
Permutation-Free High-Order Interaction Tests

Zhaolu Liu¹ Robert L. Peach^{2,3} Mauricio Barahona¹

Abstract

Kernel-based hypothesis tests offer a flexible, non-parametric tool to detect high-order interactions in multivariate data, beyond pairwise relationships. Yet the scalability of such tests is limited by the computationally demanding permutation schemes used to generate null approximations. Here we introduce a family of permutation-free high-order tests for joint independence and partial factorisations of d variables. Our tests eliminate the need for permutation-based approximations by leveraging V-statistics and a novel cross-centring technique to yield test statistics with a standard normal limiting distribution under the null. We present implementations of the tests and showcase their efficacy and scalability through synthetic datasets. We also show applications inspired by causal discovery and feature selection, which highlight both the importance of high-order interactions in data and the need for efficient computational methods.

1. Introduction

Many complex real-world systems involve interactions beyond pairwise relationships, necessitating an explicit understanding of interactions among three or more entities (Battiston et al., 2020; 2021; Rosas et al., 2022; Thibeault et al., 2024). Such high-order interactions amongst $d > 2$ variables have been shown to affect processes in social (Benson et al., 2016), ecological (Grilli et al., 2017) and biological systems (Luff et al., 2024; Arnaudon et al., 2022), often in ways that cannot be reduced to combinations of pairwise effects (Schaub et al., 2020; Bick et al., 2023). Given the relative scarcity of relational datasets that record high-order phenomena, it is important to develop unsupervised, mathematically grounded methods for their detection in ob-

servational data. A growing body of work has approached this challenge using tools from information theory (Rosas et al., 2020; 2024), persistent homology (Santoro et al., 2023; 2024), Bayesian inference (Young et al., 2021) and Hamiltonian-based functions (Schneidman et al., 2003).

Within the statistical and machine learning literature, kernel-based tests provide a flexible, non-parametric framework for identifying relationships in data. The detection of pairwise independence between two variables using HSIC (Gretton et al., 2007) has found broad applicability in tasks such as generative adversarial networks (Bińkowski et al., 2018), feature selection (Song et al., 2012), and transfer learning (Long et al., 2015). Beyond pairwise independence, kernel-based tests have been extended to assess joint independence among d variables via dHSIC (Pfister et al., 2018), as well as Lancaster interactions among three variables (Sejdinovic et al., 2013a), which were later generalised to d -order Streitberg interactions (Liu et al., 2023b). However, these tests typically rely on computationally expensive permutation or bootstrap schemes (involving $100 \leq p \leq 1000$ permutations), reducing their applicability to datasets with larger number of variables, where the subsets of variables to be tested for interactions explode combinatorially.

Here, we tackle this limitation on scalability by introducing a family of permutation-free high-order tests that retain the flexibility of kernel-based methods while eliminating the computational burden of null-distribution approximation. To note, our approach applies to the general settings where one seeks to determine factorisation properties among subsets of variables (e.g., Lancaster and Streitberg interactions), rather than merely rejecting or accepting joint independence.

Contributions Our main contributions can be summarised as follows: (1) We define the permutation-free joint independence test, $\bar{x}d$ HSIC for any d . (2) To address the combinatorial challenges of testing for high-order factorisation, we introduce the cross-centring technique to develop the permutation-free d -order Lancaster factorisation test, $\bar{x}LI$. (3) Building on $\bar{x}LI$, we derive a permutation-free Streitberg interaction test, $\bar{x}SI$. (4) We show that, for the special case $d = 2$, our high-order tests have a more concise mathematical formulation than the existing $\bar{x}HSIC$ (Shekhar et al., 2023). (5) Through empirical evaluations in machine learning applications, we demonstrate that the permutation-free

¹Department of Mathematics, Imperial College London, United Kingdom ²Department of Neurology, University Hospital Würzburg, Germany ³Department of Brain Sciences, Imperial College London, United Kingdom. Correspondence to: Mauricio Barahona <m.barahona@imperial.ac.uk>.

tests achieve substantial computational speedup, exceeding 100-fold, over their permutation-based counterparts.

Paper Structure Section 2 introduces joint independence and partial factorisation as two different generalisations of pairwise relationships, and briefly reviews kernel-based tests. In Section 3, we develop our permutation-free joint independence test for any d , \bar{x} dHSIC, and our permutation-free factorisation tests for any d , \bar{x} LI and \bar{x} SI. We relate our proposed methods to existing work in Section 4, and provide numerical validations in Section 5. We conclude by discussing directions for future research in Section 6.

2. Preliminaries

2.1. Statistical Interactions

Throughout this paper, we consider multivariate probability distributions of a set of d variables $D = \{X^1, \dots, X^d\}$ denoted $\mathbb{P}_{X^1, \dots, X^d} =: \mathbb{P}_{1 \dots d}$.

Pairwise Interaction The simplest statistical interaction emerges for $d = 2$. In this case, a lack of statistical interaction is equivalent to pairwise independence, which is determined by the difference between the joint distribution \mathbb{P}_{12} and the product of the marginals $\mathbb{P}_1 \mathbb{P}_2$:

$$\Delta^2 \mathbb{P} = \mathbb{P}_{12} - \mathbb{P}_1 \mathbb{P}_2. \quad (1)$$

Hence $\Delta^2 \mathbb{P} = 0 \iff X^1 \perp\!\!\!\perp X^2$ (pairwise independence).

Joint Independence One way to generalise the above notion to $d > 2$ variables is to focus on joint independence:

$$\Delta_J^d \mathbb{P} = \mathbb{P}_{1 \dots d} - \prod_{i=1}^d \mathbb{P}_i. \quad (2)$$

Clearly, $\Delta_J^d \mathbb{P}$ vanishes if and only if the d variables are jointly independent, a property that has been employed in Independent Component Analysis (Hyvärinen & Oja, 2000), causal discovery (Pfister et al., 2018; Laumann et al., 2023), and variational autoencoders (Lopez et al., 2018). However, for $d \geq 3$, joint independence is uninformative as a criterion for the presence of high-order interactions. For example, three of the four factorisations of \mathbb{P}_{123} ($\mathbb{P}_1 \mathbb{P}_{23}$, $\mathbb{P}_2 \mathbb{P}_{13}$ and $\mathbb{P}_3 \mathbb{P}_{12}$) are neither full 3-way interactions nor jointly independent like $\mathbb{P}_1 \mathbb{P}_2 \mathbb{P}_3$.

Lancaster Interaction An alternative generalisation interprets the vanishing condition of $\Delta^2 \mathbb{P}$ as a factorisation of the joint distribution into disjoint components. To capture this perspective, the Lancaster interaction for $d = 3$ variables is defined as:

$$\Delta_L^3 \mathbb{P} = \mathbb{P}_{123} - \mathbb{P}_1 \mathbb{P}_{23} - \mathbb{P}_2 \mathbb{P}_{13} - \mathbb{P}_3 \mathbb{P}_{12} + 2\mathbb{P}_1 \mathbb{P}_2 \mathbb{P}_3. \quad (3)$$

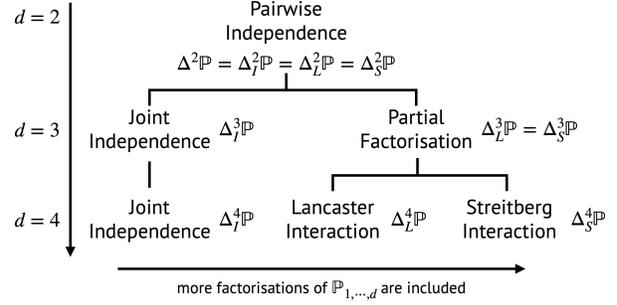


Figure 1. Generalisations of pairwise dependence to high-order interactions. All three measures of statistical interaction (joint independence, Lancaster, Streitberg) are the same for $d = 2$. Joint independence captures less information for $d \geq 3$. Only Streitberg interaction considers all factorisations for $d \geq 4$.

If \mathbb{P}_{123} can be factorised into *any* of $\mathbb{P}_1 \mathbb{P}_{23}$, $\mathbb{P}_2 \mathbb{P}_{13}$, $\mathbb{P}_3 \mathbb{P}_{12}$ or $\mathbb{P}_1 \mathbb{P}_2 \mathbb{P}_3$, then $\Delta_L^3 \mathbb{P} = 0$ (see Appendix B). Lancaster (1969) rewrote and generalised this measure for d -order interactions as

$$\Delta_L^d \mathbb{P} = \prod_{i=1}^d (\mathbb{P}_i^* - \mathbb{P}_i), \quad (4)$$

where the \mathbb{P}_i^* are defined implicitly by $\mathbb{P}_i^* \mathbb{P}_j^* \dots \mathbb{P}_k^* = \mathbb{P}_{ij \dots k}$ and $\mathbb{P}_i^* \mathbb{P}_j \dots \mathbb{P}_k = \mathbb{P}_i \mathbb{P}_j \dots \mathbb{P}_k$.

However, for $d \geq 4$, $\Delta_L^d \mathbb{P}$ fails to capture certain factorisations, e.g., if $\mathbb{P}_{1234} = \mathbb{P}_{12} \mathbb{P}_{34}$ then $\Delta_L^4 \mathbb{P} = (\mathbb{P}_{12} - \mathbb{P}_1 \mathbb{P}_2)(\mathbb{P}_{34} - \mathbb{P}_3 \mathbb{P}_4)$ which is not zero in general (Streitberg, 1990). Despite this limitation, Liu et al. (2023b) proved that for $d \geq 4$, $\Delta_L^d \mathbb{P} = 0$ only if the factorisation of $\mathbb{P}_{1 \dots d}$ includes singleton variables, highlighting its partial informativeness in capturing structural dependencies.

Streitberg Interaction To account for *all* factorisations of $\mathbb{P}_{1 \dots d}$, the Streitberg interaction was introduced (Streitberg, 1990). Let $\pi = b_1 | b_2 | \dots | b_{|\pi|}$ denote a partition of the set D of d variables into $|\pi|$ blocks b_j , and let $\mathbb{P}_\pi := \prod_{j=1}^{|\pi|} \mathbb{P}_{b_j}$. The Streitberg interaction for d variables is:

$$\Delta_S^d \mathbb{P} = \sum_{\pi \in \Pi(D)} (|\pi| - 1)! (-1)^{|\pi|-1} \mathbb{P}_\pi, \quad (5)$$

where the sum extends over $\Pi(D)$, the set of all partitions of the set D . Crucially, $\Delta_S^d \mathbb{P} = 0$ only if $\mathbb{P}_{1 \dots d}$ can be factorised in any way (Streitberg, 1990).

