Robust Independence Tests With Finite Sample Guarantees for Synchronous Stochastic Linear Systems

Ambrus Tamás, Dániel Ágoston Bálint, and Balázs Csanád Csáji, *Member, IEEE*

***Abstract*—The letter introduces robust independence tests with non-asymptotically guaranteed significance levels for stochastic linear time-invariant systems, assuming that the observed outputs are synchronous, which means that the systems are driven by jointly i.i.d. noises. Our method provides bounds for the type I error probabilities that are distribution-free, i.e., the innovations can have arbitrary distributions. The algorithm combines confidence region estimates with permutation tests and general dependence measures, such as the Hilbert–Schmidt independence criterion and the distance covariance, to detect any nonlinear dependence between the observed systems. We also prove the consistency of our hypothesis tests under mild assumptions and demonstrate the ideas through the example of autoregressive systems.**

***Index Terms*—Independence tests, stochastic linear systems, distribution-free guarantees, dependence measures.**

I. Introduction

STATISTICAL independence is a key notion in several areas of statistics and probability theory, including system identification [1], time-series analysis, signal processing and machine learning. In this letter we present a non-asymptotic framework to construct hypothesis tests for the independence of two simultaneous linear systems or time series. Our setup is distribution-free, i.e., the process noises can follow any probability law. The presented hypothesis tests might be used for instance to identify (conditionally) dependent price returns in the stock market or to find interconnections between two systems in a network. Other potential applications include biological systems, where the independence of cell mechanisms may be tested, or one could analyze whether social phenomena which occur simultaneously are dependent.

Given an i.i.d. sample of random pairs with some joint distribution, there already exists several hypothesis tests for independence, e.g., the celebrated χ^2 test, Hoeffding's test based on the factorization of the joint distribution function, and Hilbert–Schmidt independence criterion (HSIC) based tests [2]. These methods typically use the limiting distribution of some test statistic to calculate the p-values for a given sample size. Usually it is challenging or even impossible to calculate the exact distribution of these test statistics, but permutation and Monte Carlo tests offer viable options. For time series, it is even more challenging to test independence. For this, HSIC [3] and distance correlation [4] based approaches were proposed, which are only supported by asymptotic guarantees.

Manuscript received 17 March 2023; revised 19 May 2023; accepted 3 June 2023. Date of publication 20 June 2023; date of current version 11 July 2023. This work was supported in part by the European Union Project RRF-2.3.1-21-2022-00004 within the framework of the Artificial Intelligence National Laboratory; in part by the National Research, Development and Innovation Office (NRDIO), Hungary, under Grant TKP2021-NKTA-01; in part by the Cooperative Doctoral Program of the Ministry of Innovation and Technology; and in part by the NRDIO Fund. The work of Dániel Ágoston Bálint was supported by the Swissblock Technologies. Recommended by Senior Editor S. Dey.

(Corresponding author: Ambrus Tamás.)

Ambrus Tamás and Balázs Csanád Csáji are with the Eötvös Loránd Research Network, Institute for Computer Science and Control (SZTAKI), 1111 Budapest, Hungary, and also with the Institute of Mathematics, Eötvös Loránd University, 1053 Budapest, Hungary (e-mail: tamasamb@sztaki.hu; csaji@sztaki.hu).

Dániel Ágoston Bálint resides in Switzerland (e-mail: bdagoston@gmail.com).

Digital Object Identifier 10.1109/LCSYS.2023.3287797

II. Problem Setting

We construct robust permutation-based independence tests [5], [6], for linear systems in a general setting [1]. Let us consider two scalar (discrete-time, time-invariant) stochastic linear systems with general dynamics:

$$\begin{aligned} Y_t &= G_1(q^{-1};\theta_*)U_t + H_1(q^{-1};\theta_*)E_t,\\ Z_t &= G_2(q^{-1};\gamma_*)V_t + H_2(q^{-1};\gamma_*)N_t, \end{aligned} \tag{1}$$

for $t \in \mathbb{Z}$, where U_t and V_t are (exogenous) inputs; Y_t and Z_t are (observable) outputs; q^{-1} is the backshift operator given by $q^{-1}X_t \doteq X_{t-1}$ for any time series $\{X_t\}$; E_t and N_t are possibly dependent process noises; G_1, H_1, and G_2, H_2 are rational transfer functions determined by finite dimensional parameter spaces Θ and Γ, respectively. The unknown true parameters are θ_* and γ_*.

Our main assumptions are as follows, cf. [1]:

- A1 The true systems generating outputs $\{Y_t\}$ and $\{Z_t\}$ are in the model classes, i.e., $\theta_* \in \Theta$ and $\gamma_* \in \Gamma$.
- A2 Rational (causal) transfer functions G_1, G_2, H_1 and H_2 have known orders.
- A3 H_1 and H_2 are invertible for all $\theta \in \Theta$ and $\gamma \in \Gamma$.
- A4 The systems are initialized in zero, i.e., $Y_t = Z_t = U_t = V_t = E_t = N_t = 0$, for all $t \leq 0$.

A5 The systems are driven by an i.i.d. innovation sequence $\{(E_t, N_t)\}_{t=1}^{\infty}$ from the distribution of (E, N).
A6 The systems operate in open-loop: the inputs $\{U_t\}$, $\{V_t\}$ are independent of the noises $\{E_t\}$, $\{N_t\}$.

ARMAX models, e.g., can satisfy these conditions. For simplicity, we treat the inputs $\{U_t\}$, $\{V_t\}$ as deterministic sequences. This is w.l.o.g. as we can always condition on the inputs, as they are independent of the innovations. In case the inputs are stochastic, the obtained results should be interpreted as conditional to the inputs, i.e., we test whether $\{Y_t\}$ and $\{Z_t\}$ are conditionally independent given the inputs $\{U_t\}$, $\{V_t\}$. Also, we can assume that the parameterization is unique, e.g., by assuming w.l.o.g. that $H_1(0;\theta) = 1$ and $H_2(0;\gamma) = 1$ for all $\theta \in \Theta, \gamma \in \Gamma$.

