



Efficient computation of N -point correlation functions in D dimensions

Oliver H. E. Philcox^{ab,1} and Zachary Slepian^{cd}

Edited by Leslie Greengard, New York University, New York, NY; received June 19, 2021; accepted January 20, 2022

We present efficient algorithms for computing the N -point correlation functions (NPCFs) of random fields in arbitrary D -dimensional homogeneous and isotropic spaces. Such statistics appear throughout the physical sciences and provide a natural tool to describe stochastic processes. Typically, algorithms for computing the NPCF components have $\mathcal{O}(n^N)$ complexity (for a dataset containing n particles); their application is thus computationally infeasible unless N is small. By projecting the statistic onto a suitably defined angular basis, we show that the estimators can be written in a separable form, with complexity $\mathcal{O}(n^2)$ or $\mathcal{O}(n_g \log n_g)$ if evaluated using a Fast Fourier Transform on a grid of size n_g . Our decomposition is built upon the D -dimensional hyperspherical harmonics; these form a complete basis on the $(D - 1)$ sphere and are intrinsically related to angular momentum operators. Concatenation of $(N - 1)$ such harmonics gives states of definite combined angular momentum, forming a natural separable basis for the NPCF. As N and D grow, the number of basis components quickly becomes large, providing a practical limitation to this (and all other) approaches: However, the dimensionality is greatly reduced in the presence of symmetries; for example, isotropic correlation functions require only states of zero combined angular momentum. We provide a Julia package implementing our estimators and show how they can be applied to a variety of scenarios within cosmology and fluid dynamics. The efficiency of such estimators will allow higher-order correlators to become a standard tool in the analysis of random fields.

correlation functions | spherical harmonics | clustering statistics | computational physics | cosmology

Random fields are ubiquitous in the physical sciences. Perhaps the most powerful tool for their analysis is the set of N -point correlation functions (hereafter NPCFs), defined as statistical averages over N copies of the field at different spatial or temporal locations. If the field is Gaussian-random, only the first two correlators (the mean and two-point function) contain useful information, though this assumption is rarely true in practice. Examples of higher-order NPCFs populate many fields of study; a brief search will reveal their use in molecular physics (1), materials science (2), field theory (3), diffusive systems (4, 5), and cosmology (6), among other topics.

Correlation functions have found extensive use in the analysis of spectroscopic galaxy surveys (e.g., ref. 7). While the majority of information is contained within the two-point correlation function (2PCF), inclusion of the higher-order functions is expected to significantly tighten constraints on cosmological parameters, particularly those pertaining to phenomena such as extensions to General Relativity (8). Despite this, statistics beyond the 2PCF have been scarcely used in practice; in fact, almost no modern analyses have included correlators with $N > 3$. The reason is simple: Higher-order NPCFs are expensive to compute and analyze.

Consider a D -dimensional space \mathbb{M}^D (e.g., a Euclidean space) with an associated complex-valued random field $X : \mathbb{M}^D \rightarrow \mathbb{C}$. The NPCF, $\zeta : (\mathbb{M}^D)^{\otimes(N-1)} \rightarrow \mathbb{C}$ (for tensor product \otimes), is formally defined as

$$\zeta(\mathbf{r}^1, \dots, \mathbf{r}^{N-1}; \mathbf{s}) = \mathbb{E}_X [X(\mathbf{s})X(\mathbf{s} + \mathbf{r}^1) \cdots X(\mathbf{s} + \mathbf{r}^{N-1})], \quad [1]$$

where \mathbb{E}_X represents the statistical average over realizations of X , \mathbf{s} and \mathbf{r}^i are absolute and relative positions on the manifold, and we have assumed $N \geq 2$. If the random field is statistically homogeneous, all correlators must be independent of the absolute position \mathbf{s} ; this leads to the well-known NPCF estimator

$$\hat{\zeta}(\mathbf{r}^1, \dots, \mathbf{r}^{N-1}) = \frac{1}{V_D} \int_{\mathbb{M}^D} d^N \mathbf{s} [X(\mathbf{s})X(\mathbf{s} + \mathbf{r}^1) \cdots X(\mathbf{s} + \mathbf{r}^{N-1})], \quad [2]$$

Significance

Stochastic processes appear throughout the physical sciences, and their properties are usually described by correlation functions. For discrete data, the N -point correlation function (NPCF) encodes the distribution of N -tuplets of points in space; estimation of the NPCF basis coefficients from a set of n particles scales as n^N . As N increases, this measurement becomes prohibitively expensive; thus, statistics with $N > 3$ are rarely used. Here, we show that NPCF components may be estimated in n^2 time, by first expanding the statistics in separable angular bases. This approach has already found substantial application in quantifying galaxy clustering; here, we show it to be applicable to any homogeneous and isotropic space, regardless of dimension, and provide a practical implementation in Julia.

Author affiliations: ^aDepartment of Astrophysical Sciences, Princeton University, Princeton, NJ 08540; ^bSchool of Natural Sciences, Institute for Advanced Study, Princeton, NJ 08540; ^cDepartment of Astronomy, University of Florida, Gainesville, FL 32611; and ^dPhysics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94709

Author contributions: O.H.E.P. and Z.S. designed research; O.H.E.P. performed research; O.H.E.P. analyzed data; O.H.E.P. and Z.S. wrote the paper; and O.H.E.P. created the accompanying code package.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2022 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹To whom correspondence may be addressed. Email: ohep2@cantab.ac.uk.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2111366119/-DCSupplemental>.

Published August 8, 2022.

averaging \mathbf{s} over a volume V_D , and noting that the NPCF depends on only $(N - 1)$ positions. In praxis, we cannot estimate the continuous NPCF of Eq. 1; rather, we consider only the quantity projected onto some basis functions, $B(\mathbf{r}^1, \dots, \mathbf{r}^{N-1})$, for example, a set of radial and angular bins. In this case, the problem reduces to estimating the coefficients $\zeta_B \equiv \int d\mathbf{r}^1 \dots d\mathbf{r}^{N-1} \zeta(\mathbf{r}^1, \dots, \mathbf{r}^{N-1}) B(\mathbf{r}^1, \dots, \mathbf{r}^{N-1})$. Strictly, an infinite set of basis functions is required to fully specify the NPCF: Conventionally, one restricts to a finite number and applies the same basis projection to both data and theory, eliminating bias.

The computational difficulties become clear if one considers measuring the NPCF components from a discrete field containing n particles. In this instance, the random field X can be represented as a sum over n Dirac delta functions, i.e., $X(\mathbf{r}) = \sum_{j=1}^n w^j \delta^D(\mathbf{r} - \mathbf{y}^j)$, where $\{\mathbf{y}^j\}$ are the particle positions and $\{w^j\}$ are weights. Projecting onto a basis function B , the discrete version of Eq. 2 becomes

$$\hat{\zeta}_B = \frac{1}{V_D} \sum_{j_0=1}^n w^{j_0} \sum_{j_1=1}^n w^{j_1} \dots \sum_{j_{N-1}=1}^n w^{j_{N-1}} B(\mathbf{y}^{j_1} - \mathbf{y}^{j_0}, \dots, \mathbf{y}^{j_{N-1}} - \mathbf{y}^{j_0}). \quad [3]$$

This is a sum over N -tuplets of particles and, since the total number of N -tuplets scales as n^N , has complexity $\mathcal{O}(n^N)$ (cf Section 3B). Unless n is very small, direct application of Eq. 3 is infeasible for all but the smallest N .

Ref. 11 presented a new technique to measure the three-dimensional (3D) Euclidean three-point correlation function (3PCF) more efficiently, building on ref. 12. By representing $\zeta(\mathbf{r}^1, \mathbf{r}^2)$ in a factorizable angular basis of Legendre polynomials, the former work obtained an algorithm for estimating 3PCF coefficients with $\mathcal{O}(n^2)$ complexity. This algorithm facilitated a number of analyses, both in cosmology (e.g., refs. 13 and 14) and magnetohydrodynamics (15). Our companion paper (16) showed that the approach can be generalized to the computation of rotationally invariant NPCF coefficients in 3D Euclidean space, which is of particular relevance for galaxy surveys. Here, we show that similar estimators may be constructed in any homogeneous and isotropic space; in particular, if the basis is carefully chosen, an $\mathcal{O}(n^2)$ estimator for each basis component is possible, regardless of N , D , and the spatial curvature. Furthermore, if the data can be mapped to a grid of dimension n_g , Fast Fourier Transforms (FFTs) can be used to reduce this to $\mathcal{O}(n_g \log n_g)$.

Our pathway to obtaining an efficient NPCF estimator is the following:

1. Construct a set of basis functions for the angular part of a homogeneous and isotropic space \mathbb{M}^D (Section 1). A natural choice is the set of *hyperspherical harmonics*, which generalize the spherical harmonics and arise in the D -dimensional theory of angular momentum.
2. Combine $(N - 1)$ hyperspherical harmonics to create an $(N - 1)$ -particle angular basis on the manifold \mathbb{M}^D (Section 2). In particular, we form states of definite combined angular momentum, denoted \mathcal{P}_Λ^L . Since the (statistically

homogeneous) NPCF depends only on $(N - 1)$ coordinates, its angular part can be decomposed into this basis. Explicitly:

$$\zeta(\mathbf{r}^1, \dots, \mathbf{r}^{N-1}) = \sum_{\mathbf{L}} \sum_{\mathbf{\Lambda}} \zeta_{\mathbf{\Lambda}}^{\mathbf{L}}(r^1, \dots, r^{N-1}) \mathcal{P}_{\mathbf{\Lambda}}^{\mathbf{L}}(\hat{\mathbf{r}}^1, \dots, \hat{\mathbf{r}}^{N-1}), \quad [4]$$

where $\zeta_{\mathbf{\Lambda}}^{\mathbf{L}}$ are the basis components, the sets of indices $\mathbf{\Lambda}$ and \mathbf{L} represent internal and external angles of the NPCF, $\hat{\mathbf{r}}$ represents the unit vector parallel to \mathbf{r} , and $r \equiv |\mathbf{r}|$. In the language of Eq. 3, our basis functions are $B(\mathbf{r}^1, \dots, \mathbf{r}^{N-1}) = \Theta^b(r^1, \dots, r^{N-1}) \mathcal{P}_{\mathbf{\Lambda}}^{\mathbf{L}}(\hat{\mathbf{r}}^1, \dots, \hat{\mathbf{r}}^{N-1})$, where Θ^b is some separable set of radial bins indexed by the vector \mathbf{b} (Eq. 38).

3. Using Eq. 2, construct an estimator for the NPCF basis components $\zeta_{\mathbf{\Lambda}}^{\mathbf{L}}$ (Section 3). Since the basis functions are separable in $\mathbf{r}^1, \mathbf{r}^2, \dots$, the estimator can be factorized and takes the schematic form (cf Eq. 37)

$$\hat{\zeta}_{\mathbf{\Lambda}}^{\mathbf{L}}(r^1, \dots, r^{N-1}) = \frac{1}{V_D} \sum_{j=1}^n w^j \times \sum_{\ell^1 \dots \ell^{N-1}} [\text{coupling}] \times a_{\ell^1}(\mathbf{y}^j; r^1) \dots a_{\ell^{N-1}}(\mathbf{y}^j; r^{N-1}), \quad [5]$$

where ℓ specify angular momentum indices, and each $a_{\ell}(\mathbf{y}; r)$ function involves a further sum over n particles, weighted by a hyperspherical harmonic (see Eq. 37). Since each $a_{\ell}(\mathbf{s}; r)$ component can be estimated independently, the algorithm has complexity $\mathcal{O}(n^2)$, or $\mathcal{O}(n_g \log n_g)$ using an FFT with n_g grid points. Eq. 5 can be applied to each basis component separately; we caution that the total number of components (for a fixed angular and radial resolution) becomes large as N and D increase, providing a practical limitation to any NPCF algorithm.

