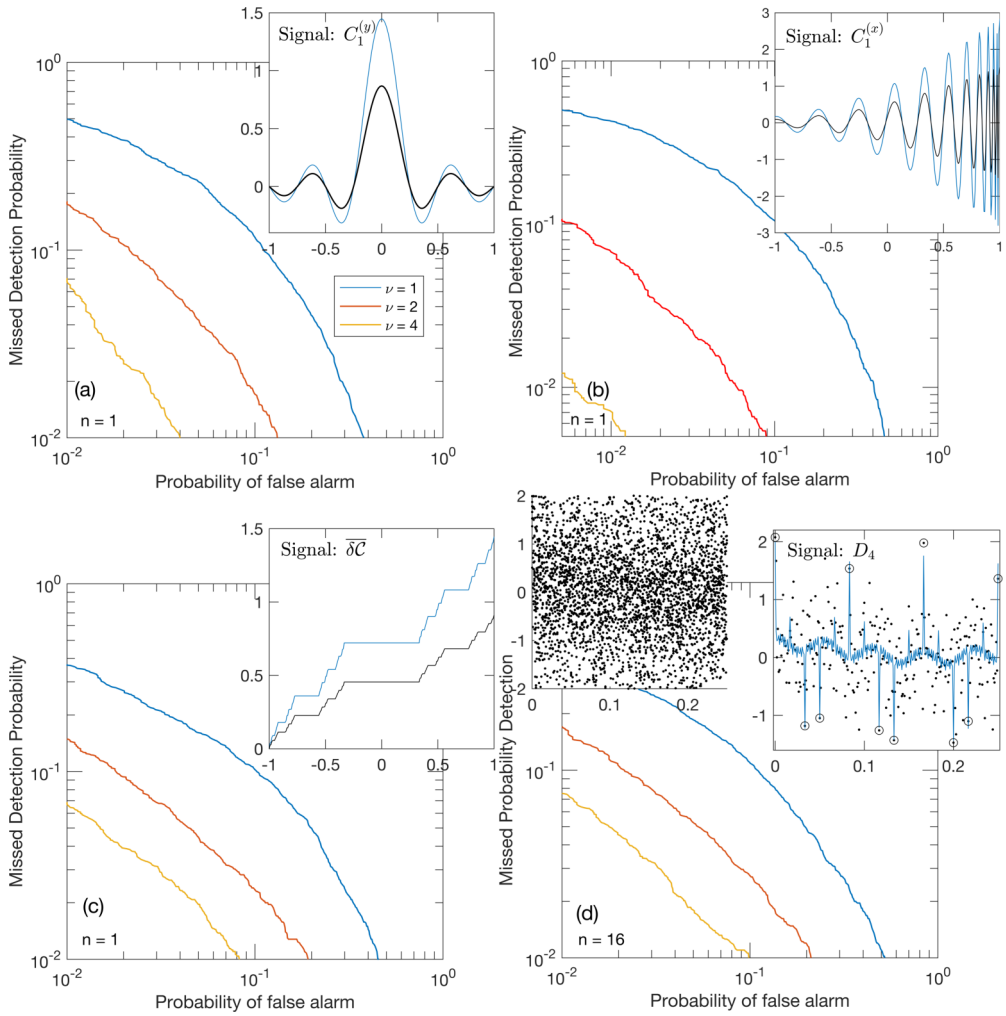

FIG. 3. General method of detecting an arbitrary signal of *unknown form* in any noise: a demonstration on four disparate signals, with $N = 256$. Noise is generated via (13) with the color-coded ν values for all panels as indicated in the legend of panel (a). The optimal symmetry for each signal is shown in the inset. In blue is the signal used for all three DET curves. The same signal form in black [(a)–(c)] shows the amplitude for a DET curve for Gaussian noise of unit variance that matches the Cauchy result ($\nu = 1$). Infinite variance of noise for $\nu \leqslant 2$ renders SNR ambiguous but $\sigma^2 = 2$ for $\nu = 2$ with SNR values of [0.113, 0.525, 0.053, 0.888] for [(a)–(d)]. (a) The sinc function, $\sin(x)/x$. (b) A chirp signal, e.g., gravity waves recorded by LIGO (Appendix C explains the use of $C_1^{(x)}$). (c) The Devil's staircase, an example of extreme intermittency, e.g., Ref. [28]. (d) The Weierstrass function with $a = 0.99$ and $b = 6$, an example of a continuous but nowhere differentiable signal. The second inset (upper left) shows all 16 traces for the case of $\nu = 1$. The vertical scale is clipped drastically down to signal level, yet none is discernible by eye. Overlain on the blue inset signal is the mean rank (black dots), linearly rescaled to match the signal. Note how all 10 prominent signal peaks map to extremal values of mean rank (circles).