In this letter we aim at constructing consistent hypothesis tests with finite sample guarantees for the independence of output sequences $\{Y_t\}$ and $\{Z_t\}$. For this we observe that if $\{Y_t\}$ and $\{Z_t\}$ are driven by a jointly i.i.d. noise (E, N), see A5, with joint distribution $Q_{E,N}$ and marginals Q_E respectively Q_N, then the independence of $\{E_t\}$ and $\{N_t\}$ is equivalent to the independence of $\{Y_t\}$ and $\{Z_t\}$ conditional on the inputs. Therefore, it is sufficient to test the null hypothesis

$$H_0 : Q_{E,N} = Q_E \otimes Q_N \quad H_1 : Q_{E,N} \neq Q_E \otimes Q_N \tag{2}$$

The main challenge is that the parameters θ^*, γ^* are unknown, henceforth the noise terms are not observable.

For simplicity, we assume that the finite sample of inputs, $\{U_t\}$, $\{V_t\}$, and outputs, $\{Y_t\}$, $\{Z_t\}$, available for estimation is large enough to compute n prediction errors $\{(E_t(\theta), N_t(\gamma))\}_{t=1}^{n}$ for any values of θ and γ.

We construct the hypothesis tests in several steps. First we estimate the system parameters with non-asymptotic confidence regions, then we reconstruct the residuals on the set of possible parameters and apply permutation tests on them. Our main assumption is that the linear systems are driven simultaneously, see assumption $A5$, and that the noise terms could be recovered if the system parameters were known, see assumption $A3$. We only rely on the i.i.d. assumption to quantify the user–chosen probability of type I error and prove that the probability of type II error vanishes asymptotically.

III. Permutation Tests for the I.I.D. Case

First, for simplicity, assume that the parameters are known. In this case we can compute the noise terms as

$$\begin{aligned} E_t &= E_t(\theta^*) = H_1^{-1}(q^{-1};\theta_*)(Y_t - G_1(q^{-1};\theta_*)U_t),\\ N_t &= N_t(\gamma_*) = H_2^{-1}(q^{-1};\gamma_*)(Z_t - G_1(q^{-1};\gamma_*)V_t), \end{aligned} \tag{3}$$

using $A3, A4$ and $A6$. Then the independence of E and N can be tested based on the i.i.d. sample, $\{(E_t, N_t)\}_{t=1}^{n}$.

A. Resampling

Let $\mathcal{D}_0 \doteq \{(E_t, N_t)\}_{t=1}^{n}$ be the known noise terms. We propose a rank test which is based on empirical dependence measure values calculated from perturbed samples. The idea is to generate new, perturbed datasets which have the same distributional properties to the original observations in case the null hypothesis is true.

We apply the permutation test which was first presented in [5]; proof of consistency and further ramifications have been provided in [6]. We choose an arbitrary (rational) significance level α in advance and integer hyperparameters $1 \le r \le p \le m$ such that

$$\alpha = 1 - \frac{p - r + 1}{m}. \tag{4}$$

Let S_n be the set of permutations on $[n] \doteq \{1, \dots, n\}$ and $\{\pi_j\}_{j=1}^{m-1}$ be uniformly randomly generated from S_n. We construct $m - 1$ new alternative samples by

$$\mathcal{D}_j = \pi_j \mathcal{D}_0 \doteq \{(E_i, N_{\pi_j(i)})\}_{i=1}^{n} \tag{5}$$

for $j = 1, \dots, m-1$. Altogether we end up having m datasets to compare. Observe that if H_0 holds true then $\mathcal{D}_0 = \big((E_1, N_1), \dots, (E_n, N_n)\big)$ has the same distribution as $\mathcal{D}_j = \big((E_1, N_{\pi_j(1)}), \dots, (E_n, N_{\pi_j(n)})\big)$ for any permutation π_j, whereas if H_1 holds then the distribution of $\mathcal{D}_j$ is different from that of $\mathcal{D}_0$. Our goal is to quantify this difference whenever the null hypothesis does not hold.

B. Exact Coverage

The comparison of the datasets is carried out with the help of ranking functions, [7], [8]. Let ψ be a ranking function that orders the datasets in a total order, i.e.,

Definition 1: Let A be a set. We say that $\psi : A^m \to [m]$ is a ranking function if it has the two properties below:

1) For all $a_1, \dots, a_m \in A$ and for all permutation τ: $\{2, \dots, m\} \to \{2, \dots, m\}$ we have that

$$\psi(a_1, \dots, a_m) = \psi(a_1, a_{\tau(2)}, \dots, a_{\tau(m)}). \tag{6}$$

2) If $a_i \neq a_j$, then $\psi(a_i, \{a_k\}_{k\neq i}) \neq \psi(a_j, \{a_k\}_{k\neq j})$.

Our main tool for quantifying the probability of type I error will be the following lemma, see [8, Lemma 1].

Lemma 1: Let $\xi_1, \dots, \xi_m$ be a.s. pairwise different exchangeable variables and let ψ be a ranking function. Then $\psi(\xi_1, \dots, \xi_m)$ has a uniform distribution on $[m]$.

A technical challenge is posed by identical datasets. We use a random permutation σ on $[m-1]_0 \doteq \{0, \dots, m-1\}$ independently generated from everything else to resolve this issue, i.e., let $\mathcal{D}_j^{\sigma} \doteq (\mathcal{D}_j, \sigma(j))$ for $j \in [m-1]_0$.

