MORE EFFICIENT EXACT PERMUTATION TESTS

By Nick W. Koning[1], Jesse Hemerik[2]

[1] *Econometric Institute, Erasmus University Rotterdam, n.w.koning@ese.eur.nl*

[2] *Biometris, Wageningen University & Research, jesse.hemerik@wur.nl*

Abstract

Non-parametric tests based on permutation, rotation or sign-flipping are examples of so-called group-invariance tests. These tests rely on invariance of the null distribution under a set of transformations that has a group structure, in the algebraic sense. Such groups are often huge, which makes it computationally infeasible to use the entire group. Hence, it is standard practice to test using a randomly sampled set of transformations from the group. This random sample still needs to be substantial to obtain good power and replicability. We improve upon the standard practice by using a well-designed subgroup of transformations instead of a random sample. We show this can yield a more powerful and fully replicable test with the same number of transformations. For a normal location model and a particular design of the subgroup, we show that the power improvement is equivalent to the power difference between a Monte Carlo Z-test and Monte Carlo t-test. In our simulations, we find that our test has the same power as a test based on sampling that uses double the number of random transformations. These benefits come entirely 'for free', as our methodology relies on an assumption of invariance under the subgroup, which is implied by invariance under the entire group.

1. Introduction. Permutation tests, randomization tests and related testing procedures are ubiquitous in modern-day statistical research (Onghena, 2018), for example in genomics (Tusher, Tibshirani and Chu, 2001; Li and Tibshirani, 2013; Debeer and Strobl, 2020), neuroimaging (Eklund, Nichols and Knutsson, 2016) and economics (Young, 2019). Such non-parametric (or semi-parametric) tests are useful in part because they require few assumptions on the data distribution (Anderson and Robinson, 2001; Hemerik, Thoresen and Finos, 2021). Additionally, they have seen recent popularity in the simultaneous testing of many hypotheses, as they are often able to take into account the dependence structure of the data in an exact way, leading to relatively high power (Westfall and Young, 1993; Tusher, Tibshirani and Chu, 2001; Meinshausen, 2006; Pesarin and Salmaso, 2010; Meinshausen et al., 2011; Hemerik and Goeman, 2018a; Blanchard, Neuvial and Roquain, 2020). For example, under strong positive dependence in the data, Bonferroni's multiple testing correction is very conservative and is greatly improved by a permutation method (Westfall and Young, 1993; Westfall and Troendle, 2008).

These non-parametric tests often rely on an invariance assumption under the null hypothesis, based on a set $\mathcal{G}$ of transformations that is a *group*, in the algebraic sense (Lehmann and Romano, 2005; Hemerik and Goeman, 2018b). That is, every element in $\mathcal{G}$ has an inverse in $\mathcal{G}$ and $\mathcal{G}$ is closed under composition. We will refer to tests based on a group-invariance assumption as *group-invariance tests*. A prominent example is a permutation test, but tests based other groups of transformations, e.g. rotation or sign-flipping, are also used (Langsrud, 2005; Solari, Finos and Goeman, 2014; Hemerik, Goeman and Finos, 2020). The requirement

*February 21, 2022

MSC2020 subject classifications: 62G10, 62G09.

Keywords and phrases: Group-invariance test, subgroup, permutation test, randomization test, resampling.

that $\mathcal{G}$ is a group is fundamental; using a set of transformations that is not a group can lead to a very conservative or anti-conservative test (Southworth, Kim and Owen, 2009; Hemerik and Goeman, 2018b, 2021).

1.1. *Current practice.* For moderate or large sample sizes, the cardinality $|\mathcal{G}|$ is typically huge, so that it is often computationally infeasible to use the whole group. For example, the order of the permutation and sign-flipping groups is $n!$ and 2^n, respectively, where n is the number of observations. As a solution, it is universal practice among researchers to use a set of random transformations (Eden and Yates, 1933; Dwass, 1957; Phipson and Smyth, 2010). This can be done in such a way that the test is still exact (Hemerik and Goeman, 2018b). We will henceforth refer to such a test as a Monte Carlo group-invariance test.

Using a small number of random transformations results in power loss compared to using the full group of transformations. Moreover, it leads to reduced replicability, since the test depends on the random transformations that happen to be sampled. For a Monte Carlo group-invariance test to have good power and to obtain replicable results, it is therefore important that a large number of random transformations is used. Typically, this number is several times larger than α^{-1}, where α is the nominal level of the test. For example, for a nominal level $\alpha = 0.05$, it is common to use 100-5000 random transformations. Unfortunately, using a large number of random transformations can remain prohibitive, as tests and multiple testing methods based on permutations (or other transformations) can be highly computationally intensive (Gao et al., 2010; Kofler and Schlötterer, 2012; Hemerik, Solari and Goeman, 2019; Vesely, Finos and Goeman, 2021).

A few solutions have been proposed to improve the performance for a fixed number of transformations. For example, Good (2005) approximates the permutation reference distribution using moment matching. Furthermore, Winkler et al. (2016) review and propose additional methods for obtaining high-resolution p-values based on a limited number of random permutations. However, although the resolution of permutation p-values is improved, these approaches are not exact. Moreover, the p-values will still depend on the particular random sample of permutations that has been drawn, which also reduces the replicability of the results. Further, it is not clear how to generally combine the methods in e.g. Winkler et al. (2016) with permutation-based multiple testing methods, which are often not p-value based. Thus, most of the drawbacks of the use of a limited set of random transformations have remained unresolved.

1.2. *Main contribution.* In this paper, we propose a novel approach to improve the power of a group-invariance test: we replace the set of random transformations with a carefully chosen (deterministic) *subgroup* of the full group of transformations with respect to which invariance is tested.

Throughout the paper, we illustrate this idea by focusing on a generalized location-shift model, which also contains the important two-sample comparison of means as a special case. As the group $\mathcal{G}$ we consider (subgroups of) the orthogonal group. This group contains the rotation group, the permutation group and the sign-flipping group as subgroups, and can be conveniently represented as a collection of orthonormal matrices.

Intuitively speaking, we find that it is desirable to select a subgroup with the property that all its non-identity elements are 'sufficiently different' from the identity element. We show that such a subgroup can yield a 'subgroup-invariance test' with more power than a Monte Carlo group-invariance test based on the full group. More precisely, we can find subgroups of size M, say, which lead to a test with higher power than using M random transformations uniformly sampled from the full group. Equivalently, we can obtain the same power as a test with M random transformations by using a subgroup with a cardinality smaller than

M (roughly $M/2$ in our simulation experiments). Thus, our method is an effective way to improve power or reduce the computation time; these are two sides of the same coin.

A crucial observation about subgroup-invariance tests is that they are exact without the need of any additional assumptions. In fact, the assumptions of the test are weakened to invariance under the subgroup. An additional benefit is that our approach eliminates the reduced replicability that comes from random transformations, as the subgroup can be chosen deterministically.

1.3. *Overview of the paper.* In Section 2, we describe the the group-invariance test in the generalized location-shift model. We present an overview of some properties of group-invariance tests. In particular, we compare the group-invariance test to a hypothetical test that derives its critical value from information that is not available under the alternative. We relate the difference between the tests to a 'leak' from signal to noise, observed by Dobriban (2021). We show that if this leak is small, the group-invariance test performs similar to the hypothetical test, using appropriate definitions of 'similar' and 'small'. Furthermore, we illustrate the difference between the tests under orthogonal invariance by establishing a link between group-invariance tests and classical parametric tests. In particular, we include a result that shows group-invariance tests can be viewed as a generalization of the t-test, which turns out to be a special case.

In Section 3, we proceed to explore 'oracle' subgroups that eliminate the leak so that the resulting subgroup-invariance test coincides with the hypothetical test. We find that, under normality, these oracle subgroups yield a test that has the same size and power properties as a Monte Carlo Z-test. We establish that a Monte Carlo orthogonal group-invariance test has the same size and power properties as a Monte Carlo t-test. We use this to conclude that oracle subgroup-invariance tests are more powerful than Monte Carlo orthogonal group-invariance tests in the normal location model. In addition, we explore the structure and existence of oracle subgroups.

As oracle subgroups can be limited in their existence, we also consider 'near-oracle' subgroups with a small leak in Section 4, for which we show the resulting test is close to the hypothetical test. We discuss the relationship between finding 'near-oracle' subgroups and group code problems (Slepian, 1968; Conway and Sloane, 1998), which relate to the spreading of points on a unit hypersphere.

As an example, we apply our methodology to the sign-flipping group in Section 5, and show how to construct oracle subgroups and 'near-oracle' subgroups. Such subgroups are conveniently available for direct used in our R-package NOSdata (Koning, 2022a). We continue with simulation experiments in Section 6 to study the difference in power between subgroup-invariance tests and Monte Carlo group-invariance tests in the case of sign-flipping. We find that our subgroup-invariance test substantially outperforms the competing Monte Carlo group-invariance tests.

Finally, we apply the methodology to fMRI data, to illustrate that our methodology can be combined with permutation-based (multiple) testing procedures.

2. Group-invariance tests.

In this section, we first describe the group-invariance test in a 'generalized location-shift' model. We then treat properties of group-invariance tests, some of which are known, to lay down the foundations for subsequent sections.

2.1. *Generalized location-shift model.* We consider a realization of the random n-vector $\boldsymbol{x} \neq \mathbf{0}$, $n \in \mathbb{N}_+$, which can be decomposed as

$$\boldsymbol{x} = \boldsymbol{\iota}\mu + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\iota}$ is some unit n-vector, $\mu \geq 0$ is the location parameter and $\boldsymbol{\varepsilon}$ is a random n-vector. We are interested in testing the hypothesis $H_0 : \mu = 0$ against $H_1 : \mu > 0$ at some level $\alpha \in (0, 1)$. In case we consider standard permutations and the two-sample model, the parameter μ is the difference between the means, see Appendix A.

As a test statistic, we consider

$$\boldsymbol{\iota}'\boldsymbol{x} = \mu + \boldsymbol{\iota}'\boldsymbol{\varepsilon},$$

where $\boldsymbol{\iota}$ is a unit n-vector. For example, if $\boldsymbol{\iota} = (n^{-1/2}, n^{-1/2}, \dots, n^{-1/2})'$ then $\boldsymbol{\iota}'\boldsymbol{x}$ is the $\sqrt{n}$-scaled sample mean of $\boldsymbol{x}$. Or, if we replace the second half of $\boldsymbol{\iota}$ with $-n^{1/2}$ then $\boldsymbol{\iota}'\boldsymbol{x}$ is the $\sqrt{n}$-scaled sample mean difference between the first and second half of $\boldsymbol{x}$, which is a test statistic often used for two-group comparison tests (see Appendix A for more details).

2.2. *Group-Invariance.* We consider a setting in which the distribution of $\boldsymbol{\varepsilon}$ is invariant under a group of transformations. We restrict ourselves to groups $\mathcal{G}$ that can be represented as a collection of $n \times n$ orthogonal matrices. Based on such a group, we make the following group-invariance assumption.

ASSUMPTION 1. *For all $\boldsymbol{G} \in \mathcal{G}$, we have $\boldsymbol{\varepsilon} \overset{\mathrm{d}}{=} \boldsymbol{G}\boldsymbol{\varepsilon}$.*

Note that under H_0 this implies $\boldsymbol{x} \overset{\mathrm{d}}{=} \boldsymbol{G}\boldsymbol{x}$, for all $\boldsymbol{G} \in \mathcal{G}$.

EXAMPLE 1. *An important example of a group is the collection of permutation matrices $\mathcal{P}$, for which the assumption is sometimes referred to as 'exchangeability'. Other examples are the collection of 'sign-flipping' matrices $\mathcal{R}$, which are diagonal matrices with diagonal elements in $\{-1, 1\}$, and the collection of all orthonormal matrices $\mathcal{H}$, for which the invariance assumption is sometimes referred to as 'sphericity'.*

Let $\mathcal{O}$ be the partitioning of the sample space induced by the group $\mathcal{G}$ into sets ('orbits') $O_{\boldsymbol{a}}^{\mathcal{G}} := \{\boldsymbol{G}\boldsymbol{a} \mid \boldsymbol{G} \in \mathcal{G}\}$, where $\boldsymbol{a}$ is some arbitrary element of the sample space used to represent the orbit. In the remainder, we suppress the superscript and write $O_{\boldsymbol{a}}$ unless ambiguity could arise. The group-invariance assumption allows us to disintegrate the distribution of $\boldsymbol{\varepsilon}$ into two components: a distribution over the partitioning $\mathcal{O}$ and a uniform distribution over each orbit $O \in \mathcal{O}$ (see e.g. Theorem 4.3 in Eaton (1989)). We will henceforth refer to the latter distribution as 'the O-conditional distribution' of $\boldsymbol{\varepsilon}$. The O-conditional distribution of $\boldsymbol{\varepsilon}$ induces distributions of $\boldsymbol{\iota}'\boldsymbol{\varepsilon}$ and $\boldsymbol{\iota}'\boldsymbol{x}$ given $\boldsymbol{\varepsilon} \in O$, which we will refer to as the O-conditional distribution of $\boldsymbol{\iota}'\boldsymbol{\varepsilon}$ and $\boldsymbol{\iota}'\boldsymbol{x}$, respectively.