In Section 4, we discuss a numerical implementation of the NPCF estimator,[†] alongside a variety of applications.

1. Single-Particle Basis

We begin by discussing the angular basis for functions of one position in \mathbb{M}^D , hereafter referred to as the “single-particle” basis. In Section 2, the basis will allow construction of a joint basis of $(N - 1)$ positions onto which the NPCF can be projected.

A. Constant-Curvature Metric. To form an efficient angular basis, we require the underlying manifold to be 1) homogeneous and 2) isotropic. This leads to the well-known line-element

$$d\Sigma_{\mathbb{M}^D}^2 = dr^2 + \chi_k^2(r) \{ d\theta_{D-1}^2 + \sin^2 \theta_{D-1} [d\theta_{D-2}^2 + \sin^2 \theta_{D-2} (d\theta_{D-3}^2 + \dots)] \}, \quad [6]$$

(e.g., ref. 17), adopting hyperspherical coordinates $\mathbf{r} \equiv \{r, \theta_1, \dots, \theta_{D-1}\}$.[‡] In this parametrization, r is a radial coordinate, while the θ_i are angular variables, with $\theta_1 \in [0, 2\pi)$ (often denoted by ϕ in $D = 3$) and $\theta_j \in [0, \pi)$ for $j > 1$. A sketch of the coordinates is shown in Fig. 1 for a Euclidean geometry. Eq. 6 is a *constant-curvature* metric, specified by $\chi_k(r) = \sin(r\sqrt{k})/\sqrt{k}$ if $|k| > 0$ and $\chi_k(r) = r$ else. Here, $k > 0$ gives the D -sphere, \mathbb{S}^D , $k < 0$ leads to the hyperbolic geometry \mathbb{H}^D , and $k = 0$ results in a

^{*}See ref. 9 for an efficient, but inexact, tree-based approach in 3D, as well as ref. 10 for a solution involving graph databases.

[†]Available at <https://github.com/oliverphilcox/NPCFs.jl>.

[‡]Note that some conventions label the θ coordinates in the opposite order.

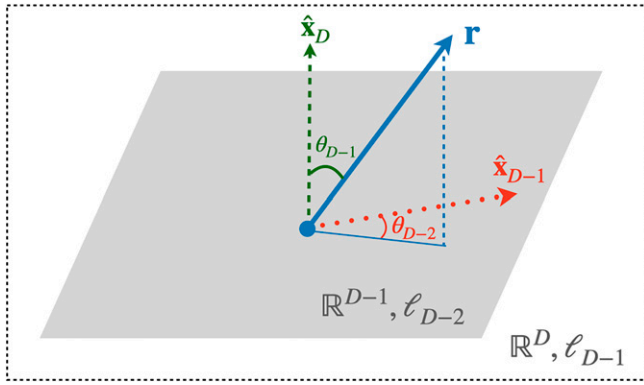


Fig. 1. Cartoon illustrating the coordinate system used in this work, assuming a Euclidean geometry. The outer dashed box indicates \mathbb{R}^D , while the gray plane represents \mathbb{R}^{D-1} , containing the Cartesian coordinates $\{x_1, \dots, x_{D-1}\}$. The position vector $\mathbf{r} \in \mathbb{R}^D$ can be expressed in hyperspherical coordinates $\{r, \theta_1, \dots, \theta_{D-1}\}$, where r is a radial coordinate. θ_{K-1} is defined as angle between the $\hat{\mathbf{x}}_K$ axis and the projection of \mathbf{r} into the subspace \mathbb{R}^K , as shown for θ_{D-1} and θ_{D-2} . The angles obey the restrictions $\theta_K \in [0, \pi)$ for $K > 1$, and $\theta_1 \in [0, 2\pi)$. Each angle can be associated with an angular momentum index ℓ_{K-1} , encoding the orbital angular momentum in the subspace \mathbb{R}^K in which only the first K Cartesian coordinates are varied (Section 1). ℓ_1 gives the azimuthal angular momentum in the subspace containing $\{x_1, x_2, x_3\}$, and ℓ_{D-1} gives the total angular momentum.

Euclidean geometry \mathbb{R}^D . Such manifolds are ubiquitous in the physical sciences; for instance, this describes the spatial part of the Friedmann–Lemaître–Robertson–Walker metric for an expanding Universe (18).

B. Hyperspherical Harmonics. A convenient basis for the constant-curvature space \mathbb{M}^D is formed from the set of *harmonic* functions $H: \mathbb{M}^D \rightarrow \mathbb{C}$. These satisfy the *Laplace–Beltrami* equation

$$\Delta H(\mathbf{r}) \equiv \frac{1}{\sqrt{|g|}} \partial_i \left(\sqrt{|g|} g^{ij} \partial_j H(\mathbf{r}) \right) = 0, \quad [7]$$

where H is twice continuously differentiable, g is the metric (with $|g| = \det [g_{ij}]$), $i \in \{1, 2, \dots, D\}$, and we have assumed the Einstein summation convention. Assuming the metric of Eqs. 6 and 7 permits a separable solution of the form $H(\mathbf{r}) = R(r)Y(\hat{\mathbf{r}})$, where $\hat{\mathbf{r}} = \{\theta_1, \dots, \theta_{D-1}\}$. The angular part of this must satisfy the following eigenfunction equation for constant λ_{D-1} :

$$\Delta_{\mathbb{S}^{D-1}} Y(\hat{\mathbf{r}}) = -\lambda_{D-1} Y(\hat{\mathbf{r}}), \quad [8]$$

where $\Delta_{\mathbb{S}^{D-1}}$ is the Laplace–Beltrami operator on the $(D-1)$ -sphere, given explicitly in *SI Appendix, Eq. S1*. The corresponding solutions are the *hyperspherical harmonics* on \mathbb{S}^{D-1} (e.g., refs. 19–21). Since the harmonic functions are separable, the hyperspherical harmonics form an angular basis for *any* function on \mathbb{M}^D , regardless of the spatial curvature k .⁵ Physically, this decomposition is guaranteed since we have assumed the metric to be homogeneous and isotropic, enforcing invariance under the rotation group $SO(D)$ about any origin.

For a given dimension D , the hyperspherical harmonics, denoted $Y_{\ell_1 \dots \ell_{D-1}}(\hat{\mathbf{r}})$, may be obtained by solving Eq. 8 recursively,

⁵This is additionally seen by noting that Eq. 6 can be written $d\Sigma_D^2 = dr^2 + \chi_k^2(r) d\Omega_{D-1}^2$, where $d\Omega_{D-1}^2$ is the line element on the $(D-1)$ -sphere.

as detailed in *SI Appendix*. These depend on a set of $(D-1)$ integers, $\{\ell_k\}$, which are related to angular momentum (cf Section 1C), and satisfy the selection rules

$$-\ell_2 \leq \ell_1 \leq \ell_2, \quad \ell_{K-1} \leq \ell_K \leq \ell_{K+1} \quad (2 \leq K \leq D-1). \quad [9]$$

In two and three dimensions, the hyperspherical harmonics take a simple form:

$$Y_{\ell_1}(\theta_1) = \frac{1}{\sqrt{2\pi}} e^{i\ell_1\theta_1},$$

$$Y_{\ell_1\ell_2}(\theta_1, \theta_2) = \sqrt{\frac{2\ell_2 + 1}{4\pi} \frac{(\ell_2 + \ell_1)!}{(\ell_2 - \ell_1)!}} e^{i\ell_1\theta_1} P_{\ell_2}^{-\ell_1}(\cos\theta_2), \quad [10]$$

for associated Legendre polynomials P_{ℓ}^m . The $D=3$ functions are the usual spherical harmonics (in the Condon–Shortley convention), made clear by the identification $(\ell_1, \ell_2) \rightarrow (m, \ell)$, $(\theta_1, \theta_2) \rightarrow (\phi, \theta)$, where ϕ is the azimuthal angle. The explicit form of the hyperspherical harmonics for general D is given in *SI Appendix, Eq. S3*.

C. Connection to Angular Momentum Eigenstates. In 3D, the theory of angular momentum is centered around two operators, \hat{L}^2 and \hat{L}_3 , which, respectively, give the total angular momentum and that projected onto the $\hat{\mathbf{x}}_3$ axis. Both may be constructed from the vector operator $\hat{\mathbf{L}} = \mathbf{r} \times \mathbf{p}$, where $p_j = -i(\partial/\partial x^j)$ is the linear momentum in \mathbb{R}^3 . For $D \neq 3$, we cannot define a cross-product; thus, we instead start from the *tensorial* angular momentum operator, following (22):

$$\hat{L}_{ij} = x_i p_j - x_j p_i, \quad i \neq j \in \{1, 2, \dots, D\}, \quad [11]$$

where $\{x_1, \dots, x_D\}$ is a Cartesian coordinate chart.[¶] These operators are antisymmetric ($\hat{L}_{ij} = -\hat{L}_{ji}$) and form the Lie algebra $\mathfrak{so}(D)$, i.e., that of the rotation group in D dimensions. Of particular interest is \hat{L}_{12} ; when applied to some state, this gives the azimuthal angular momentum in the subspace containing $\{x_1, x_2, x_3\}$, just as for \hat{L}_3 in $D=3$.

To fully define the rotational properties of a single-particle function in D -dimensional space, we must specify not only its total angular momentum, but also that projected into lower-dimensional subspaces. These can be obtained from the operators

$$\hat{\mathcal{L}}_K^2 \equiv \sum_{i=1}^K \sum_{j=i+1}^K (\hat{L}_{ij})^2, \quad [12]$$

for $K \geq 3$. Here, $\hat{\mathcal{L}}_3^2 \equiv (\hat{L}_{12})^2 + (\hat{L}_{13})^2 + (\hat{L}_{23})^2$ gives the angular momentum in the subspace $\{x_1, x_2, x_3, 0, \dots, 0\}$ of \mathbb{R}^D , $\hat{\mathcal{L}}_4^2$ gives that in the subspace $\{x_1, x_2, x_3, x_4, 0, \dots, 0\}$, etc. $\hat{\mathcal{L}}_D^2$ is the total angular momentum operator, analogous to \hat{L}^2 in the 3D theory.

As shown in ref. 22 (Ch. 3), a complete set of commuting angular momentum operators is given by $\{\hat{L}_{12}, \hat{\mathcal{L}}_3^2, \dots, \hat{\mathcal{L}}_D^2\}$.

[¶]This occurs since the K -dimensional Laplace–Beltrami operator of *SI Appendix, Eq. S1* (which the hyperspherical harmonics are eigenfunctions of) is related to the K -dimensional total angular momentum operator by $\Delta_{\mathbb{S}^{K-1}} = -\hat{\mathcal{L}}_K^2$.