Theorem 1: Assume $A1 - A6$ and that θ_*, γ_* and ψ are given. If H_0 holds true then for any $1 \le r \le p \le m$:

$$\mathbb{P}\left(r \le \psi(\mathcal{D}_0^{\sigma}, \dots, \mathcal{D}_{m-1}^{\sigma}) \le p\right) = \frac{p - r + 1}{m}. \tag{7}$$

Proof: Notice that $\{\mathcal{D}_j^{\sigma}\}_{j=0}^{m-1}$ are almost surely pairwise different exchangeable variables under H_0, therefore Theorem 1 follows from Lemma 1. ∎

Observe that Theorem 1 is completely distribution-free and provides us finite sample guarantees for the type I error probabilities. In addition, the significance level of the proposed scheme is exact and user-chosen (rational). In the following section we construct ranking functions that ensure the consistency of the proposed test.

C. Dependence Measures

Dependence measures are used for assessing dependency between random variables E, N with joint distribution $Q_{E,N}$, or, in the empirical case, i.i.d. datasets $\mathcal{D}_0 = \{(E_i, N_i)\}_{i=1}^{n}$ generated from $Q_{E,N}$. A dependence measure, Δ, has two properties.

First, it needs to be *characteristic*, i.e., $\Delta(E, N) = 0$ if and only if E and N are independent. Second, it needs to exhibit a *consistent estimator* $\widehat{\Delta}$, that is (as the sample size n increases)

$$\widehat{\Delta}(\mathcal{D}_0) \doteq \widehat{\Delta}(\{(E_i, N_i)\}_{i=1}^n) \longrightarrow p\Delta(E, N). \tag{8}$$

We consider the following two dependence measures.

1) Hilbert–Schmidt Independence Criterion: Let $\mathcal{H}_k$ be a reproducing kernel Hilbert space (RKHS) and $k : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be its reproducing kernel. For a random variable E with $\mathbb{E}\sqrt{k(E, E)} < \infty$, the distribution Q_E of E can be embedded into $\mathcal{H}_k$ by $\mu(Q_E) \doteq \mathbb{E}[k(E, \cdot)]$, where the expectation is a Bochner integral; $\mu(Q_E)$ is called the *kernel mean embedding* of Q_E [9]. Kernel k is called *characteristic* if μ is injective.

In what follows, let k and ℓ be positive definite kernels. It is well-known that the (tensor) product kernel

$$\begin{aligned} &k \otimes \ell : (\mathbb{R} \times \mathbb{R})^2 \to \mathbb{R} \\ &((x_1, x_2), (y_1, y_2)) \mapsto k(x_1, x_2) \cdot \ell(y_1, y_2) \end{aligned} \tag{9}$$

is also positive definite. With this object, the (centered) cross–covariance operator is defined as

$$\begin{aligned} \mathbf{C}_{E,N} &\doteq \mathbb{E}[k \otimes \ell((E, N), (\cdot, \cdot))] - \mathbb{E}[k(E, \cdot)] \otimes \mathbb{E}[\ell(N, \cdot)] \\ &= \mu(Q_{E,N}) - \mu(Q_E \otimes Q_N). \end{aligned}$$

Note that one can recover the standard covariance by applying linear kernels, $k(x, y) = \ell(x, y) = x \cdot y$.

With this notation, we are now able to recall the definition of Hilbert–Schmidt independence criterion [2]:

$$\mathrm{HSIC}(Q_{E,N}, \mathcal{H}_k, \mathcal{H}_l) \doteq \left\|\mathbf{C}_{E,N}\right\|_{\otimes}^2, \tag{10}$$

where $\|\cdot\|_{\otimes}$ denotes the norm of the product RKHS. If k, ℓ and $Q_{E,N}$ are fixed, we may write $\mathrm{HSIC}(E, N) \doteq \mathrm{HSIC}(Q_{E,N}, \mathcal{H}_k, \mathcal{H}_l)$ and use a more intuitive form:

$$\begin{aligned} &\mathrm{HSIC}(E, N) = \mathbb{E}[k(E, E')\ell(N, N')] \\ &+ \mathbb{E}[k(E, E')]\mathbb{E}[\ell(N, N')] - 2\mathbb{E}[k(E, E')\ell(N, N'')], \end{aligned} \tag{11}$$

where (E, N), (E', N') and (E'', N'') are i.i.d. copies from $Q_{E,N}$. HSIC is characteristic if $k \otimes \ell$ is characteristic.

Formula (11) motivates the empirical estimate:

$$\begin{aligned} &\mathrm{HSIC}_n(\mathcal{D}_0) \doteq \frac{1}{n^2} \sum_{(i,j)\in[n]^2} k(E_i, E_j)\ell(N_i, N_j) \\ &+ \frac{1}{n^4} \sum_{(i,j,r,s)\in[n]^4} k(E_i, E_j)\ell(N_r, N_s) - \frac{1}{n^3} \sum_{(i,j,s)\in[n]^3} k(E_i, E_j)\ell(N_i, N_s). \end{aligned} \tag{12}$$

The consistency of HSIC_n is proved in [5].

2) Distance Covariance: Distance covariance was first introduced in [10, Definition 2] and the definition below is due to [4, Th. 8]. For E, N with finite expectations, the distance covariance is defined as

$$\begin{aligned} &\mathrm{dCov}^2(E, N) \doteq \mathbb{E}\left\|E - E'\right\|\left\|N - N'\right\| \\ &+ \mathbb{E}\left\|E - E'\right\|\mathbb{E}\left\|N - N'\right\| - 2\mathbb{E}\left\|E - E'\right\|\left\|N - N''\right\|, \end{aligned}$$

where (E, N), (E', N') and (E'', N'') are i.i.d. copies from $Q_{E,N}$. Distance covariance has some excellent properties in terms of measuring independence, most importantly, it is characteristic [10, Th. 3].