EXAMPLE 2. *For the orthogonal group $\mathcal{H}$, which can be represented by the collection of all orthonormal matrices, the set $\mathcal{O}$ is the collection of all spheres in dimension n. Using the group-invariance assumption, the distribution of $\boldsymbol{\varepsilon}$ can therefore be decomposed into a distribution over all spheres in dimension n, and a uniform distribution over each sphere. For the orthogonal group $\mathcal{H}$, the distribution of $\boldsymbol{\iota}'\boldsymbol{\varepsilon}$ does not depend on the choice of $\boldsymbol{\iota}$, as, for any unit vector $\boldsymbol{b}$, $\boldsymbol{\iota}'\boldsymbol{\varepsilon} \overset{\mathrm{d}}{=} \boldsymbol{\iota}'\boldsymbol{H}^*\boldsymbol{\varepsilon} = \boldsymbol{b}'\boldsymbol{\varepsilon}$, for some $\boldsymbol{H}^* \in \mathcal{H}$. It can be shown that the O^r-conditional distribution of $\boldsymbol{\iota}'\boldsymbol{\varepsilon}$ is the Beta$(\frac{n-1}{2}, \frac{n-1}{2}, -r, r)$-distribution, where O^r is the sphere in dimension $n > 1$ with radius $r = \|\boldsymbol{\varepsilon}\|_2$ (see Lemma 4 for a proof). An example of an orthogonal-invariant distribution is the multivariate standard normal distribution.*

2.3. *Group-Invariance test.* The group-invariance test is based on Assumption 1. The key idea behind a group-invariance test is that it only uses the O-conditional distributions. That is, the test does not require knowledge of the distribution over the partition $\mathcal{O}$. In particular, let $q_\alpha(O)$ denote the α upper-quantile of the O-conditional distribution of $\boldsymbol{\iota}'\boldsymbol{\varepsilon}$ (see Remark 1 for details). A group-invariance test $\phi_\alpha^{\mathcal{G}}$ based on the group $\mathcal{G}$ (henceforth, $\mathcal{G}$-invariance test) with level α is then defined as

$$\phi_\alpha^{\mathcal{G}}(x) := \mathbb{I}\{\boldsymbol{\iota}'\boldsymbol{x} > q_\alpha(O_{\boldsymbol{x}})\}, \tag{1}$$

where we will suppress the subscript and superscript unless ambiguity could arise. That is, the test compares the test statistic $\boldsymbol{\iota}'\boldsymbol{x}$ to the α upper-quantile of the $O_{\boldsymbol{x}}$-conditional distribution of $\boldsymbol{\iota}'\boldsymbol{\varepsilon}$. Notice that this test is 'feasible' in practice, as $\boldsymbol{x}$ is observed and the orbit $O_{\boldsymbol{x}}$ can be reconstructed from a single one of its elements (see Remark 2).

We will compare the group-invariance to an 'infeasible' group-invariance test, which essentially compares $\boldsymbol{\iota}'\boldsymbol{x}$ to the orbit we would have observed under the null: $O_{\boldsymbol{\varepsilon}}$. In particular, it compares $\boldsymbol{\iota}'\boldsymbol{x}$ to the α upper-quantile of the $O_{\boldsymbol{\varepsilon}}^*$-conditional distribution of $\boldsymbol{\iota}'\boldsymbol{\varepsilon}$,

$$\overline{\phi}_\alpha^{\mathcal{G}}(\boldsymbol{x}) := \mathbb{I}\{\boldsymbol{\iota}'\boldsymbol{x} > q_\alpha(O_{\boldsymbol{\varepsilon}}^*)\}, \tag{2}$$

where $O_{\boldsymbol{\varepsilon}}^*$ is $O_{\boldsymbol{\varepsilon}}$ but with the element $\boldsymbol{\varepsilon}$ replaced by $\boldsymbol{x}$. The reason we use of $O_{\boldsymbol{\varepsilon}}^*$ instead of $O_{\boldsymbol{\varepsilon}}$ is because this will allow us to later consider groups $\mathcal{G}$ for which $\phi_\alpha^{\mathcal{G}} = \overline{\phi}_\alpha^{\mathcal{G}}$. The inclusion of $\boldsymbol{x}$ is necessary to do this, as $\boldsymbol{x} \in O_{\boldsymbol{x}}^{\mathcal{G}}$ for any group $\mathcal{G}$ due to the presence of the identity transformation.

Notice that the tests coincide under the null hypothesis, where $\boldsymbol{x} = \boldsymbol{\varepsilon}$.

PROPOSITION 1. *Under H_0 we have $\phi = \overline{\phi}$.*

As $\overline{\phi}$ uses knowledge of $O_{\boldsymbol{\varepsilon}}$, one may expect $\phi \leq \overline{\phi}$, so that $\overline{\phi}$ serves as an upper bound for ϕ. Here, we mean $\phi(\boldsymbol{x}) \leq \overline{\phi}(\boldsymbol{x})$, for all $\boldsymbol{x}$. Proposition 2 shows that this is indeed true if $\mathcal{G} = \mathcal{H}$. We believe that it is possible to establish similar results for other common situations. We elaborate on the difference between $\overline{\phi}^{\mathcal{H}}$ and $\phi^{\mathcal{H}}$ in Section 2.4. The proof of the result can be found in Appendix B, together with the proofs of all other results not proven in the main text.

PROPOSITION 2. *Suppose $\alpha < .5$, $\mu > 0$ and $\mathcal{G} = \mathcal{H}$. Then $\phi_\alpha^{\mathcal{H}} \leq \overline{\phi}_\alpha^{\mathcal{H}}$.*

REMARK 1. *More formally, the quantile is equal to*

$$q_\alpha(O_{\boldsymbol{x}}^{\mathcal{G}}) = \inf\{z \in \mathbb{R} \mid \int_{\mathcal{G}} \mathbb{I}\{\boldsymbol{\iota}'\boldsymbol{G}\boldsymbol{x} > z\} dQ(\boldsymbol{G}) \leq \alpha\} \tag{3}$$

where Q is the Haar measure on $\mathcal{G}$.

REMARK 2. *In practice, the quantile $q_\alpha(O_{\boldsymbol{x}})$ is typically approximated by the empirical quantile based on a large number of draws uniformly sampled from $O_{\boldsymbol{x}}$. With regard to (3), this means the Haar measure is replaced by an appropriate random measure. We refer to such a Monte Carlo group-invariance test based on a sample of size, say, M (including the identity) as a $\mathcal{G}^M$-invariance test.*

2.4. *Orthogonal invariance and the t-test.* In this section, we consider ϕ and $\overline{\phi}$ under orthogonal invariance. We do this to provide the reader with some stronger intuitions about these tests, by establishing a relationship to 'conventional' parametric tests.

The following result shows that group-invariance tests can be viewed as a natural generalization of the t-test to subgroups of the orthogonal group.

THEOREM 1. *The t-test is the orthogonal group-invariance test. That is, let $\widehat{\sigma} = \sqrt{\boldsymbol{x}'(\boldsymbol{I} - \boldsymbol{\iota}\boldsymbol{\iota}')\boldsymbol{x}/(n-1)}$, then*

$$\phi_\alpha^{\mathcal{H}}(\boldsymbol{x}) = \mathbb{I}\{\boldsymbol{\iota}'\boldsymbol{x}/\widehat{\sigma} > t_\alpha^{n-1}\},$$

where t_α^{n-1} is the α upper-quantile of the t-distribution with $(n-1)$ degrees of freedom.

To our surprise, we were unable to find Theorem 1 in the literature or textbooks in this form, although several strongly related results exist: see Chmielewski (1981). For example, the result does not appear in Lehmann and Romano (2005), who extensively discuss group-invariance tests and their relationship to the t-test (see Chapter 15.2 and in particular Example 15.2.4). The result is straightforward to generalize to F-tests to test parameters of higher dimension.

In addition, we can quantify the difference between $\phi^{\mathcal{H}}$ and $\overline{\phi}^{\mathcal{H}}$, as suggested in Proposition 2, by comparing Theorem 1 to Proposition 3. This comparison tells us the following. Consider the normal location model $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\iota}\mu, \sigma^2\boldsymbol{I})$. In this normal location model, $\overline{\phi}_\alpha^{\mathcal{H}}$ is equivalent to a test based on the estimator $\widetilde{\sigma} = n^{-1/2}\|\boldsymbol{\varepsilon}\|_2$ for σ, while $\phi_\alpha^{\mathcal{H}}$ is based on the sample standard deviation $\widehat{\sigma} = \sqrt{\boldsymbol{x}'(\boldsymbol{I} - \boldsymbol{\iota}\boldsymbol{\iota}')\boldsymbol{x}/(n-1)}$. Hence, the power loss when using $\phi_\alpha^{\mathcal{H}}$ instead of $\overline{\phi}_\alpha^{\mathcal{H}}$ can attributed to the loss of oracle knowledge of μ in the estimation of σ.

PROPOSITION 3. *Suppose $\boldsymbol{\varepsilon} \neq \boldsymbol{0}$ and $n > 1$. Let $\widetilde{\sigma} = \sqrt{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}/n}$, then*

$$\overline{\phi}_\alpha^{\mathcal{H}}(\boldsymbol{x}) = \mathbb{I}\{\boldsymbol{\iota}'\boldsymbol{x}/\widetilde{\sigma} > \sqrt{n}\beta_\alpha^{n-1,n-1}\},$$

where $\beta_\alpha^{n-1,n-1}$ is the α upper-quantile of the Beta$(\frac{n-1}{2}, \frac{n-1}{2}, -1, 1)$-distribution.

2.5. *Exactness of the group-invariance test.* A large part of popularity of group-invariance test can be attributed to their finite sample properties under H_0. In particular, ϕ has size at most α as captured by Theorem 2. A proof of this result, as well as all other proofs not included in the text, can be found in the appendix.

THEOREM 2. *If Assumption 1 holds, then ϕ rejects with probability at most α, under H_0.*

Moreover, under mild assumptions one can guarantee exactness (see Remark 3). As this is not the focus of this article, we will simply impose exactness as an assumption. We make this assumption without mention in the remainder.

ASSUMPTION 2. *The nominal level α is such that $\phi_\alpha^{\mathcal{G}}$ rejects with probability α under H_0.*

REMARK 3. *Due to discreteness issues, especially in case of finite groups, Assumption 2 may not hold for every $\alpha \in (0,1)$. In such cases it is common to: restrict the values of α (it typically suffices to pick α as a multiple of $1/|\mathcal{G}|$ if $\boldsymbol{x}$ follows a continuous distribution), accept a small size distortion, or to mix the test with a data independent 'trivial test' that rejects with probability α (see e.g. Hoeffding 1952; Hemerik and Goeman 2018b).*

2.6. *The leak.*

In this section, we study the difference between ϕ and $\overline{\phi}$ as a function of $\boldsymbol{\iota}$ and $\mathcal{G}$.

Observe from the definitions in equations (1) and (2) that the tests ϕ and $\overline{\phi}$ can be represented to share the same test statistic $\boldsymbol{\iota}'\boldsymbol{x}$. Hence, the difference between the tests is due entirely to the critical values: $q_\alpha(O_{\boldsymbol{x}})$ and $q_\alpha(O_{\varepsilon}^*)$. In turn, the difference between these critical values is due to the distributions of which they are quantiles. Fix $\boldsymbol{x}$ and let $\widetilde{\boldsymbol{G}}$ be uniformly distributed on $\mathcal{G}$. Then $q_\alpha(O_{\boldsymbol{x}})$ is a quantile of the distribution of $\boldsymbol{\iota}'\widetilde{\boldsymbol{G}}\boldsymbol{\iota}\mu + \boldsymbol{\iota}'\widetilde{\boldsymbol{G}}\boldsymbol{\varepsilon}$ and $q_\alpha(O_{\varepsilon}^*)$ a quantile of the distribution of $\mathbb{I}\{\widetilde{\boldsymbol{G}} = \boldsymbol{I}\}\mu + \boldsymbol{\iota}'\widetilde{\boldsymbol{G}}\boldsymbol{\varepsilon}$. Hence, the difference between the tests depends entirely on the term

$$\mu\boldsymbol{\iota}'\boldsymbol{G}\boldsymbol{\iota}, \tag{4}$$

for $\boldsymbol{G} \in \mathcal{G} \setminus \{\boldsymbol{I}\}$.

Dobriban (2021) calls the term in (4) a "leak". As this leak $\boldsymbol{\iota}'\widetilde{\boldsymbol{G}}\boldsymbol{\iota}\mu$ depends on μ but is independent of $\boldsymbol{\iota}'\widetilde{\boldsymbol{G}}\boldsymbol{\varepsilon}$ (see e.g. Lemma 5.3 in Dobriban (2021)), they describe it as a spillover from signal into noise that is "hopefully small". The reason one would hope for the leak to be small is because one would then expect that ϕ is close to $\overline{\phi}$. This is desirable because $\overline{\phi}$ uses knowledge of $O_{\boldsymbol{\varepsilon}}$ and one would therefore expect it to be more powerful than ϕ, as we established the special case of the orthogonal group $\mathcal{H}$ in Proposition 2.

We consider the following quantification of the size of the leak:

$$\delta_{\mathcal{G}} := \sup_{\boldsymbol{G} \in \mathcal{G} \setminus \{\boldsymbol{I}\}} \boldsymbol{\iota}'\boldsymbol{G}\boldsymbol{\iota}. \tag{5}$$

Notice here that $-1 \leq \boldsymbol{\iota}'\boldsymbol{G}\boldsymbol{\iota} \leq 1$.

REMARK 4. *Analogously to (5), one can consider maximum of the absolute value* $\delta_{\mathcal{G}}^{abs} := \sup_{\boldsymbol{G} \in \mathcal{G} \setminus \{\boldsymbol{I}\}} |\boldsymbol{\iota}'\boldsymbol{G}\boldsymbol{\iota}|$. *However, in case of one-sided hypotheses, we do not mind large negative values of the leak.*

The following result shows that a group $\mathcal{G}$ with a sufficiently small $\delta_{\mathcal{G}}$ results in a test $\phi^{\mathcal{G}}$ that is close to $\overline{\phi}^{\mathcal{G}}$.

THEOREM 3. *We have* $\phi_\alpha^{\mathcal{G}}(\mu\boldsymbol{\iota} + \boldsymbol{\varepsilon}) \geq \overline{\phi}_\alpha^{\mathcal{G}}((1 - \delta_{\mathcal{G}})\mu\boldsymbol{\iota} + \boldsymbol{\varepsilon})$.

The interpretation of Theorem 3 is as follows. The test $\phi_\alpha^{\mathcal{G}}$ applied to data $\boldsymbol{x}$ is at least as powerful as $\overline{\phi}_\alpha^{\mathcal{G}}$ applied to data $\boldsymbol{x}^{\delta_{\mathcal{G}}} := (1 - \delta_{\mathcal{G}})\mu\boldsymbol{\iota} + \boldsymbol{\varepsilon}$, which is $\boldsymbol{x}$ but with the parameter μ shrunk by a factor of $(1 - \delta_{\mathcal{G}})$. If $\delta_{\mathcal{G}}$ is small, we have $\boldsymbol{x}^{\delta_{\mathcal{G}}} \approx \boldsymbol{x}$ and so we expect $\overline{\phi}_\alpha^{\mathcal{G}}(\boldsymbol{x}^{\delta_{\mathcal{S}}}) \approx \overline{\phi}_\alpha^{\mathcal{G}}(\boldsymbol{x})$. Hence, we would expect $\phi_\alpha^{\mathcal{G}}$ to be nearly as powerful as $\overline{\phi}_\alpha^{\mathcal{G}}$.

3. Subgroup-invariance tests.

Once invariance has been established under a group $\mathcal{G}$, one would typically like to perform the test $\phi^{\mathcal{G}}$. Unfortunately, this is often computationally infeasible in practice. Therefore, it is standard practice to use a Monte Carlo group-invariance test $\phi^{\mathcal{G}^M}$. The key idea in this paper is to instead use a 'subgroup-invariance test' $\phi^{\mathcal{S}}$, where $\mathcal{S}$ is a carefully selected subgroup of $\mathcal{G}$.