Moreover, the hyperspherical harmonics of Section 1B are simultaneous eigenfunctions of these, satisfying[#]

$$\begin{aligned}\widehat{L}_{12} Y_{\ell_1 \dots \ell_{D-1}}(\hat{\mathbf{r}}) &= \ell_1 Y_{\ell_1 \dots \ell_{D-1}}(\hat{\mathbf{r}}), \\ \widehat{L}_K Y_{\ell_1 \dots \ell_{D-1}}(\hat{\mathbf{r}}) &= \ell_{K-1}(\ell_{K-1} + K - 2) Y_{\ell_1 \dots \ell_{D-1}}(\hat{\mathbf{r}}) \\ &\quad (3 \leq K \leq D).\end{aligned}\quad [13]$$

We may thus associate ℓ_{K-1} with the orbital angular momentum in the subspace comprising the first K Cartesian coordinates (for $K \geq 3$) and ℓ_1 with that projected onto the $\hat{\mathbf{x}}_3$ axis. Such an interpretation also justifies the conditions in Eq. 9; projections of the angular momentum into lower-dimensional spaces must have equal or lesser magnitudes than the total angular momentum in D dimensions.

For later use, we introduce abstract notation for the angular momentum basis functions, writing $Y_{\ell_1 \dots \ell_{D-1}}(\hat{\mathbf{r}}) \rightarrow |\ell_1 \dots \ell_{D-1}\rangle$ in Dirac (or bra-ket) notation.^{||} In total, there are $\binom{\ell_{D-1} + D - 3}{D - 3} (2\ell_{D-1} + D - 2) / (D - 2)$ eigenstates corresponding to a given total angular momentum ℓ_{D-1} if $D \geq 3$, and one if $D = 2$ (ref. 22, equation 3.37). For a suitably defined inner product, the basis functions are orthonormal (20, 21, 23):

$$\begin{aligned}\int_{\mathbb{S}^{D-1}} d\Omega_{D-1} Y_{\ell_1 \dots \ell_{D-1}}^*(\hat{\mathbf{r}}) Y_{\ell'_1 \dots \ell'_{D-1}}(\hat{\mathbf{r}}) \\ \equiv \langle \ell_1 \dots \ell_{D-1} | \ell'_1 \dots \ell'_{D-1} \rangle = \delta_{\ell_1 \ell'_1}^K \dots \delta_{\ell_{D-1} \ell'_{D-1}}^K,\end{aligned}\quad [14]$$

where the Kronecker delta, δ_{ij}^K , is unity if $i = j$ and zero otherwise. Furthermore, they form a complete basis on \mathbb{S}^{D-1} , such that, for any $h : \mathbb{S}^{D-1} \rightarrow \mathbb{C}$

$$\begin{aligned}|h\rangle &= \sum_{\ell_1 \dots \ell_{D-1}} \langle \ell_1 \dots \ell_{D-1} | h \rangle |\ell_1 \dots \ell_{D-1}\rangle \leftrightarrow, \\ h(\hat{\mathbf{r}}) &= \sum_{\ell_1 \dots \ell_{D-1}} h_{\ell_1 \dots \ell_{D-1}} Y_{\ell_1 \dots \ell_{D-1}}(\hat{\mathbf{r}}),\end{aligned}\quad [15]$$

where the summation runs over all angular momentum indices allowed by the selection rules of Eq. 9. The basis coefficients $h_{\ell_1 \dots \ell_{D-1}} = \langle \ell_1 \dots \ell_{D-1} | h \rangle$ may be obtained via orthonormality. Since the angular part of \mathbb{M}^D is just the $(D - 1)$ -sphere \mathbb{S}^{D-1} (Section 1A), any one-particle function on \mathbb{M}^D can be decomposed into this basis; in general, the coefficients retain dependence on the radial coordinate r .

2. (N - 1)-Particle Basis

We now utilize the mathematics of angular momentum addition to generalize the angular basis of Section 1 to functions of $(N - 1)$ positions, i.e., $f : (\mathbb{M}^D)^{\otimes(N-1)} \rightarrow \mathbb{C}$.

A. Angular Momentum Addition. To begin, we consider the combination of two angular basis functions on the $(D - 1)$ -sphere. For convenience, we will work in the Dirac representation and denote the set of angular momentum indices by $\ell \equiv \{\ell_1, \dots, \ell_{D-1}\}$, with superscripts used to distinguish between

particles. Given single-particle states $|\ell^1\rangle$ and $|\ell^2\rangle$, the simplest two-particle state is $|\ell^1, \ell^2\rangle \equiv |\ell^1\rangle |\ell^2\rangle$, which exists in the product space $\mathbb{S}^{D-1} \otimes \mathbb{S}^{D-1}$. This is a simultaneous eigenstate of angular momentum operators for the first and second particles, $\widehat{L}_{ij}^{(1)}$ and $\widehat{L}_{ij}^{(2)}$ (cf Section 1C). While the product states $|\ell^1, \ell^2\rangle$ do form a basis on $\mathbb{S}^{D-1} \otimes \mathbb{S}^{D-1}$ [sometimes called the “uncoupled basis” (24)], it is not an efficient one, since 1) we require $2(D - 1)$ angular momentum indices to specify the state, which, as shown below, is considerably more than necessary; and 2) the indices are not straightforwardly connected to the joint rotation properties of the two-particle state.

A more appropriate basis is wrought by considering the combined angular momentum operator $\widehat{L}_{ij}^{(12)} \equiv \widehat{L}_{ij}^{(1)} + \widehat{L}_{ij}^{(2)}$, which specifies the properties of some two-particle function, $f(\mathbf{r}^1, \mathbf{r}^2)$, under joint rotations of \mathbf{r}^1 and \mathbf{r}^2 about a common origin. As in Section 1C, $\widehat{L}_{ij}^{(12)}$ may be used to construct a set of commuting angular momentum operators, whose eigenstates can be written $|\mathbf{L}\rangle \equiv |L_1 \dots L_{D-1}\rangle$. Here, L_{K-1} specifies the combined angular momentum projected into the K -dimensional subspace, in which the first K Cartesian coordinates are varied (for $K \geq 3$), and L_1 gives that projected onto the $\hat{\mathbf{x}}_3, \hat{\mathbf{x}}_3^2$ axes. Similarly to the single-particle state, the \mathbf{L} indices must obey the selection rules of Eq. 9.

Two-particle states of definite combined angular momentum are formed by summing over products of one-particle states, just as in the 3D case (e.g., refs. 25 and 26; see also refs. 24, 27, and 28) for a discussion with more general D). Explicitly, they are given by

$$|\ell_{D-1}^1 \ell_{D-1}^2; \mathbf{L}\rangle = \sum_{\ell_1^1 \dots \ell_{D-2}^1} \sum_{\ell_2^2 \dots \ell_{D-2}^2} \langle \ell^1, \ell^2 | \mathbf{L} \rangle |\ell^1\rangle |\ell^2\rangle, \quad [16]$$

where $\langle \ell^1, \ell^2 | \mathbf{L} \rangle$ is a Clebsch–Gordan (hereafter CG) coefficient (e.g., ref. 29). This is often referred to as the “coupled basis” (24). To uniquely define the state, we must specify 1) the combined angular momentum eigenvalues \mathbf{L} , and 2) the total angular momentum of the first and second particle, ℓ_{D-1}^1 and ℓ_{D-1}^2 . Importantly, Eq. 16 involves a sum over $\{\ell_1^i, \dots, \ell_{D-2}^i\}$, i.e., the projection of the particles’ angular momentum into lower-dimensional subspaces. Due to this, the combined state is specified by only $(D + 1)$ indices, which, for $D > 3$, is significantly fewer than the $2(D - 1)$ required for $|\ell^1, \ell^2\rangle$.

Practically, one must know the CG coefficients $\langle \ell^1, \ell^2 | \mathbf{L} \rangle$ in order to form the coupled basis of Eq. 16.^{**} These have been studied in depth (e.g., refs. 25, 26, and 29–32) and may be simplified by techniques such as Racah’s factorization lemma (33). One route to their computation is via Eq. 16; starting from a state of maximal combined angular momentum, $|\ell_{D-1}^1 \ell_{D-1}^2; \ell^1 + \ell^2\rangle \equiv |\ell^1\rangle |\ell^2\rangle$ (wherein the two angular momentum vectors in \mathbb{R}^D are aligned), ladder operators may be applied iteratively to obtain states of lower angular momentum, whose weightings give the CG coefficients. As in refs. 25 and 26, the explicit forms for the CG coefficients in $D = 2$ and $D = 3$ are given by

[#]Since \mathbb{M}^D and \mathbb{R}^D share the same angular parametrization, and angular momentum is independent of radial coordinates, we may work in a Euclidean space for the purposes of this section.

^{||}The Hilbert space formed from these states is an infinite-dimensional representation of the rotation group $SO(D)$. Practically, the eigenvalue ℓ_{D-1} represents the behavior in $SO(D)$, while the values of ℓ_{K-1} with $K < D$ specify the rotation’s action in a lower-dimensional subgroup $SO(K)$.

^{**}The CG coefficients used in this work are those of the $SO(D) \supset SO(D - 1) \supset \dots \supset SO(2)$ reduction.

$$\begin{aligned} \langle \ell^1, \ell^2 | L \rangle &= \delta_{(\ell^1 + \ell^2)L}^K, \\ \langle \ell^1 \ell^1, \ell^2 \ell^2 | L_1 L_2 \rangle \\ &= (-1)^{-\ell^1 + \ell^2 - L_1} \sqrt{2L_2 + 1} \begin{pmatrix} \ell^1_2 & \ell^2_2 & L_2 \\ \ell^1_1 & \ell^2_1 & -L_1 \end{pmatrix}, \end{aligned} \quad [17]$$

where the 2×3 matrix is a Wigner 3- j symbol (e.g., ref. 34, section 34). The $D = 4$ case is similar (see refs. 29, 30, and 35), but includes a Wigner 9- j symbol.

CG coefficients satisfy certain orthogonality conditions, including

$$\begin{aligned} \sum_{\ell^1 \dots \ell^{D-2}} \sum_{\ell^1 \dots \ell^{D-2}} \langle \ell^1, \ell^2 | \mathbf{L} \rangle \langle \ell^1, \ell^2 | \mathbf{L}' \rangle \\ = \delta_{L_1 L'_1}^K \times \dots \times \delta_{L_{D-1} L'_{D-1}}^K. \end{aligned} \quad [18]$$

Coupled with the orthonormality of the one-particle states $|\ell\rangle$ (Eq. 24), this ensures that the combined-angular-momentum basis is orthonormal, i.e.,

$$\begin{aligned} \langle l^1_{D-1} l^2_{D-1}; \mathbf{L}' | \ell^1_{D-1} \ell^2_{D-1}; \mathbf{L} \rangle \\ = \left(\delta_{L_1 L'_1}^K \dots \delta_{L_{D-1} L'_{D-1}}^K \right) \times \left(\delta_{\ell^1_{D-1} l^1_{D-1}}^K \delta_{\ell^2_{D-1} l^2_{D-1}}^K \right). \end{aligned} \quad [19]$$

Furthermore, $\langle \ell^1, \ell^2 | \mathbf{L} \rangle$ is nonzero only if the following conditions are satisfied:

$$\begin{aligned} L_1 = \ell^1 + \ell^2, \quad |\ell^1_K - \ell^2_K| \leq L_K \leq \ell^1_K + \ell^2_K, \\ (2 \leq K \leq D-1). \end{aligned} \quad [20]$$

The first constraint occurs since L_1 is the eigenvalue corresponding to $\widehat{L}_{12}^{(12)}$, which is linear in $\widehat{L}_{12}^{(1)}$ and $\widehat{L}_{12}^{(2)}$ (cf addition of m indices in 3D), and the second is due to the triangle inequality, recalling that L_K corresponds to the magnitude of the angular momentum in the subspace \mathbb{R}^K . In particular, the constraints fix the total combined angular momentum, L_{D-1} , to be no greater than $\ell^1_{D-1} + \ell^2_{D-1}$.