Figure 1 summarises these three perspectives of high-order interactions, showing how they all coincide for $d = 2$; begin to diverge for $d = 3$; and Streitberg interactions are unique in capturing all possible factorisations for $d \geq 4$. For details on derivations via lattice theory, see Liu et al. (2025).

2.2. Kernel-Based Tests

To transform the interaction measures from Section 2.1 into non-parametric test statistics, we embed the underlying distributions into reproducing kernel Hilbert spaces (RKHS). Recall that a symmetric, positive-definite kernel $k^i : \mathcal{X}^i \times \mathcal{X}^i \rightarrow \mathbb{R}$ induces an associated RKHS \mathcal{H}^i with the reproducing property. For any $X^i \in \mathcal{X}^i$, let $\phi^i(X^i) = k^i(X^i, \cdot)$ denote the canonical feature map. The kernel mean embedding of a distribution \mathbb{P}_i is then given by $\mu_{\mathbb{P}_i} = \mathbb{E}_{X^i}[\phi^i(X^i)]$. This embedding satisfies $\mathbb{E}_{X^i} f(X^i) = \langle f, \mu_{\mathbb{P}_i} \rangle_{\text{HS}}$ for all f in \mathcal{H}^i , where $\langle \cdot, \cdot \rangle_{\text{HS}}$ is the Hilbert-Schmidt inner product. If k^i is *characteristic*, the map $\mu_{\mathbb{P}_i}$ is injective, ensuring that the norm of the signed measure is zero if and only if the measure itself is zero (Gretton et al., 2007; Muandet et al., 2017).

Using such embeddings, all high-order interaction measures in Section 2.1 can be naturally converted into kernel-based test statistics by constructing kernel mean embeddings of the relevant distributions, and computing their Hilbert-Schmidt norms. Adopting this methodology, the measures in Eq. (2), Eq. (4) and Eq. (5) have been developed into, respectively, the dHSIC (Pfister et al., 2018), the Lancaster interaction test (LI) (Sejdicovic et al., 2013a; Liu et al., 2023b), and the Streitberg interaction test (SI) (Liu et al., 2023b).

A key practical issue is the fact that all of these test statistics are degenerate under their respective null hypotheses. To obtain valid rejections, one typically employs a permutation procedure to approximate the null distribution, or uses heuristic strategies such as a Gamma approximation (as in HSIC (Gretton et al., 2007) and dHSIC (Pfister et al., 2018)). Permutation-based approaches involve p runs (typically $p > 100$) to achieve an accurate approximation of the null distribution, and are thus computationally expensive for large sample sizes or multiple variables. This computational burden underscores the need for alternative methods that are both theoretically principled and computationally efficient.

Recently, Shekhar et al. (2023) introduced $\bar{\text{x}}\text{dHSIC}$, a variant of HSIC for $d = 2$ that avoids the need for permutations by establishing that its null distribution converges to a standard normal. Inspired by this approach, we develop a family of permutation-free high-order tests in Section 3.

3. Permutation-Free High-Order Tests

3.1. Joint Independence

To extend the permutation-free pairwise independence measure of Shekhar et al. (2023) to the multivariate setting, we first define the unnormalised permutation-free dHSIC using the data-splitting technique:

$$\text{xdHSIC} = \left\langle \hat{\mu}_{\mathbb{P}_{1\dots d}}^1 - \hat{\mu}_{\mathbb{P}_1 \dots \mathbb{P}_d}^1, \hat{\mu}_{\mathbb{P}_{1\dots d}}^2 - \hat{\mu}_{\mathbb{P}_1 \dots \mathbb{P}_d}^2 \right\rangle_{\text{HS}},$$

where $\hat{\mu}_{(\cdot)}^1$ and $\hat{\mu}_{(\cdot)}^2$ are the empirical embeddings in dHSIC, estimated from two disjoint sample sets: the first half $\mathcal{S}_1^d = \{(x_i^1, \dots, x_i^d) : 1 \leq i \leq n\}$ and the second half $\mathcal{S}_2^d = \{(x_i^1, \dots, x_i^d) : n+1 \leq i \leq 2n\}$ of the *i.i.d.* samples.

Unlike $\bar{\text{x}}\text{dHSIC}$ in (Shekhar et al., 2023), which was based on U-statistics, here we adopt V-statistics to reduce computational overhead, while retaining robust estimation accuracy (Pfister et al., 2018). Specifically, we form the $2n \times 2n$ kernel matrices K^i with entries $K_{ab}^i = k^i(x_a^i, x_b^i)$ for $1 \leq a, b \leq 2n$ where $\{x_a^i\}_{a=1}^{2n}$ are *i.i.d.* samples for the i -th variable. In our construction, the $n \times n$ submatrices on the diagonal of K^i (with entries K_{ab}^i where both samples a, b come from either \mathcal{S}_1^d or from \mathcal{S}_2^d) do not contribute to the estimate. Instead, we only use the $n \times n$ upper-right off-diagonal block (lower-left block is redundant due to symmetry), which measures distances between samples from the two disjoint subsets. We then define $\bar{\text{x}}\text{dHSIC}$, the normalised permutation-free dHSIC.

Definition 3.1. Let $\mathcal{K}^j := K_{[1:n, n+1:2n]}^j$ be the off-diagonal $n \times n$ submatrix of K^j . Then $\bar{\text{x}}\text{dHSIC}$ is defined as:

$$\begin{aligned} \bar{\text{x}}\text{dHSIC} &= \frac{\sqrt{n} \cdot \text{xdHSIC}}{s_I^2}, \quad \text{with} \\ \text{xdHSIC} &= \frac{1}{n^2} \mathbf{1}^T \left(\bigcirc_{j=1}^d \mathcal{K}^j \right) \mathbf{1} + \frac{1}{n^{2d}} \left(\bigcirc_{j=1}^d \mathbf{1}^T \mathcal{K}^j \mathbf{1} \right) \\ &\quad - \frac{1}{n^{d+1}} \mathbf{1}^T \left(\bigcirc_{j=1}^d \mathcal{K}^j \mathbf{1} \right) - \frac{1}{n^{d+1}} \left(\bigcirc_{j=1}^d \mathbf{1}^T \mathcal{K}^j \right) \mathbf{1}, \\ s_I^2 &= \frac{1}{n} \left\| \frac{1}{n} \left(\bigcirc_{j=1}^d \mathcal{K}^j \right) \mathbf{1} + \frac{1}{n^{2d-1}} \mathcal{K}^1 \mathbf{1} \left(\bigcirc_{j=2}^d \mathbf{1}^T \mathcal{K}^j \mathbf{1} \right) \right. \\ &\quad \left. - \frac{1}{n^d} \left(\bigcirc_{j=1}^d \mathcal{K}^j \mathbf{1} \right) - \frac{1}{n^d} \left(\bigcirc_{j=1}^d \mathcal{K}^j \mathbf{1}^T \right) - \text{xdHSIC} \mathbf{1} \right\|^2, \end{aligned}$$

where the operator \bigcirc represents the Hadamard product and $\mathbf{1}$ is the $n \times 1$ vector of ones.

$\bar{\text{x}}\text{dHSIC}$ can be used to test d -order joint independence:

Hypothesis 3.2. (Joint Independence)

$$\begin{aligned} H_0 &: \mathbb{P}_{1, \dots, d} = \prod_{i=1}^d \mathbb{P}_i \\ H_1 &: \mathbb{P}_{1, \dots, d} \neq \prod_{i=1}^d \mathbb{P}_i. \end{aligned}$$

Then we have: $H_0 \iff \Delta_I^d P = 0 \iff \bar{\text{x}}\text{dHSIC} = 0$.

The construction of $\bar{\text{x}}\text{dHSIC}$ ensures it follows $\mathcal{N}(0, 1)$ under the null (Shekhar et al., 2023), thus enabling the test with a single computation in contrast with p permutations for standard dHSIC. Importantly, its time complexity matches that of a single dHSIC computation.

Proposition 3.3. $\bar{\text{x}}\text{dHSIC}$ can be computed in $\mathcal{O}(dn^2)$.

To prove this, it is sufficient to show that both the numerator, xdHSIC, and the denominator, s_I^2 , can be computed

in $\mathcal{O}(dn^2)$, which follows from an argument in Pfister et al. (2018) (see Appendix A).

3.2. Partial Factorisation

A limitation of the joint independence test is that, if the null hypothesis is rejected, it does not distinguish between *partial factorisation*, where disjoint subsets of variables indicate low-order interactions, and *true non-factorisation*, where the joint distribution cannot be factorised in any way, implying that all variables interact. To address this limitation, more sophisticated criteria have been developed (Streitberg, 1990), and the corresponding kernel-based tests have been introduced (Sejdinovic et al., 2013a; Liu et al., 2023b). However, these methods encounter degenerate cases, which necessitate permutation-based resampling—a step that is eliminated in the permutation-free Lancaster and Streitberg interaction tests introduced below.

Centring: Our approach relies on *matrix centring*, which is widely used in both kernel-based (Gretton et al., 2007; Sejdinovic et al., 2013a) and distance-based (Chakraborty & Zhang, 2019; Böttcher, 2020) high-order interaction tests to improve computational efficiency and simplify notation. For a $n \times n$ kernel matrix K , its centred version is $\tilde{K} = CKC$, where $C = I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T$, where I_n is the $n \times n$ identity matrix.

Equivalence: Two kernels are said to be equivalent if they induce the same semimetric on the domain (Sejdinovic et al., 2013b). In particular, the centred kernel is translation-invariant and thus is equivalent to the original kernel.

Cross-centring: To extend the permutation-free scheme to a broader class of measures, we introduce an analogous *cross-centring*, which arises naturally under data splitting.

Definition 3.4. (Cross-centring) Let $\mathbf{1}_u = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \mathbf{1}$ and $\mathbf{1}_\ell = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \otimes \mathbf{1}$, where \otimes denotes the Kronecker product, and define the matrices $C_u = I_{2n} - \frac{1}{n}\mathbf{1}_u\mathbf{1}_u^T$, and $C_\ell = I_{2n} - \frac{1}{n}\mathbf{1}_\ell\mathbf{1}_\ell^T$, where I_{2n} is the $2n \times 2n$ identity matrix. For a $2n \times 2n$ kernel matrix K with indices $1 : n$ associated with \mathcal{S}_1^d and indices $n + 1 : 2n$ associated with \mathcal{S}_2^d , its cross-centred version is given by $\overline{K} = C_u K C_\ell$.