Notice that invariance under $\mathcal{G}$, implies invariance under every subgroup $\mathcal{S}$ of $\mathcal{G}$. As a consequence, the 'subgroup-invariance test' $\phi^{\mathcal{S}}$ controls size by Theorem 2 for any subgroup $\mathcal{S}$ of $\mathcal{G}$ (and is even exact under the modest condition that α is such that Assumption 2 still holds for $\mathcal{S}$). This allows us to attempt to 'maximize' the power of $\phi^{\mathcal{S}}$ over a class of subgroups of $\mathcal{G}$. In particular, as was described in Section 2.6, the power of $\phi^{\mathcal{S}}$ depends on the leak. Therefore, we consider subgroups that minimize the impact of the leak. Under normality, we show that we can find subgroups this way that outperform $\phi^{\mathcal{G}^M}$.

3.1. *Oracle subgroup-invariance tests.* A promising candidate subgroup $\mathcal{S}$ is one that eliminates the leak, so that $\delta_{\mathcal{S}}^{\text{abs}} = 0$, which implies $\phi^{\mathcal{S}} = \overline{\phi}^{\mathcal{S}}$. We call such subgroups *oracle subgroups*.

DEFINITION 1 (Oracle Subgroup). *A subgroup $\mathcal{S}$ of $\mathcal{G} \subset \mathcal{H}$ is an oracle subgroup with respect to a given $\boldsymbol{\iota}$, if*

$$\boldsymbol{\iota}'\boldsymbol{S}\boldsymbol{\iota} = 0, \textit{ for all } \boldsymbol{S} \in \mathcal{S} \setminus \{\boldsymbol{I}\}.$$

For an oracle subgroup $\mathcal{S}$, we have $\boldsymbol{s}'\boldsymbol{x} = \boldsymbol{\iota}'\boldsymbol{S}'\boldsymbol{\iota}\mu + \boldsymbol{\iota}'\boldsymbol{S}'\boldsymbol{\varepsilon} = \boldsymbol{\iota}'\boldsymbol{S}'\boldsymbol{\varepsilon} = \boldsymbol{s}'\boldsymbol{\varepsilon}$, with $\boldsymbol{s} = \boldsymbol{S}\boldsymbol{\iota}$, for all $\boldsymbol{S} \in \mathcal{S} \setminus \{\boldsymbol{I}\}$. That is, the leak is entirely eliminated.

We now first explore some of the properties of oracle subgroups, which turn out to be quite elegant. Next, we present the properties of the resulting oracle subgroup-invariance tests, from which the chosen adjective 'oracle' will become clear. In Section 4, we consider subgroups that 'nearly' eliminate the leak.

3.2. *Properties of the oracle subgroups.* In this section, we study subgroups $\mathcal{S}$ of $\mathcal{H}$ through the orbit $O_{\boldsymbol{\iota}}^{\mathcal{S}} := \{\boldsymbol{S}\boldsymbol{\iota} \mid \boldsymbol{S} \in \mathcal{S}\}$ they induce with respect to $\boldsymbol{\iota}$ on the unit sphere. Some subgroups, such as oracle subgroups, have a one-to-one correspondence to their orbit $O_{\boldsymbol{\iota}}^{\mathcal{S}}$. In particular, we show that oracle subgroups correspond to orbits with orthogonal elements, which allows us to immediately derive some of their properties. Furthermore, we discuss the existence of oracle subgroups of the orthogonal group.

Notice that $O_{\boldsymbol{\iota}}$ has unit n-vectors as elements, as $\boldsymbol{\iota}$ is unit n-vector and each $\boldsymbol{S} \in \mathcal{S}$ is an orthonormal matrix. Furthermore, $\boldsymbol{\iota}$ is an element of $O_{\boldsymbol{\iota}}$ as $\boldsymbol{I} \in \mathcal{S}$. For oracle subgroups, post-multiplication of an element of $\mathcal{S}$ by $\boldsymbol{\iota}$ defines a bijection between $\mathcal{S}$ and the $O_{\boldsymbol{\iota}}^{\mathcal{S}}$. This is not true for all subgroups of $\mathcal{H}$. Take, for example, the permutation group $\mathcal{P}$ and $\boldsymbol{\iota} = (n^{-1/2}, n^{-1/2}, \dots)$, where $O_{\boldsymbol{\iota}}^{\mathcal{P}}$ consists has only one element while $\mathcal{P}$ has $n!$ elements.

PROPOSITION 4. *If $\mathcal{S}$ is an oracle subgroup of $\mathcal{H}$, then the map $\boldsymbol{S} \mapsto \boldsymbol{S}\boldsymbol{\iota}$ is a bijection from $\mathcal{S}$ to $O_{\boldsymbol{\iota}}^{\mathcal{S}}$*

Using Proposition 4, we can show that oracle subgroups correspond exactly to the subgroups that are represented by unit orbits that have orthogonal elements.

THEOREM 4. *A subgroup $\mathcal{S}$ of $\mathcal{H}$ is an oracle subgroup of $\mathcal{H}$ if and only if $O_{\boldsymbol{\iota}}^{\mathcal{S}}$ has orthogonal elements and $|\mathcal{S}| = |O_{\boldsymbol{\iota}}^{\mathcal{S}}|$.*

Theorem 4 immediately yields the following insightful result regarding the maximum order of oracle subgroups.

COROLLARY 1. *An oracle subgroup $\mathcal{S}$ of $\mathcal{H}$ has at most n elements.*

PROOF. The orbit $O_{\boldsymbol{\iota}}^{\mathcal{S}}$ has orthogonal n-dimensional elements by Proposition 4, so it has at most n elements. Furthermore, $\mathcal{S}$ and $O_{\boldsymbol{\iota}}^{\mathcal{S}}$ have the same number of elements by Proposition 4. Hence, $\mathcal{S}$ has at most n elements. □

REMARK 5. *As a corollary to Corollary 1, the largest oracle subgroup of $\mathcal{P}$ or $\mathcal{R}$ (or any other subgroup of the orthogonal group) for any choice of $\boldsymbol{\iota}$ is also at most of order n.*

Finally, we include a result regarding the existence of oracle subgroups of the orthogonal group.

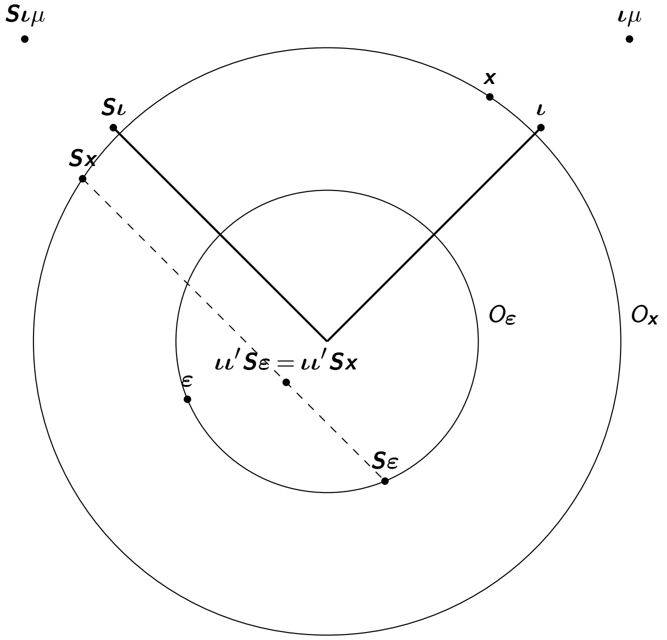

Figure 1: Two-dimensional example of the geometry of an oracle subgroup $\{\boldsymbol{I}, \boldsymbol{S}\}$, where $\boldsymbol{S}$ has first and second column $(0, 1)'$ and $(-1, 0)'$, respectively. We have that $\boldsymbol{\iota}'\boldsymbol{S\iota} = 0$ and so $\boldsymbol{\iota}$ and $\boldsymbol{S\iota}$ are orthogonal. Hence, the projection of $\boldsymbol{Sx}$ on $\boldsymbol{\iota}$ coincides with the projection of $\boldsymbol{S}\varepsilon$ on $\boldsymbol{\iota}$.

THEOREM 5. *There exists an oracle subgroup $\mathcal{S}$ of $\mathcal{H}$ with respect to any unit vector $\boldsymbol{\iota}$, of any order p, with $1 \leq p \leq n$.*

Unfortunately, even for quite 'large' subgroups $\mathcal{G}$ of $\mathcal{H}$, there often exist values $p \leq n$ such that $\mathcal{G}$ has no oracle subgroups of order p. In addition, the existence of oracle subgroups of $\mathcal{G}$ of $\mathcal{H}$ depends intimately on the choice of $\boldsymbol{\iota}$. This will be seen in Section 5, where we fully characterize oracle subgroups of the sign-flipping group for $\boldsymbol{\iota} = (n^{-1/2}, n^{-1/2}, \ldots, n^{-1/2})'$.

Figure 1 displays the geometry of an oracle subgroup. In particular, it can be seen that $\boldsymbol{S}$ rotates $\boldsymbol{x}$ and ε in such a way that the resulting projections on $\boldsymbol{\iota}$ coincide. As a consequence, the projection of $\boldsymbol{Sx}$ on $\boldsymbol{\iota}$ is as if ε is observed.

3.3. *Properties of an oracle subgroup-invariance test.* As the size of an oracle subgroup-invariance test is guaranteed by Theorem 2, it remains to discuss its power.

If a subgroup $\mathcal{S}$ of $\mathcal{H}$ is finite, we define the matrix representation $\mathfrak{S} := (\boldsymbol{\iota}, \boldsymbol{S}_1\boldsymbol{\iota}, \boldsymbol{S}_2\boldsymbol{\iota}, \ldots)$, $\boldsymbol{S}_1, \boldsymbol{S}_2, \dots \in \mathcal{S}$. Suppose that $\mathcal{S}$ is an oracle subgroup of $\mathcal{H}$ with respect to $\boldsymbol{\iota}$. We consider the situation where the distribution of ε is invariant under pre-multiplication by the matrix $\mathfrak{S}'$ and the elements of ε are independent. This holds, for example, if $\varepsilon \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$, where $\sigma^2 > 0$, as $\mathfrak{S}'$ is an orthonormal matrix by Theorem 4 and the normal distribution is orthogonal invariant. We can then establish an equivalence between an $\mathcal{S}$-invariance test and a Monte Carlo test that draws from the unconditional null distribution in the following result. Here, we define an M-Monte Carlo test as a test that appends the test statistic to $(M-1)$ samples of the test statistic from the null distribution, and compares the test statistic to the α upper-quantile of (a uniform distribution over) this (multi)set.

THEOREM 6. *Let $\mathcal{S}$ be an oracle subgroup of $\mathcal{H}$ of order $|\mathcal{S}| = n$, with matrix representation $\mathfrak{S}$. Suppose that $\varepsilon \overset{\mathrm{d}}{=} \mathfrak{S}' \varepsilon$ and that the elements of ε are i.i.d.. Then, an $\mathcal{S}$-invariance test has the same size and power as an n-Monte Carlo test, i.e. a test based on independent draws from the null distribution of the elements of $\boldsymbol{x}$.*

PROOF. Note that $\boldsymbol{\iota}$ is the first column of $\mathfrak{S}$. From Theorem 4, we know that all columns of $\mathfrak{S}$ are orthogonal. Hence, $\mathfrak{S}'\boldsymbol{\iota} = \boldsymbol{e}_1 = (1, 0, \ldots 0)'$. This yields

$$\mathfrak{S}' \boldsymbol{x} = \mu \mathfrak{S}' \boldsymbol{\iota} + \mathfrak{S}' \varepsilon = \mu \boldsymbol{e}_1 + \mathfrak{S}' \varepsilon \overset{\mathrm{d}}{=} \mu \boldsymbol{e}_1 + \varepsilon,$$

where the final step uses invariance of the distribution of ε under pre-multiplication by $\mathfrak{S}'$. The critical value $q_\alpha(O_{\boldsymbol{x}}^{\mathcal{S}})$ is the α upper-quantile of the empirical distribution over the elements of the vector $\mathfrak{S}' \boldsymbol{x}$, which is the α upper-quantile of the empirical distribution over the elements of an observation of $\mu \boldsymbol{e}_1 + \varepsilon$. The relation to a Monte Carlo test follows from noticing that the elements of the vector ε are assumed to be i.i.d.. □

REMARK 6. *The result is easily extended to an oracle subgroup of order, say, $M \leq n$ to obtain a relation to a Monte Carlo test based on M draws from the null distribution. This can be done by padding the resulting $n \times M$ matrix representation with zero-columns until it is $n \times n$.*

As normality implies the conditions of the theorem, we include the following simpler corollary.

COROLLARY 2. *Let $\mathcal{S}$ be an oracle subgroup of $\mathcal{H}$ of order $|\mathcal{S}| = n$, with matrix representation $\mathfrak{S}$. Let $\varepsilon \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$, where $\sigma^2 > 0$. Then, an $\mathcal{S}$-invariance test has the same size and power as an n-Monte Carlo Z-test.*

Intuitively, the result states that if the analyst only knows that ε is $\mathcal{S}$-invariant, and ε happens to have distribution $\varepsilon \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$, $\sigma^2 > 0$, unbeknownst to the analyst, then the $\mathcal{S}$-invariance test has the same size and power as a Monte Carlo Z-test. The subgroup therefore essentially allows the analyst to sample from the unknown null distribution: hence the name *oracle*.

REMARK 7. *In fact, normality is almost necessary for the conditions of Theorem 6, for the following reason. Let $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ denote two columns of $\mathfrak{S}$. Then $(\boldsymbol{s}_1, \boldsymbol{s}_2)'\varepsilon \overset{\mathrm{d}}{=} (\boldsymbol{e}_1, \boldsymbol{e}_2)'\varepsilon$. As $\boldsymbol{e}_1'\varepsilon$ and $\boldsymbol{e}_2'\varepsilon$ are independent, so are $\boldsymbol{s}_1'\varepsilon$ and $\boldsymbol{s}_2'\varepsilon$. This brings us very close to the conditions of the Darmois-Skitovich Theorem (Darmois, 1953; Skitovitch, 1953). This theorem states the following: $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ have non-zero elements, ε has i.i.d. elements and $\boldsymbol{s}_1'\varepsilon$ and $\boldsymbol{s}_2'\varepsilon$ are independent if and only if $\varepsilon \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$, for some $\sigma^2 > 0$. Therefore, if $\mathfrak{S}$ has two columns with non-zero elements, then the conditions of Theorem 6 imply that ε is i.i.d. normally distributed.*

Analogously to Theorem 1, we can derive the following proposition regarding Monte Carlo tests. Recall from Remark 2 that a $\mathcal{G}^M$-invariance test the test statistic is compared to a quantile of a (multi)set including the test statistic and $(M - 1)$ draws (without replacement) from the $O_{\boldsymbol{x}}$-conditional distribution of the test statistic.