B. Combined Angular Momentum Basis. By repeated application of the angular momentum addition rule (Eq. 16), we may build up an $(N-1)$ -particle state of definite combined angular momentum. First, we combine $|\ell^1\rangle$ and $|\ell^2\rangle$ to form the two-particle state $|\ell^1_{D-1} \ell^2_{D-1}; \ell^{12}\rangle$, which is then combined with ℓ^3 to form the three-particle state $|\ell^1_{D-1} \ell^2_{D-1} \ell^3_{D-1}; \ell^{123}\rangle$, etc. In full, we obtain the $(N-1)$ -particle basis function

$$\boxed{|\mathbf{\Lambda}; \mathbf{L}\rangle = \sum_{[\ell^1][\ell^2] \dots [\ell^{N-1}]} C_{\ell^1 \ell^2 \dots \ell^{N-1}}^{\mathbf{\Lambda}; \mathbf{L}} |\ell^1\rangle |\ell^2\rangle \dots |\ell^{N-1}\rangle,} \quad [21]$$

summing over all intermediate angular momenta ℓ^i_k with $k < (D-1)$, as denoted by $[\ell^i]$.^{††} The combined state in Eq. 21 is specified by the set of total angular momentum indices $\mathbf{\Lambda} \equiv \{\ell^1_{D-1}, \ell^2_{D-1}, \ell^3_{D-1}, \ell^4_{D-1}, \dots, \ell^{12 \dots (N-2)}_{D-1}, \ell^{N-1}_{D-1}\}$ and involves the coupling coefficients

$$\begin{aligned} C_{\ell^1 \ell^2 \dots \ell^{N-1}}^{\mathbf{\Lambda}; \mathbf{L}} = \sum_{[\ell^{12}] \dots [\ell^{12 \dots (N-2)}]} \langle \ell^1, \ell^2 | \ell^{12} \rangle \langle \ell^{12}, \ell^3 | \ell^{123} \rangle \dots \\ \langle \ell^{12 \dots (N-2)}, \ell^{N-1} | \mathbf{L} \rangle, \end{aligned} \quad [22]$$

which is a product of CG symbols. We have additionally set $\ell^{12 \dots (N-1)} \equiv \mathbf{L}$, representing the combined angular momentum eigenvalues. Note that Eq. 21 contains a sum over both the primary indices ℓ^i (which define the single-particle states) and intermediates, e.g., $\ell^{12 \dots}$ (within the coupling definitions), but not those corresponding to total angular momenta, i.e., ℓ_{D-1} .

The meaning of Eq. 21 is straightforward; an $(N-1)$ -particle state with combined angular momentum \mathbf{L} can be obtained as a sum of product states, weighted by $(N-2)$ CG coefficients. To define the state uniquely, we must specify not only the total angular momentum ℓ_{D-1} of each single-particle state, but also the total angular momentum of the intermediate states, e.g., ℓ^1_{D-1} arising from the coupling of $|\ell^1\rangle$ and $|\ell^2\rangle$. In total, the state is specified by $(2N + D - 5)$ indices; again significantly fewer than the $(N-1)(D-1)$ required for the product state $|\ell^1\rangle \dots |\ell^{N-1}\rangle$.

A particularly interesting state is that of zero combined angular momentum, i.e., $\mathbf{L} = \mathbf{0}$.^{‡‡} From Eq. 21, this is simply

$$\boxed{|\mathbf{\Lambda}; \mathbf{0}\rangle = \sum_{[\ell^1] \dots [\ell^{N-1}]} C_{\ell^1 \dots \ell^{N-1}}^{\mathbf{\Lambda}; \mathbf{0}} |\ell^1\rangle |\ell^2\rangle \dots |\ell^{N-1}\rangle.} \quad [23]$$

Unlike the general state (Eq. 21), this involves only $(N-4)$ intermediate couplings $\ell^{12 \dots}$, since the final CG coefficient in Eq. 22 fixes $\ell^{12 \dots (N-2)} = -\ell^{N-1}$ and $\ell^{12 \dots (N-2)} = \ell^{N-1}$ for $K > 1$. In total, this requires $(2N-5)$ indices to fully specify, regardless of dimension.

C. $(N-1)$ -Particle Basis Properties. The $(N-1)$ -particle basis functions of Eqs. 21 and 23 have analogous properties to those of the single-particle basis (cf Section 1C). Using Eq. 18, we can show orthonormality:

$$\langle \mathbf{\Lambda}'; \mathbf{L}' | \mathbf{\Lambda}; \mathbf{L} \rangle = \delta_{\mathbf{\Lambda} \mathbf{\Lambda}'}^K \delta_{\mathbf{L} \mathbf{L}'}^K, \quad [24]$$

requiring equality of both the combined angular momentum vectors (\mathbf{L} and \mathbf{L}') and all components of $\mathbf{\Lambda}$ and $\mathbf{\Lambda}'$. Since the angular momentum states form a complete basis, any $(N-1)$ -particle function $h : (\mathbb{S}^{D-1})^{\otimes(N-1)}$ can be decomposed into a sum of basis states:

$$\begin{aligned} |h\rangle &= \sum_{\mathbf{L}} \sum_{\mathbf{\Lambda}} \langle \mathbf{\Lambda}; \mathbf{L} | h \rangle |\mathbf{\Lambda}; \mathbf{L}\rangle \leftrightarrow h(\hat{\mathbf{r}}^1, \dots, \hat{\mathbf{r}}^{N-1}) \\ &= \sum_{\mathbf{L}} \sum_{\mathbf{\Lambda}} h_{\mathbf{\Lambda}}^{\mathbf{L}} \mathcal{P}_{\mathbf{\Lambda}}^{\mathbf{L}}(\hat{\mathbf{r}}^1, \dots, \hat{\mathbf{r}}^{N-1}), \end{aligned} \quad [25]$$

summing over both combined angular momentum indices \mathbf{L} and the indices contained within $\mathbf{\Lambda}$ (analogously to Eq. 15). The basis components are denoted by $\langle \mathbf{\Lambda}; \mathbf{L} | h \rangle \equiv h_{\mathbf{\Lambda}}^{\mathbf{L}}$. For the second equality, we switch to wavefunction notation, with the basis functions defined as (cf Eq. 21)

^{††}For $D = 3$, our basis functions match the coupled representation of $SU(2)$ discussed in ref. 24 in the context of quantum chemistry.

^{‡‡}For $\mathbb{M}^D = \mathbb{R}^3$, our treatment of the $L = 0$ states exactly follows that of ref. 36.

$$\mathcal{P}_\Lambda^L(\hat{\mathbf{r}}^1, \dots, \hat{\mathbf{r}}^{N-1}) = \sum_{[\ell^1][\ell^2]\dots[\ell^{N-1}]} C_{\ell^1 \ell^2 \dots \ell^{N-1}}^{\Lambda; L} Y_{\ell^1}(\hat{\mathbf{r}}^1) Y_{\ell^2}(\hat{\mathbf{r}}^2) \dots Y_{\ell^{N-1}}(\hat{\mathbf{r}}^{N-1}), \quad [26]$$

where $Y_\ell(\hat{\mathbf{r}})$ are the hyperspherical harmonics of Section 1B. Since \mathbb{S}^{D-1} is the angular part of \mathbb{M}^D , the directional dependence of any $(N-1)$ -particle function in \mathbb{M}^D can be expanded in the separable form of Eq. 25.

If the function h appearing in Eq. 25 has rotational symmetry, a simpler decomposition is possible. In particular, we assume it to be invariant under rotations drawn from the subspace $\{x_1, \dots, x_K, 0, \dots, 0\}$ of $SO(D)$. An example of this would be azimuthal symmetry in $D=3$; here, the system is invariant under rotations about one axis only. To obey rotational symmetry in K dimensions, the basis functions must satisfy

$$\widehat{\mathcal{L}}_K^2 |\Lambda; \mathbf{L}\rangle = 0 \quad \Rightarrow \quad L_i = 0, \quad \forall i < K, \quad [27]$$

where $\widehat{\mathcal{L}}_K^2$ is the combined angular momentum operator (Eq. 12).^{§§} In this instance, only basis functions of the form $|\Lambda; 0 \dots 0 L_K \dots L_{D-1}\rangle$ enter into Eq. 25, reducing the number of basis coefficients from approximately $(\ell_{D-1}^{\max})^{2N+D-5}$ to $(\ell_{D-1}^{\max})^{2N+D-K-5}$, for maximum multipole ℓ_{D-1}^{\max} . If h is invariant under spatial rotations about *any* axis (i.e., it is isotropic), only the $\mathbf{L} = \mathbf{0}$ state is required. In this case,

$$\begin{aligned} |h\rangle &= \sum_{\Lambda} \langle \Lambda; \mathbf{0} | h \rangle |\Lambda; \mathbf{0}\rangle \quad \leftrightarrow \quad h(\hat{\mathbf{r}}^1, \dots, \hat{\mathbf{r}}^{N-1}) \\ &= \sum_{\Lambda} h_{\Lambda}^0 \mathcal{P}_{\Lambda}^0(\hat{\mathbf{r}}^1, \dots, \hat{\mathbf{r}}^{N-1}), \end{aligned} \quad [28]$$

for components $h_{\Lambda}^0 \equiv \langle \Lambda; \mathbf{0} | h \rangle$, which may be determined via orthonormality. The directional dependence of any isotropic $(N-1)$ -particle function in \mathbb{M}^D can be expanded in the separable form of Eq. 28.

For $N=3$ and $N=4$, the isotropic basis functions take the explicit forms:

$$\begin{aligned} \mathcal{P}_{\ell_{D-1}}^0(\hat{\mathbf{r}}^1, \hat{\mathbf{r}}^2) &= \sum_{\ell_1 \dots \ell_{D-2}} Y_{\ell_1 \ell_2 \dots \ell_{D-1}}(\hat{\mathbf{r}}^1) Y_{(-\ell_1) \ell_2 \dots \ell_{D-1}}(\hat{\mathbf{r}}^2), \\ \mathcal{P}_{\ell_{D-1}^1 \ell_{D-1}^2 \ell_{D-1}^3}^0(\hat{\mathbf{r}}^1, \hat{\mathbf{r}}^2, \hat{\mathbf{r}}^3) &= \sum_{\ell_1^1 \dots \ell_{D-2}^1} \sum_{\ell_1^2 \dots \ell_{D-2}^2} \sum_{\ell_1^3 \dots \ell_{D-2}^3} \langle \ell^1, \ell^2 | (-\ell_1^3) \ell_2^3 \dots \ell_{D-1}^3 \rangle \\ &Y_{\ell_1^1 \dots \ell_{D-1}^1}(\hat{\mathbf{r}}^1) Y_{\ell_1^2 \dots \ell_{D-1}^2}(\hat{\mathbf{r}}^2) Y_{\ell_1^3 \dots \ell_{D-1}^3}(\hat{\mathbf{r}}^3), \end{aligned} \quad [29]$$

noting that the final CG coefficient is of the form $\langle \ell^{12 \dots (N-2)}, \ell^{N-1} | \mathbf{0} \rangle$, which enforces $\ell_1^{12 \dots (N-2)} + \ell_1^{N-1} = 0$ and $\ell_K^{12 \dots (N-2)} = \ell_K^{N-1}$ for $K > 1$. This is a natural extension of the Legendre polynomials to D dimensions; indeed, the $D=3$ case recovers the Legendre polynomial $\mathcal{L}_{\ell_2}(\hat{\mathbf{r}}^1 \cdot \hat{\mathbf{r}}^2)$ rescaled by $(-1)^{\ell_2} (4\pi) / \sqrt{2\ell_2 + 1}$ (cf ref. 36).