Lemma 3.5. *The cross-centred kernel*

$$\overline{k}(x, x') = \left\langle \phi(x) - \frac{1}{n} \sum_{i=1}^n \phi(x_i), \phi(x') - \frac{1}{n} \sum_{j=n+1}^{2n} \phi(x'_j) \right\rangle,$$

where $\phi(\cdot)$ is the canonical feature map, is equivalent to $k(x, x')$ as $n \rightarrow \infty$. See Appendix A.

Having established this asymptotic equivalence, we express our test statistics in terms of cross-centred kernel matrices.

3.2.1. LANCASTER INTERACTION TEST

We now derive the d -order permutation-free Lancaster interaction test. The Lancaster criterion can be used to detect

factorisations involving at least one singleton variable (Liu et al., 2023b). To formalise the test, we introduce the following null hypothesis:

Hypothesis 3.6. (Lancaster Factorisation)

$$\begin{aligned} H_0 &: \bigvee_{\pi_L} H_{\pi_L} & \forall |\pi_L| = 2 \\ H_1 &: \bigwedge_{\pi_L} \neg H_{\pi_L} & \forall |\pi_L| = 2 \end{aligned}$$

where H_{π_L} is the event such that the joint distribution factorises according to a partition π_L containing exactly one singleton, whereas $\neg H_{\pi_L}$ indicates that such a factorisation does not hold. If any of the d subhypotheses in H_0 is true, then $\Delta_L^d \mathbb{P} = 0$. Hence if all d subhypotheses are rejected, a d -order Lancaster interaction is detected.

Note that even though H_0 contains only d subhypotheses, which correspond to the factorisations with one singleton, all other factorisations that satisfy the Lancaster vanishing condition are actually subsumed by them (Liu et al., 2023b).

We can now define the normalised permutation-free statistic associated with a subhypothesis H_{π_L} of H_0 .

Definition 3.7. Let $H_{\pi_L} : \mathbb{P}_{1\dots d} = \mathbb{P}_m \mathbb{P}_{1,\dots,m-1,m+1,\dots,d} = \mathbb{P}_{\pi_L}$ for some $1 \leq m \leq d$. Under the null of H_{π_L} , the normalised permutation-free Lancaster interaction, $\overline{\text{xLI}}$, is:

$$\begin{aligned} \overline{\text{xLI}} &= \frac{\sqrt{n} \cdot \text{xLI}}{s_L}, \quad \text{with} \\ \text{xLI} &= \frac{1}{n^2} \mathbf{1}_u^T \left(\overline{K}^m \circ \bigcirc_{i \in D/m} \overline{K}^i \right) \mathbf{1}_\ell, \\ s_L^2 &= \frac{1}{n^3} \left\| \text{diag}(\mathbf{1}_u) \left(\overline{K}^m \circ \bigcirc_{i \in D/m} \overline{K}^i \right) \mathbf{1}_\ell \right\|^2 - \text{xLI}^2, \end{aligned}$$

where D/m denotes all d variables except X^m , the operators \circ, \bigcirc represent Hadamard products, and the overline $\overline{(\cdot)}$ indicates cross-centring.

Remark: In our permutation-free procedure we have a different test statistic for each subhypothesis, all with a *fixed* null distribution (standardised to $\mathcal{N}(0, 1)$).

To reject the null H_0 for Lancaster factorisation, we carry out d subtests (one for each of the singleton factorisations), each with complexity $\mathcal{O}(dn^2)$.

Proposition 3.8. *The time complexity of $\overline{\text{xLI}}$ for all H_{π_L} across any d is $\mathcal{O}(dn^2)$.*

The composite hypothesis H_0 is rejected only if all subhypotheses are individually rejected, and, as this has been proven to be less conservative than the Bonferroni correction (Rubenstein et al., 2016), we use this correction for our tests. When the subtests are executed sequentially rather

than in parallel, the procedure can terminate early if any subtest fails to reject its corresponding null, thereby saving computational time. We provide a detailed description of the test procedure of $\bar{x}\text{LI}$ for $d = 3$ in Appendix B.

3.2.2. STREITBERG INTERACTION TEST

As discussed, the Lancaster interaction is unable to detect factorisations that do not involve singletons. The measure that does vanish for all factorisations of the joint distribution is the Streitberg interaction in Eq. (5). To test if $\mathbb{P}_{1\dots d}$ can be factorised in any way, we have the following hypothesis:

Hypothesis 3.9. (Complete Factorisation)

$$\begin{aligned} H_0 &: \bigvee_{\pi} H_{\pi} & \forall |\pi| = 2 \\ H_1 &: \bigwedge_{\pi} \neg H_{\pi} & \forall |\pi| = 2 \end{aligned}$$

where H_{π} denotes the event such that the joint distribution $\mathbb{P}_{1\dots d}$ is factorised as $\mathbb{P}_{\pi} = \mathbb{P}_{b_1}\mathbb{P}_{b_2}$.

Note that only partitions with exactly two blocks need to be considered in the composite tests, as all further factorisations $\mathbb{P}_{1\dots d} = \mathbb{P}_{b_1}\mathbb{P}_{b_2}\dots\mathbb{P}_{b_{|\pi|}}$ are subsumed in bipartition subtests. The partitions with two blocks correspond to the second level of the d -order partition lattice (Liu et al., 2023b).

The Streitberg (complete) factorisation hypothesis includes the Lancaster factorisation hypothesis plus, additionally, the H_{π_S} where π_S are partitions into two blocks with no singletons. For example, for $d = 6$, the Streitberg subtests must check specifically for $\mathbb{P}_{123}\mathbb{P}_{456}$, whereas, in contrast, $\mathbb{P}_1\mathbb{P}_{23456}$ is already included in the Lancaster factorisation tests (due to the singleton), and $\mathbb{P}_{12}\mathbb{P}_{34}\mathbb{P}_{56}$ does not need to be checked, since it has cardinality 3 and is subsumed in 2-block factorisations $\mathbb{P}_{1234}\mathbb{P}_{56}$, $\mathbb{P}_{1256}\mathbb{P}_{34}$ and $\mathbb{P}_{12}\mathbb{P}_{3456}$.

Definition 3.10. Under the null subhypothesis of H_{π_S} , where $\pi_S = b_1|b_2$, the permutation-free Streitberg interaction is defined as

$$\begin{aligned} \bar{x}\text{SI} &= \frac{\sqrt{n} \cdot \text{xSI}}{s_S}, \quad \text{with} \\ \text{xSI} &= \frac{1}{n^2} \mathbf{1}_u^T \left(\bigcirc_{p \in b_1} \overline{K^p} \circ \bigcirc_{q \in b_2} \overline{K^q} \right) \mathbf{1}_{\ell}, \\ s_S^2 &= \frac{1}{n^3} \left\| \text{diag}(\mathbf{1}_u) \left(\bigcirc_{p \in b_1} \overline{K^p} \circ \bigcirc_{q \in b_2} \overline{K^q} \right) \mathbf{1}_{\ell} \right\|^2 - \text{xSI}^2. \end{aligned}$$

Unlike the permutation-based Streitberg statistic, whose time complexity grows combinatorially with d (Liu et al., 2023b), $\bar{x}\text{SI}$ grows only quadratically with n .

Proposition 3.11. *The time complexity of $\bar{x}\text{SI}$ for all H_{π_S} across any d is $\mathcal{O}(dn^2)$.*

In our numerics, we test the complete factorisation hypothesis (3.9) via a two-step process we denote $\bar{x}\text{LI} + \bar{x}\text{SI}$: we first use $\bar{x}\text{LI}$ to test the d subhypotheses involving π_L (bipartitions with a singleton), followed by applying $\bar{x}\text{SI}$ for the subhypotheses involving the non-singleton bipartitions, π_S . Together, these two sets cover all the $(2^{d-1} - 1)$ bipartitions needed to test for the complete factorisation, i.e., $\pi_{|\pi|=2} = \pi_L \cup \pi_S$. Although $\bar{x}\text{SI}$ could, in principle, be used to perform all subtests in Hypothesis (3.9) directly, we found this approach to be computationally less advantageous. Indeed, given that Liu et al. (2023b) showed that the permutation-based SI test involves only non-singleton partitions, if we were to use $\bar{x}\text{SI}$, the terms in the permutation-free $\bar{x}\text{SI}$ exceed this scope, further contributing to its higher computational cost if it were applied for all subtests. As a result, using $\bar{x}\text{SI}$ alone does not always guarantee improved computational efficiency. In contrast, $\bar{x}\text{LI} + \bar{x}\text{SI}$ ensures efficiency, by maintaining quadratic complexity with the number of samples n regardless of order d . This advantage becomes increasingly significant as d grows, making our method more suitable for practical applications involving high-order data.

3.3. Computational Complexity

Table 1 provides a summary of the time complexity of the permutation-based and permutation-free tests, highlighting the computational advantage of the tests proposed in this paper. For the joint independence (dHSIC) and Lancaster interaction (LI) tests, the permutation-free tests offer a p -fold reduction in complexity by eliminating the need for resampling. For the complete factorisation test (SI), our strategy of combining $\bar{x}\text{SI}$ and $\bar{x}\text{LI}$ is far more efficient than the permutation-based test in two ways: (i) p -fold reduction for each each subtest; (ii) n^2 scaling decoupled from d . Empirical evaluation of these scalings is shown in Section 5 using synthetic data with ground truths.

Table 1. Time complexity of high-order tests. Here p is the number of permutations, n is the number of samples, d is the number of variables and F_d is the number of partitions without singletons for d variables, a combinatorial number. The permutation-free tests developed in this paper are highlighted in blue.

Permutation-based	Permutation-free
dHSIC: $\mathcal{O}(pdn^2)$	$\bar{x}\text{dHSIC}$: $\mathcal{O}(dn^2)$
LI: $\mathcal{O}(pd^2n^2)$	$\bar{x}\text{LI}$: $\mathcal{O}(d^2n^2)$
SI: $\mathcal{O}(pF_d^2 2^d n^{d/2})$	$\bar{x}\text{SI} + \bar{x}\text{LI}$: $\mathcal{O}(2^d dn^2)$

4. Related Work

Our work is inspired by Shekhar et al. (2023) who introduced $\bar{x}\text{HSIC}$ to make HSIC (Gretton et al., 2007) permutation-free by approximating HSIC using the data-

splitting technique and U-statistic (Kim & Ramdas, 2024; Shekhar et al., 2022), and normalising by the empirical standard deviation, thereby achieving a standard normal null distribution. However, to the best of our knowledge, these methods do not extend readily to more than two variables. In contrast, our kernel-based formulation leverages V-statistics and cross-centring to define permutation-free tests for Lancaster and Streitberg interactions for $d > 2$ variables.