PROPOSITION 5. *An $\mathcal{H}^M$-invariance test has the same size and power as an M-Monte Carlo t_{n-1}-test.*

Combining Corollary 2 with Proposition 5 allows us to show that an oracle subgroup-invariance test has higher power than a Monte Carlo group-invariance test based on draws from the orthogonal group, in a normal location model. This is captured by the following result.

THEOREM 7. *Let $\varepsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I})$, $\sigma^2 > 0$ and $\mathcal{S} \subset \mathcal{H}$ be an oracle subgroup with $|\mathcal{S}| = M$. Then, an $\mathcal{S}$-invariance test is more powerful than an $\mathcal{H}^M$-invariance test.*

PROOF. An $\mathcal{S}$-invariance test has the same power as an M-Monte Carlo Z-test by Theorem 6 and Remark 6, which is more powerful than an M-Monte Carlo t_{n-1}-test in a normal location model, which in turn has power that coincides with the power of an $\mathcal{H}^M$-invariance test by Proposition 5. □

4. Near-oracle subgroups.

In the previous section, we showed that oracle subgroups have attractive statistical properties. However, it is not guaranteed that oracle subgroups are the best choice in practice. Of particular concern is that the maximum order of oracle subgroups is small, even for the orthogonal group as shown in Corollary 1. Furthermore, if we only have invariance under some subgroup $\mathcal{G}$ of $\mathcal{H}$, then the maximum order of oracle subgroups of $\mathcal{G}$ may be even smaller, for the desired choice of $\boldsymbol{\iota}$. Under the conditions for Theorem 6 the resulting oracle subgroup invariance test would be analogous to a Monte Carlo test based on a small sample, which we would expect to have low power. Unfortunately, larger subgroups necessarily introduce a leak. This motivates the search for 'near-oracle' subgroups: subgroups that are larger than oracle subgroups but only introduce a small leak.

In particular, for a given $\boldsymbol{\iota}$ we define an 'M-optimal' near-oracle subgroup to be a subgroup of $\mathcal{G}$ of a given order M, such that the $\delta_{\mathcal{S}}$ is minimized over all subgroups of order M. Such a subgroup may not be unique. From Theorem 3, we obtain the guarantee that if $\delta_{\mathcal{S}}$ is sufficiently small, the resulting test still has good power. In the next section, we discuss the equivalence between finding M-optimal near-oracle subgroups and spherical (group) code problems.

An illustration of a near-oracle subgroup is given in Figure 2, which differs from Figure 1 because it features a rotation that is (slightly) less than orthogonal. As a result, the projections are no longer identical, but they are still 'close' to each other.

4.1. *Group codes.*

In order to understand M-optimal near-oracle subgroups, we establish a link to a problem related to spreading out points on the unit hypersphere in dimension n. Recall that matrix $\mathfrak{S}$ has columns $\boldsymbol{S\iota}$ where $\boldsymbol{S} \in \mathcal{S}$, which is well-defined as we only consider finite subgroups $\mathcal{S}$: those of order M.

PROPOSITION 6. *We have $\delta_{\mathcal{S}} = \max_{i,j,i \neq j} \boldsymbol{e}_i' \mathfrak{S}' \mathfrak{S} \boldsymbol{e}_j$.*

Notice that $\mathfrak{S}' \mathfrak{S}$ contains all the inner-products between the columns of $\mathfrak{S}$, which are all n-dimensional unit vectors and so are located on the unit hypersphere in dimension n. The value $\delta_{\mathcal{S}}$ can therefore be interpreted as the maximum inner-product between two of such points on the unit hypersphere, which has a one-to-one correspondence with the minimum angle, $\arccos(\delta_{\mathcal{S}})$, between any two points. A collection of M, say, unit vectors is also called a *spherical code* with parameters $(n, M, \delta_{\mathcal{S}})$. If a spherical code is induced by a subgroup of $\mathcal{H}$, as is the case here, it is also known as a *group code* (Slepian, 1968; Conway and Sloane, 1998). The problem of finding an M-optimal near-oracle subgroup is therefore equivalent to the problem of finding the group code $(n, M, \delta_{\mathcal{S}})$, for which $\delta_{\mathcal{S}}$ is minimized for fixed n and M.

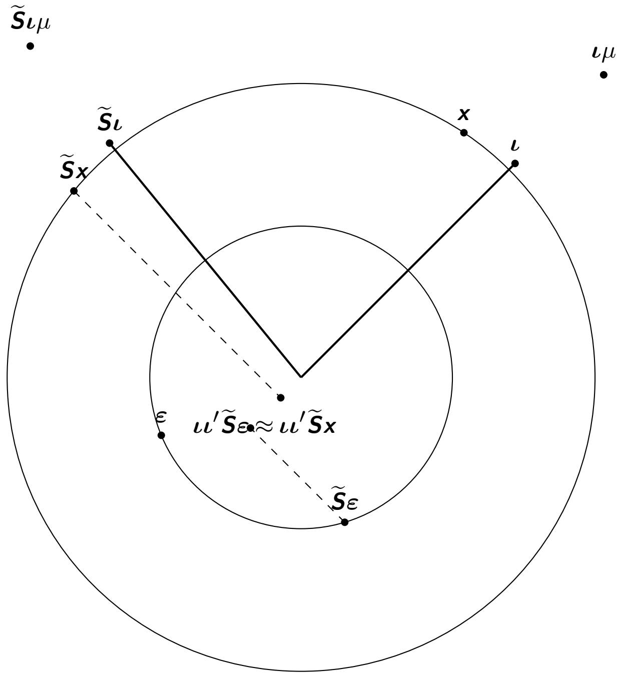
$\widetilde{\boldsymbol{S}}\boldsymbol{\iota}\mu$
$\boldsymbol{\iota}\mu$
$\widetilde{\boldsymbol{S}}\boldsymbol{\iota}$
$\boldsymbol{x}$
$\boldsymbol{\iota}$
$\widetilde{\boldsymbol{S}}\boldsymbol{x}$
$\boldsymbol{\varepsilon}$
$\boldsymbol{\iota\iota}'\widetilde{\boldsymbol{S}}\boldsymbol{\varepsilon} \approx \boldsymbol{\iota\iota}'\widetilde{\boldsymbol{S}}\boldsymbol{x}$
$\widetilde{\boldsymbol{S}}\boldsymbol{\varepsilon}$

Figure 2: Two-dimensional example of the geometry of a near-oracle subgroup $\mathcal{S} = \{\boldsymbol{I}, \widetilde{\boldsymbol{S}}\}$, where $\widetilde{\boldsymbol{S}}$ is chosen such that $\delta_{\mathcal{S}} = \boldsymbol{\iota}'\widetilde{\boldsymbol{S}}\boldsymbol{\iota} = 0.1$. Hence, the difference in length of the projections is $\boldsymbol{\iota}'\widetilde{\boldsymbol{S}}\boldsymbol{x} - \boldsymbol{\iota}'\widetilde{\boldsymbol{S}}\boldsymbol{\varepsilon} = .1\mu$.

It is well established that a moderately large value of M often yields a small minimum value of $\delta_{\mathcal{S}}$. For example, Sloane et al. (1996) lists a group code for $n = 16$ and $M = 256$ that is induced by a subgroup of $\mathcal{R}$, for which $\delta_{\mathcal{S}} = .25$ (as a comparison, we find an average 'δ' $\approx .92$ from 10^5 samples of size M from $\mathcal{R}$). Unfortunately, we were unable to find suitable algorithms to construct relevant M-optimal group codes or 'good' group codes of order M, nor could we find libraries that contain them. For example, Sloane et al. (1996) lists spherical codes up to $n = 24$ for few values of M, only some of which are group codes. In addition, we may not be satisfied with any 'good' group code: the group code needs to be induced by a subgroup of the group under which invariance is established. Therefore, we propose a simple algorithm in Section 5.3 for the case of the sign-flipping group.

5. Example: sign-flipping. In order to demonstrate the ideas presented in this paper, we consider invariance under the sign-flipping group $\mathcal{R}$ and choose the vector $\boldsymbol{\iota} = n^{-1/2}(1, 1, \dots, 1)'$. If we would additionally assume independence, this would mean we assume that the elements of $\boldsymbol{\varepsilon}$ are marginally symmetrically distributed about the origin. The resulting test is often used in paired data, as was already proposed in (Fisher, 1935). The idea is then to sign-flip differences between paired observations. For additional discussions and applications, see Efron 1969; Bekker and Lawford 2008; Davidson and Flachaire 2008; Winkler et al. 2014; Andreella et al. 2020. This setting is highly convenient to work with, which allows us to illustrate our proposed methodology.

The sign-flipping group is convenient to work with for the following reasons. First, the group is finite. The sign-flipping group is a subgroup of the orthogonal group $\mathcal{H}$, and can be represented by a collection of diagonal matrices with diagonal elements in $\{-1, 1\}$. The fact that it can be represented as a group of diagonal matrices, means that the group is abelian:

$\boldsymbol{R}_1\boldsymbol{R}_2 = \boldsymbol{R}_2\boldsymbol{R}_1$ for all $\boldsymbol{R}_1, \boldsymbol{R}_2 \in \mathcal{R}$. Furthermore, each element is its own inverse as it is symmetric and orthonormal, which means the group is a Boolean group.

We denote the matrix representation of $\mathcal{R}$ by $\mathfrak{R} := (\boldsymbol{\iota}, \boldsymbol{R}_1\boldsymbol{\iota}, \boldsymbol{R}_2\boldsymbol{\iota}, \dots)$, $\boldsymbol{R}_1, \boldsymbol{R}_2, \dots \in \mathcal{R}$. The group $\mathcal{R}$ is isomorphic to $n^{1/2}\,\mathfrak{R}$ under element-wise multiplication of its columns, which correspond to the diagonals of the elements in $\mathcal{R}$. The same holds for a subgroup $\mathcal{S} \subset \mathcal{R}$ and its analogous matrix representation $\mathfrak{S}$. Due to the isomorphism and the fact that $\boldsymbol{\iota}$ is fixed throughout this section, we will henceforth also refer to the sign-flipping group and an arbitrary subgroup as $n^{1/2}\,\mathfrak{R}$ and $n^{1/2}\,\mathfrak{S}$, respectively.

Notice that the distribution of $n\boldsymbol{\iota}'\widetilde{\boldsymbol{S}}\boldsymbol{\iota}$, where $\widetilde{\boldsymbol{S}}$ is uniformly distributed on $\mathcal{S}$, coincides with the empirical distribution over $n\boldsymbol{\iota}'\,\mathfrak{S}$. We refer to this distribution as the 'leak distribution'.

EXAMPLE 3. *If $n = 2$, then*

$$n^{1/2}\,\mathfrak{R} = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \end{bmatrix}.$$

The leak distribution of $n^{1/2}\,\mathfrak{R}$ is then the empirical distribution over

$$n\boldsymbol{\iota}'\,\mathfrak{R} = \begin{bmatrix} 2 & 0 & 0 & -2 \end{bmatrix},$$

which assigns .25 mass to both 2 and -2, and .5 mass to 0. Notice that $n^{1/2}\,\mathfrak{R}$ is not an oracle subgroup, as $n\boldsymbol{\iota}'\,\mathfrak{R}$ contains non-zero elements besides the first element.

EXAMPLE 4. *If $n = 2$, then an example of an oracle subgroup $n^{1/2}\,\mathfrak{S}$ of $n^{1/2}\,\mathfrak{R}$ is*

$$n^{1/2}\,\mathfrak{S} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Furthermore,

$$n\,\mathfrak{S}'\,\mathfrak{S} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

which confirms this is an oracle subgroup by Theorem 4. The values taken on by the leak are the elements of

$$n\boldsymbol{\iota}'\,\mathfrak{S} = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

5.1. *Subgroups of the sign-flipping group.* In this section, we describe some (well-known) properties of the sign-flipping group, as well as their induced leak distributions. As $n^{1/2}\,\mathfrak{R}$ is a boolean group, its subgroups are of order 2^p, for some $p \leq n$, $p \in \mathbb{N}$, where p the called the rank of the subgroup. The subgroups of the sign-flipping group are abundant. The number of subgroups of rank p is equal to the pth element of the nth row of the 2-binomial coefficient triangle listed as entry A022166 in the OEIS (2021a). The total number of subgroups is equal to the sum of the nth row of this triangle, which can be found in entry A006116 of the OEIS (2021b). This means that if $n = 9$, say, then we have 3309747 subgroups of rank $p = 4$, and 8283458 subgroups in total. Thus, the subgroups of $n^{1/2}\,\mathfrak{R}$ are quite abundant even if n is small.

While the number of different subgroups is enormous, many of them yield the same vector $n^{1/2}\boldsymbol{\iota}'\,\mathfrak{R}$, and hence the same leak distribution. In particular, the number of different leak distributions corresponding to a subgroup of rank p is equal to the pth element of the nth row of the triangle listed as entry A076831 of the OEIS (2021c). The total number of different distributions is equal to the sum of the nth row of this triangle, which can be found in entry A076766 of the OEIS (2021d). For example, for $n = 9$ there are 240 different leak distributions corresponding to subgroups of rank $p = 4$, and 848 different leak distributions in total.

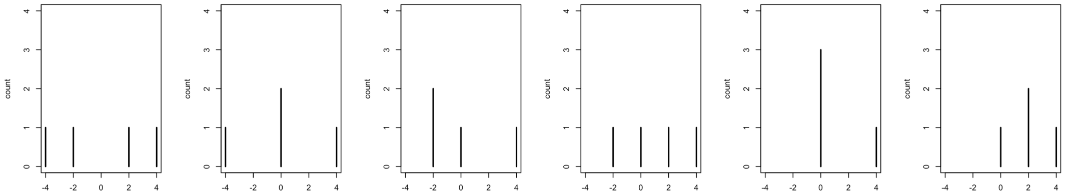

Figure 3: Histograms of leak distributions for all subgroups of $\mathcal{R}$ of order 4 for $n = 4$. The fifth histogram corresponds to an oracle subgroup, as all its mass it at zero, except for the mass at n produced by the identity element. Notice that the leak distributions are quite diverse.