Finally, we note the properties of the basis functions under complex conjugation and parity inversion:

$$\begin{aligned} [\mathcal{P}_{\Lambda}^L(\hat{\mathbf{r}}^1, \dots, \hat{\mathbf{r}}^{N-1})]^* &= (-1)^{L_{D-1} - L_1} (-1)^{\ell_{D-1}^1 + \dots + \ell_{D-1}^{N-1}} \\ &\mathcal{P}_{\Lambda}^0(\hat{\mathbf{r}}^1, \dots, \hat{\mathbf{r}}^{N-1}) \\ \mathbb{P} [\mathcal{P}_{\Lambda}^L(\hat{\mathbf{r}}^1, \dots, \hat{\mathbf{r}}^{N-1})] &= (-1)^{\ell_{D-1}^1 + \dots + \ell_{D-1}^{N-1}} \\ &\mathcal{P}_{\Lambda}^0(\hat{\mathbf{r}}^1, \dots, \hat{\mathbf{r}}^{N-1}), \end{aligned} \quad [30]$$

using *SI Appendix, Eq. S4*, noting that the CG coefficients enforce $\ell_1^1 + \ell_1^2 + \dots + \ell_1^{N-1} + L_1 = 0$. For $\mathbf{L} = \mathbf{0}$, this implies that even-(odd-)parity basis functions are purely real (imaginary).

3. An Efficient Correlation Function Estimator

Armed with the $(N-1)$ -particle angular basis of Section 2, we now proceed to construct an efficient estimator for the NPCF components. For full generality, we do not assume the NPCF to have any rotational symmetry; such symmetries set various basis components to zero, as discussed in Section 2C.

A. Derivation. Assuming statistical homogeneity, the NPCF defined in Eq. 1 is a function of $(N-1)$ points on \mathbb{M}^D and may thus be expanded in the combined angular momentum basis of Section 2 (cf Eq. 4):

$$\zeta(\mathbf{r}^1, \dots, \mathbf{r}^{N-1}) = \sum_{\mathbf{L}} \sum_{\Lambda} \zeta_{\Lambda}^{\mathbf{L}}(r^1, \dots, r^{N-1}) \mathcal{P}_{\Lambda}^{\mathbf{L}}(\hat{\mathbf{r}}^1, \dots, \hat{\mathbf{r}}^{N-1}), \quad [31]$$

where the basis states, $\mathcal{P}_{\Lambda}^{\mathbf{L}}$, are defined in Eq. 26. As before, we sum both over \mathbf{L} , which specifies the properties of the NPCF under joint rotations of all $(N-1)$ direction vectors (with only $\mathbf{L} = \mathbf{0}$ required if the NPCF is isotropic), and $\Lambda \equiv \{\ell_{D-1}^1, \ell_{D-1}^2, \ell_{D-1}^{12}, \ell_{D-1}^3, \ell_{D-1}^{123}, \ell_{D-1}^4, \dots, \ell_{D-1}^{N-1}\}$, which defines the relative orientations of the direction vectors. In this form, the NPCF is fully specified by the basis coefficients $\zeta_{\Lambda}^{\mathbf{L}}$, which are functions only of the radial parameters r^i .

Due to the parity properties of the basis functions (Eq. 30), basis coefficients with even (odd) $\sum_{i=1}^{N-1} \ell_{D-1}^i$ represent even-parity (odd-parity) NPCF contributions; furthermore, they are purely real (imaginary) if the random field X is real-valued and $\mathbf{L} = \mathbf{0}$. Parity-odd isotropic basis functions occur only for $N > D$; at lower orders, a parity transformation is equivalent to a rotation, under which the basis functions are invariant.

The basis coefficients can be extracted from Eq. 31 via an inner product:

$$\begin{aligned} \zeta_{\Lambda}^{\mathbf{L}}(r^1, \dots, r^{N-1}) &\equiv \langle \Lambda; \mathbf{L} | \zeta \rangle \\ &= \int_{(\mathbb{S}^{D-1})^{\otimes (N-1)}} d\Omega_{D-1}^1 \dots d\Omega_{D-1}^{N-1} \\ &\left[\zeta(\mathbf{r}^1, \dots, \mathbf{r}^{N-1}) \mathcal{P}_{\Lambda}^{\mathbf{L},*}(\hat{\mathbf{r}}^1, \dots, \hat{\mathbf{r}}^{N-1}) \right], \end{aligned} \quad [32]$$

where the integral is over $(N-1)$ copies of the angular space. As in Eq. 2, the NPCF may be estimated as a product of N

^{§§}This occurs since any basis function with $L_K \neq 0$ has nonzero combined angular momentum in the K -dimensional subspace, violating rotational invariance.

random fields, integrated over space; inserted into Eq. 32, this yields

$$\hat{\zeta}_{\Lambda}^L(r^1, \dots, r^{N-1}) = \frac{1}{V_D} \int_{\mathbb{M}^D} d^D \mathbf{s} \int_{(\mathbb{S}^{D-1})^{\otimes (N-1)}} d\Omega_{D-1}^1 \cdots d\Omega_{D-1}^{N-1} \left[X(\mathbf{s}) X(\mathbf{s} + \mathbf{r}^1) \cdots X(\mathbf{s} + \mathbf{r}^{N-1}) \mathcal{P}_{\Lambda}^{L,*}(\hat{\mathbf{r}}^1, \dots, \hat{\mathbf{r}}^{N-1}) \right]. \quad [33]$$

Finally, we insert the explicit forms of the $(N-1)$ -particle basis functions (Eq. 26), which yields

$$\hat{\zeta}_{\Lambda}^L(r^1, \dots, r^{N-1}) = \frac{1}{V_D} \sum_{[\ell^1] \dots [\ell^{N-1}]} C_{\ell^1 \dots \ell^{N-1}}^{\Lambda; L} \int_{\mathbb{M}^D} d^D \mathbf{s} \left[X(\mathbf{s}) a_{\ell^1}(\mathbf{s}; r^1) \cdots a_{\ell^{N-1}}(\mathbf{s}; r^{N-1}) \right], \quad [34]$$

defining the functions

$$a_{\ell}(\mathbf{s}; r) \equiv \int_{\mathbb{S}^{D-1}} d\Omega_{D-1} X(\mathbf{s} + \mathbf{r}) Y_{\ell}^*(\hat{\mathbf{r}}). \quad [35]$$

Importantly, the angular integrals are now fully decoupled.

Usually, the NPCF statistic is binned in radius via a set of $(N-1)$ top-hat filters, $\Theta^{b^K}(r^K)$, which are equal to one if r^K is in bin b^K and zero else. In this case, $\hat{\zeta}_{\Lambda}^L(r^1, \dots, r^{N-1})$ is replaced by its binned form $\hat{\zeta}_{\Lambda}^{L,b}$, where $\mathbf{b} \equiv \{b^1, \dots, b^{N-1}\}$. This is again estimated using Eq. 34, defining the bin-integrated functions

$$a_{\ell}^b(\mathbf{s}) = \frac{1}{v_b} \int_{\mathbb{M}^D} d^D \mathbf{r} X(\mathbf{s} + \mathbf{r}) Y_{\ell}^*(\hat{\mathbf{r}}) \Theta^b(r), \quad [36]$$

where $v_b \equiv \int_{\mathbb{M}^D} d^D \mathbf{r}$ is the bin volume. Assuming a fixed maximum multipole ℓ_{D-1}^{\max} and some number of bins N_b , this ensures that only a finite number of $a_{\ell}^b(\mathbf{s})$ coefficients (asymptotically, $N_b \times (\ell_{D-1}^{\max})^{N-1}$) need to be estimated at each position \mathbf{s} . If one wished to reconstruct the full correlator $\zeta(\mathbf{r}^1, \dots, \mathbf{r}^{N-1})$ from the set of measured basis coefficients, this truncation would lead to an approximation error. In practice, this can be avoided by projecting the theory model in the same manner as the data; then, using a low ℓ_{D-1}^{\max} will lead only to a slight loss of information, depending on the model in question.

Estimation of the NPCF basis components reduces to two operations: 1) computing $a_{\ell}^b(\mathbf{s})$ for each radial bin b and angular momentum eigenvalue set ℓ of interest, and 2) performing a spatial integral over \mathbf{s} , alongside a sum over the lower-dimensional angular momentum eigenvalues $[\ell^i] \equiv \{\ell_1^i, \dots, \ell_{D-2}^i\}$, subject to the coupling rules of Eqs. 9 and 20. Computationally, this is much more efficient than a direct implementation of Eq. 33. A cartoon indicating this procedure for $N=4$ is shown in Fig. 2.

B. Application to Discrete Data. For discrete data, the random field X can be represented as a (weighted) sum of Dirac deltas, as

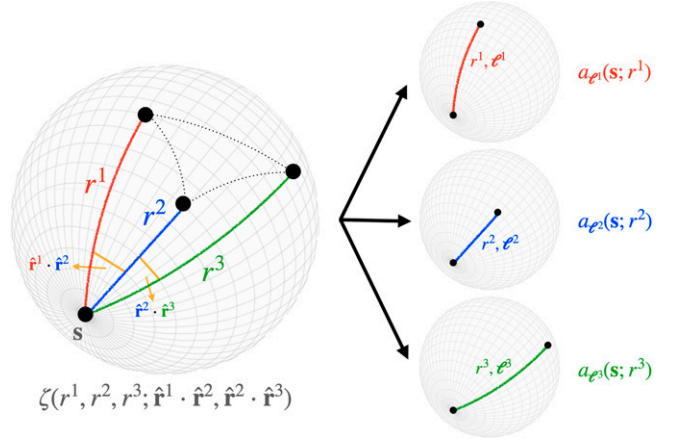


Fig. 2. Sketch of the decomposition underlying our NPCF algorithm, visualized for $N=4$ in a two-dimensional spherical geometry. On the left, we show the (rotation-averaged) 4PCF defined by three distances, r^1, r^2, r^3 , and two angles $\hat{\mathbf{r}}^1 \cdot \hat{\mathbf{r}}^2, \hat{\mathbf{r}}^2 \cdot \hat{\mathbf{r}}^3$, relative to a primary position \mathbf{s} ; naïve 4PCF estimation from n particles proceeds by summing over each of the possible n^4 sets of points. On the right, we show our decomposition, factorizing the 4PCF into three functions of two positions, $a_{\ell}(\mathbf{s}; r)$, which may be independently estimated from the dataset, requiring consideration of only n^2 pairs of particles. Each function depends on a side length r^i and a set of angular momentum indices ℓ^i ; the latter specifies $\hat{\mathbf{r}}^i$ in the hyperspherical harmonic basis. This permits the NPCF components to be estimated by an algorithm with $\mathcal{O}(n^2)$ complexity.

in Eq. 3. Inserting this definition into Eq. 34 leads to the following estimator for the NPCF basis coefficients in bins \mathbf{b} :

$$\hat{\zeta}_{\Lambda}^{L,b} = \frac{1}{V_D} \sum_{j=1}^n w^j \sum_{[\ell^1] \dots [\ell^{N-1}]} C_{\ell^1 \dots \ell^{N-1}}^{\Lambda; L} \left[a_{\ell^1}^{b^1}(\mathbf{y}^j) \cdots a_{\ell^{N-1}}^{b^{N-1}}(\mathbf{y}^j) \right], \quad [37]$$

$$a_{\ell}^b(\mathbf{y}^j) = \frac{1}{v_b} \sum_{k=1}^n w^k Y_{\ell}^*(\widehat{\mathbf{y}^k - \mathbf{y}^j}) \Theta^b(|\mathbf{y}^k - \mathbf{y}^j|).$$

Practically, the $a_{\ell}^b(\mathbf{y}^j)$ functions may be computed by summing over n points, $\{\mathbf{y}^k\}$, weighted by a hyperspherical harmonic and a binning function in the separation vector $\mathbf{y}^k - \mathbf{y}^j$. Since the functions must be estimated at the location of each primary particle, \mathbf{y}^j , the algorithm has complexity $\mathcal{O}(n^2)$ (with respect to n) for any choice of D or N . This is significantly faster than the naïve NPCF estimator of Eq. 3, inserting the basis functions

$$B(\mathbf{r}^1, \dots, \mathbf{r}^{N-1}) \equiv \mathcal{P}_{\Lambda}^L(\hat{\mathbf{r}}^1, \dots, \hat{\mathbf{r}}^{N-1}) \Theta^{b^1}(r^1) \cdots \Theta^{b^{N-1}}(r^{N-1}), \quad [38]$$

showing the utility of our hyperspherical harmonic decomposition.