Algorithm 1 Independence Test for an I.I.D. Sample

Inputs: i.i.d. sample $\mathcal{D}_0$, desired significance level α, dependence measure estimator $\widehat{\Delta}$ tie breaking permutation σ on $[m-1]_0$

1: Choose integers $1 \leq r \leq m$ such that $\alpha = r/m$.
2: Generate $m-1$ random permutation $\{\pi_j\}_{j=1}^{m-1}$ uniformly from S_n.
3: Construct $m-1$ new alternative sample by $\mathcal{D}_j = \{(E_i, N_{\pi_j(i)})\}_{i=1}^n$ for $j \in [m-1]$ and let $\mathcal{D}_j^\sigma \doteq (\mathcal{D}_j, \sigma(j))$ for $j \in [m-1]_0$.
4: Calculate the dependence measure estimates $\widehat{\Delta}_n^{(j)} = |\widehat{\Delta}(\mathcal{D}_j)|$ for $j = 0, 1, \ldots, m-1$.
5: Compute the rank statistic:
$$\psi_\Delta(\mathcal{D}_0^\sigma, \ldots, \mathcal{D}_{m-1}^\sigma) = 1 + \sum_{j=1}^{m-1} \mathbb{I}\left(\widehat{\Delta}_n^{(0)} \prec_\sigma \widehat{\Delta}_n^{(j)}\right).$$
6: Reject H_0 if and only if
$$\psi_\Delta(\mathcal{D}_0^\sigma, \ldots, \mathcal{D}_{m-1}^\sigma) \leq r.$$

The doubly centered distances are $A_{j,k} = a_{j,k} - a_{j\cdot} - a_{\cdot k} + a_{\cdot\cdot}$ where $a_{j,k} = \left\|E_j - E_k\right\|$, $a_{j\cdot} = \sum_{k=1}^n a_{j,k}/n$, $a_{\cdot k} = \sum_{j=1}^n a_{j,k}/n$ and $a_{\cdot\cdot} = \sum_{j,k=1}^n a_{j,k}/(n^2)$; and let B be analogously for $\{N_i\}_{i=1}^n$. Then, the empirical distance covariance [10, Definition 4] can be computed as:

$$\mathrm{dCov}_n^2(\mathcal{D}_0) \doteq \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n A_{j,k}B_{j,k}. \tag{13}$$

This empirical estimate is consistent [10, Th. 2].

D. Hypothesis Test for Independence

Our hypothesis test compares empirical dependence measure estimates via a ranking function ψ. To resolve (the unlikely event of) ties during comparison, we amend the order function $\prec_\sigma$; for technical details, see [11]. Let

$$\begin{aligned} &\widehat{\Delta}_n^{(j)} = |\widehat{\Delta}(\mathcal{D}_j)| \quad \text{for } j = 0, 1, \ldots, m-1 \ \text{ and} \\ &\psi_\Delta(\mathcal{D}_0^\sigma, \ldots, \mathcal{D}_{m-1}^\sigma) \doteq 1 + \sum_{j=1}^{m-1} \mathbb{I}\left(\widehat{\Delta}_n^{(0)} \prec_\sigma \widehat{\Delta}_n^{(j)}\right). \end{aligned} \tag{14}$$

We choose an integer r for significance level α in such a way that $\alpha = r/m$ and reject H_0 if and only if $\psi_\Delta(\mathcal{D}_0^\sigma, \ldots, \mathcal{D}_{m-1}^\sigma) \leq r$. If H_0 holds then $\{\mathcal{D}_j^\sigma\}_{j=0}^{m-1}$ are exchangeable, hence the rank statistic is uniform. If H_0 does not hold, then $\Delta(E, N) \neq 0$ and by (8), variable $\widehat{\Delta}(\mathcal{D}_0)$ tends to a positive number. However, the perturbed samples $\{\mathcal{D}_j^\sigma\}_{j=1}^{m-1}$ are almost i.i.d., because the pairs are shuffled, thus one expects that $\widehat{\Delta}(\mathcal{D}_j)$ tends to 0 for $j \in [m-1]$. In conclusion asymptotically $\widehat{\Delta}(\mathcal{D}_0)$ dominates $\widehat{\Delta}(\mathcal{D}_j)$ for every $j \in [m-1]$. The hypothesis test is summarised in Algorithm 1. Our exact coverage result below is a direct consequence of Theorem 1.

Theorem 2: Assume $A1-A6$. Let $\widehat{\Delta}$ be a dependence measure estimator, ψ_Δ be the corresponding ranking function and

$r \le m$ be integers. If H_0 holds, then

$$\mathbb{P}(\psi_\Delta(\mathcal{D}_0^\sigma, \dots, \mathcal{D}_{m-1}^\sigma) \le r) = \frac{r}{m}. \tag{15}$$

E. Strong Consistency

In this section we give asymptotic bounds for the type II error probability. Assume:

A7 A characteristic dependence measure Δ and a consistent estimator is given such that for $j \in [m-1]$

$$\widehat{\Delta}_n^{(0)} \longrightarrow p|\Delta(E,N)| \quad \text{and} \quad \widehat{\Delta}_n^{(j)} \longrightarrow p0.$$

Assumption $A7$ ensures that the empirical estimates of the used dependence measure vanish for the permuted samples. This assumption is satisfied, for example, by the HSIC-based ranking, see [6, Lemma 1].

Theorem 3: Assume that $A1 - A7$ hold and θ_*, γ_* are given. If H_1 holds, then for any $r \ge 1$:

$$\mathbb{P}(\psi_\Delta(\mathcal{D}_0^\sigma, \dots, \mathcal{D}_{m-1}^\sigma) \le r) \xrightarrow{n\to\infty} 1. \tag{16}$$

Proof: One can bound the probability in (16) as

$$\begin{aligned}
&\mathbb{P}(\psi_\Delta \le r) \ge \mathbb{P}(\psi_\Delta \le 1) \ge \mathbb{P}(\forall j \in [m-1] : \widehat{\Delta}_n^{(0)} > \widehat{\Delta}_n^{(j)}) \\
&= 1 - \mathbb{P}(\exists j \in [m-1] : \widehat{\Delta}_n^{(0)} \le \widehat{\Delta}_n^{(j)}) \\
&\ge 1 - \sum_{j=1}^{m-1} \mathbb{P}(\widehat{\Delta}_n^{(0)} \le \widehat{\Delta}_n^{(j)}) \ge 1 - m \cdot \mathbb{P}(\widehat{\Delta}_n^{(0)} \le \widehat{\Delta}_n^{(1)}),
\end{aligned}$$

where $\mathbb{P}\big(\widehat{\Delta}_n^{(0)} \le \widehat{\Delta}_n^{(1)}\big)$ goes to zero, because of $A7$. ■

By Theorem 3 the probability of type II error tends to zero as the sample size goes to infinity, assuming that the applied dependence measure is characteristic.