EXAMPLE 5. *For $n = 4$ and subgroups of order 4 (so $p = 2$), there exist 6 different leak distributions. These distributions are illustrated in Figure 3. As we can see in the figure, the leak distributions are quite diverse. The leak corresponding to the oracle subgroups is given in the fifth image: except for the identity element, all mass is at 0. Notice that the second image may also be of interest in the case of the one-sided test: except for the identity element, the leak is non-positive. Subgroups that induce such leak distributions are discussed in Section 5.4. Both the second, third and fifth image correspond to 4-optimal near-oracle subgroups, and $\delta_{\mathcal{S}} = 0$.*

5.2. *Oracle and near-oracle subgroups.* The structure of the sign-flipping group $\mathcal{R}$ and choice $\boldsymbol{\iota} = (n^{-1/2}, n^{-1/2}, \dots)$ allow us to easily characterize the order of the oracle subgroups. See the appendix for a constructive proof.

THEOREM 8. *Let $k \leq l$, $k \in \mathbb{N}_+$, where l is the number of 2s in the prime factorization of n. Then $\mathcal{R}$ has an oracle subgroup with respect to $\boldsymbol{\iota} = (n^{-1/2}, n^{-1/2}, \dots)$ of order 2^k. Furthermore, if $\mathcal{S}$ is an oracle subgroup of $\mathcal{R}$, then it is of order 2^k for some $k \in \mathbb{N}_+$.*

Theorem 8 implies that the number of 2's in the prime factorization of n determines the maximum cardinality of its oracle subgroups. In particular if $n = 2^k$, for some $k \in \mathbb{N}$, then there exists an oracle subgroup of order n, which is the largest order that exists by Corollary 1. However, if n is an odd number, then the only oracle subgroup of $\mathcal{R}$ that exists is the trivial group containing only the identity element. This shows the importance of near-oracle subgroups of $\mathcal{R}$ with respect to $\boldsymbol{\iota}$. We discuss an algorithm to construct near-oracle subgroups in the Section 5.3, and in Section 5.4 we discuss a type of near-oracle subgroup of $\mathcal{R}$ that is of particular interest for one-sided testing.

5.3. *Near-oracle subgroup algorithm and database.* In order to discuss the construction of near-oracle subgroups of $\mathcal{R}$, we first provide a result regarding the expansion of subgroups of $\mathcal{R}$ to larger subgroups. The result can be summarized as follows: subgroups of $\mathcal{R}$ are easy to expand, and an expansion at least doubles the order of the subgroup.

PROPOSITION 7. *Let $\mathcal{A}$ be a subgroup of $\mathcal{R}$ and $\boldsymbol{R}$ an element in $\mathcal{R}$. Then $\mathcal{B} = \mathcal{A} \cup \boldsymbol{R}\mathcal{A}$ is a subgroup of $\mathcal{R}$, where $\boldsymbol{R}\mathcal{A} := \{\boldsymbol{R}\boldsymbol{A} \mid \boldsymbol{A} \in \mathcal{A}\}$. In addition, if $\boldsymbol{R} \notin \mathcal{A}$, then $|\mathcal{B}| = 2|\mathcal{A}|$.*

Based on this result, we propose a simple 'greedy' algorithm that constructs near-oracle subgroups of $\mathcal{R}$. This algorithm is presented in Algorithm 1. The idea of the algorithm is to iteratively expand a subgroup $\mathcal{S}$ of $\mathcal{R}$. It starts by setting $\mathcal{S}$ equal to some 'good' initial subgroup $\mathcal{S}^{\text{init}}$ of $\mathcal{R}$, such as an oracle subgroup that can be found using the constructive proof of Theorem 8. It then considers the expanded subgroup of the form $\mathcal{S} \cup \boldsymbol{R}\mathcal{S}$ of $\mathcal{R}$ for each $\boldsymbol{R} \in \mathcal{R} \setminus \mathcal{S}$, where $\boldsymbol{R}\mathcal{S} := \{\boldsymbol{RS} \mid \boldsymbol{S} \in \mathcal{S}\}$, which is indeed a subgroup by the first part of Proposition 7. Next, it updates the current subgroup $\mathcal{S}$ to an expanded subgroup that minimizes $\delta_{\mathcal{S}}$ among the candidate subgroups. Note that this minimum is not necessarily unique. Finally, the algorithm terminates when the current subgroup $\mathcal{S}$ is of the desired order. If $M \leq |\mathcal{R}|$, this algorithm is guaranteed to terminate by the second part of Proposition 7, as it will eventually expand to the entire group $\mathcal{R}$.

While the algorithm is based on minimizing $\delta_{\mathcal{S}}$, there is no guarantee that it terminates at an M-optimal subgroup, but it does guarantee a method of constructing a subgroup of desired order. In addition, the algorithm is not optimal in terms of time complexion, as many of the expansions may be duplicates of each other, but it suffices for our purposes. We leave the improvement of the algorithm for future work. The performance of group-invariance tests based on subgroups that were found using this algorithm is assessed in Section 6.

Algorithm 1 has been implemented in the R-package NOS (Near-Oracle Subgroups) (Koning, 2022b). Furthermore, the R-package NOSdata (Koning, 2022a) contains a database of approximately 2500 subgroups for $n \in \{1, \dots, 256\}$ and order $2^{\{0,\dots,10\}}$. For the construction of these subgroups, we drew 100 000 times without replacement (whenever possible) from $\mathcal{R} \setminus \mathcal{S}$ in line 4 of the algorithm. Next to minimizing using $\delta_{\mathcal{S}}$, it also contains subgroups minimized using $\delta_{\mathcal{S}}^{\text{abs}}$ for use in two-sided tests. For the latter case, we used oracle subgroups as initial subgroups, and for the minimization of $\delta_{\mathcal{S}}$ we used negative oracle subgroups, which are discussed in Section 5.4.

Algorithm 1 Near-oracle subgroups

```
1: $\mathcal{S} \leftarrow \mathcal{S}^{\text{init}}$,  where $\mathcal{S}^{\text{init}}$ is a subgroup of $\mathcal{R}$
2: $\mathcal{C} \leftarrow \emptyset$
3: while $|\mathcal{S}| < M$ do
4:     for $\boldsymbol{R} \in \mathcal{R} \setminus \mathcal{S}$ do
5:         $\mathcal{S}^* \leftarrow \mathcal{S} \cup \boldsymbol{R}\mathcal{S}$
6:         $\mathcal{C} \leftarrow \mathcal{C} \cup \{\mathcal{S}^*\}$
7:     $\mathcal{S} \leftarrow \arg\min_{\mathcal{S}^* \in \mathcal{C}} \delta_{\mathcal{S}^*}$
8: return $\mathcal{S}$
```

5.4. *Negative subgroups.* In this section, we study a particular type of expansion of a subgroup. We use this expansion to produce subgroups with a 'smaller' leak than an oracle subgroup, and an order twice as large as the largest oracle subgroup.

DEFINITION 2. *For a subgroup $\mathcal{S}$ of $\mathcal{R}$, $\mathcal{S}^- := \mathcal{S} \cup \{-\boldsymbol{I}\mathcal{S}\}$ is its corresponding negative subgroup.*

Notice that a negative subgroup is still a subgroup of $\mathcal{R}$ by Proposition 7, as $-\boldsymbol{I} \in \mathcal{R}$. Furthermore, it is twice the size of the original subgroup by the second part of Proposition 7, if the original subgroup did not contain $-\boldsymbol{I}$. This is the case if $\mathcal{S}$ is an oracle subgroup of $\mathcal{R}$ with respect to $\boldsymbol{\iota}$, in which case we call $\mathcal{S}^-$ a negative oracle subgroup with respect to $\boldsymbol{\iota}$. As the order of oracle subgroups is quite restrictive, this doubling of the order is desirable, assuming that the leak of $\mathcal{S}^-$ is as 'small' as that of $\mathcal{S}$.

It turns out that the leak of a negative oracle subgroup is indeed 'small'. This follows from the fact that a negative subgroup symmetrizes the original leak distribution around 0. In particular, the leak distribution of $\mathcal{S}^-$ with respect to $\boldsymbol{\iota}$ is the empirical distribution over

$n(\mathfrak{S}, -\mathfrak{S})'\boldsymbol{\iota} = n(\mathfrak{S}'\boldsymbol{\iota}, -\mathfrak{S}'\boldsymbol{\iota})$. This distribution assigns $1/|\mathcal{S}^-|$ mass to $-n$, $1/|\mathcal{S}^-|$ mass to n and the remaining mass to 0. In contrast, an oracle subgroup of order $|\mathcal{S}^-|$ induces a leak distribution that assigns $1/|\mathcal{S}^-|$ mass to n and the remaining mass to 0. This can be seen in the second and fifth image in Figure 3. This means that the leak distribution of $\mathcal{S}^-$ is stochastically dominated by that of an oracle subgroup of order $|\mathcal{S}^-|$. Hence, the leak is 'smaller' for the negative oracle subgroup. Therefore, we would expect an $\mathcal{S}^-$-invariance test to be more powerful than an invariance test based on an oracle subgroup of order $|\mathcal{S}^-|$. This is indeed what we find in the simulation results presented in Section 6.

6. Simulation results. In this section, we present the simulation results. We simulated data $\boldsymbol{x}$ using the standard normal location model

$$\boldsymbol{x} = \boldsymbol{\iota}\mu + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n),$$

with $\boldsymbol{\iota} = (n^{-1/2}, n^{-1/2}, \dots)'$. We tested the hypothesis $H_0 : \mu = 0$ against $H_1 : \mu > 0$, using the sign-flipping invariance assumption $\boldsymbol{\varepsilon} \overset{\mathrm{d}}{=} \boldsymbol{R}\boldsymbol{\varepsilon}$, $\boldsymbol{R} \in \mathcal{R}$. Notice that this assumption indeed holds for the standard normal location model, so that Assumption 1 is satisfied.

We used the following tests, where the abbreviation between brackets corresponds to the column names in the simulation tables. The tests are grouped by the computational effort: the number of Monte Carlo draws and the order of the subgroup, both indicated with the notation M.

- Several benchmark tests to provide upper bounds on the power:
  - A Z-test, which exploits the knowledge that $\boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$ (Z).
  - A t-test, which exploits knowledge about the orthogonal invariance of the distribution of $\boldsymbol{\varepsilon}$ (t).
  - A group-invariance test based on $\mathcal{R}$ (if computationally feasible) or a Monte Carlo sign-flipping test based on 1000 draws ($\mathcal{R}$ or MC $\mathcal{R}$).
- Tests with $M = n$:
  - An oracle subgroup-invariance test (Oracle).
  - An n-Monte Carlo Z-test (MC Z).
  - An order n negative oracle subgroup-invariance test (Neg.).
  - An n-Monte Carlo $\mathcal{R}$-invariance test (MC $\mathcal{R}$).
- Tests with $M = 2n$:
  - An order $2n$ negative oracle subgroup-invariance test (Neg.).
  - A $2n$-Monte Carlo $\mathcal{R}$-invariance test (MC $\mathcal{R}$).
- Tests with $M = 4n$:
  - A near-oracle subgroup-invariance test based on a subgroup produced by Algorithm 1 (NOS).
  - A $4n$-Monte Carlo $\mathcal{R}$-invariance test (MC $\mathcal{R}$).

We considered $n \in \{8, 16, 32, 64, 128\}$, to guarantee the existence of oracle subgroups of $\mathcal{R}$ of order n, by Theorem 8. We chose the level α such that αn is integer and $\alpha \approx .05$. This ensures Assumption 2 holds, so that all tests we considered have size α. The parameter μ was chosen such that the power is sufficiently far away from α and 1. For each setting and test, we generated $\boldsymbol{x}$ 10^6 times, independently across iterations, tests and settings, and recorded the proportion of times the tests rejected the null hypothesis. The results are reported in Tables 1 to 5.

Our findings are as follows. As expected, the rejection proportion under H_0 is approximately α for all tests, because all tests are exact. The benchmark Z- and t-tests outperform the other tests, which is unsurprising as they make explicit use of information about normality and orthogonal invariance, respectively, to which the other tests do not have access. The

1000-MC $\mathcal{R}$ test and $\mathcal{R}$-invariance tests perform similarly to the t-test. In line with Theorem 6, the oracle subgroup-invariance tests have the same power as the MC Z-tests. Furthermore, they are outperformed by the negative oracle subgroup tests.

We now turn to the comparison of the (sub)group-invariance tests and the Monte Carlo group-invariance tests. All order M subgroup-invariance tests we consider outperform the corresponding M-MC $\mathcal{R}$ tests. This is especially the case if M and n are small, where we find a power gap of over 6 percentage points between the negative oracle subgroup-invariance and n-MC tests in the most extreme case where $\mu = .7$, $n = 8$.

Notice that as M increases, both the MC sample (without replacement) from $\mathcal{R}$ and subgroup converge to the entire group $\mathcal{R}$. Hence, the M-MC $\mathcal{R}$ and order M subgroup-invariance tests both converge towards the $\mathcal{R}$-invariance test as M increases. Therefore, we expect the power gap between the two tests to decrease as M increases. This is indeed what is observed.

While the power gap between the order M subgroup-invariance tests and M-MC $\mathcal{R}$ tests narrows with M, the power of the subgroup-invariance tests appears to converge much more rapidly to the power of $\mathcal{R}$-invariance tests. In particular, we find that in each case we consider, the power of the M-MC $\mathcal{R}$ test is closer to that of the best order $M/2$ subgroup-invariance test than the power of the corresponding order M subgroup-invariance test. Furthermore, in many cases the M-MC $\mathcal{R}$ test is outperformed by the best order $M/2$ subgroup-invariance test. This means we could halve M if we switch from a MC $\mathcal{R}$ test to a subgroup-invariance test, and retain a similar power. An analogous pattern is found for fixed M, when comparing the relative power difference to the t-test, which serves as an upper bound of the tests and proxy for the $\mathcal{R}$-invariance test. In particular, we find that in all cases the power difference between the M-MC $\mathcal{R}$ test to the t-test is 1.5 to 5 times larger than power difference between the best order M subgroup-invariance test and the t-test.