Although the scaling with n is independent of D and N , we do expect the computational cost of a full NPCF measurement to increase in higher-dimensional scenarios. In part, this occurs since the number of intermediate $[\ell^i]$ summations in Eq. 37 is a strong function of D and N . In total, we must sum over approximately $(D-1)(N-1)$ indices, each of which is bounded by ℓ_{D-1}^{\max} ; thus, the computation time is asymptotically $\propto (\ell_{D-1}^{\max})^{(D-1)(N-1)}$. Since this summation must be done once per primary particle \mathbf{y}^j , we expect the algorithm to scale linearly with n if this process dominates over a_{ℓ}^b computation. Secondly, the number of basis vectors at fixed ℓ_{D-1}^{\max} is exponential in D

and N . Asymptotically, this scales as $N_b^{N-1} \times (\ell_{D-1}^{\max})^{2N-4} \times (\ell_{D-1}^{\max})^{D-1}$ (counting elements of \mathbf{b} , $\mathbf{\Lambda}$ and \mathbf{L} , respectively, using N_b radial bins per dimension). Although this is a strong scaling, it is generic to any higher-point basis (and usually referred to as the ‘‘curse of dimensionality’’). In the presence of certain symmetries, the number of basis functions is significantly reduced: For example, isotropy demands that $\mathbf{L} = \mathbf{0}$, reducing the scaling to $N_b^{N-1} \times (\ell_{D-1}^{\max})^{2N-5}$, independent of D . In the below, we will always compare naïve and efficient estimators projected onto the same basis; thus, this factor appears in the ratio of computation times.

C. Application to Gridded Data. Our estimators may also be applied to continuous data discretized on some regular grid, which is of use for the analysis of hydrodynamic simulations, for example. For this, we first rewrite $a_\ell^b(\mathbf{s}; r)$ (Eq. 36) as

$$\begin{aligned} a_\ell^b(\mathbf{s}) &= \int_{\mathbb{M}^D} d^D \mathbf{r} X(\mathbf{s} + \mathbf{r}) Y_\ell^*(\hat{\mathbf{r}}) \Theta^b(r) \\ &\equiv (-1)^{\ell_{D-1}} \int_{\mathbb{M}^D} d^D \mathbf{r} X(\mathbf{s} - \mathbf{r}) [Y_\ell(\hat{\mathbf{r}}) \Theta^b(r)], \end{aligned} \quad [39]$$

relabeling variables and utilizing the conjugate properties of hyperspherical harmonics (SI Appendix, Eq. S4). For gridded data in Euclidean space, i.e., with $\mathbb{M}^D = \mathbb{R}^D$, this may be straightforwardly computed using the D -dimensional FFT. Explicitly:

$$a_\ell^b(\mathbf{s}) = (-1)^{\ell_{D-1}} \text{FFT}^{-1} [\text{FFT}(X) \text{FFT}(Y_\ell \Theta^b)], \quad [40]$$

where FFT^{-1} is the inverse FFT. These operations have complexity $\mathcal{O}(n_g \log n_g)$ for n_g grid points. Following computation of the various $a_\ell^b(\mathbf{s})$ terms, the estimator for the NPCF components can be constructed from Eq. 34 as a simple sum in D dimensions. The full estimator has complexity $\mathcal{O}(n_g \log n_g)$, which is again much faster than the naïve $\mathcal{O}(n_g^N)$ result. The above procedure may also be applied to the discrete data-sets discussed in Section B, via a nonuniform FFT (37).

4. Applications

We now consider a number of physical scenarios in which the above methods can be employed and give numerical examples. For this purpose, we provide a Julia implementation of the two main algorithms discussed above (3,37).^{¶¶} This can compute the NPCF of discrete particles with $N \in \{2, 3, 4, 5\}$, using Cartesian geometries with $D \in \{2, 3, 4\}$ or spherical geometries with $D = 2$, and is fully parallelized.

A. Cosmic Microwave Background. Cosmic Microwave Background (CMB) radiation encodes a snapshot of the Universe at the epoch of recombination, around 380,000 y after the Big Bang. This may be probed using microwave satellites such as *WMAP* and *Planck*, which map the CMB temperature fluctuations as a function of direction; such observations have been used to place strong constraints on cosmological parameters such as the matter density and Universe’s expansion rate (e.g., refs. 38 and 39).

In this setting, the random field in question is the fractional temperature fluctuation on the 2-sphere, $\Theta : \mathbb{S}^2 \rightarrow \mathbb{R}$, i.e., $\mathbb{M}^D = \mathbb{S}^2$. Conventionally, Θ is expanded in $D = 3$ spherical harmonics

as $\Theta(\phi, \theta; \hat{\mathbf{n}}) = \sum_{\lambda=0}^{\infty} \sum_{\mu=-\lambda}^{\lambda} \Theta_{\lambda\mu} Y_{\lambda\mu}(\phi, \theta)$, using spherical polar coordinates $\phi \in [0, 2\pi)$, $\theta \in [0, \pi)$ relative to some pole on the sphere at position vector $\hat{\mathbf{n}}$. The statistical properties of Θ are then characterized in terms of the spherical harmonic coefficients $\Theta_{\lambda\mu}$ (often denoted $a_{\lambda\mu}^T$). To apply the techniques of this work, we instead expand the temperature fluctuations using the $D = 2$ hyperspherical harmonics (Section 1B), i.e.,

$$\Theta(\mathbf{r}; \hat{\mathbf{n}}) = \sum_{\ell=0}^{\infty} \Theta_\ell(\theta) Y_\ell(\phi) \equiv \frac{1}{\sqrt{2\pi}} \sum_{\ell=0}^{\infty} \Theta_\ell(\theta) e^{i\ell\phi}, \quad [41]$$

identifying ℓ as the total angular momentum and θ as the radial coordinate relative to $\hat{\mathbf{n}}$, which acts as an origin on \mathbb{S}^2 .^{##} This is a convenient basis for computing higher-order clustering statistics on the 2-sphere, since 1) it avoids the need for an embedding space, and 2) it provides a natural split into isotropic and anisotropic correlators.

As in Eq. 1, the temperature NPCF is defined as a statistical average over Θ :

$$\zeta(\mathbf{r}^1, \dots, \mathbf{r}^{N-1}) = \mathbb{E}_\Theta [\Theta(\mathbf{0}; \hat{\mathbf{n}}) \Theta(\mathbf{r}^1; \hat{\mathbf{n}}) \dots \Theta(\mathbf{r}^{N-1}; \hat{\mathbf{n}})]; \quad [42]$$

by statistical homogeneity, this is independent of the choice of origin $\hat{\mathbf{n}}$. Eq. 42 may be expanded in the basis of Eq. 25, where the $D = 2$ basis functions take the form

$$\begin{aligned} \mathcal{P}_{\ell^1 \dots \ell^{N-1}}^L(\phi^1, \dots, \phi^{N-1}) &= (2\pi)^{-(N-1)/2} \exp \\ &[i(\ell^1 \phi^1 + \dots + \ell^{N-1} \phi^{N-1})], \end{aligned} \quad [43]$$

with $\ell^1 + \dots + \ell^{N-1} = L$, where $L = 0$ for isotropic correlators. Note that no intermediate angular momenta need to be specified due to the coupling rules of Eq. 20. As in Eq. 34, we may form an $\mathcal{O}(n^2)$ estimator for the NPCF coefficients:

$$\begin{aligned} \zeta_{\ell^1 \dots \ell^{N-1}}^L(\theta^1, \dots, \theta^{N-1}) &= \frac{1}{4\pi} \int_{\mathbb{S}^2} d\Omega_2 [\Theta(\hat{\mathbf{n}}) a_{\ell^1}(\hat{\mathbf{n}}; \theta^1) \\ &\dots a_{\ell^{N-1}}(\hat{\mathbf{n}}; \theta^{N-1})], \\ a_\ell(\hat{\mathbf{n}}; \theta) &\equiv \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} d\phi \Theta(\phi, \theta; \hat{\mathbf{n}}) e^{i\ell\phi}. \end{aligned} \quad [44]$$

The first integral is over all possible choices of origin $\hat{\mathbf{n}}$, while the second is over a circle centered at $\hat{\mathbf{n}}$ with radial parameter θ (which may be discretized into bins, as before).

Such estimators are straightforward to implement and allow efficient computation of the higher-order CMB NPCFs, albeit in a basis somewhat different to that usually adopted. We caution that the ℓ indices play a different role to those of the λ indices appearing in the standard spherical harmonic expansion of Θ . In our basis, ℓ represents angular momentum around an origin *on* the 2-sphere, while λ is with reference to the origin *of* the 2-sphere in the embedding space \mathbb{R}^3 . In practice, this allows us to restrict

^{¶¶} <https://github.com/oliverphilcox/NPCFs.jl>.

^{##} Here, θ measures the arc-lengths of great-circles through two points on the 2-sphere, as viewed in \mathbb{R}^3 .