IV. Robust Independence Tests

We now turn our attention to the general problem when the true parameters are unknown constants. In this case, the exact noise terms cannot be computed, but only estimated. Assume we can build non-asymptotically guaranteed confidence sets for the true parameters. The idea is then to use a two-step algorithm: first, we construct these (distribution-free) confidence regions, and then apply a parameter-dependent version of the above hypothesis test on each parameter in the confidence set.

A. Parameter-Dependent Hypothesis Test

We present a meta-algorithm that can work with any confidence region construction for θ_* and γ_*. We assume that we have confidence sets $\widehat{\Theta}_n$ and $\widehat{\Gamma}_n$ such that

$$\mathbb{P}(\theta_* \in \widehat{\Theta}_n) \ge 1-\beta \quad \text{and} \quad \mathbb{P}(\gamma_* \in \widehat{\Gamma}_n) \ge 1-\beta \tag{17}$$

hold for all sample size $n \in \mathbb{N}$ and for some significance level $\beta \in (0,1)$. For simplicity, we will omit n from the notation. By $A3$ we can obtain $E_t(\theta)$ and $N_t(\theta)$ for any parameter-pair candidate $(\theta,\gamma) \in \widehat{\Theta} \times \widehat{\Gamma}$ by

$$\begin{aligned}
E_t(\theta) &\doteq H_1^{-1}(q^{-1},\theta)(Y_t - G_1(q^{-1},\theta)U_t), \\
N_t(\gamma) &\doteq H_2^{-1}(q^{-1},\gamma)(Z_t - G_2(q^{-1},\gamma)V_t),
\end{aligned} \tag{18}$$

for $t = 1, \dots, n$. These quantities can be perturbed as before to construct parameterized alternative datasets

$$\mathcal{D}_j(\theta,\gamma) = \{(E_i(\theta), N_{\pi_j(i)}(\gamma))\}_{i=1}^n \tag{19}$$

for $j = 1, \dots, m-1$ and extended by σ as before to create $\mathcal{D}_j^\sigma(\theta,\gamma)$ for $j = 0, \dots, m-1$. Finally, for any ranking function ψ one can define parameter-dependent ranks as

$$\psi(\theta,\gamma) \doteq \psi(\mathcal{D}_0^\sigma(\theta,\gamma), \dots, \mathcal{D}_{m-1}^\sigma(\theta,\gamma)). \tag{20}$$

Theorem 4: Assume that $A1 - A6$ hold. Let ψ be any ranking function, $\widehat{\Theta}$ and $\widehat{\Gamma}$ conservative confidence sets with significance level at most β. If H_0 holds true then

$$\mathbb{P}(\max_{(\theta,\gamma)\in\widehat{\Theta}\times\widehat{\Gamma}} \psi(\theta,\gamma) \le r) \le \frac{r}{m} + 2\beta. \tag{21}$$

Proof: Using the union bound and (17), one can show that $\widehat{\Theta} \times \widehat{\Gamma}$ is a confidence region for (θ_*, γ_*), i.e.,

$$\mathbb{P}((\theta_*,\gamma_*) \notin \widehat{\Theta} \times \widehat{\Gamma}) \le \mathbb{P}(\theta_* \notin \widehat{\Theta}) + \mathbb{P}(\gamma_* \notin \widehat{\Gamma}) \le 2\beta.$$

Then, under H_0, we have

$$\begin{aligned}
&\mathbb{P}\left(\max_{\theta\in\widehat{\Theta},\gamma\in\widehat{\Gamma}} \psi(\theta,\gamma) \le r\right) \\
&= \mathbb{P}\left(\{(\theta_*,\gamma_*) \notin \widehat{\Theta} \times \widehat{\Gamma}\} \cap \{\max_{\theta\in\widehat{\Theta},\gamma\in\widehat{\Gamma}} \psi(\theta,\gamma) \le r\}\right) \\
&\quad + \mathbb{P}\left(\{(\theta_*,\gamma_*) \in \widehat{\Theta} \times \widehat{\Gamma}\} \cap \{\max_{\theta\in\widehat{\Theta},\gamma\in\widehat{\Gamma}} \psi(\theta,\gamma) \le r\}\right) \\
&\le \mathbb{P}((\theta_*,\gamma_*) \notin \widehat{\Theta} \times \widehat{\Gamma}) + \mathbb{P}(\psi(\theta_*,\gamma_*) \le r),
\end{aligned}$$

where the first term is less than 2β because of (21) and the second term is r/m because of Theorem 1. ■

B. Dependence Measure Ranking

Ranking functions can be defined similarly to the i.i.d. case via dependence measures. The idea is to compute dependence measure estimates w.r.t. plausible parameters. Let Δ be some characteristic dependence measure and $\widehat{\Delta}$ be its (consistent) estimator as above. Let us define the dependence measure estimate functions as

$$\widehat{\Delta}_n^{(j)}(\theta,\gamma) \doteq |\widehat{\Delta}(\mathcal{D}_j(\theta,\gamma))|, \tag{22}$$

for $j = 0, 1, \dots, m-1$ and the ranking function as

$$\psi_\Delta(\theta,\gamma) \doteq 1 + \sum_{j=1}^{m-1} \mathbb{I}(\widehat{\Delta}_n^{(0)}(\theta,\gamma) \prec_\sigma \widehat{\Delta}_n^{(j)}(\theta,\gamma)). \tag{23}$$

If H_0 does not hold, then $\widehat{\Delta}_n^{(0)}(\theta,\gamma)$ tends to be the greatest around (θ_*,γ_*), henceforth, we reject the null hypothesis if $\psi_\Delta(\theta,\gamma)$ is at most some user-chosen rank value r on $\widehat{\Theta} \times \widehat{\Gamma}$. The step–by–step method is presented in the pseudocode of Algorithm 2. Note that the test exhibits finite sample guarantees for the significance level, as it is showed by Theorem 4.