				$M = n = 8$				$M = 2n = 16$		$M = 4n = 32$	
μ	Z	t	$\mathcal{R}$	Oracle	MC Z	Neg.	MC $\mathcal{R}$	Neg.	MC $\mathcal{R}$	NOS	MC $\mathcal{R}$
.0	.12587	.12543	.12491	.12483	.12531	.12511	.12533	.12463	.12494	.12460	.12539
.3	.38145	.36669	.36486	.33627	.33643	.34395	.32267	.36149	.34346	.36228	.35488
.5	.60173	.57755	.57593	.52532	.52276	.53296	.49510	.56675	.53448	.57301	.55591
.7	.79648	.76768	.76700	.70453	.70425	.71680	.65678	.75625	.71144	.76114	.74068

TABLE 1

$n = 8$, 10^6 *simulations*, $\alpha = 1/8 = .125$ *and* $|\mathcal{R}| = 2^8 = 256$.

			$M = 1000$	$M = n = 16$				$M = 2n = 32$		$M = 4n = 64$	
μ	Z	t	MC $\mathcal{R}$	Oracle	MC Z	Neg.	MC $\mathcal{R}$	Neg.	MC $\mathcal{R}$	NOS	MC $\mathcal{R}$
.0	.06223	.06274	.06290	.06261	.06292	.06221	.06276	.06282	.06259	.06274	.06225
.3	.36921	.35105	.35172	.32162	.32090	.32429	.30469	.34416	.32563	.34858	.33771
.5	.67925	.65038	.65130	.59473	.59520	.59972	.55808	.63810	.60127	.64282	.62578
.7	.89667	.87576	.87485	.82502	.82513	.82940	.77991	.86511	.83035	.86929	.85370

TABLE 2

$n = 16$, 10^6 *simulations*, $\alpha = 1/16 = .0625$.

7. Application: fMRI data. In this article we have mainly focused on the problem of testing a single hypothesis. However, the idea of using near-oracle subgroups directly extends to permutation-based multiple testing methods (Westfall and Young, 1993; Tusher, Tibshirani and Chu, 2001; Meinshausen, 2006; Pesarin and Salmaso, 2010; Meinshausen et al., 2011; Hemerik and Goeman, 2018a; Blanchard, Neuvial and Roquain, 2020). Such methods allow testing a large number of hypotheses simultaneously. A main advantage of permutation-based

			$M = 1000$	$M = n = 32$				$M = 2n = 64$		$M = 4n = 128$	
μ	Z	t	MC $\mathcal{R}$	Oracle	MC Z	Neg.	MC $\mathcal{R}$	Neg.	MC $\mathcal{R}$	NOS	MC $\mathcal{R}$
.0	.06206	.06252	.06302	.06289	.06257	.06264	.06280	.06225	.06251	.06216	.06299
.3	.56387	.55103	.55208	.52535	.52520	.52789	.51110	.54648	.53152	.54867	.54139
.4	.76734	.75309	.75342	.72365	.72298	.72674	.70666	.74676	.72930	.74997	.74191
.5	.90212	.89323	.89239	.86899	.86932	.87126	.85343	.88734	.87414	.89016	.88319

TABLE 3

$n = 32$. 10^6 *simulations.* $\alpha = 2/32 = .0625$.

			$M = 1000$	$M = n = 64$				$M = 2n = 128$		$M = 4n = 256$	
μ	Z	t	MC $\mathcal{R}$	Oracle	MC Z	Neg.	MC $\mathcal{R}$	Neg.	MC $\mathcal{R}$	NOS	MC $\mathcal{R}$
.0	.04695	.04653	.04687	.04680	.04681	.04675	.04689	.04675	.04668	.04703	.04720
.2	.46994	.46242	.46191	.44659	.44765	.44811	.43941	.45895	.45183	.46121	.45681
.3	.76499	.75655	.75592	.73785	.73748	.73901	.72858	.75187	.74250	.75503	.75005
.4	.93598	.93154	.93132	.91993	.92019	.92045	.91379	.92932	.92312	.92988	.92818

TABLE 4

$n = 64$. 10^6 *simulations.* $\alpha = 3/64 = .046875$.

			$M = 1000$	$M = n = 128$				$M = 2n = 256$		$M = 4n = 512$	
μ	Z	t	MC $\mathcal{R}$	Oracle	MC Z	Neg.	MC $\mathcal{R}$	Neg.	MC $\mathcal{R}$	NOS	MC $\mathcal{R}$
.0	.04650	.04691	.04695	.04674	.04668	.04673	.04651	.04696	.04661	.04684	.04695
.15	.50894	.50498	.50331	.49668	.49663	.49662	.49247	.50319	.49862	.50365	.50190
.2	.72134	.71750	.71603	.70804	.70760	.70884	.70309	.71548	.70957	.71620	.71327
.25	.87495	.87278	.87163	.86480	.86444	.86452	.86116	.87065	.86611	.87102	.86942

TABLE 5

$n = 128$. 10^6 *simulations.* $\alpha = 6/128 = .046875$.

multiple testing method as compared to other multiple testing methods, is that they take into account the dependence structure in the data, leading to good power properties (Westfall and Young, 1993; Hemerik and Goeman, 2018a; Hemerik, Solari and Goeman, 2019). Like permutation tests, these methods require using a group of transformations or transformations randomly sampled from a group. In the present data analysis example, we will use a near-oracle subgroup within a permutation-based multiple testing method from Hemerik and Goeman (2018a).

Permutation-based multiple testing methods can be computationally demanding. The first reason is that for each permutation we need to compute a large number of test statistics, equal to the total number of hypotheses. The second reason is that some multiple testing methods apply sophisticated combinatorical algorithms to the matrix of computed test statistics.

The multiple testing method that we will use here is the *approximate closed testing method* from Hemerik and Goeman (2018a, p.144). An implementation has been available in Hemerik and Goeman (2018c). Recently a faster implementation (of the closed testing method) has become available, which we use here, see Koning (2022c).

We applied the multiple testing metod to a high-dimensional functional magnetic resonance imaging (fMRI) dataset. The original data are available at `https://openfmri.org` and we used the pre-prosessed data from Andreella (2021) (for details see Smeets et al., 2013; Andreella et al., 2020). The dataset contains measurements for $n = 29$ subjects who interchangeably looked a images of food and non-food. Thus, the subjects were exposed to two experimental conditions. For 152472 voxels making up the brain, the activity was recored while the subjects looked at the images. For each subject and for voxel i, a difference statistic was computed with mean μ_i say. For voxel i, we define the corresponding null hypothesis to be $H_i : \mu_i = 0$, which mean that there is no difference in mean response between the two conditions.

For each voxel we computed a one-sample t-statistic based on the 29 measurements. We assumed that under H_i, the correponding t-statistic was symmetric about 0. This framework allows us to use the permutation-based multiple testing method from Hemerik and Goeman (2018a), where rather than permutation, we used sign-flipping (we took $|\mathcal{S}| = 10^3$, see Hemerik and Goeman, 2018a). The method requires the user to set a rejection threshold. All hypotheses with test statistics exceeding this threshold are rejected. For our illustration purposes, we simply set the the threshold to 3, so that all hypotheses with t-statistic above 3 or below -3 were rejected. This led to 10580 rejected hypotheses, i.e. 10580 voxels were selected as seemingly 'activated'.

The multiple testing method provides a (median unbiased) estimate and a 95%-confidence upper bound for the *false discovery proportion* (FDP), which is the fraction of incorrect rejections among all 10580 rejected null hypotheses (Hemerik and Goeman, 2018a). As our simulations show (see Section 6), by using a near-oracle subgroup within a sign-flip test, we only require about half the number of sign-flips compared to using random sign-flips, to achieve the same power. This extends to permutation-based multiple testing methods, which are based on the same group invariance principle.

This means that where one would use perhaps 2000 random transformations, we can instead use a near-oracle subgroup of cardinality 1024 and often still have equal or higher power, as well as improved replicability. We did the latter in this example. Note that this reduces the computation time with about 50% on average, compared to using 2000 random sign-flips. We conveniently obtained the sign-flipping matrix encoding the near-oracle subgroup from Koning (2022a). Thus, it was straightforward to implement the near-oracle subgroup within the multiple testing method.

In our example the computation time on a laptop was 6 minutes, counting from the moment that the multiple testing method was called. The method estimated the FDP to be $360/10580 \approx .034$ and provided a 95%-confidence upper bound for the FDP of $2306/10580 \approx .217$. This means that we can be confident that most of the 10580 selected voxels are truly activated, i.e. that their activity depends on the experimental condition (looking at food vs. looking at non-food). Interpreting the results further is beyond the scope of this paper, but we have illustrated that near-oracle subgroups can easily be used within permutation-based multiple testing procedures.

8. Discussion.

Group invariance tests, such as tests based on rotation, permutation or sign-flipping, require the set of transformations used to have a group structure. If there is no group structure, these tests tend to be conservative or anti-conservative (Southworth, Kim and Owen, 2009). Consequently, conditional on a random subset of transformations sampled from a group, a group invariance test is usually not exact. Nevertheless, on average over all possible sets of random transformations, the level is exactly α (Hemerik and Goeman, 2018b). It is in this sense that testing with random transformations is valid. However, as is known (Dwass, 1957; Vesely, Finos and Goeman, 2021), the power of such tests tends to be lower than when the full group of transformations is used if the number of random transformations is small, and the tests are computationally demanding if the number is large.

In this paper, we have proposed using a *fixed* set of transformations that constitutes a subgroup of the group with respect to which invariance is tested. As a group-invariance test based on any such subgroup is still exact (not just on average), this allows us to choose a subgroup of transformations that possesses desirable properties for the test statistic that is considered. In a generalized location model, we consider appropriately designed subgroups. Compared to random sampling, these are more representative of the full group, in the sense that the resulting power is closer to that of the test using the full group.

We have found that by taking the subgroup to be an *oracle subgroup*, a phenomenon of signal leaking into the reference distribution (Dobriban, 2021) is entirely avoided. This tends to lead to improved power, compared to using random transformations. Oracle subgroups cannot have more than n elements, so in case a larger subgroup is required, we propose using a *near-oracle subgroup*. It turns out that this subgroup is such that it generates points on an $(n-1)$-sphere that are spread out as much as possible, in a way that no two of its elements are very close to each other (see also the literature on group codes and spherical codes, Slepian, 1968; Conway and Sloane, 1998).

As an illustration, we have focused on sign-flipping transformations. These are often used in practice when the data entries are symmetric about 0 under the null, as is often the case for paired data (Fisher, 1935) and fMRI data (Winkler et al., 2014; Andreella et al., 2020). We have illustrated in detail how to construct (near-)oracle subgroups in this case. For many different sample sizes and subgroup sizes, one can directly download such (near-)oracle subgroups from Koning (2022a). In addition, (near-)oracle subgroups can be constructed using the R-package available in Koning (2022b).

Simulations suggest that in case of sign-flipping, to attain a given power we usually require at most half the number of transformations required when using random transformations. Since group invariance-based methods for high-dimensional data can be computationally very time-demanding, this allows for a substantial gain in computation speed (Westfall and Troendle, 2008; Hemerik, Solari and Goeman, 2019; Koning, 2019).

There are many applications, for example two-sample comparison testing (see Appendix A), where regular permutations are used, rather than sign-flipping. In that case, constructing near-oracle subgroups of a given size is more challenging than in case of sign-flipping. We plan to address this issue in future work.

APPENDIX A: TWO-SAMPLE COMPARISON

In this section, we explain how the two-sample comparison problem fits into the location model described in Section 2. In particular, we consider testing the equality of means μ_1 and μ_2 of two samples which we will denote by $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. The samples contain m_1 and m_2 observations respectively, where $m_1 + m_2 = n$. Define $\mathbf{1}_m = n^{-1/2}(1, 1, \dots, 1)'$. We represent the two samples as the first and second part of a vector $\boldsymbol{x}$, without loss of generality:

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{1}_{m_1}\mu_1 \\ \mathbf{1}_{m_2}\mu_2 \end{bmatrix} + \widetilde{\boldsymbol{\varepsilon}},$$

where $\widetilde{\boldsymbol{\varepsilon}}$ is an exchangeable random vector, that is, $\widetilde{\boldsymbol{\varepsilon}} \overset{d}{=} \boldsymbol{P}\widetilde{\boldsymbol{\varepsilon}}$ for every permutation matrix $\boldsymbol{P}$. We want to test the null hypothesis $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 > \mu_2$.

We now show how this fits into the location model described in Section 2. Define $\mu = \frac{1}{2}(\mu_1 - \mu_2)$, and let $\boldsymbol{\iota} = (\mathbf{1}_{m_1}, -\mathbf{1}_{m_2})$ for the remainder of this section. Notice that the hypotheses are equivalent to $H_0 : \mu = 0$ and $H_1 : \mu > 0$. Define $\boldsymbol{\varepsilon} = \widetilde{\boldsymbol{\varepsilon}} + \mathbf{1}_n \frac{1}{2}(\mu_1 + \mu_2)$, so that

$$\boldsymbol{x} = \begin{bmatrix} \mathbf{1}_{m_1}\mu_1 \\ \mathbf{1}_{m_2}\mu_2 \end{bmatrix} - \mathbf{1}_n \tfrac{1}{2}(\mu_1 + \mu_2) + \boldsymbol{\varepsilon} = \boldsymbol{\iota}\mu + \boldsymbol{\varepsilon}.$$

Finally, observe that $\boldsymbol{\varepsilon} \overset{d}{=} \boldsymbol{P}\boldsymbol{\varepsilon}$, for every permutation matrix $\boldsymbol{P}$, as addition of a vector with equal elements does not affect exchangeability.

A.1. Oracle subgroups for two-sample comparison.

In this section, we describe how oracle subgroups of $\mathcal{P}$ with respect to $\boldsymbol{\iota}$ relate to oracle subgroups of $\mathcal{R}$ with respect to $\mathbf{1}_n$. In particular, we show how an oracle subgroup of $\mathcal{R}$ with respect to $\mathbf{1}_n$ can be used to construct the matrix representation of an oracle subgroup of $\mathcal{P}$ with respect to $\boldsymbol{\iota}$, as captured by Theorem 9. This is useful because it is sufficient to have the matrix representation in order to perform the associated group-invariance test. Furthermore, such oracle subgroups of $\mathcal{R}$ with respect to $\mathbf{1}_n$ can be found in the database of Koning (2022a) and constructed using Koning (2022b).