the vector defining each signal, and hence the results plotted are only a *conservative lower bound* on detection performance with $\delta\mathcal{C}$.

Although the odd or even expansion was motivated by signal forms, symmetries of pure noise are of broader significance and can be used to detect overfitting and instrumental errors and test postextraction residual data, for randomness, e.g., in metrology [29]. Sometimes parity itself is the signal of interest, e.g., parity of the wave function ψ associated with odd or even number of photons stored in a cavity [30]. The five terms in the symmetry group expansion detect subtle yet distinct deviations of the postextraction residuals from the independent and identically distributed noise.

We have confined the detection to signals in independent and identically distributed noise but correlated (colored) noise is ubiquitous so how is the proposed detection method affected by correlation? Traditionally, the knowledge of the correlation function is used to whiten the data and then proceed with the detection algorithms suited to white noise [6,31]. However, if the correlation function is not known, then the main practical effect is that the *effective* number of independent samples N is reduced. Note also that correlation itself becomes a "signal," subject to detection. Indeed, insofar as correlation breaks the uniformity of $p(r, t)$ tending to $U(1, N) \times U(1, N)$, it is a signal, albeit a random one as was explored in a different context in Ref. [26] where it was remarkably effective in distinguishing chaos from noise and

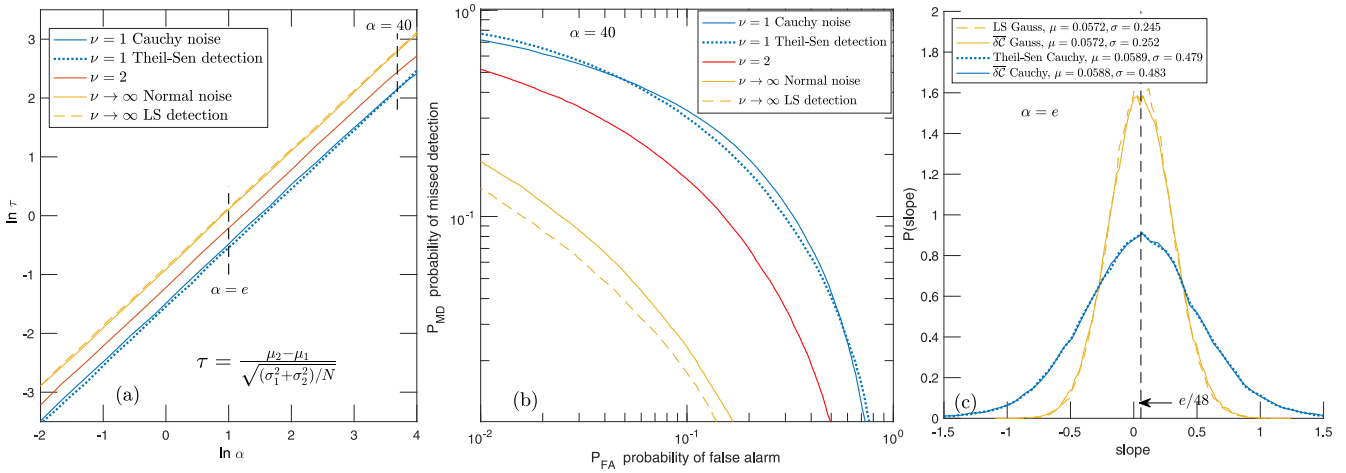

FIG. 4. Detection and extraction of a weak linear signal buried in noise: Even for single realizations ($n = 1$), the proposed method is nearly as efficient for heavy-tailed as for normal noise. The signal $f(x) = \alpha\, x/N$ is sampled uniformly at N points on $[-1, 1]$. Three noise distributions are used: $\nu = 1$ (Cauchy noise, blue), $\nu = 2$ (red), and $\nu \to \infty$ (Gaussian noise, gold), with solid lines for the proposed method and broken lines for conventional methods: dashed for least squares (LS) and dotted for Theil-Sen. (a) The means and standard deviations of $\overline{\delta\mathcal{C}}$ determine performance of the detection metric τ, a scaled difference of means (Welch statistic), defined in the legend, with subscripts 1 for signal absent and 2 for signal present. The ln-ln plot of τ vs. the (scaled) slope α shows $\tau = \alpha + c$ for all cases plotted, varying only in the intercept c. For normal noise, LS is slightly better. For Cauchy noise, the present method edges out Theil-Sen everywhere except for $\alpha > 3$. (b) Detection error trade-off curves for $\alpha = 40$. Operating at $P_{\rm FA} = P_{\rm MD}$, detection error ranges from 20% at $\nu = 1$ to 5% as $\nu \to \infty$. The companion broken curves are almost identical. (c) Slope extraction for $f(x) = \exp(1)\, x/48 \approx 0.0566\, x$. We determine slope by requiring $\overline{\delta\mathcal{C}} = 0$ (see text) and match the Theil-Sen histogram as well as the narrower LS result.

in detecting various types of random processes. We have confirmed that the far simpler $\delta\mathcal{C}$ method proposed here matches these results, reported in Ref. [26].