Remark: For $d = 2$, our formulation recovers pairwise independence (Fig. 1). Yet by using cross-centring and V-statistics we obtain a permutation-free statistic $\bar{x}\text{HSIC}_V := \bar{x}\text{SI}_{d=2} = \bar{x}\text{LI}_{d=2}$, which is simpler, more compact, and with higher power than the original $\bar{x}\text{HSIC}$ in Shekhar et al. (2023). See derivation and numerics in Appendix C.

5. Experiments

We evaluate our proposed permutation-free tests using a range of experiments. For all cases, under the null hypothesis of each test, we confirm standard normality and controlled type-I errors of the corresponding $\bar{x}\text{dHSIC}$, $\bar{x}\text{LI}$ & $\bar{x}\text{SI}$ statistics (see Appendix D, Figs. 10–11 for examples). Below we focus on the statistical power and computational efficiency of these methods compared to their permutation-based counterparts. Unless noted otherwise, the significance level is set to $\alpha = 0.05$, and we use Gaussian kernels with bandwidth given by the median heuristic. Additional details of each experiment are available in Appendix E.

5.1. Validation on Synthetic Examples

Joint Independence Test We first demonstrate the $\bar{x}\text{dHSIC}$ statistic for permutation-free joint independence testing. Figure 2a-c shows the application to a $d = 4$ multivariate Gaussian (MVG) distribution (covariance matrix in the inset of (a), with $\beta = 0.5$). The power of the $\bar{x}\text{dHSIC}$ test has similar sensitivity compared to the permutation-based dHSIC test with $p = 100$ permutations, but $\bar{x}\text{dHSIC}$ reaches the maximum power roughly 100 times faster (Fig. 2b-c), as expected from the p -fold reduction in complexity. The same reduction in computation time is obtained when applied to d -variable MVGs of varying d (Fig. 2d).

Factorisation Tests Next, we demonstrate that our permutation-free high-order tests successfully preserve the vanishing conditions for different factorisation tests. In Figure 3a, we consider data from a MVG with $d = 5$ variables and covariance matrix such that $\mathbb{P}_1\mathbb{P}_{2345}$ (see inset). As expected, $\bar{x}\text{dHSIC}$ fails to detect this partial (singleton) factorisation, which is correctly detected by both $\bar{x}\text{LI}$ and $\bar{x}\text{SI}+\bar{x}\text{LI}$. Figure 3b shows that, when analyse data from a MVG that is factorisable as $\mathbb{P}_{12}\mathbb{P}_{345}$ (no singletons), only $\bar{x}\text{SI}+\bar{x}\text{LI}$ succeeds in detecting this interaction.

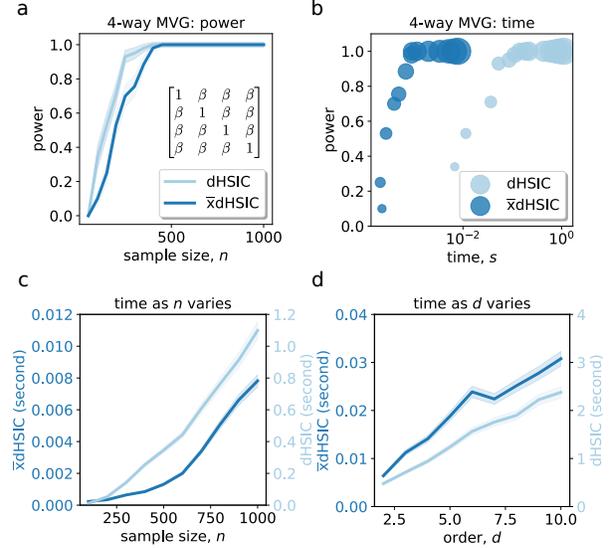


Figure 2. Joint independence test. (a)-(c) Sampling from a $d = 4$ variable MVG (covariance matrix, inset of (a)), we compare dHSIC (with $p = 100$ permutations) and $\bar{x}\text{dHSIC}$: (a) statistical power as a function of the number of samples, n ; (b) trade-off between power and computational time (circle diameter proportional to sample size n); (c) CPU time as n increases. (d) CPU time for MVGs of the same form as in (a) but with increasing number of variables, d (with $n = 500$ samples). Note that the right scale of the CPU time in (c) and (d) is 100 times larger, thus reflecting the p -fold computational reduction over dHSIC.

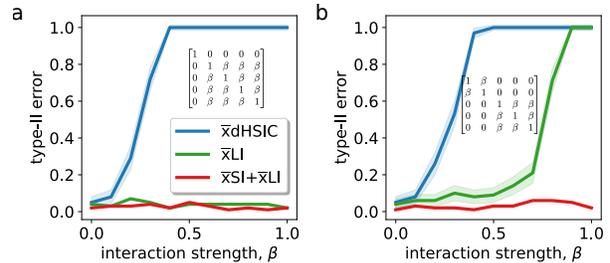


Figure 3. Partial factorisations of MVG with $d = 5$ variables. (a) $\bar{x}\text{dHSIC}$ fails to detect the partial factorisation with a singleton, $\mathbb{P}_1\mathbb{P}_{2345}$, regardless of interaction strength. (b) Only $\bar{x}\text{SI}+\bar{x}\text{LI}$ identifies the factorisation with no singletons, $\mathbb{P}_{12}\mathbb{P}_{345}$. MVG covariance matrices are insets in each case.

We then simulate data from a XOR gate with $d = 5$ variables, in which the interaction is exclusively of order 5 without 2-, 3-, or 4-order interactions. Fig. 4a-c shows that the power of the permutation-free $\bar{x}\text{SI}+\bar{x}\text{LI}$ is superior to the permutation-based SI with a substantial reduction in computation time. For example, for $n = 500$ samples, $\bar{x}\text{SI}+\bar{x}\text{LI}$ can be computed within 0.5 seconds whereas SI requires roughly 6 minutes. In Figure 4d, we consider a series of XOR gates with increasing number of variables, d . Whereas $\bar{x}\text{SI}+\bar{x}\text{LI}$ can be computed in less than 1 second for $d = 10$,

SI becomes already computational infeasible for $d = 6$, due to the rapid combinatorial increase in the number of non-singleton partitions (F_d) (Sloane et al., 2018), which makes the computation quickly intractable as d increases (Table 1).

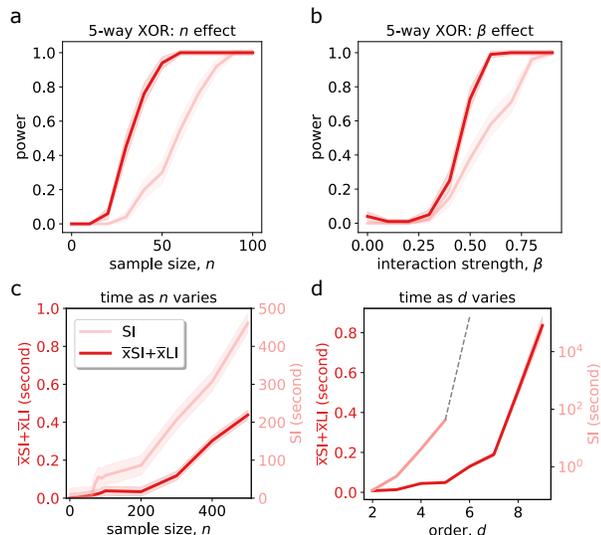


Figure 4. 5-way XOR factorisation. Power of $\bar{x}SI+\bar{x}LI$ and SI tests for an XOR gate with $d = 5$ variables as a function of: (a) sample size, n ; (b) interaction strength, β . Computation time of both tests for: (c) increasing sample size, n (for $d = 5$); (d) increasing number of variables, d (for $n = 100$). The SI test is computationally infeasible above $d = 5$ (dashed grey line). Note the different left/right scales for CPU times in (c) and (d), highlighting the large reduction in computational cost achieved by $\bar{x}SI+\bar{x}LI$.

5.2. Applications to Causal Discovery

V-Structure Detection A V-structure (or *collider*) is a fundamental component in constraint-based causal discovery, where two variables X and Y both causally influence a third variable Z . If $X \perp\!\!\!\perp Y$, rejecting the Lancaster factorisation hypothesis for $d = 3$ is equivalent to rejecting the conditional independence $X \perp\!\!\!\perp Y|Z$, thereby indicating the presence of a V-structure (Sejdinovic et al., 2013a).

Here, we use two datasets (A and B) from Sejdinovic et al. (2013a), in which $X \perp\!\!\!\perp Y$, and we increase the noise dimension to make the causal relationship more difficult to detect. Figures 5a-b shows the comparison of our permutation-free $\bar{x}LI$ to the permutation-based LI and kernel-based conditional independence KCI (Zhang et al., 2011) tests. Each experiment uses $n = 500$ samples and the time complexity is $\mathcal{O}(n^2)$, $\mathcal{O}(pn^2)$ and $\mathcal{O}(pn^3)$, respectively. KCI rapidly fails as the noise dimension grows, whereas Lancaster-based approaches remain more robust. Although $\bar{x}LI$ is marginally less powerful than LI, it runs approximately 100 times faster (0.082s vs. 7.6s for $p = 100$ permutations), providing a trade-off between statistical power and computational efficiency,

making $\bar{x}LI$ an attractive choice for detecting V-structures when conventional conditional independence tests become unreliable or too slow.

Furthermore, in Appendix B we thoroughly evaluate the sensitivity of the permutation-free method by applying it to V-structures encoded by three nonlinear forms (sinc, log, polynomial) using three different kernel functions (RQ, Laplace, RBF). We show that regardless of the non-linear form or kernel, our proposed method achieves similar accuracy to the permutation-based counterpart (Fig. 8).

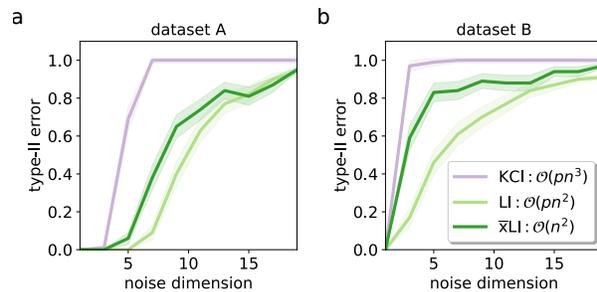


Figure 5. V-structure detection. Type-II errors of $\bar{x}LI$, LI and KCI for detecting the conditional dependence as the noise dimensions vary in dataset (a) A and (b) B from Sejdinovic et al. (2013a).