THEOREM 9. *Let $\mathcal{S}$ be an oracle subgroup of $\mathcal{R}$ with respect to $\mathbf{1}_n$ and assume that there exists some $\boldsymbol{R}^* \in \mathcal{S}$ such that $\boldsymbol{R}^*\mathbf{1}_n = \boldsymbol{\iota}$. Let $\mathcal{S}^*$ be a subgroup of $\mathcal{R}$ with elements in $\mathcal{S} \setminus \{\boldsymbol{R}^*\}$, such that its order is maximal. Then, the matrix representation $\mathfrak{S}^*$ of $\mathcal{S}^*$ is also the matrix representation of an oracle subgroup of $\mathcal{P}$ with respect to $\boldsymbol{\iota}$.*

PROOF. Observe that $\boldsymbol{R}^*\boldsymbol{R}\boldsymbol{R}^* \in \mathcal{S}$ for each $\boldsymbol{R} \in \mathcal{S}$, due to group closure under composition. Hence, $\boldsymbol{\iota}'\boldsymbol{R}\boldsymbol{\iota} = \mathbf{1}_n'\boldsymbol{R}^*\boldsymbol{R}\boldsymbol{R}^*\mathbf{1}_n = 0$ for all $\boldsymbol{R} \in \mathcal{S} \setminus \{\boldsymbol{I}\}$, so that $\mathcal{S}$ is also an oracle subgroup with respect to $\boldsymbol{\iota}$.

Observe that the diagonal elements of $\boldsymbol{R} \in \mathcal{S} \setminus \{\boldsymbol{I}\}$ are all permutations of each other. Taking $\boldsymbol{\iota} = \boldsymbol{R}^*\mathbf{1}_n$ as a reference point, multiplication of $\boldsymbol{\iota}$ with any $\boldsymbol{R} \in \mathcal{S}^*$ coincides with permutation of its elements. Notice that $\mathcal{S}^*$ acts as a group of permutations on the elements of $\boldsymbol{\iota}$. This means there exists a subgroup $\widehat{\mathcal{S}}$ of $\mathcal{P}$ that is isomorphic to $\mathcal{S}^*$, such that for each $\boldsymbol{R} \in \mathcal{S}^*$ there exists a unique $\boldsymbol{P} \in \widehat{\mathcal{S}}$ such that $\boldsymbol{R}\boldsymbol{\iota} = \boldsymbol{P}\boldsymbol{\iota}$. As a consequence, $\boldsymbol{\iota}'\boldsymbol{P}\boldsymbol{\iota} = 0$, for all $\boldsymbol{P} \in \widehat{\mathcal{S}} \setminus \{\boldsymbol{I}\}$, so that $\widehat{\mathcal{S}}$ is an oracle subgroup of $\mathcal{P}$ with respect to $\boldsymbol{\iota}$. In addition, the matrix representation $\mathfrak{S}^*$ of $\mathcal{S}^*$ coincides with the matrix representation of $\widehat{\mathcal{S}}$. □

The inclusion of $\boldsymbol{R}^*$ in $\mathcal{S}$ is without loss of generality in terms of the applicability, as the elements of $\boldsymbol{\iota}$ (and $\boldsymbol{x}$) can simply be rearranged to ensure that $\boldsymbol{R}^* \in \mathcal{S}$, since each element of $\mathcal{S} \setminus \{\boldsymbol{I}\}$ has an equal amount of 1s and -1s on the diagonal. In addition, note that $2|\mathcal{S}^*| = |\mathcal{S}|$ by Proposition 7.

Finally, we include an example of an oracle subgroup of $\mathcal{P}$ with respect to $\boldsymbol{\iota}$.

EXAMPLE 6. *An example of an oracle subgroup of $\mathcal{P}$ with respect to $\boldsymbol{\iota} = (\mathbf{1}_{n/2}, -\mathbf{1}_{n/2})$ for $n = 2m_1 = 4$, is $\mathcal{S} = \{\boldsymbol{I}, \boldsymbol{P}_1\}$, where*

$$\boldsymbol{P}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Notice that $\boldsymbol{P}_1$ swaps exactly half of the elements from each sample to the other sample. This pattern is a necessary condition for $\mathcal{S}$ to be an oracle subgroup that generalizes to larger n. This means we require that $m_1 = m_2$ and m_1 is even for an oracle subgroup to exist.

APPENDIX B: PROOFS

We first present the following Lemma, which is used in later proofs.

LEMMA 1. *Let $\widetilde{\boldsymbol{G}}$ be uniformly (Haar) distributed over a subgroup $\mathcal{G}$ of the orthogonal group. Define $\boldsymbol{g} = \widetilde{\boldsymbol{G}}\boldsymbol{\iota}$. Let $\boldsymbol{a} \in \mathbb{R}^n$. And use $q_\alpha(O^{\mathcal{G}}_{\boldsymbol{a}})$ to denote the α upper-quantile of the distribution of $\boldsymbol{g}'\boldsymbol{a}$. Then,*

$$q_\alpha(O^{\mathcal{G}}_{\boldsymbol{a}}) = \|\boldsymbol{a}\|_2 q_\alpha(O^{\mathcal{G}}_{\boldsymbol{a}/\|\boldsymbol{a}\|_2}).$$

Moreover, if $\mathcal{G} = \mathcal{H}$, then

$$q_\alpha(O^{\mathcal{G}}_{\boldsymbol{a}}) = \|\boldsymbol{a}\|_2 q_\alpha(O^{\mathcal{H}}_{\boldsymbol{\iota}}).$$

PROOF. We have that $q^{\mathcal{H}}_\alpha(O_{\boldsymbol{a}})$ is the α upper quantile of the distribution of

$$\boldsymbol{g}'\boldsymbol{a} = \|\boldsymbol{a}\|_2(\boldsymbol{g}'\boldsymbol{a}/\|\boldsymbol{a}\|_2) = \|\boldsymbol{a}\|_2\boldsymbol{g}'(\boldsymbol{a}/\|\boldsymbol{a}\|_2).$$

The first statement follows from noticing that $q_\alpha(\|\boldsymbol{a}\|_2 O^{\mathcal{G}}_{\boldsymbol{a}/\|\boldsymbol{a}\|_2}) = \|\boldsymbol{a}\|_2 q_\alpha(O^{\mathcal{G}}_{\boldsymbol{a}/\|\boldsymbol{a}\|_2})$. The second statement follows from the fact that $O^{\mathcal{H}}_{\boldsymbol{a}/\|\boldsymbol{a}\|_2} = O^{\mathcal{H}}_{\boldsymbol{u}}$, for any unit vector $\boldsymbol{u}$. □

B.1. Proof of Proposition 2.

PROOF. From the second part of the proof of Proposition 3, we learn that $q_\alpha(O_{\boldsymbol{x}}^{*\mathcal{H}}) = q_\alpha(O_{\boldsymbol{x}}^{\mathcal{H}})$. Hence, $\overline{\phi}_\alpha^{\mathcal{H}} = \mathbb{I}\{\boldsymbol{\iota}'\boldsymbol{x} > q_\alpha(O_{\boldsymbol{x}}^{\mathcal{H}})\}$. Notice that as $\alpha \leq .5$ then $q_\alpha(O_{\boldsymbol{x}}^{\mathcal{H}}) \geq 0$, because the distribution of $\boldsymbol{\iota}'\boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ orthogonal-invariant is symmetric about the origin.

We now consider two cases: $\boldsymbol{\iota}'\boldsymbol{x} \leq 0$ and $\boldsymbol{\iota}'\boldsymbol{x} > 0$. If $\boldsymbol{\iota}'\boldsymbol{x} \leq 0$, then $\phi_\alpha^{\mathcal{H}} = \overline{\phi}_\alpha^{\mathcal{H}}$. It remains to consider the case $\boldsymbol{\iota}'\boldsymbol{x} > 0$. From Lemma 1, we have $q_\alpha(O_{\boldsymbol{x}}) = \|\boldsymbol{x}\|_2 q_\alpha(O_{\boldsymbol{\iota}})$. As $q_\alpha(O_{\boldsymbol{\iota}}) \geq 0$ and the derivative of $\|\boldsymbol{x}\|_2$ with respect to μ is $\boldsymbol{\iota}'\boldsymbol{x}/\|\boldsymbol{x}\|_2 > 0$, we have that $q_\alpha(O_{\boldsymbol{x}})$ is increasing in μ. Hence, $q_\alpha(O_{\boldsymbol{x}}) \geq q_\alpha(O_{\boldsymbol{\varepsilon}})$. As a consequence, $\phi_\alpha^{\mathcal{H}} \leq \overline{\phi}_\alpha^{\mathcal{H}}$ □

B.2. Proof of Theorem 1. In order to present the proof of Theorem 1, we first prove two lemmas. Let the map $f : (-1, 1) \to \mathbb{R}$ be defined as

$$f(y) = \frac{y}{\sqrt{1-y^2}}.$$

Notice that f is a strictly increasing function.

LEMMA 2. *If $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\iota}\mu, \boldsymbol{I})$, then*

$$\sqrt{n-1} f\left(\frac{\boldsymbol{\iota}'\boldsymbol{x}}{\|\boldsymbol{x}\|_2}\right) \sim t_{n-1,\mu},$$

where $t_{n-1,\mu}$ is a non-central t-distribution with $(n-1)$ degrees of freedom and non-centrality parameter μ.

PROOF. This follows immediately from observing that

$$\sqrt{n-1} f\left(\frac{\boldsymbol{\iota}'\boldsymbol{x}}{\|\boldsymbol{x}\|_2}\right) = \sqrt{n-1}\frac{\boldsymbol{\iota}'\boldsymbol{x}}{\sqrt{\boldsymbol{x}'(\boldsymbol{I}-\boldsymbol{\iota}\boldsymbol{\iota}')\boldsymbol{x}}},$$

and noting that the right hand side follows a $t_{n-1,\mu}$ distribution. □

The following lemma was already proven by Efron (1969) by generalizing geometric arguments made by Fisher (1925) for the normal distribution. We provide a short proof using Lemma 2.

LEMMA 3. *If $\boldsymbol{\varepsilon} \neq \boldsymbol{0}$ has an orthogonal invariant distribution, then*

$$\sqrt{n-1} f\left(\frac{\boldsymbol{\iota}'\boldsymbol{\varepsilon}}{\|\boldsymbol{\varepsilon}\|_2}\right) \sim t_{n-1},$$

where t_{n-1} is a (central) t-distribution with $(n-1)$ degrees of freedom.

PROOF. Note by Lemma 2 that if $\boldsymbol{\varepsilon}$ is normally distributed, then

$$\sqrt{n-1} f\left(\frac{\boldsymbol{\iota}'\boldsymbol{\varepsilon}}{\|\boldsymbol{\varepsilon}\|_2}\right) \sim t_{n-1}.$$

The left-hand side only depends on $\boldsymbol{\varepsilon}$ through $\boldsymbol{\varepsilon}/\|\boldsymbol{\varepsilon}\|_2$. As the normal distribution is orthogonal invariant, $\boldsymbol{\varepsilon}/\|\boldsymbol{\varepsilon}\|_2$ is uniformly distributed over the unit-sphere in dimension n. This holds for any orthogonal invariant distribution. □

Using Lemma 3 and fact that f is strictly increasing, we obtain Theorem 1 as follows.

PROOF OF THEOREM 1. Exclude the case that $\boldsymbol{x} = \mathbf{0}$, where the t-test is poorly defined. We can show the equivalence of the tests by comparing the rejection events:

$$\begin{aligned}\{\boldsymbol{\iota}'\boldsymbol{x} > q_\alpha(O_{\boldsymbol{x}}^{\mathcal{H}})\} &= \left\{\boldsymbol{\iota}'\boldsymbol{x}/\|\boldsymbol{x}\|_2 > q_\alpha(O_{\boldsymbol{\iota}}^{\mathcal{H}})\right\} \\ &= \left\{\sqrt{n-1}f\left(\boldsymbol{\iota}'\boldsymbol{x}/\|\boldsymbol{x}\|_2\right) > \sqrt{n-1}f\left(q_\alpha(O_{\boldsymbol{\iota}}^{\mathcal{H}})\right)\right\} \\ &= \left\{\sqrt{n-1}\frac{\boldsymbol{\iota}'\boldsymbol{x}}{\sqrt{\boldsymbol{x}'(\boldsymbol{I}-\boldsymbol{\iota}\boldsymbol{\iota}')\boldsymbol{x}}} > q_\alpha^{t_{n-1}}\right\}.\end{aligned}$$

where $q_\alpha(O_{\boldsymbol{x}}^{\mathcal{H}}) = \|\boldsymbol{x}\|_2 q_\alpha(O_{\boldsymbol{\iota}}^{\mathcal{H}})$ from Lemma 1, $q_\alpha^{t_{n-1}}$ is the α upper-quantile of t_{n-1}-distribution, and $q_\alpha^{t_{n-1}} = \sqrt{n-1}f\left(q_\alpha(O_{\boldsymbol{\iota}}^{\mathcal{H}})\right)$ follows from Lemma 3 and the fact that f is strictly increasing. □

B.3. Proof of Proposition 3.

To prove Proposition 3 we first prove a lemma.

LEMMA 4. *Let $\boldsymbol{\iota}$ be a unit vector and suppose that $\boldsymbol{h}$ is uniformly distributed on the unit sphere. Then, $\boldsymbol{\iota}'\boldsymbol{h} \sim Beta(\frac{n-1}{2}, \frac{n-1}{2})$.*

PROOF. We first show that $\boldsymbol{\iota}'\boldsymbol{h}$ has the same distribution as $\boldsymbol{b}'\boldsymbol{h}$, where $\boldsymbol{b}$ is uniformly distributed on the unit sphere. Let $\widetilde{\boldsymbol{H}}$ be uniformly distributed on $\mathcal{H}$. Observe that $\boldsymbol{h}$ is orthogonal-invariant, so that $\widetilde{\boldsymbol{H}}\boldsymbol{h} \overset{\mathrm{d}}{=} \boldsymbol{h}$, and so $\boldsymbol{\iota}'\widetilde{\boldsymbol{H}}\boldsymbol{h} \overset{\mathrm{d}}{=} \boldsymbol{\iota}'\boldsymbol{h}$. Finally, observe that $\widetilde{\boldsymbol{H}}'\boldsymbol{\iota}$ is uniformly distributed on the unit sphere.

Next, we establish that $\boldsymbol{\iota}'\widetilde{\boldsymbol{H}}\boldsymbol{h}$ shares its distribution with a sample correlation coefficient. Suppose $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent multivariate standard normally distributed n-vectors. Then, $\boldsymbol{x}/\|\boldsymbol{x}\|_2$ and $\boldsymbol{y}/\|\boldsymbol{y}\|_2$ are uniformly distributed over the unit sphere, so $\boldsymbol{x}'\boldsymbol{y}/(\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2) \overset{\mathrm{d}}{=} \boldsymbol{\iota}'\widetilde{\boldsymbol{H}}\boldsymbol{h}$. Notice that $\boldsymbol{x}'\boldsymbol{y}/(\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2)$ is the sample correlation coefficient between $\boldsymbol{x}$ and $\boldsymbol{y}$.