So how is the correlation revealed, in the language of $p(r, t)$? Whereas for uncorrelated noise all positions in the $N \times N$ matrix are indistinguishable, correlation breaks it and boundary effects appear, distinguishing among interior points with four nearest neighbors, edge points with three, and the four corners with two. The limiting form of $p(r, t)$ is, therefore, perturbed and hence also the background xy form for the equilibrium $\mathcal{C}$. That departure induces a nonzero equilibrium $\delta\mathcal{C}$ resulting from correlated pure noise but solely of R_2 symmetry. Should one wish to proceed with the signal detection rms metric suggested here, one recalculates the group deviation cdfs of Fig. 2 and resulting thresholds.

In summary, a simple signal-noise decomposition is proposed based on mapping raw data onto the rank-time plane and observing that only independent and identically distributed (pure) noise possesses a jointly uniform pdf, is invariant with respect to reshuffling, and satisfies rank-time duality. Mapping the raw data to rank-time space requires no original noise modeling at all in order to obtain distributions of $\overline{\delta\mathcal{C}}$ and the group-based rms metrics based on the formulation in terms of relative cdf $\delta\mathcal{C}$. Further, insofar as ranking is invariant to monotonic (nonlinear) transformations, multichannel and, indeed, multisensor data can be readily "fused" and "apples can be compared to oranges." Mapping pure noise data onto the rank-time plane led to the recognition that symmetries of noise are essential and the five symmetry group components are more selective, yielding a robust detection of rather complex *a priori* unknown signals in heavy-tailed arbitrary noise, even in single time series.

ACKNOWLEDGMENTS

This work was supported by the NSF Grant No. AGS-1639868. We thank two anonymous reviewers for helpful comments.

APPENDIX A: LINEAR REGRESSION WITH $\overline{\delta\mathcal{C}}$

Using $\overline{\delta\mathcal{C}}$ to fit a line to noisy data or to detect a ramp signal (linear trend) is described here. While our results for detecting step functions (e.g., breaks in time series) or detecting rectangular pulses are nearly identical, the linear case enables a companion discussion on slope extraction. We focus on the practically important case of $n = 1$, a single time series. Considering the vast literature on the question [1,5], it is surprising to learn that linear regression in heavy-tailed noise remains challenging [17,32]. To that end, Fig. 4 shows that the proposed *general-purpose* method works for all noise distributions.

For detection, framed as a binary decision, one asks: Given a time series of N points, is there a signal? The figure of merit, quantifying detection performance is

$$\tau = \frac{\mu_2 - \mu_1}{\sqrt{\left(\sigma_1^2 + \sigma_2^2\right)/N}}\,, \tag{A1}$$

the scaled difference of the random variable $\overline{\delta\mathcal{C}}(N)$ averages with and without the signal. Computation of $\overline{\delta\mathcal{C}}(N)$ is $\mathcal{O}(N)$, following the reduced form (4) in the main text.