C. Strong Consistency of the Robust Test

In this section we quantify the asymptotic behaviour of rejection probability when H_1 is true, thus let us consider the case when E and N are dependent. We prove that the power of the suggested test tends to 1 as the sample size goes to infinity. Let $B(\theta,\varepsilon)$ denote the Euclidean ball around θ with radius ε. We assume that:

Algorithm 2 Independence Test for Synchronous Systems

Inputs: observations $\{U_t\}$, $\{Y_t\}$, $\{V_t\}$ and $\{Z_t\}$,
transfer functions G_1, G_2, H_1 and H_2 parameterized by $\theta \in \Theta$ and $\gamma \in \Gamma$,
user-chosen significance level $\alpha \in (0, 1)$,
confidence sets $\widehat{\Theta}$ and $\widehat{\Gamma}$ for θ_* and γ_* respectively with confidence level at least $1 - \beta$,
dependence measure estimator $\widehat{\Delta}$,
tie breaking permutation σ on $[m-1]_0$

1: Choose integers $1 \le r \le m$ such that $r/m \le \alpha - 2\beta$.
2: Generate $m-1$ random permutation $\{\pi_j\}_{j=1}^{m-1}$ uniformly from S_n.
3: Construct noise term functions for $t = 1, \dots, n$ by
$E_t(\theta) \doteq H_1^{-1}(q^{-1}, \theta)\big(Y_t - G_1(q^{-1}, \theta)U_t\big)$,
$N_t(\gamma) \doteq H_2^{-1}(q^{-1}, \gamma)\big(Z_t - G_2(q^{-1}, \gamma)V_t\big)$.
4: Build $m-1$ new alternative sample functions by
$\mathcal{D}_j(\theta, \gamma) = \{(E_i(\theta), N_{\pi_j(i)}(\gamma))\}_{i=1}^n$ for $j \in [m-1]$
and let $\mathcal{D}_j^\sigma(\theta, \gamma) = (\mathcal{D}_j(\theta, \gamma), \sigma(j))$ for $j \in [m-1]_0$.
5: Formulate the dependence measure estimates
$\widehat{\Delta}_n^{(j)}(\theta, \gamma) = |\widehat{\Delta}(\mathcal{D}_j(\theta, \gamma))|$ for $j = 0, 1, \dots, m-1$.
6: Construct the parameter-dependent ranking function
$$\psi_\Delta(\theta, \gamma) \doteq \psi_\Delta\big(\mathcal{D}_0^\sigma(\theta, \gamma), \dots, \mathcal{D}_{m-1}^\sigma(\theta, \gamma)\big) = 1 + \sum_{j=1}^{m-1} \mathbb{I}\left(\widehat{\Delta}_n^{(0)}(\theta, \gamma) \prec_\sigma \widehat{\Delta}_n^{(j)}(\theta, \gamma)\right).$$
7: Reject the null hypothesis if and only if
$$\max_{\theta \in \widehat{\Theta}, \gamma \in \widehat{\Gamma}} \psi_\Delta(\theta, \gamma) \le r.$$

A8 Control inputs $\{U_t\}$, $\{V_t\}$ and driving noises $\{E_t\}$, $\{N_t\}$ are a.s. included in a Césaro space for $p = \infty$, i.e., for $\{W_t\} \in \{\{U_t\}, \{V_t\}, \{E_t\}, \{N_t\}\}$ we have
$$\|W\|_{c(\infty)} \doteq \sup_{n \in \mathbb{N}} \frac{1}{n} \sum_{t=1}^n |W_t| < \infty.$$
A9 There a.s. exist $K, \tilde{\varepsilon} > 0$ such that for $\theta \in B(\theta_*, \tilde{\varepsilon})$:
$$\|E(\theta_*) - E(\theta)\|_{c(\infty)} \le K \cdot \|\theta_* - \theta\|,$$
and respectively for $N(\gamma)$, where $\gamma \in B(\gamma^*, \tilde{\varepsilon})$.
A10 The confidence sets are uniformly consistent, i.e., for all $\varepsilon > 0$ there a.s. exists an $N_0 \in \mathbb{N}$ such that for all $n > N_0$ both $\widehat{\Theta}_n \subseteq B(\theta_*, \varepsilon)$ and $\widehat{\Gamma}_n \subseteq B(\gamma_*, \varepsilon)$.
A11 Dependence measure estimator $\widehat{\Delta}$ is Lipschitz continuous around (θ_*, γ_*), i.e., $\exists C, \tilde{\tilde{\varepsilon}} > 0$ such that
$$|\widehat{\Delta}(\mathcal{D}_j(\theta_*, \gamma_*)) - \widehat{\Delta}(\mathcal{D}_j(\theta, \gamma))| \le C \cdot (\|E(\theta_*) - E(\theta)\|_{c(\infty)} + \|N(\gamma_*) - N(\gamma)\|_{c(\infty)})$$
for $\theta \in B(\theta_*, \tilde{\tilde{\varepsilon}})$, $\gamma \in B(\gamma_*, \tilde{\tilde{\varepsilon}})$ and $j = 0, \dots, m-1$.