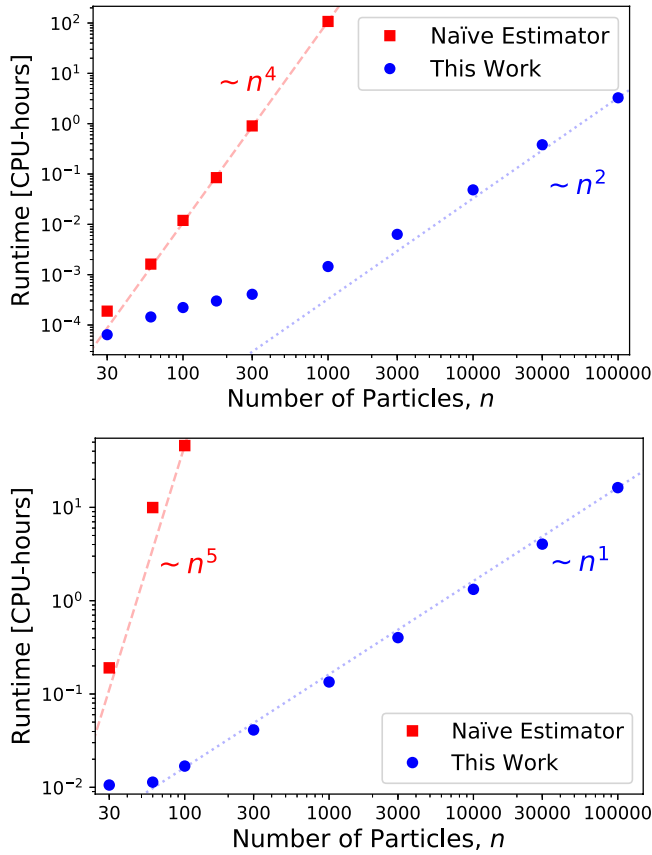


Fig. 3. Timing comparison of the naïve NPCF estimator with that introduced in this work. In the first case, the NPCF is estimated by counting sets of N particles, via Eq. 3, weighting by the relevant basis functions (Section 2). The new estimators exploit hyperspherical harmonic decompositions to reduce the estimator to a sum over pairs. Here, we show results for a variety of choices of n , for both 4PCF estimation on the 2-sphere (*Upper*) and 5PCF estimation in 3D Cartesian space (*Lower*). As expected, the runtime of naïve estimator scales as n^N (as indicated by the red dotted lines), but the new estimators scale as n^2 for the 4PCF, or n for the 5PCF. The latter scaling arises since we are dominated by the sum over intermediate momenta $[\ell']$, rather than the sum over pairs; at larger n , we expect a quadratic scaling with n . All computations were performed in Julia using 16 CPUs, and we have verified that the measured NPCF components agree to machine precision.

to much smaller ℓ than used conventionally.^{|||} We further note that the CMB contains also polarization fluctuations. These may be analyzed using an extension of the above formalism, replacing the hyperspherical harmonics with spin-weighted hyperspherical harmonics (e.g., refs. 40 and 41).

To give a sense of how the above algorithms work in CMB contexts, we consider a simple problem: estimating the isotropic four-point correlation function (4PCF) of randomly placed points on the 2-sphere. This corresponds to the above scenarios with $N = 4$, $D = 2$, and a spherical geometry. For this test, we generate a set of n evenly distributed points and compute the coefficients $\zeta_{\ell^1 \ell^2 \ell^3}^{0, b^1 b^2 b^3}$, using 10 radial bins per dimension with $\cos \theta \in [-0.5, 0.5]$ and $\ell_{\max} = 4$. This leads to a total of 35 (120) angular (radial) components. Timing results for the computation using

both the naïve and quadratic estimators (projecting the 4PCF onto the same basis functions in both cases) are shown in Fig. 3, *Upper* for a variety of choices of n , using our public Julia code. As expected, the runtime scales as n^4 for the naïve estimator, which leads to unwieldy computation times, even for a few thousand particles. For the estimators introduced in this work, the runtime scales as n^2 for large n , as expected from the algorithm's complexity.

B. Hydrodynamic Turbulence. NPCFs have found significant use in the study of hydrodynamical turbulence. Being a chaotic process, the evolution of the velocity and density fields in a turbulent flow cannot be treated deterministically; rather, they must be analyzed statistically. Furthermore, the density fields of turbulent media are known to be close to log-normal (42), implying that the higher-order NPCF functions contain a wealth of information, particularly concerning the sonic and Alfvénic Mach numbers (e.g., refs. 15 and 43–45).

One of the simplest observables is the turbulent density field, $\rho: \mathbb{R}^3 \rightarrow \mathbb{R}$, whose N -point function is defined in Eq. 1. In the absence of any external forcing, we expect the NPCF to be statistically isotropic; thus, it can be expanded via Eq. 28 in terms of the $D = 3$ basis functions with $\mathbf{L} = \mathbf{0}$. As shown in Section 1B, the corresponding one-particle basis functions are just the usual spherical harmonics, $Y_{m\ell}(\hat{\mathbf{r}})$, and their coupling can be expressed in terms of Wigner 3- j symbols.

As before, we may form $\mathcal{O}(n^2)$ estimators for the NPCF coefficients via Eq. 34. As an example, the isotropic five-point correlation function (5PCF) estimator becomes (ignoring radial binning for clarity)

$$\begin{aligned} & \zeta_{\ell^1 \ell^2 \ell^3 \ell^4}^0(r^1, r^2, r^3) \\ &= \frac{1}{V_3} (-1)^{\ell^1 + \ell^2 + \ell^3 + \ell^4} \sum_{m^1 m^2 m^3 m^4} (-1)^{\ell^{12} - m^{12}} \sqrt{2\ell^{12} - 1} \\ & \quad \begin{pmatrix} \ell^1 & \ell^2 & \ell^{12} \\ m^1 & m^2 & -m^{12} \end{pmatrix} \begin{pmatrix} \ell^{12} & \ell^3 & \ell^4 \\ m^{12} & m^3 & m^4 \end{pmatrix} \\ & \quad \times \int_{\mathbb{R}^3} d\mathbf{s} [\rho(\mathbf{s}) a_{m^1 \ell^1}(\mathbf{s}; r^1) \\ & \quad a_{m^2 \ell^2}(\mathbf{s}; r^2) a_{m^3 \ell^3}(\mathbf{s}; r^3) a_{m^4 \ell^4}(\mathbf{s}; r^4)], \end{aligned} \quad [45]$$

where V_3 is the volume of the space, $m^{12} \equiv m^1 + m^2$, and $a_{m\ell}(\mathbf{s}; r) = \int_{\mathbb{S}^2} d\Omega_2 \rho(\mathbf{s} + \mathbf{r}) Y_{m\ell}^*(\hat{\mathbf{r}})$. Introducing spin-weighted (or vector) spherical harmonics, the approach may be extended to tensorial correlators, such as those of the velocity field.

Fig. 3, *Lower* presents a practical demonstration of the isotropic 5PCF estimator, applied to n discrete points in 3D. We consider 10 radial bins in $[0.1, 0.4]$ for uniformly distributed data in a periodic cube of length one and fix $\ell_{\max} = 4$. In this case, the 5PCF is specified by four radial bin indices and five angular multiplets, as in Eq. 45. This gives a total of 210 radial and 585 angular components. As before, we find that the runtime of the naïve estimator scales as n^N , which quickly becomes computationally prohibitive. For the NPCF estimator of this work, the runtime appears to be linear in n , rather than quadratic: This occurs since the work is dominated by the m^i summations (cf Section 1B), though we expect the n^2 scaling to dominate for denser samples. In all cases, however, our approach is significantly faster than that of the naïve estimator. We note that our algorithm can be further accelerated by gridding the data and making use of Fourier transforms.

^{|||}CMB fluctuations have characteristic angular scale $r_s(z_*)/d_A(z_*)$, where r_s is the sound horizon in comoving coordinates, d_A is the angular diameter distance, and z_* is the redshift at the end of the baryon drag epoch. This imprints a characteristic angular momentum scale $L \sim \pi d_A(z_*)/r_s(z_*) \gg 1$. In our basis, ℓ corresponds to the ratio of two polygon sides on \mathbb{S}^2 and is thus $\mathcal{O}(1)$.

C. Large-Scale Structure. The large-scale distribution of matter in the late Universe follows a weakly non-Gaussian distribution and is commonly analyzed using N -point statistics (e.g., ref. 46). The underlying space is expected to be flat, homogeneous, and isotropic and is thus described by the metric of Eq. 6 with $k = 0$ and $D = 3$.*** A common task in cosmology is the estimation of isotropic NPCFs for the galaxy overdensity field, $\delta_g : \mathbb{R}^3 \rightarrow \mathbb{R}$; this proceeds identically to Section 4B, except that the data are discrete. Full discussion of this (including implementation in the `encore` code^{†††}) is presented in our companion work (16) and allows information to be extracted from the high-order NPCFs, which are otherwise computationally prohibitive to measure.

Due to the effects of redshift-space distortions (e.g., ref. 47), observed galaxy distributions are not isotropic, implying that the decomposition of Eq. 28 does not capture all possible NPCF information. However, the statistics are invariant under rotations about an (assumed fixed) line-of-sight, here set to $\hat{\mathbf{x}}^3$. For a full treatment, we must instead expand the NPCF using Eq. 25, keeping only terms with $L_1 = 0$ (cf Section 2C):

$$\begin{aligned} \zeta(\mathbf{r}^1, \dots, \mathbf{r}^{N-1}) \\ = \sum_{L=0}^{\infty} \sum_{\Lambda} \zeta_{\Lambda}^{0L}(r^1, \dots, r^{N-1}) \mathcal{P}_{\Lambda}^{0L}(\hat{\mathbf{r}}^1, \dots, \hat{\mathbf{r}}^{N-1}), \end{aligned} \quad [46]$$

writing $L \equiv L_2$. As an example, the 3PCF becomes

$$\begin{aligned} \zeta(\mathbf{r}^1, \mathbf{r}^2) &= \sum_{L=0}^{\infty} \sum_{\ell^1 \ell^2} \zeta_{\ell^1 \ell^2}^{0L}(r^1, r^2) \mathcal{P}_{\ell^1 \ell^2}^{0L}(\hat{\mathbf{r}}^1, \hat{\mathbf{r}}^2), \\ \mathcal{P}_{\ell^1 \ell^2}^{0L}(\hat{\mathbf{r}}^1, \hat{\mathbf{r}}^2) &= (-1)^{\ell^2 - \ell^1} \sqrt{2L+1} \\ &\sum_m \begin{pmatrix} \ell^1 & \ell^2 & L \\ m & -m & 0 \end{pmatrix} Y_{\ell^1 m}(\hat{\mathbf{r}}^1) Y_{\ell^2 -m}(\hat{\mathbf{r}}^2), \end{aligned} \quad [47]$$

(cf refs. 48 and 49). Such statistics may be estimated via Eq. 34, as before. The above decomposition provides a complete basis for the redshift-space 3PCF (analogous to ref. 49) and extends naturally to higher orders, which have not previously been discussed.

5. Summary

Many areas of research require computation of clustering statistics from continuous or discrete random fields. Perhaps the most prevalent statistic is the NPCF, defined as the statistical average

***Our methodology applies similarly to $k \neq 0$, though there is significant evidence implying that the Universe is flat (39). Additionally, our methods can be used to compute projected correlation functions in \mathbb{R}^2 , requiring the $D = 2$ basis functions, as demonstrated in ref. 11 for the 3PCF.

^{†††}<https://github.com/oliverphilcox/encore>.

1. S. Garrett-Roe, P. Hamm, Three-point frequency fluctuation correlation functions of the OH stretch in liquid water. *J. Chem. Phys.* **128**, 104507 (2008).
2. J. G. Berryman, Interpolating and integrating three-point correlation functions on a lattice. *J. Comput. Phys.* **75**, 86–102 (1988).
3. V. S. Dotseko, Three-point correlation functions of the minimal conformal theories coupled to 2D gravity. *Mod. Phys. Lett. A* **06**, 3601–3612 (1991).
4. K. Hwang, B. Schmittmann, R. K. P. Zia, Three-point correlation functions in uniformly and randomly driven diffusive systems. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* **48**, 800–809 (1993).
5. F. Sanda, S. Mukamel, Multipoint correlation functions for continuous-time random walk models of anomalous diffusion. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **72**, 031108 (2005).
6. A. W. Moore *et al.*, "Fast algorithms and efficient statistics: N-point correlation functions" in *Mining the Sky: Proceedings of the MPA/ESO/MPE Workshop Held at Garching, Germany, July 31 - August 4, 2000*, A. J. Banday, S. Zaroubi, M. Bartelmann, Eds. (ESO Astrophysics Symposia, Springer, Berlin, 2001), pp. 71–82.
7. P. J. E. Peebles, "Large scale clustering in the universe" in *Large Scale Structures in the Universe*, M. S. Longair, J. Einasto, Eds. (International Astronomical Union/Union Astronomique Internationale, Springer, Dordrecht, Netherlands, 1978), vol. 79, pp. 217–227.

over N fields in different spatial locations. If the random field is Gaussian-distributed, only the 2PCF is of interest; in the general case, all correlators have nontrivial forms. Given a set of n particles, a nave estimator for the NPCF components in some basis has $\mathcal{O}(n^N)$ complexity with respect to n . As N increases, this rapidly becomes computationally infeasible to apply: Alternative methods must be sought if one wishes to unlock the information contained within higher-order NPCFs.