Here we use the same convenient one-parameter (ν) noise distribution as given by (13). The Cauchy variance, albeit formally infinite, is $\sim 10^6 - 10^8$ for a Monte Carlo (MC) sample in

Fig. 4(a) with $N = 48$, $n = 10^5$. The N scaling of the normalized difference between the mean slopes [y axis in panel (a)] is such that the plot is independent of N given $f(x) = \alpha x/N$.

The logarithmic plots of τ versus the parameter α in Fig. 4(a) demonstrate that the proposed rank-time approach is robust and all purpose, performing nearly as well for heavy-tailed noise as for Gaussian. The Cauchy result overlaps with the Gaussian when shifted left by ≈ -0.61. Comparable detection for the two hence results if the signal in Cauchy noise is amplified by a factor of $\exp(0.61) \approx 1.82$, a modest penalty for changing to such heavy-tailed noise, where LS fails. Comparison plots of τ are shown for LS and Theil-Sen methods, roughly matching our results. The latter is a median-based special-purpose method, designed specifically for heavy-tailed noise and used exclusively for linear regression [26,32].

The probability of missed detection vs. the probability of false alarm is plotted in Fig. 4(b) for various values of ν. As in Fig. 4(a), the proposed method closely approximates the LS result for Gaussian noise and the Theil-Sen result for Cauchy. Illustrating the interpretation of these curves, for a weak signal in a single time series in heavy-tailed noise, at the cost of ≈50% false alarms, one can reduce the probability of a miss to only ≈1%. The signal could be, say, a decrease in global cloud coverage (so-called global dimming) during the past 20 years when reliable satellite observations became available [33].

The distribution of extracted slopes is shown in Fig. 4(c). For each time series consisting of signal plus noise, a slope is subtracted such that the residual rank data yield $\overline{\delta\mathcal{C}} = 0$. Slopes determined from this "detrending" by annulling of $\overline{\delta\mathcal{C}}$ closely match the Theil-Sen results, peaking at the same theoretical mean, and with nearly equal variance. Similar agreement obtains for Gaussian noise. For that case with unit noise variance, $\mathrm{SNR} = (\alpha/N)^2\,(1/3 + (N - 1/3)/(N - 1)^2)$. For $\alpha = \exp(1)$ and $N = 48$, $\mathrm{SNR} \approx 1.1 \times 10^{-3}$, that is, a weak signal.[8]

The $\delta\mathcal{C}$-based linear trend detection is a simple and flexible all-purpose approach for single time series, matching the standard techniques not only for both heavy-tailed noise and normal noise but also for all values of ν and any other distribution. To reiterate, such generality stems from the simple fact that *all* stationary white noise reduces to a universal cdf in the rank-time plane. Any departure therefrom as measured by $\overline{\delta\mathcal{C}}$ is the indicator of a trend.

The conventional slope estimate from the LS fit is in units of standard error, interpreted as a likelihood via the Student distribution. While this and all such comparisons inevitably invoke statistical ensembles, the present formulation offers a discrete binary test of consistency for such estimates *without reference to any parent distribution*: Evaluate $\overline{\delta\mathcal{C}}^{(0)}$ of the raw data and compare it to $\overline{\delta\mathcal{C}}^{(1)}$ for the detrended data. A telling case is linear regression of pure Gaussian noise, for which LS returns residual test values *greater* in magnitude than those for the raw data approximately 12% of the time. The fraction increases with increasing skewness or kurtosis of the noise. For the Theil-Sen method with Cauchy noise, such an outcome occurs ∼6% of the time. Any slope estimate disobeying $|\,\overline{\delta\mathcal{C}}^{(1)}\,| < |\,\overline{\delta\mathcal{C}}^{(0)}\,|$ would seem to constitute *a fortiori* grounds for its rejection. None of this contradicts LS being a maximum likelihood estimator meaning, for example, that no linear estimator can beat the LS standard deviation in Fig. 4(c). But the focus here is on a single trace: Is *this slope* for *this data* acceptable? Any single realization includes a transient odd-parity component either opposing or reinforcing the trend due to signal. The binary test simply addresses consistency in estimating the effective linear trend arising from *the sum* of these two; not merely the latter. For any single realization, the two sources of trend are indistinguishable.