Theorem 5: Assume $A1 - A11$. If H_1 holds true, then
$$\mathbb{P}\left(\max_{(\theta,\gamma) \in \widehat{\Theta}_n \times \widehat{\Gamma}_n} \psi_\Delta(\theta, \gamma) \le r\right) \xrightarrow{n \to \infty} 1. \tag{24}$$

Proof: We use a characteristic Δ, thus under H_1 we have $\widehat{\Delta}_n^{(0)}(\theta_*, \gamma_*) \xrightarrow{p} \kappa \doteq |\Delta(E, N)| > 0$. In addition, $\widehat{\Delta}_n^{(j)}(\theta_*, \gamma_*) \xrightarrow{p} 0$ for $j \in [m-1]$ because of $A7$. Let us fix a positive ε that is smaller than $\tilde{\varepsilon}$ and $\tilde{\tilde{\varepsilon}}$. By $A10$, there exists almost surely an $N_0 \in \mathbb{N}$ such that $\widehat{\Theta}_n \in B(\theta_*, \varepsilon)$ and $\widehat{\Gamma}_n \in B(\gamma_*, \varepsilon)$ for all $n > N_0$. We prove that for n large enough the rank values equal to 1 uniformly on $B(\theta_*, \varepsilon) \times B(\gamma_*, \varepsilon)$ with large probability. We know that $\widehat{\Delta}_n^{(j)}(\theta_*, \gamma_*)$ is closer than ε to κ with large probability if n is large enough. Then, condition $A9$ and $A10$ yield
$$|\widehat{\Delta}_n^{(0)}(\theta_*, \gamma_*) - \widehat{\Delta}_n^{(0)}(\theta, \gamma)| \le C \cdot (\|E(\theta_*) - E(\theta)\|_{c(\infty)} + \|N(\gamma_*) - N(\gamma)\|_{c(\infty)}) \le C \cdot K \cdot (\|\theta_* - \theta\| + \|\gamma_* - \gamma\|),$$
which proves that there exists $C_1 > 0$ such that a.s.
$$\sup_{\theta \in B(\theta_*, \varepsilon), \gamma \in B(\gamma_*, \varepsilon)} |\widehat{\Delta}_n^{(0)}(\theta_*, \gamma_*) - \widehat{\Delta}_n^{(0)}(\theta, \gamma)| \le C_1 \cdot \varepsilon.$$
Thus, the infimum of $\widehat{\Delta}_n^{(0)}(\theta, \gamma)$ on confidence set $\widehat{\Theta}_n \times \widehat{\Gamma}_n$ tends to κ in probability. Similarly one can prove that $\sup \widehat{\Delta}_n^{(j)}(\theta, \gamma)$ on $\widehat{\Theta}_n \times \widehat{\Gamma}_n$ tends to 0 for $j = 1, \dots, m-1$, implying that the probability of rejection goes to 1. ■

V. Illustrative Example: AR(1) Systems

We considered two AR(1) systems: $Y_t = \alpha_* Y_{t-1} + E_t$, and $Z_t = \beta_* Z_{t-1} + N_t$ with $\alpha_* = 0.5$ and $\beta_* = 0.3$. For parameters (α, β) the residuals $\{E_t(\alpha)\}$ and $\{N_t(\beta)\}$ can be computed from $\{Y_t\}$, $\{Z_t\}$ as $E_t(\alpha) = Y_t - \alpha Y_{t-1}$ and similarly for $\{N_t(\beta)\}$. For AR(1) systems, the required Lipschitz continuity can be easily satisfied. If $|\alpha| < 1$, then under $A8$ one can prove $\|Y\|_{c(\infty)} < \infty$ and $A9$. For HSIC estimates, presented in (12), $A7$ holds and if $k \otimes \ell$ is Lipschitz and characteristic, then $A11$ is also satisfied.

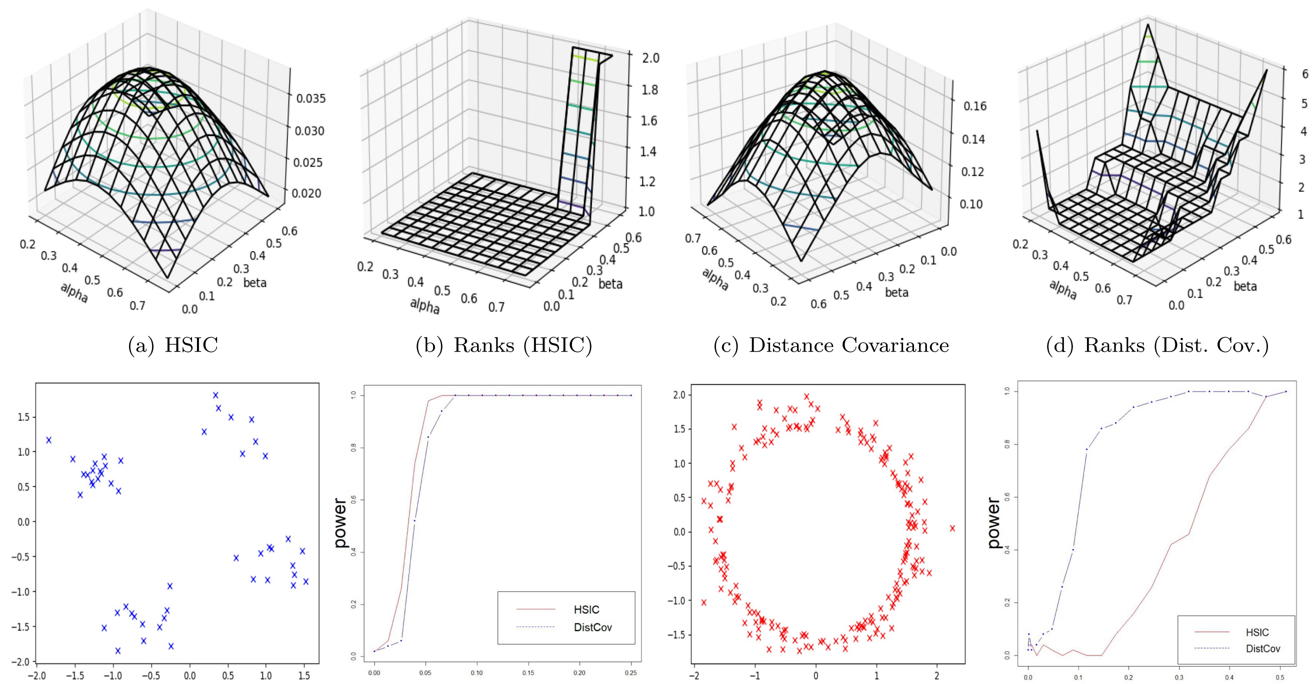
A. Sign-Perturbed Sums

For ARX systems, confidence sets satisfying $A10$ can be constructed, e.g., by the Sign-Perturbed Sums (SPS) method [11]. Standard SPS assumes independent and symmetric noises. Under the i.i.d. assumption, symmetry is no more required [12]. An instrumental variable-based extension of SPS was proposed in [13], which is uniformly consistent, even for ARX systems with feedback control.