DAG Causal Structural Learning An alternative route for causal learning is provided by score-based methods (Peters et al., 2014; Laumann et al., 2023), whereby each variable is modelled as a function of its direct causes (i.e., parent nodes in a directed acyclic graph (DAG)) plus an additive noise term. If the noise terms for all regressions are jointly independent, then the proposed DAG cannot be rejected.

Following closely Simulation 4 from Pfister et al. (2018), we simulate n samples from one DAG with $d = 4$ variables (see Appendix D). We then compare the accuracy of our permutation-free $\bar{x}dHSIC$ against the permutation-based $dHSIC$ (using $p = 200$) to recover the ground truth DAG. Figure 6a shows that $\bar{x}dHSIC$ identifies the correct DAG more frequently than $dHSIC$ for $n > 500$, and it reaches perfect accuracy using roughly two-thirds of the samples needed by $dHSIC$, all of it with a 200-fold computational speed-up. Similarly, the structural Hamming distance (SHD) of $\bar{x}dHSIC$ reaches zero faster than $dHSIC$ (Fig. 6b).

In real-world causal discovery, one often needs to search over all possible DAGs (see Appendix D). For 3, 4, 5, and 6 nodes, there are 25, 543, 29281, and 3781503 DAGs, respectively. Since the ratio of DAG counts between consecutive node sets $\{3, 4, 5, 6\}$ is less than 200, the computation time required for $dHSIC$ (with $p = 200$ permutations) to test joint independence in DAGs of 3 (or 4, 5) nodes exceeds that of $\bar{x}dHSIC$ applied to DAGs with an additional node, i.e., 4 (or 5, 6). This efficiency gain can facilitate experiments to uncover causal relationships more effectively.

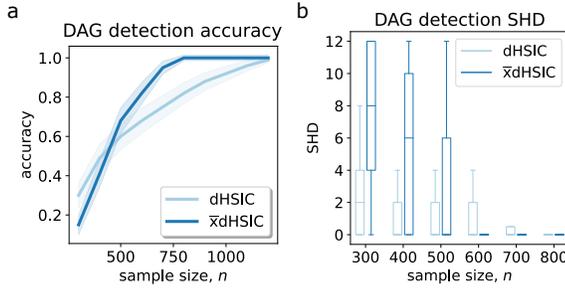


Figure 6. **DAG causal discovery.** (a) Accuracies of the correct DAG being detected among all fully connected DAGs of four nodes and (b) the distributions of the SHD of the detected DAG using \bar{x} dHNSIC and dHNSIC out of 100 experiments.

5.3. Application to Feature Selection

Feature selection is a crucial step in many machine learning pipelines. The task is to identify the most informative features to enhance model performance and interpretability. Most techniques, however, focus solely on *pairwise* relationships between a feature and the target variable (Póczos et al., 2012; Song et al., 2012), overlooking high-order interactions (Fumagalli et al., 2024; Muschalik et al., 2024).

To illustrate the application to feature selection, we construct a dataset ($n = 500$) with nine features $\{X^1, \dots, X^9\}$, of which $\{X^1, \dots, X^7\}$ are *i.i.d.* uniform variables interacting through an XOR gate to generate a target variable Y , and $\{X^8, X^9\}$ are MVG’s with covariance $[[1, 0.9], [0.9, 1]]$. The ground truth interaction between the variables is thus $\{Y, X^1, \dots, X^7\} \perp\!\!\!\perp \{X^8, X^9\}$; hence no single variable or pair of variables is predictive of Y . In Table 2, we compare the performance of feature selection methods based on high-order tests with popular methods based on univariate and pairwise measures, including the sequential method, recurrent feature selection (RFE) (Pedregosa et al., 2011) and SHAP (Lundberg & Lee, 2017). The permutation-free test \bar{x} SI+ \bar{x} LI is the only method able to detect the ground truth feature set with lowest mean squared error (MSE). All other methods, including pairwise independence, joint independence and Lancaster tests, suffer from errors in selecting the correct features due to their inability to identify the complex high-order interactions between the input features and the target variable in this example.

5.4. Application to Dataset of Stock Daily Returns

To demonstrate the scalability and applicability of our permutation-free tests, we consider a large real-world financial dataset: daily returns of stocks in the S&P 500 from 2020 to 2024. This dataset comprises 504 variables (stocks) and 1004 samples (returns), spanning 11 sectors (Fig. 7). As is standard in the finance literature, we assume the daily returns to be *i.i.d.* (Ali & Giaccotto, 1982).

Table 2. **Feature selection.** Only \bar{x} SI+ \bar{x} LI identifies the ground truth variables (●) that generate Y . Other methods either include confounding variables (●) or miss generating variables (○).

Method	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	MSE
Variance	●	●	●	●	●	●	●	●	●	5.01
Univariate	●	●	●	●	●	●	●	●	●	5.01
Sequential	●	●	●	●	○	○	○	○	○	4.69
RFE	●	●	●	○	○	○	○	○	○	4.47
SHAP	●	●	●	●	●	○	○	○	○	4.67
\bar{x} HNSIC	●	●	●	●	●	○	○	○	○	4.67
\bar{x} dHNSIC	●	●	●	●	●	○	○	○	○	5.01
\bar{x} LI	●	●	●	●	●	●	●	○	○	5.01
\bar{x} SI+ \bar{x} LI	●	●	●	●	●	●	●	○	○	4.35
Ground truth	●	●	●	●	●	●	●	○	○	4.35

Figure 7 shows the percentage of 2-, 3-, 4-, and 5-way interactions for all sectors, where we compute interactions for 500 sets of stocks within the same sector and, as a comparison, for 5000 sets of stocks randomly drawn from different sectors. As expected, 2-, 3-, 4-, and 5-way interactions are more common within a sector than across sectors, in concordance with the GICS industrial taxonomy that underpins the definition of the sectors. Interestingly, both Utilities and Energy exhibit very high 3-, 4-, and 5-way within-sector interactions. This likely stems from the highly regulated nature of these sectors, resulting in stocks with similar returns and, consequently, high (within-sector) redundancy. Conversely, Information Technology, Consumer Discretionary and Health Care display lower high-order interactions. This suggests that companies within these sectors may be more closely connected to firms in other sectors as they are likely influenced by external factors, indicating (cross-sector) synergistic relationships. A summary heatmap of the high-order interaction profiles of all sectors is shown in Figure 13 in Appendix E. These high-order interactions could be of help when considering portfolio diversification.

6. Discussion

The use of fine-tuned permutation schemes has been shown to be an essential ingredient for interaction tests to produce reliable results, but this process can be both time-consuming and challenging (Rindt et al., 2021). The permutation-free methods introduced here, \bar{x} dHNSIC, \bar{x} LI and \bar{x} SI, offer a more efficient approach to handling complex data structures circumventing the need for permutations and the tuning of associated hyperparameters, thus simplifying the workflow and ensuring consistent performance.

Regarding when to use permutation-free vs. permutation-based tests, our experiments show that for small sample size, permutation-based methods perform comparably, and occasionally slightly better. This is likely due to the need for data splitting in permutation-free methods, which limits their ability to capture the full data structure in low-sample regimes as test statistics rely on inner products between

Permutation-Free High-Order Interaction Tests

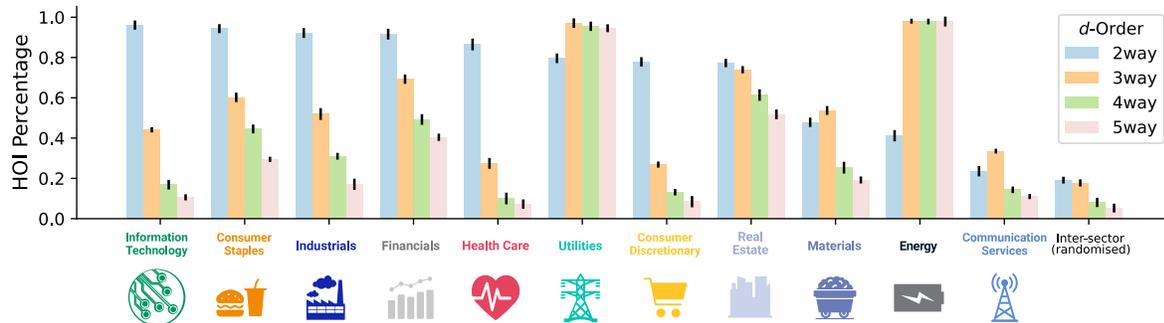


Figure 7. **Stock interactions in S&P500.** Utilities and Energy show high 3-, 4- and 5-way interactions, likely due to regulation-driven redundancy. In contrast, Information Technology, Health Care, and Consumer Discretionary exhibit lower high-order interactions, suggesting stronger cross-sector connectivity.

embeddings estimated from the two separate halves. Hence, as a rule of thumb, for low number of samples n and low order d the trade-off between computation time and statistical power would not be unfavourable for permutation-based methods, e.g., for $n \leq 50$ and $d \leq 5$ permutation-based methods perform well with just $p = 100$ permutations. Outside this regime, however, we strongly recommend the permutation-free methods introduced here.

Nevertheless, several challenges and limitations remain. The data-splitting technique that underpins our permutation-free strategy relies on the assumption of *i.i.d.* samples, which can be restrictive when dependencies exist across samples (e.g., time series, network data, spatial data). Although permutation-based techniques have been developed to address stationary (Chwialkowski & Gretton, 2014; Rubenstein et al., 2016) and non-stationary time-series data (Liu et al., 2023a), the development of permutation-free methods for such data remains an open challenge. Additionally, while we focus here on joint independence and factorisation hypotheses, other forms of higher-order structure, such as conditional independence or more intricate composite hypotheses, might benefit from similar permutation-free schemes. Finally, although the permutation-free approach substantially improves the efficiency of high-order tests, there remains potential to further reduce computational overhead by incorporating scalable kernel approximation techniques such as random Fourier features and the Nyström method (Zhang et al., 2018). Investigating these directions could further broaden the applicability of kernel-based high-order interaction tests in large-scale machine learning and statistical settings.

Code Availability Code to implement the permutation-free tests in this paper available at <https://github.com/barahona-research-group/PermFree-HOI.git>.

Acknowledgements

MB acknowledges support by EPSRC through grants EP/W024020/1, funding the project “Statistical physics of cognition”, and EP/N014529/1, funding the EPSRC Centre for Mathematics of Precision Healthcare, and by the Nuffield Foundation, under the project “The Future of Work and Well-being: The Pissarides Review”. RP acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG) Project-ID 424778381-TRR 295.

The authors thank Felix Laumman, Jianxiong Sun, Shubhanshu Shekhar and Zhuoyu Li for valuable discussions.