This work considers NPCF estimation on isotropic and homogeneous manifolds in D dimensions. Under these assumptions (which encompass spherical, flat, and hyperbolic geometries), we show that any function of one position can be expanded in hyperspherical harmonics; a D -dimensional analog of the conventional spherical harmonics. These are also eigenstates of the angular momentum operators; utilizing the mathematics of angular momentum addition, we can construct basis functions involving $(N - 1)$ points on \mathbb{S}^{D-1} as a sum over products of $(N - 1)$ hyperspherical harmonics. This forms a natural angular basis for the NPCF, particularly if the random field is statistically isotropic. The decomposition allows construction of an NPCF estimator that separates into a product of $(N - 1)$ spatial integrals; this has $\mathcal{O}(n^2)$ complexity (with respect to n), or $\mathcal{O}(n_g \log n_g)$ using FFTs with n_g grid points. The algorithms have been validated numerically using a Julia implementation; in all scenarios tested, we find our approach to yield significantly faster measurements of the NPCF coefficients in our angular momentum basis.

Such techniques will allow high-order correlation functions to be computed from data, allowing more complete analysis of phenomena ranging from fluid turbulence to galaxy clustering. Furthermore, since the algorithm can be applied to scenarios with $D \neq 3$, we may consider also the computation of NPCFs on the surface of spheres (relevant, e.g., for atmospheric physics) or in higher-dimensional atomic treatments with $D = 6$ (e.g., ref. 19). These ideas may be extended further; a case of particular interest is in the correlation functions of random fields with nonzero spin; these are required to describe the statistics of tensor fields, such as turbulent velocities and CMB polarization.

Data Availability. Code Package NPCFs.jl data have been deposited in GitHub (<https://github.com/oliverphilcox/NPCFs.jl>) (50).

ACKNOWLEDGMENTS. We thank David Spergel for encouraging us to write this paper, as well as Robert Cahn, Simone Ferraro, Jiamin Hou, Bhuv Jain, Adam Lidz, Moritz Munchmeyer, Shivam Pandey, Ue-Li Pen, Cristiano Sabiu, Frederik Simons, Sauro Succi, and Wen Yan for insightful discussions. We are additionally indebted to the anonymous referees for insightful feedback. O.H.E.P. acknowledges funding from the Wide Field Infrared Survey Telescope program through Grants NNG26PJ30C and NNN12AA01C and thanks the Simons Foundation for additional support.

8. S. Alam *et al.*, Testing the theory of gravity with DESI: estimators, predictions and simulation requirements. *arXiv [Preprint]* (2020). <https://arxiv.org/abs/2011.05771> (Accessed 23 July 2022).
9. L. L. Zhang, U. L. Pen, Fast n-point correlation functions and three-point lensing application. *New Astron.* **10**, 569–590 (2005).
10. C. G. Sabiu, B. Hoyle, J. Kim, X. D. Li, Graph database solution for higher-order spatial statistics in the era of big data. *Astrophys. J.* **242** (suppl.), 29 (2019).
11. Z. Slepian, D. J. Eisenstein, Computing the three-point correlation function of galaxies in $\mathcal{O}(N^2)$ time. *Mon. Not. R. Astron. Soc.* **454**, 4142–4158 (2015).
12. I. Szapudi, Three-point statistics from a new perspective. *Astrophys. J.* **605**, L89–L92 (2004).
13. Z. Slepian *et al.*, Detection of baryon acoustic oscillation features in the large-scale three-point correlation function of SDSS BOSS DR12 CMASS galaxies. *Mon. Not. R. Astron. Soc.* **469**, 1738–1751 (2017).
14. Z. Slepian *et al.*, The large-scale three-point correlation function of the SDSS BOSS DR12 CMASS galaxies. *Mon. Not. R. Astron. Soc.* **468**, 1070–1083 (2017).
15. S. K. N. Portillo, Z. Slepian, B. Burkhardt, S. Kahraman, D. P. Finkbeiner, Developing the 3-point correlation function for the turbulent interstellar medium. *Astrophys. J.* **862**, 119 (2018).
16. O. H. E. Philcox *et al.*, ENCORE: Estimating galaxy N -point correlation functions in $\mathcal{O}(N_g^2)$ time. *arXiv [Preprint]* (2021). <https://arxiv.org/abs/2105.08722> (Accessed 23 July 2022).

17. S. Chatterjee, B. Bhui, Homogeneous cosmological model in higher dimension. *Mon. Not. R. Astron. Soc.* **247**, 57 (1990).
18. R. M. Wald, *General Relativity* (University of Chicago Press, Chicago, 1984).
19. J. W. Cooper, U. Fano, F. Prats, Classification of two-electron excitation levels of helium. *Phys. Rev. Lett.* **10**, 518–521 (1963).
20. H. Cohl, Fundamental solution of Laplace's equation in hyperspherical geometry. *SIGMA* **7**, 108, (2011).
21. C. Frye, C. J. Efthimiou, Spherical harmonics in p dimensions. *arXiv [Preprint]* (2012). <https://arxiv.org/abs/1205.3548> (Accessed 23 July 2022).
22. J. D. Louck, "Theory of angular momentum in n -dimensional space" (OSTI Tech. Rep., Office of Scientific and Technical Information, Department of Energy, Oak Ridge, TN, 1960).
23. A. Higuchi, Symmetric tensor spherical harmonics on the N -sphere and their application to the de Sitter group $SO(N,1)$. *J. Math. Phys.* **28**, 1553–1566 (1987).
24. A. Tichai, R. Wirth, J. Ripoché, T. Duguet, Symmetry reduction of tensor networks in many-body theory I. Automated symbolic evaluation of $SU(2)$ algebra. *arXiv [Preprint]* (2020). <https://arxiv.org/abs/2002.05011> (Accessed 23 July 2022).
25. L. C. Biedenharn, J. D. Louck, P. A. Carruthers, *Angular Momentum in Quantum Physics: Theory and Application*, Encyclopedia of Mathematics and its Applications (Cambridge University Press, Cambridge, UK, 1984).
26. D. A. Varshalovich, A. N. Moskalev, V. K. Khersonskii, *Quantum Theory of Angular Momentum* (World Scientific, Singapore, 1988).
27. S. Szalay *et al.*, Tensor product methods and entanglement optimization for ab initio quantum chemistry. *arXiv [Preprint]* (2014). <https://arxiv.org/abs/1412.5829> (Accessed 23 July 2022).
28. A. Tichai, R. Schutski, G. E. Scuseria, T. Duguet, Tensor-decomposition techniques for ab initio nuclear structure calculations: From chiral nuclear potentials to ground-state energies. *Phys. Rev. C* **99**, 034320 (2019).
29. P. Van Isacker, "O(4) symmetry and angular momentum theory in four dimensions" in *Symmetries in Science V*, B. Gruber, L. C. Biedenharn, H. D. Doebner, Eds. (Springer US, Boston, 1991), pp. 323–340.
30. L. C. Biedenharn, Wigner coefficients for the R_4 group and some applications. *J. Math. Phys.* **2**, 433–441 (1961).
31. J. Louck, "Angular momentum theory" in *Springer Handbook of Atomic, Molecular, and Optical Physics*, G. Drake, Ed. (Springer, New York, 2006), pp. 9–74.
32. M. A. Caprio, K. D. Sviratcheva, A. E. McCoy, Racah's method for general subalgebra chains: Coupling coefficients of $SO(5)$ in canonical and physical bases. *J. Math. Phys.* **51**, 093518 (2010).
33. G. Racah, "Group theory and spectroscopy" in *Springer Tracts in Modern Physics*, G. Höhler, Ed. (Springer, Berlin, Heidelberg, 1965), vol. 37, pp. 28–84.
34. F. W. J. Olver *et al.*, NIST digital library of mathematical functions (2021). <https://dlmf.nist.gov/> (Release 1.1.2 of 2021-06-15). Accessed 23 July 2022.
35. A. V. Meremianin, Multipole expansions in four-dimensional hyperspherical harmonics. *J. Phys. Math. Gen.* **39**, 3099–3112 (2006).
36. R. N. Cahn, Z. Slepian, Isotropic N -point basis functions and their properties. *arXiv [Preprint]* (2020). <https://arxiv.org/abs/2010.14418> (Accessed 23 July 2022).
37. A. Dutt, V. Rokhlin, Fast Fourier transforms for nonequispaced data. *SIAM J. Sci. Comput.* **14**, 1368–1393 (1993).
38. E. Komatsu *et al.*, Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Cosmological interpretation. *Astrophys. J.* **192** (suppl.), 18 (2011).
39. Planck Collaboration *et al.*, Planck 2018 results. VI. Cosmological parameters. *Astron. Astrophys.* **641**, A6 (2020).
40. L. Dai, M. Kamionkowski, D. Jeong, Total angular momentum waves for scalar, vector, and tensor fields. *Phys. Rev. D Part. Fields Gravit. Cosmol.* **86**, 125013 (2012).
41. M. Boyle, How should spin-weighted spherical functions be defined? *J. Math. Phys.* **57**, 092504 (2016).
42. E. Vazquez-Semadeni, Hierarchical structure in nearly pressureless flows as a consequence of self-similar statistics. *Astrophys. J.* **423**, 681 (1994).
43. G. Kowal, A. Lazarian, A. Beresnyak, Density fluctuations in MHD turbulence: Spectra, Intermittency, and topology. *Astrophys. J.* **658**, 423–445 (2007).
44. B. Burkhart, D. Falceta-Gonçalves, G. Kowal, A. Lazarian, Density studies of MHD interstellar turbulence: Statistical moments, correlations and bispectrum. *Astrophys. J.* **693**, 250–266 (2009).
45. B. Burkhart, S. Stanimirović, A. Lazarian, G. Kowal, Characterizing magnetohydrodynamic turbulence in the small magellanic cloud. *Astrophys. J.* **708**, 1204–1220 (2010).
46. P. J. E. Peebles, "The galaxy and mass N -point correlation functions: A blast from the past" in *Historical Development of Modern Cosmology*, V. J. Martínez, V. Trimble, M. J. Pons-Bordería, Eds. (Astronomical Society of the Pacific Conference Series, Astronomical Society of the Pacific, San Francisco, 2001), vol. 252, pp. 201–218.
47. N. Kaiser, Clustering in real space and in redshift space. *Mon. Not. R. Astron. Soc.* **227**, 1–21 (1987).
48. Z. Slepian, D. J. Eisenstein, A practical computational method for the anisotropic redshift-space three-point correlation function. *Mon. Not. R. Astron. Soc.* **478**, 1468–1483 (2018).
49. N. S. Sugiyama, S. Saito, F. Beutler, H. J. Seo, A complete FFT-based decomposition formalism for the redshift-space bispectrum. *Mon. Not. R. Astron. Soc.* **484**, 364–384 (2019).
50. O. H. E. Philcox, NPCFs.jl. GitHub. <https://github.com/oliverphilcox/NPCFs.jl>. Deposited 23 July 2022.