B. Numerical Simulations

We simulated two AR(1) systems with nonlinearly dependent i.i.d. innovations $\{(E_t, N_t)\}_{i=1}^n$. The processes were initialized in zero ($A4$). First, a rotated mixture of Gaussian distributions [2] was considered for innovations, see Figure 1(e). We generated a sample with $n = 50$ elements from a zero mean two-dimensional Gaussian distribution with covariance matrix $1/4 \cdot \mathbb{I}$, then we shifted each data-point with a pair of random signs and rotated the obtained points around the origin with an angle of 0.1 (radian). We used SPS to construct confidence sets for α_* and β_* with significance level $1/80$ and tested independence with $m = 40$ datasets. We maximized the ranking function on a fine grid of the confidence region.

Reference functions $\widehat{\Delta}_n^{(0)}$ and ranking functions ψ_Δ are plotted on Figure 1(a), 1(c), 1(b) and 1(d) for HSIC



(a) HSIC (b) Ranks (HSIC) (c) Distance Covariance (d) Ranks (Dist. Cov.)

(e) Rotated Gaussian Mixture (f) Power Functions (g) Extinct Gaussian (h) Power Functions

Fig. 1. Reference functions, ranking functions, innovation data and power functions for AR(1) systems.

and distance covariance. At significance level 0.15 based on Figure 1(b) we reject the null hypothesis, but we accept H_0 at this level based on Figure 1(d), because ψ_Δ exceeds 5 at some points. The (estimated) power functions, i.e., the rejection probabilities, are plotted for $n = 200$ and significance level 0.15 on Figure 1(f) w.r.t. the rotation angle, which served as a factor inducing dependence.

Second, a sample from an extinct multivariate Gaussian distribution with covariance $1/4 \cdot \mathbb{I}$ was used to generate innovations, see Figure 1(g). That is, we introduced dependence between E and N by throwing away the pairs that lied in a circle around the origin with radius r. Figure 1(h) shows the (estimated) power functions w.r.t. the distinction rate increasing with r for $n = 500$ using HSIC and distance covariance with significance level 0.15.

VI. Conclusion

In this letter we introduced hypothesis tests for the independence of synchronous general linear systems with non-asymptotically guaranteed significance levels. The main idea was to apply permutation tests over a confidence region for the system parameters. We combined these ideas with characteristic dependence measures to detect any nonlinear dependence between the innovations of the systems. We proved consistency under general assumptions and demonstrated the method on AR(1) systems.

References

[1] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Hoboken, NJ, USA: Prentice-Hall, 1999.

[2] A. Gretton, K. Fukumizu, C. Teo, L. Song, B. Schölkopf, and A. Smola, “A kernel statistical test of independence,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 585–592.

[3] K. Chwialkowski and A. Gretton, “A kernel independence test for random processes,” in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 1422–1430.

[4] G. J. Székely and M. L. Rizzo, “Brownian distance covariance,” *Ann. Appl. Statist.*, vol. 3, no. 4, pp. 1235–1265, 2009.

[5] N. Pfister, P. Bühlmann, B. Schölkopf, and J. Peters, “Kernel-based tests for joint independence,” *J. Roy. Statist. Soc. Ser. B*, vol. 80, no. 1, pp. 5–31, 2018.

[6] D. Rindt, D. Sejdinovic, and D. Steinsaltz, “Consistency of permutation tests of independence using distance covariance, HSIC and dHSIC,” *Stat*, vol. 10, no. 1, 2021, Art. no. e364.

[7] V. Vovk, A. Gammerman, and G. Shafer, *Algorithmic Learning in a Random World*. New York, NY, USA: Springer, 2005.

[8] B. Cs. Csáji and A. Tamás, “Semi-parametric uncertainty bounds for binary classification,” in *Proc. 58th IEEE Conf. Decis. Control (CDC)*, 2019, pp. 4427–4432.

[9] A. Smola, A. Gretton, L. Song, and B. Schölkopf, “A Hilbert space embedding for distributions,” in *Proc. Int. Conf. Algor. Learn. Theory*, 2007, pp. 13–31.

[10] G. J. Székely, M. L. Rizzo, and N. K. Bakirov, “Measuring and testing dependence by correlation of distances,” *Ann. Statist.*, vol. 35, no. 6, pp. 2769–2794, 2007.

[11] B. Cs. Csáji, M. C. Campi, and E. Weyer, “Sign-Perturbed Sums: A new system identification approach for constructing exact non–asymptotic confidence regions in linear regression models,” *IEEE Trans. Signal Process.*, vol. 63, no. 1, pp. 169–181, Jan. 2015.

[12] S. Kolumbán, I. Vajk, and J. Schoukens, “Perturbed datasets methods for hypothesis testing and structure of corresponding confidence sets,” *Automatica*, vol. 51, pp. 326–331, Jan. 2015.

[13] V. Volpe, B. C. Csáji, A. Care, E. Weyer, and M. C. Campi, “Sign-Perturbed Sums (SPS) with instrumental variables for the identification of ARX systems,” in *Proc. 54th IEEE Conf. Decis. Control (CDC)*, 2015, pp. 2115–2120.