[8]Note that LS a local quadratic minimization problem while $\overline{\delta\mathcal{C}}$ seeks a zero crossing.

APPENDIX B: ANALYTIC APPROACH TO DETECTION-ERROR TRADE-OFF (DET) CURVES

The five DET curves in Fig. 4(b) expand on the five τ loci lying on the vertical dashed line at $\alpha = 40$ in Fig. 4(a). These "trade-off" curves show the balance for any selected threshold between the fraction of false alarms and that for missed detections. Depending on the context, one picks the "operating" point on the trade-off curve and a common choice is the intersection of the trade-off curve with the main diagonal where $P_{\mathrm{FA}} = P_{\mathrm{MD}}$. Various practical detection algorithms often give rise to fairly intricate curves. Here, however, the curves are all smooth and almost perfectly symmetric when reflected about the diagonal. Such structure can be readily understood on the simplifying assumption that both the pdf of the chosen test statistic, e.g., $\overline{\delta\mathcal{C}}$, for noise alone and the pdf for that of noise plus signal are Gaussian (normal), with equal variance, and the former is zero mean.

On those assumptions it follows after integration that

$$P_{\mathrm{FA}}(s) = \tfrac{1}{2}\,[1 - \mathrm{erf}(s/\sqrt{2})]$$
$$P_{\mathrm{MD}}(s) = \tfrac{1}{2}\,[1 - \mathrm{erf}(\tau/\sqrt{N} - s/\sqrt{2})],$$

yielding a parametric representation of the curves in Fig. 4(b). Each DET curve then is completely prescribed once its τ value is read from Fig. 4(a). For the DET curve associated with least-squares detection in Gaussian noise, the parametric result is graphically indistinguishable from the dashed line. For the other four DET curves, the parametric result does not exactly overlie the plotted one but remains quite close, with the tiny deviations due the departure from the perfect normal pdf. The operating point follows from the above representation as:

$$P_{\mathrm{FA}} = P_{\mathrm{MD}} = \tfrac{1}{2}\,[1 - \mathrm{erf}(\tau/2\sqrt{N})].$$

APPENDIX C: OPTIMAL SIGNALS

What types of signals disturb which of the five $\delta\mathcal{C}$ symmetry components the most? Alternatively, are there signals such that, in the absence of noise, the five-term expansion of $\delta\mathcal{C}$ in Eqs. (9)–(11) collapses to a single term? It turns out that there is such a signal for R_2 and, for the remaining four cases, it is almost the case as only a small residual remains in a companion term.

Returning to rms as the metric, here we complete the main text characterization of signals exciting various $\delta\mathcal{C}$ patterns. $C_1^{(x)}$ and R_2 components form an orthogonal complement to D_4 (odd signals) and $C_1^{(y)}$ (even signals), respectively. Specifically, $C_1^{(x)}$ pairs with a signal oscillating at the grid scale frequency inside a linear envelope (optimizing variance and mean, respectively) and R_2 by the same oscillation in a quadratic envelope. In particular, the chirp signal in Fig. 3(b) can be viewed as a rapid oscillation inside a linear (monotone) envelope, leading to $C_1^{(x)}$ as the preferred metric for its detection.

The maximum for the D_2 rms value is attained simultaneously with a minimum of the D_4 contribution and resists a simple characterization. The signal has odd parity and its form must be sought in series form with coefficients determined numerically.

In parallel to (5), for these optimal signals, the explained variation captured by each of three of the four elementary maximizing forms is of the form

$$\left(\frac{\delta\mathcal{C}_{\rm rms}^{\{D_4,C_1^{(x,y)}\}}}{\delta\mathcal{C}_{\rm rms}}\right)^2 \sim \frac{15}{16}\left(1-\frac{c}{N^2}\right), \tag{C1}$$

while the fourth is

$$\left[\frac{\delta\mathcal{C}_{\rm rms}^{(R_2)}}{\delta\mathcal{C}_{\rm rms}}\right]^2 \sim 1-\frac{c}{N^2}. \tag{C2}$$

For the D_2 limited results suggest the asymptotic form $\rho^2_{\delta\mathcal{C}}\{D_2\} \approx 0.9$.