Impact Statement

The societal impact of our work stems from its ability to address the critical need for efficient and scalable methods to detect high-order interactions in multivariate data. By eliminating the reliance on computationally intensive permutation schemes, our permutation-free tests enable researchers to uncover complex relationships beyond pairwise interactions, which are crucial in domains like causal discovery and feature selection. These advancements facilitate more effective data analysis in fields such as healthcare, economics, and environmental science, where understanding intricate interactions can lead to breakthroughs in diagnosis, decision-making, and policy formulation.

References

- Ali, M. M. and Giaccotto, C. The identical distribution hypothesis for stock market prices—location-and scale-shift alternatives. *Journal of the American Statistical Association*, 77(377):19–28, 1982.
- Arnaudon, A., Peach, R. L., Petri, G., and Expert, P. Connecting hodge and sakaguchi-kuramoto through a mathematical framework for coupled oscillators on simplicial complexes. *Communications Physics*, 5(1):211, 2022.

- Battiston, F., Cencetti, G., Iacopini, I., Latora, V., Lucas, M., Patania, A., Young, J.-G., and Petri, G. Networks beyond pairwise interactions: structure and dynamics. *Physics Reports*, 874:1–92, 2020.
- Battiston, F., Amico, E., Barrat, A., Bianconi, G., Ferraz de Arruda, G., Franceschiello, B., Iacopini, I., Kéfi, S., Latora, V., Moreno, Y., et al. The physics of higher-order interactions in complex systems. *Nature Physics*, 17(10):1093–1098, 2021.
- Benson, A. R., Gleich, D. F., and Leskovec, J. Higher-order organization of complex networks. *Science*, 353(6295):163–166, 2016.
- Bick, C., Gross, E., Harrington, H. A., and Schaub, M. T. What are higher-order networks? *SIAM Review*, 65(3):686–731, 2023.
- Bińkowski, M., Sutherland, D. J., Arbel, M., and Gretton, A. Demystifying mmd gans. *arXiv preprint arXiv:1801.01401*, 2018.
- Böttcher, B. Dependence and dependence structures: estimation and visualization using the unifying concept of distance multivariate. *Open Statistics*, 1(1):1–48, 2020.
- Chakraborty, S. and Zhang, X. Distance metrics for measuring joint dependence with application to causal inference. *Journal of the American Statistical Association*, 2019.
- Chwialkowski, K. and Gretton, A. A kernel independence test for random processes. In *International Conference on Machine Learning*, pp. 1422–1430. PMLR, 2014.
- Fumagalli, F., Muschalik, M., Hüllermeier, E., Hammer, B., and Herbinger, J. Unifying feature-based explanations with functional anova and cooperative game theory. *arXiv preprint arXiv:2412.17152*, 2024.
- Gretton, A., Fukumizu, K., Teo, C., Song, L., Schölkopf, B., and Smola, A. A kernel statistical test of independence. *Advances in neural information processing systems*, 20, 2007.
- Grilli, J., Barabás, G., Michalska-Smith, M. J., and Allesina, S. Higher-order interactions stabilize dynamics in competitive network models. *Nature*, 548(7666):210–213, 2017.
- Hyvärinen, A. and Oja, E. Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5):411–430, 2000.
- Kim, I. and Ramdas, A. Dimension-agnostic inference using cross u-statistics. *Bernoulli*, 30(1):683–711, 2024.
- Lancaster, H. *The Chi-Squared Distribution*. Wiley, 1969.
- Laumann, F., Von Kügelgen, J., Park, J., Schölkopf, B., and Barahona, M. Kernel-based independence tests for causal structure learning on functional data. *Entropy*, 25(12):1597, 2023.
- Liu, Z., Peach, R. L., Laumann, F., Vallejo Mengod, S., and Barahona, M. Kernel-based joint independence tests for multivariate stationary and non-stationary time series. *Royal Society Open Science*, 10(11):230857, 2023a.
- Liu, Z., Peach, R. L., Mediano, P. A., and Barahona, M. Interaction measures, partition lattices and kernel tests for high-order interactions. *Advances in Neural Processing Systems*, 36, 2023b.
- Liu, Z., Barahona, M., and Peach, R. L. Information-theoretic measures on lattices for high-order interactions. *AISTATS - Proceedings of The 28th International Conference on Artificial Intelligence and Statistics*, PMLR, 2025.
- Long, M., Cao, Y., Wang, J., and Jordan, M. Learning transferable features with deep adaptation networks. In *International conference on machine learning*, pp. 97–105. PMLR, 2015.
- Lopez, R., Regier, J., Jordan, M. I., and Yosef, N. Information constraints on auto-encoding variational bayes. *Advances in neural information processing systems*, 31, 2018.
- Luff, C. E., Peach, R., Mallas, E.-J., Rhodes, E., Laumann, F., Boyden, E. S., Sharp, D. J., Barahona, M., and Grossman, N. The neuron mixer and its impact on human brain dynamics. *Cell Reports*, 43(6), 2024.
- Lundberg, S. M. and Lee, S.-I. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 2017.
- Muandet, K., Fukumizu, K., Sriperumbudur, B., and Schölkopf, B. Kernel mean embedding of distributions: A review and beyond. *Found. Trends Mach. Learn.*, 10(1–2):1–141, June 2017. ISSN 1935-8237. doi: 10.1561/22000000060. URL <https://doi.org/10.1561/22000000060>.
- Muschalik, M., Baniecki, H., Fumagalli, F., Kolpaczki, P., Hammer, B., and Hüllermeier, E. shapiq: Shapley interactions for machine learning. *Advances in Neural Information Processing Systems*, 37:130324–130357, 2024.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

- Peters, J., Mooij, J. M., Janzing, D., and Schölkopf, B. Causal discovery with continuous additive noise models. *J. Mach. Learn. Res.*, 15(1):2009–2053, January 2014. ISSN 1532-4435.
- Pfister, N., Bühlmann, P., Schölkopf, B., and Peters, J. Kernel-based tests for joint independence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1):5–31, 2018.
- Póczos, B., Ghahramani, Z., and Schneider, J. Copula-based kernel dependency measures. In *Proceedings of the 29th International Conference on International Conference on Machine Learning, ICML'12*, pp. 1635–1642, Madison, WI, USA, 2012. Omnipress. ISBN 9781450312851.
- Rindt, D., Sejdinovic, D., and Steinsaltz, D. Consistency of permutation tests of independence using distance covariance, hsc and dhsic. *Stat.*, 10(1):e364, 2021.
- Rosas, F. E., Mediano, P. A., Rassouli, B., and Barrett, A. B. An operational information decomposition via synergistic disclosure. *Journal of Physics A: Mathematical and Theoretical*, 53(48):485001, 2020.
- Rosas, F. E., Mediano, P. A., Luppi, A. I., Varley, T. F., Lizier, J. T., Stramaglia, S., Jensen, H. J., and Marinazzo, D. Disentangling high-order mechanisms and high-order behaviours in complex systems. *Nature Physics*, 18(5):476–477, 2022.
- Rosas, F. E., Gutknecht, A., Mediano, P. A., and Gastpar, M. Characterising high-order interdependence via entropic conjugation. *arXiv preprint arXiv:2410.10485*, 2024.
- Rubenstein, P. K., Chwialkowski, K. P., and Gretton, A. A kernel test for three-variable interactions with random processes. *arXiv preprint arXiv:1603.00929*, 2016.
- Santoro, A., Battiston, F., Petri, G., and Amico, E. Higher-order organization of multivariate time series. *Nature Physics*, 19(2):221–229, 2023.
- Santoro, A., Battiston, F., Lucas, M., Petri, G., and Amico, E. Higher-order connectomics of human brain function reveals local topological signatures of task decoding, individual identification, and behavior. *Nature Communications*, 15(1):10244, 2024.
- Schaub, M. T., Benson, A. R., Horn, P., Lippner, G., and Jadbabaie, A. Random walks on simplicial complexes and the normalized hodge 1-laplacian. *SIAM Review*, 62(2):353–391, 2020.
- Schneidman, E., Still, S., Berry, M. J., Bialek, W., et al. Network information and connected correlations. *Physical review letters*, 91(23):238701, 2003.
- Sejdinovic, D., Gretton, A., and Bergsma, W. A kernel test for three-variable interactions. *Advances in Neural Information Processing Systems*, 26, 2013a.
- Sejdinovic, D., Sriperumbudur, B., Gretton, A., and Fukumizu, K. Equivalence of distance-based and rkhs-based statistics in hypothesis testing. *The annals of statistics*, pp. 2263–2291, 2013b.
- Shekhar, S., Kim, I., and Ramdas, A. A permutation-free kernel two-sample test. *Advances in Neural Information Processing Systems*, 35:18168–18180, 2022.
- Shekhar, S., Kim, I., and Ramdas, A. A permutation-free kernel independence test. *Journal of Machine Learning Research*, 24(369):1–68, 2023.
- Sloane, N. J. et al. The on-line encyclopedia of integer sequences, sequence a000296. *Published electronically at https://oeis.org*, 2018.
- Song, L., Smola, A., Gretton, A., Bedo, J., and Borgwardt, K. Feature selection via dependence maximization. *Journal of Machine Learning Research*, 13(5), 2012.
- Streitberg, B. Lancaster interactions revisited. *The Annals of Statistics*, pp. 1878–1885, 1990.
- Thibeault, V., Allard, A., and Desrosiers, P. The low-rank hypothesis of complex systems. *Nature Physics*, 20(2):294–302, 2024.
- Young, J.-G., Petri, G., and Peixoto, T. P. Hypergraph reconstruction from network data. *Communications Physics*, 4(1):1–11, 2021.
- Zhang, K., Peters, J., Janzing, D., and Schölkopf, B. Kernel-based conditional independence test and application in causal discovery. In *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence, UAI'11*, pp. 804–813, Arlington, Virginia, USA, 2011. AUAI Press. ISBN 9780974903972.
- Zhang, Q., Filippi, S., Gretton, A., and Sejdinovic, D. Large-scale kernel methods for independence testing. *Statistics and Computing*, 28:113–130, 2018.

Appendices

A. Proofs for Section 3

A.1. Proof of Proposition 3.3

First observe that each term in the numerator can be computed in $\mathcal{O}(dn^2)$ since the expression only contains Hadamard product and vector matrix multiplication Pfister et al. (2018). Similarly for the denominator, the variance estimation introduces no significant computational overhead beyond simple vector norm compared to `xdHSIC`.

A.2. Proof of Lemma 3.5

The conventional centring of a matrix K is denoted by $\tilde{K} = CKC$, where $C = I - \frac{1}{n}\mathbf{1}\mathbf{1}^T$ and $\mathbf{1}$ is an $n \times 1$ vector of ones. It has been shown that the respective centred kernel function \tilde{k} is an equivalent kernel to k . Indeed, as shown in Sejdinovic et al. (2013b), a kernel \tilde{k}_f is equivalent to k if it can be expressed as:

$$\tilde{k}_f(z, z') = \langle k(\cdot, z) - f, k(\cdot, z') - f \rangle_{\text{HS}}$$

for some function f . Now for conventional centring \tilde{k} , $f = \mu_{\mathbb{P}_X}$ which can be approximated by $\hat{\mu}_{\mathbb{P}_X} = \frac{1}{n} \sum_{i=1}^n \phi(x_i)$ and thus is an equivalent kernel.

Here, the cross-centred \bar{k} is also equivalent as $\frac{1}{n} \sum_{i=1}^n \phi(x_i)$ and $\frac{1}{n} \sum_{i=n+1}^{2n} \phi(x_i)$ can be seen as approximation of $\mu_{\mathbb{P}_X}$.

A.3. Proof of Proposition 3.8

Here, consider the Hadamard product of all the kernels in D except K^k as an individual kernel, then the $\bar{\text{xLI}}$ can be seen as a special $\bar{\text{xHSIC}}$ and therefore they share the same time complexity.

A.4. Proof of Proposition 3.11

Here, consider the two Hadamard products as two kernels themselves, then the $\bar{\text{xSI}}$ can be seen as a special $\bar{\text{xHSIC}}$ and therefore they share the same time complexity.

B. Extended analysis of the $d = 3$ Lancaster interaction test

B.1. Lancaster interaction hypothesis and test

$\Delta_L^3 \mathbb{P}$ in Equation (3) is the extension as d increases from 2 to 3 in the direction of the factorisation route as illustrated in Figure 1. Now instead of testing for joint independence through the full factorisation $\mathbb{P}_{123} = \mathbb{P}_1 \mathbb{P}_2 \mathbb{P}_3$, we would like to understand if the joint distribution \mathbb{P}_{123} is partially factorisable as summarised in the composite hypothesis below:

Hypothesis B.1. (Order-3 Factorisation)

$$\begin{aligned} H_0 &: (\mathbb{P}_{123} = \mathbb{P}_{12} \mathbb{P}_3) \vee (\mathbb{P}_{123} = \mathbb{P}_{13} \mathbb{P}_2) \vee (\mathbb{P}_{123} = \mathbb{P}_{23} \mathbb{P}_1) \\ H_1 &: (\mathbb{P}_{123} \neq \mathbb{P}_{12} \mathbb{P}_3) \wedge (\mathbb{P}_{123} \neq \mathbb{P}_{13} \mathbb{P}_2) \wedge (\mathbb{P}_{123} \neq \mathbb{P}_{23} \mathbb{P}_1) \end{aligned}$$

If any of the subhypotheses in the null hypothesis H_0 is true, then $\Delta_L^3 \mathbb{P} = 0$. For example, if $\mathbb{P}_{123} = \mathbb{P}_{12} \mathbb{P}_3$, this implies $\mathbb{P}_{13} = \mathbb{P}_1 \mathbb{P}_3$ and $\mathbb{P}_{23} = \mathbb{P}_2 \mathbb{P}_3$, and, as a result, we have:

$$\mathbb{P}_{123} = \mathbb{P}_{12} \mathbb{P}_3 \implies \Delta_L^3 \mathbb{P} = \mathbb{P}_{123} - \mathbb{P}_{12} \mathbb{P}_3 - \mathbb{P}_{13} \mathbb{P}_2 - \mathbb{P}_{23} \mathbb{P}_1 + 2\mathbb{P}_1 \mathbb{P}_2 \mathbb{P}_3 = 0$$

And, similarly, $\mathbb{P}_{123} = \mathbb{P}_{13} \mathbb{P}_2 \implies \Delta_L^3 \mathbb{P} = 0$ and $\mathbb{P}_{123} = \mathbb{P}_{23} \mathbb{P}_1 \implies \Delta_L^3 \mathbb{P} = 0$. Hence $H_0 \implies \Delta_L^3 \mathbb{P} = 0$. Note that even though H_0 only considers the three partial factorisations, the factorisation $\mathbb{P}_1 \mathbb{P}_2 \mathbb{P}_3$ is actually subsumed by them, i.e. $\mathbb{P}_{123} = \mathbb{P}_1 \mathbb{P}_2 \mathbb{P}_3$ implies that $\mathbb{P}_{123} = \mathbb{P}_{12} \mathbb{P}_3$ and $\mathbb{P}_{123} = \mathbb{P}_{13} \mathbb{P}_2$ and $\mathbb{P}_{123} = \mathbb{P}_{23} \mathbb{P}_1$. In other words, $\mathbb{P}_{123} = \mathbb{P}_1 \mathbb{P}_2 \mathbb{P}_3 \implies \Delta_L^3 \mathbb{P} = 0$ also.

By the contrapositive, this means: $\Delta_L^3 \mathbb{P} \neq 0 \implies \neg H_0 \equiv H_1$. Hence, when a $d = 3$ interaction is detected via the alternative hypothesis H_1 , all three subhypotheses in H_0 , as well as the complete factorisation $\mathbb{P}_{123} = \mathbb{P}_1 \mathbb{P}_2 \mathbb{P}_3$, need to be rejected.

B.2. Construction of the permutation-free normalised test statistic

By computing the RKHS embedding and applying the permutation-free scheme, the unnormalised test statistics of three-variable Lancaster interaction can be immediately written out in terms of the cross-centred kernels:

$$\text{xLI} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=n+1}^{2n} \left[\overline{K^1} \circ \overline{K^2} \circ \overline{K^3} \right]_{ij}$$

Note that a key distinction between the permutation-based methods and the permutation-free approaches discussed here lies in how multiple tests are conducted. In permutation-based methods, the test statistic remains constant, while the permutations adapt to changes in the subhypotheses. In contrast, with the permutation-free approach, where the null distributions are all standardised as $\mathcal{N}(0, 1)$, the test statistics must instead vary.

Without loss of generality, under the subhypothesis $\mathbb{P}_{123} = \mathbb{P}_{12}\mathbb{P}_3$, xLI can be simplified into:

$$\text{xLI}_{[12|3]} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=n+1}^{2n} \left[\overline{K^1} \circ \overline{K^2} \circ \overline{K^3} \right]_{ij}$$

The normalised test statistic is

$$\begin{aligned} \overline{\text{xLI}}_{[12|3]} &= \frac{\sqrt{n} \cdot \text{xLI}_{[12|3]}}{s_{L_{[12|3]}}}, \quad \text{with} \\ s_{L_{[12|3]}}^2 &= \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{n} \sum_{j=n+1}^{2n} \left[\overline{K^1} \circ \overline{K^2} \circ \overline{K^3} \right]_{ij} - \text{xLI}_3 \right)^2 \\ &= \frac{1}{n} \sum_{i=1}^n \left(\left(\frac{1}{n} \sum_{j=n+1}^{2n} \left[\overline{K^1} \circ \overline{K^2} \circ \overline{K^3} \right]_{ij} \right)^2 + \text{xLI}_3^2 - 2\text{xLI}_3 \left(\frac{1}{n} \sum_{j=n+1}^{2n} \left[\overline{K^1} \circ \overline{K^2} \circ \overline{K^3} \right]_{ij} \right) \right) \\ &= \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{n} \sum_{j=n+1}^{2n} \left[\overline{K^1} \circ \overline{K^2} \circ \overline{K^3} \right]_{ij} \right)^2 - \text{xLI}_3^2 \end{aligned}$$

Similarly, for the other two subhypotheses, we just move the double cross-centring to the pairwise variables. The whole testing procedure is summarised in Algorithm 1.

Algorithm 1 Lancaster Interaction Test at $d = 3$

Input: Centred kernel matrices $\overline{K^1}, \overline{K^2}, \overline{K^3}$ computed from observational data
for each subhypothesis (total three subhypotheses) **do**
 Compute the test statistic $\overline{\text{xLI}}$
 if $\overline{\text{xLI}} > (1 - \alpha)$ -quantile of $\mathcal{N}(0, 1)$ **then**
 Reject H_i and proceed to the next subhypothesis
 else
 Terminate and do not reject composite hypothesis
 end if
end for
if all three subhypotheses are rejected **then**
 Reject the composite hypothesis
else
 Do not reject the composite hypothesis
end if

The composite hypothesis H_0 is rejected only if all subhypotheses are individually rejected. As this has been proven to be less conservative than the Bonferroni correction (Rubenstein et al., 2016), we use this correction for the factorisation tests for any d . When the subtests are executed sequentially rather than in parallel, the procedure can terminate early if any subtest fails to reject its corresponding subhypothesis, thereby saving computational time.

B.3. Application to additional examples: nonlinear form and different kernels

In addition to the examples from [Sejdinovic et al. \(2013a\)](#), we also apply our method to three more examples with diverse nonlinear forms. In all cases, our proposed permutation-free method achieves similar accuracy compared with the permutation-based counterpart. Moreover the results are robust with different kernel functions, Rational-Quadratic (RQ) kernel, Laplace Kernel and Radial Basis Function (RBF) kernel.

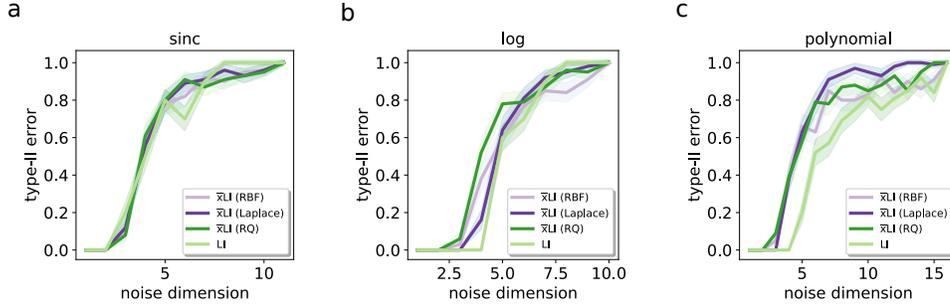


Figure 8. Structural causal models with diverse nonlinear forms and kernel functions. We constructed three V-structures using (a) sinc (b) log, and (c) polynomial nonlinear forms (see inset).