[1] P. R. Bevington and D. K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences* (McGraw–Hill, New York, 1992).

[2] D. K. C. MacDonald, *Noise and Fluctuations: An Introduction* (Courier Corporation, London, 2006).

[3] A. Van der Ziel, *Noise in Measurements* (Wiley, New York, 1976).

[4] R. Wittje, Phys. Today **73**, 42 (2020).

[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, Cambridge, UK, 2007).

[6] R. N. McDonough and A. D. Whalen, *Detection of Signals in Noise* (Academic Press, San Diego, CA, 1995).

[7] L. L. Scharf, *Statistical Signal Processing*, Vol. 98 (Addison-Wesley, Reading, MA, 1991).

[8] P. L. Krapivsky, S. Redner, and E. Ben-Naim, *A Kinetic View of Statistical Physics* (Cambridge University Press, Cambridge, UK, 2010).

[9] J. Klafter and I. M. Sokolov, *First Steps in Random Walks: From Tools to Applications* (Oxford University Press, Oxford, 2011).

[10] M. Manceau, K. Y. Spasibko, G. Leuchs, R. Filip, and M. V. Chekhova, Phys. Rev. Lett. **123**, 123606 (2019).

[11] G. Williams, B. Schäfer, and C. Beck, Phys. Rev. Res. **2**, 013019 (2020).

[12] R. C. Dalang and T. Humeau, Ann. Probab. **45**, 4389 (2017).

[13] T. Srokowski, Eur. Phys. J. B **85**, 65 (2012).

[14] F. Bardou, J.-P. Bouchaud, A. Aspect, and C. Cohen-Tannoudji, *Lévy Statistics and Laser Cooling: How Rare Events Bring Atoms to Rest* (Cambridge University Press, Cambridge, UK, 2002).

[15] R. Frieden, *Probability, Statistical Optics, and Data Testing: A Problem Solving Approach*, Vol. 10 (Springer Science & Business Media, New York, 2012).

[16] D. Sivia and J. Skilling, *Data Analysis: a Bayesian Tutorial* (Oxford University Press, Oxford, 2006).

[17] S. R. K. Vadali, P. Ray, S. Mula, and P. K. Varshney, IEEE Trans. Commun. **65**, 1061 (2017).

[18] C. Guozhong and L. Xingzhao, J. Syst. Eng. Electr. **19**, 717 (2008).

[19] Y. Edery, A. B. Kostinski, S. N. Majumdar, and B. Berkowitz, Phys. Rev. Lett. **110**, 180602 (2013).

[20] M. Hollander, D. A. Wolfe, and E. Chicken, *Nonparametric Statistical Methods*, Vol. 751 (John Wiley & Sons, New York, 2013).

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

[22] P. Akimov, F. Evstratov, and S. Zakharov, *Obnaruzhenie Radiological* (Radio i Svyaz', Moscow, 1989).

[23] M. Herrero-Collantes and J. C. Garcia-Escartin, Rev. Mod. Phys. **89**, 015004 (2017).

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

[25] I. Pinelis, Bull. Austr. Math. Soc. **102**, 104 (2020).

[26] G. Ierley and A. Kostinski, Phys. Rev. X **9**, 031039 (2019).

[27] M. Lax, *Symmetry Principles in Solid State and Molecular Physics* (Courier Corporation, London, 2001).

[28] I. Sokolović, P. Mali, J. Odavić, S. Radošević, S. Y. Medvedeva, A. E. Botha, Y. M. Shukrinov, and J. Tekić, Phys. Rev. E **96**, 022210 (2017).

[29] N. Kura and M. Ueda, Phys. Rev. Lett. **124**, 010507 (2020).

[30] J.-C. Besse, S. Gasparinetti, M. C. Collodo, T. Walter, A. Remm, J. Krause, C. Eichler, and A. Wallraff, Phys. Rev. X **10**, 011046 (2020).

[31] A. Koivunen and A. Kostinski, J. Appl. Meteorol. **38**, 741 (1999).

[32] G. Balkema and P. Embrechts, Risks **6**, 93 (2018).

[33] M. Wild, J. Geophys. Res.: Atmospheres **114**, D00D16 (2009).