C. Definition of the $\bar{x}\text{HSIC}_V$ pairwise test using V-statistics

We formulate the corresponding V-statistic version of the permutation-free pairwise independence test statistic, $\bar{x}\text{HSIC}_V$ using the cross-centring technique:

$$\bar{x}\text{HSIC}_V = \frac{\sqrt{n} \cdot x\text{HSIC}_V}{s_V}$$

$$x\text{HSIC}_V = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=n+1}^{2n} [\bar{K}^1 \circ \bar{K}^2]_{ij}$$

$$s_V^2 = \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{n} \sum_{j=n+1}^{2n} [\bar{K}^1 \circ \bar{K}^2]_{ij} - x\text{HSIC}_V \right)^2$$

This is a simpler formulation compared with the original $\bar{x}\text{HSIC}$ in [Shekhar et al. \(2023\)](#).

Figure 9 shows that not only is $\bar{x}\text{HSIC}_V$ theoretically grounded, like $x\text{HSIC}$, but it also has higher power compared with both $\bar{x}\text{HSIC}$ and HSIC .

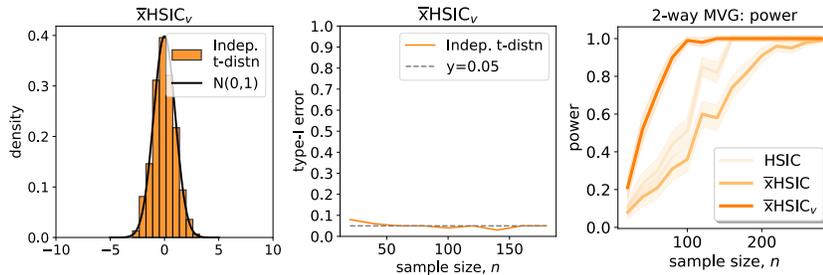


Figure 9. Experiments of pairwise independence tests: $\bar{x}\text{HSIC}_V$ follows a standard normal distribution with two independent t -distributions; controls the type-I error rate with the same data from t -distributions, and outperforms both HSIC ([Gretton et al., 2007](#)) and $\bar{x}\text{HSIC}$ ([Shekhar et al., 2023](#)) with dependent MVG from the example in Equation (13) from [Shekhar et al. \(2023\)](#) with parameters $b = 2$, $ndim = 10$, and $\epsilon = 0.5$.

D. Further numerical checks of the test statistics

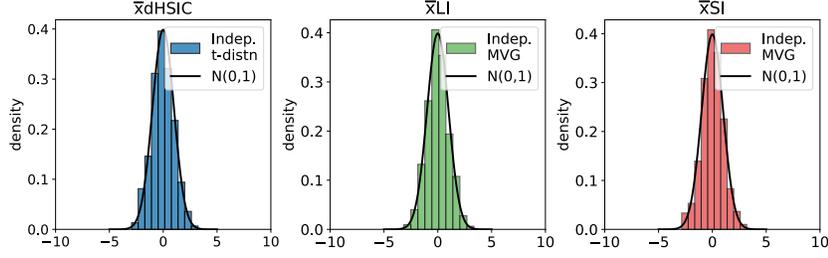


Figure 10. **Null distributions of permutation-free test statistics.** All three test statistics introduced in this paper follow a standard normal distribution $\mathcal{N}(0, 1)$ under their respective nulls.

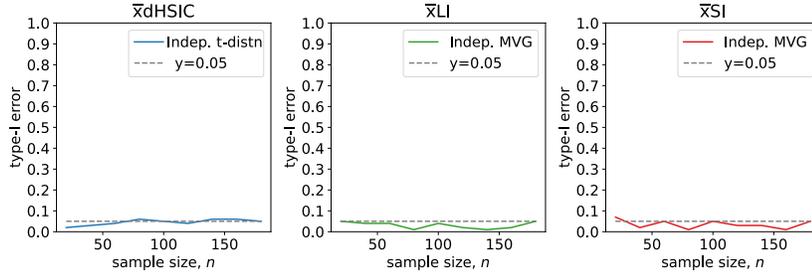


Figure 11. **Controlled type-I error of permutation-free test statistics.** All three test statistics introduced in this paper have a valid level of type-I error for $\alpha = 0.05$, indicated by the dashed line.

E. Further details of datasets used in Section 5

The code to implement the permutation-free tests introduced in this paper is available at <https://github.com/barahona-research-group/PermFree-HOI.git>.

E.1. XOR example from Liu et al. (2023b)

We generate n samples of $V, W, X, Y, Z \sim \mathcal{U}(0, 4)$, $Z_{:i} = (V_{:i} + W_{:i} + X_{:i} + Y_{:i}) \bmod 4$ and $Z_{i+1:n} \sim \mathcal{U}(0, 4)$ (where the samples $[i + 1 : n]$ act as noise). We then gradually increase the interaction proportion, $0 \leq i/n \leq 1$.

E.2. V-structure dataset A and B from Sejdinovic et al. (2013a)

Dataset A: For $\{X, Y, Z\} \in \mathbb{R}^a \times \mathbb{R}^a \times \mathbb{R}^a$, $X, Y \sim \mathcal{N}(0, I_p)$, $W \sim \text{Exp}\left(\frac{1}{\sqrt{2}}\right)$, $Z_1 = \text{sign}(X_1 Y_1) W$, and $Z_{2:p} \sim \mathcal{N}(0, I_{p-1})$.

Dataset B: For $\{X, Y, Z\} \in \mathbb{R}^a \times \mathbb{R}^a \times \mathbb{R}^a$, $X, Y \sim \mathcal{N}(0, I_p)$, $Z_{2:p} \sim \mathcal{N}(0, I_{p-1})$, and

$$Z_1 = \begin{cases} X_1^2 + \epsilon, & \text{w.p. } 1/3, \\ Y_1^2 + \epsilon, & \text{w.p. } 1/3 \\ X_1 Y_1 + \epsilon, & \text{w.p. } 1/3 \end{cases}$$

where $\epsilon \sim \mathcal{N}(0, 0.01)$ and W.P. stands for with probability.

In both datasets, the 3-way interaction becomes increasingly difficult to detect as the noise dimension p increases.

E.3. Causal discovery example from Pfister et al. (2018)

For an additive noise model over random variables X^1, \dots, X^d ,

$$X^j := \sum_{k \in \text{Parents}} f^{j,k}(X^k) + N^j, \quad j \in \{1, \dots, d\} \quad (6)$$

with corresponding DAG G . The noise variables N^j are normally distributed and are jointly independent with a standard deviation sampled uniformly between $\sqrt{2}$ and 2. Nodes without parents follow a Gaussian distribution with standard deviation sampled uniformly between $5\sqrt{2}$ and 10. The functions f_{jk} are sampled from a Gaussian process with Gaussian kernel and bandwidth one.

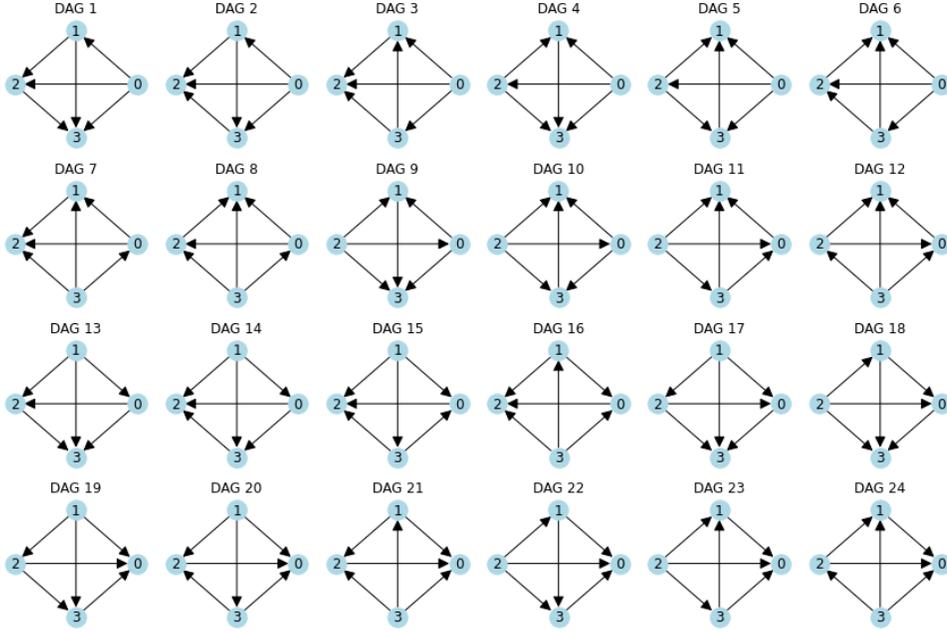


Figure 12. **Fully connected DAG of 4 nodes.** We use DAG 1 as the ground truth causal structure to simulate the data.

In this example we simulate the data from DAG 1 and compare all the possible fully connected DAGs shown in Figure 12. The procedure of the causal discovery is outlined as the following: (i) Run the generalised additive noise model for each DAG and get the residuals. (ii) Check the joint independence between the d residuals. (iii) Report the DAG with the largest p-value.

E.4. Real-World Financial Data

The CPU time for computing high-order interaction percentages took 8 hours without any parallelisation on a 2015 iMac with 4 GHz Quad-Core Intel Core i7 processor and 32 GB 1867 MHz DDR3 memory. We estimate the time taken for the permutation-based methods would have been at least 2 weeks. This highlights the scalability of our proposed method.

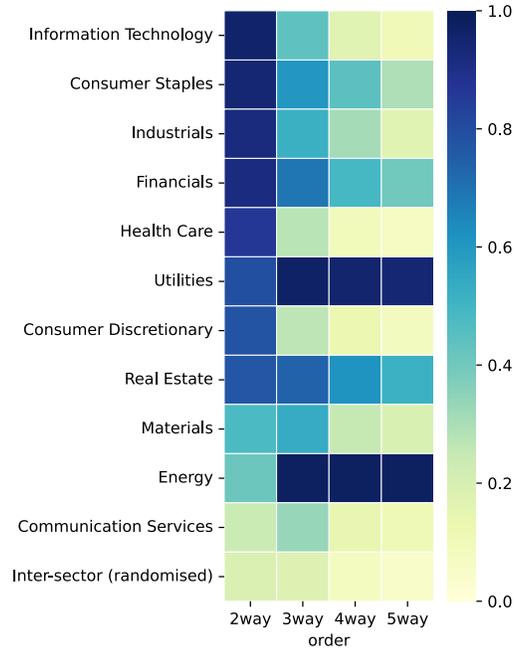


Figure 13. Percentages of high-order interactions between stocks within sectors in S&P500. Heatmap of the percentages shown in Figure 7 summarising the differences across sectors in their high-order interaction structure, also relative to randomised sets (last row).