The sample correlation coefficient between two independent multivariate standard normally distributed n-vectors is Beta$(\frac{n-1}{2}, \frac{n-1}{2}, -1, 1)$ (Fisher, 1915; Hotelling, 1953), which proves the result. The reasoning can be summarized as:

$$\boldsymbol{\iota}'\boldsymbol{h} \overset{\mathrm{d}}{=} \boldsymbol{\iota}'\widetilde{\boldsymbol{H}}\boldsymbol{h} \overset{\mathrm{d}}{=} \boldsymbol{x}'\boldsymbol{y}/(\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2) \sim \mathrm{Beta}\left(\tfrac{n-1}{2}, \tfrac{n-1}{2}, -1, 1\right).$$

□

PROOF OF PROPOSITION 3. We first show that $q_\alpha(O_{\boldsymbol{\iota}}^{\mathcal{H}}) = \beta_\alpha^{n-1,n-1}$. Observe that $q_\alpha(O_{\boldsymbol{\iota}}^{\mathcal{H}})$ is the α upper-quantile of $\boldsymbol{h}'\boldsymbol{\iota}$, where $\boldsymbol{\iota}$ is a unit n-vector and $\boldsymbol{h}$ is uniformly distributed on the unit sphere in dimension n. By Lemma 4, we have $\boldsymbol{h}'\boldsymbol{\iota} \sim \mathrm{Beta}(\frac{n-1}{2}, \frac{n-1}{2}, -1, 1)$. Hence, $q_\alpha(O_{\boldsymbol{\iota}}^{\mathcal{H}}) = \beta_\alpha^{n-1,n-1}$.

As a consequence, $\mathbb{I}\{\boldsymbol{\iota}'\boldsymbol{x}/\widetilde{\sigma} > \sqrt{n}\beta_\alpha^{n-1,n-1}\} = \mathbb{I}\{\boldsymbol{\iota}'\boldsymbol{x}/\widetilde{\sigma} > \sqrt{n}q_\alpha(O_{\boldsymbol{\iota}}^{\mathcal{H}})\} = \mathbb{I}\{\boldsymbol{\iota}'\boldsymbol{x} > q_\alpha^{\mathcal{H}}(O_{\boldsymbol{\varepsilon}}^{\mathcal{H}})\}$, where the second equality follows from Lemma 1. Finally, notice that Beta$(\frac{n-1}{2}, \frac{n-1}{2}, -1, 1)$ has a continuous CDF for $n > 1$, so $q_\alpha(O_{\boldsymbol{\varepsilon}}^{\mathcal{H}}) = q_\alpha(O_{\boldsymbol{\varepsilon}}^{*\mathcal{H}})$. Thus $\mathbb{I}\{\boldsymbol{\iota}'\boldsymbol{x} > q_\alpha(O_{\boldsymbol{\varepsilon}}^{\mathcal{H}})\} = \mathbb{I}\{\boldsymbol{\iota}'\boldsymbol{x} > q_\alpha(O_{\boldsymbol{\varepsilon}}^{*\mathcal{H}})\}$. □

B.4. Proof of Theorem 2.

PROOF. Suppose that H_0 is true, so that $\boldsymbol{x} = \boldsymbol{\varepsilon}$. Let $O \in \mathcal{O}$ and suppose $\boldsymbol{\varepsilon} \in O$. By Assumption 1, we have that $q_\alpha(O)$ is the α upper-quantile of the O-conditional distribution of $\boldsymbol{\iota}'\boldsymbol{\varepsilon}$. By definition of the upper-quantile, the O-conditional expectation of $\phi(\boldsymbol{x}) = \phi(\boldsymbol{\varepsilon})$ is at most α. As O was arbitrarily given, this holds for every $O \in \mathcal{O}$. Integrating using the distribution over $\mathcal{O}$ therefore yields the same statements, unconditionally. □

B.5. Proof of Theorem 3.

PROOF. Define the notation $\Theta_{\boldsymbol{x}} = O_{\boldsymbol{x}} \setminus \{\boldsymbol{x}\}$ and $\boldsymbol{x}^{\delta_{\mathcal{G}}} = (1-\delta_{\mathcal{G}})\mu + \boldsymbol{\varepsilon}$. Observe that $\boldsymbol{\iota}'\boldsymbol{G}\boldsymbol{x} - \boldsymbol{\iota}'\boldsymbol{G}\boldsymbol{\varepsilon} = \boldsymbol{\iota}'\boldsymbol{G}\boldsymbol{\iota} \leq \mu\delta_{\mathcal{G}}$, for all $\boldsymbol{G} \in \mathcal{G} \setminus \{\boldsymbol{I}\}$, so that $q_\alpha(\Theta_{\boldsymbol{x}}) - q_\alpha(\Theta_{\boldsymbol{\varepsilon}}) \leq \delta_{\mathcal{G}}\mu$ and so $q_\alpha(\Theta_{\boldsymbol{x}}) \leq \delta\mu + q_\alpha(\Theta_{\boldsymbol{\varepsilon}})$. Using this inequality, we obtain

$$\begin{aligned}
\overline{\phi}^{\mathcal{G}}_\alpha(\boldsymbol{x}^{\delta_{\mathcal{G}}}) &:= \mathbb{I}\{\boldsymbol{x}^{\delta_{\mathcal{G}}} > q_\alpha(O^*_{\boldsymbol{\varepsilon}})\} \\
&= \mathbb{I}\{\boldsymbol{x}^{\delta_{\mathcal{G}}} > q_\alpha(\Theta_{\boldsymbol{\varepsilon}})\} \\
&= \mathbb{I}\{(1-\delta)\mu + \boldsymbol{\iota}'\boldsymbol{\varepsilon} > q_\alpha(\Theta_{\boldsymbol{\varepsilon}})\} \\
&= \mathbb{I}\{\boldsymbol{\iota}'\boldsymbol{x} > \delta\mu + q_\alpha(\Theta_{\boldsymbol{\varepsilon}})\} \\
&\leq \mathbb{I}\{\boldsymbol{\iota}'\boldsymbol{x} > q_\alpha(\Theta_{\boldsymbol{x}})\} \\
&= \mathbb{I}\{\boldsymbol{\iota}'\boldsymbol{x} > q_\alpha(O_{\boldsymbol{x}})\} \\
&=: \overline{\phi}^{\mathcal{G}}_\alpha(\boldsymbol{x}).
\end{aligned}$$

□

B.6. Proof of Proposition 4.

PROOF. Let $\boldsymbol{S}_1, \boldsymbol{S}_2 \in \mathcal{S}$ be arbitrary. We must show that $\boldsymbol{S}_1\boldsymbol{\iota} = \boldsymbol{S}_2\boldsymbol{\iota}$ if and only if $\boldsymbol{S}_1 = \boldsymbol{S}_2$. The 'right to left' implication is obvious. We will prove the converse by contradiction. Suppose that $\boldsymbol{S}_1 \neq \boldsymbol{S}_2$ and $\boldsymbol{S}_1\boldsymbol{\iota} = \boldsymbol{S}_2\boldsymbol{\iota}$. As $\mathcal{S}$ is a group and $\boldsymbol{S}_1'$ is the inverse of $\boldsymbol{S}_1$, $\boldsymbol{S}_1'\boldsymbol{S}_2 \in \mathcal{S}$. As $\boldsymbol{S}_1\boldsymbol{\iota} = \boldsymbol{S}_2\boldsymbol{\iota}$, we have $\boldsymbol{\iota}'\boldsymbol{S}_1'\boldsymbol{S}_2\boldsymbol{\iota} = 1$. By Definition 1, this implies $\boldsymbol{S}_1'\boldsymbol{S}_2 = \boldsymbol{I}$. By uniqueness of inverses, this means $\boldsymbol{S}_1 = \boldsymbol{S}_2$, which contradicts the assertion that $\boldsymbol{S}_1 \neq \boldsymbol{S}_2$. □

B.7. Proof of Theorem 4.

PROOF. We start with the 'left-to-right' implication. First, by Proposition 4, there exists a bijection between $\mathcal{S}$ and $O^{\mathcal{S}}_{\boldsymbol{\iota}}$, and so $|\mathcal{S}| = |O^{\mathcal{S}}_{\boldsymbol{\iota}}|$ by definition of cardinality. As the elements of $O^{\mathcal{S}}_{\boldsymbol{\iota}}$ are unit vectors, we only need to show that for every pair of elements $\boldsymbol{s}_1, \boldsymbol{s}_2$ of $O^{\mathcal{S}}_{\boldsymbol{\iota}}$, with $\boldsymbol{s}_1 \neq \boldsymbol{s}_2$ we have $\boldsymbol{s}_1'\boldsymbol{s}_2 = 0$. By Proposition 4, $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ correspond to two distinct matrices $\boldsymbol{S}_1, \boldsymbol{S}_2 \in \mathcal{S}$ for which $\boldsymbol{S}_1\boldsymbol{\iota} = \boldsymbol{s}_1$ and $\boldsymbol{S}_2\boldsymbol{\iota} = \boldsymbol{s}_2$. As $\boldsymbol{S}_1, \boldsymbol{S}_2 \in \mathcal{S}$ and $\mathcal{S}$ is a group, we have $\boldsymbol{S}_1'\boldsymbol{S}_2 \in \mathcal{S}$. As $\boldsymbol{S}_1 \neq \boldsymbol{S}_2$, we have $\boldsymbol{S}_1'\boldsymbol{S}_2 \neq \boldsymbol{I}$. So, by Definition 1, $\boldsymbol{\iota}'\boldsymbol{S}_1'\boldsymbol{S}_2\boldsymbol{\iota} = 0$. Hence, $\boldsymbol{s}_1'\boldsymbol{s}_2 = 0$.

Now we prove the 'right-to-left' implication. As $|\mathcal{S}| = |O^{\mathcal{S}}_{\boldsymbol{\iota}}|$ and $O^{\mathcal{S}}_{\boldsymbol{\iota}}$ has elements $\boldsymbol{S}\boldsymbol{\iota}$, $\boldsymbol{S} \in \mathcal{S}$, the map $\boldsymbol{S} \mapsto \boldsymbol{S}\boldsymbol{\iota}$ must be a bijection. As $\boldsymbol{\iota}$ is an element of $O^{\mathcal{S}}_{\boldsymbol{\iota}}$ and $O^{\mathcal{S}}_{\boldsymbol{\iota}}$ has orthogonal elements, we have $\boldsymbol{\iota}'\boldsymbol{s} = 0$ for all $\boldsymbol{s} \in O^{\mathcal{S}}_{\boldsymbol{\iota}}$ with $\boldsymbol{s} \neq \boldsymbol{\iota}$. By the bijection, each $\boldsymbol{s} \in O^{\mathcal{S}}_{\boldsymbol{\iota}}$ corresponds to some unique element $\boldsymbol{S} \in \mathcal{S}$, for which $\boldsymbol{s} = \boldsymbol{S}\boldsymbol{\iota}$. Therefore, $\boldsymbol{\iota}'\boldsymbol{S}\boldsymbol{\iota} = \boldsymbol{\iota}'\boldsymbol{s} = 0$, for all $\boldsymbol{S} \neq \boldsymbol{I}$. Hence, $\mathcal{S}$ is an oracle subgroup of $\mathcal{H}$. □

B.8. Proof of Theorem 5.

We first prove two lemmas, which we then combine to prove Theorem 5.

LEMMA 5. *There exists a cyclic subgroup $\mathcal{S}$ of the permutation group $\mathcal{P}$ of order $1 < p \leq n$ for which $\boldsymbol{e}_1$ is not a fixed point. This subgroup $\mathcal{S}$ is an oracle subgroup of $\mathcal{H}$ of order p with respect to $\boldsymbol{e}_1$.*

PROOF. Without loss of generality, let $\boldsymbol{S}_1 = (\boldsymbol{e}_p, \boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_{p-1}, \boldsymbol{e}_{p+1}, \boldsymbol{e}_{p+2}, \dots, \boldsymbol{e}_n)$. The matrix $\boldsymbol{S}_1$ is a generator of a cyclic subgroup $\mathcal{S}$ of $\mathcal{P}$ of order p. It is easily verified that $\boldsymbol{e}_1'\boldsymbol{S}\boldsymbol{e}_1 = 0$ for all $\boldsymbol{S} \in \mathcal{S}$. Hence, $\mathcal{S}$ is an oracle subgroup of $\mathcal{H}$ of order p with respect to $\boldsymbol{e}_1$. □

LEMMA 6. *Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be unit vectors. Let $\boldsymbol{Q}$ be an orthonormal matrix such that $\boldsymbol{Q}\boldsymbol{a} = \boldsymbol{b}$, which exists. Let $\mathcal{S}$ be an oracle subgroup of $\mathcal{H}$ with respect to $\boldsymbol{a}$. Then, $\mathcal{G}$ with elements $\boldsymbol{G} = \boldsymbol{Q}\boldsymbol{S}\boldsymbol{Q}'$, $\boldsymbol{S} \in \mathcal{S}$, is an oracle subgroup of $\mathcal{H}$ with respect to $\boldsymbol{b}$, and $\mathcal{G}$ is isomorphic to $\mathcal{S}$.*

PROOF. We have that $\mathcal{G} \subset \mathcal{H}$, as the elements of $\mathcal{G}$ are compositions of orthonormal matrices. Furthermore, it is a *subgroup*, as

- $\boldsymbol{I} \in \mathcal{S}$, so $\boldsymbol{Q}\boldsymbol{I}\boldsymbol{Q}' = \boldsymbol{I} \in \mathcal{G}$,
- for any $\boldsymbol{G} \in \mathcal{G}$ and some $\boldsymbol{S} \in \mathcal{S}$, we have $\boldsymbol{G}' = (\boldsymbol{Q}\boldsymbol{S}\boldsymbol{Q}')' = \boldsymbol{Q}\boldsymbol{S}'\boldsymbol{Q}' \in \mathcal{G}$,
- for any $\boldsymbol{G}_1, \boldsymbol{G}_2 \in \mathcal{G}$ and some $\boldsymbol{S}_1, \boldsymbol{S}_2 \in \mathcal{S}$, we have $\boldsymbol{G}_1\boldsymbol{G}_2 = \boldsymbol{Q}\boldsymbol{S}_1\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{S}_2\boldsymbol{Q}' = \boldsymbol{Q}\boldsymbol{S}_1\boldsymbol{S}_2\boldsymbol{Q}' \in \mathcal{G}$.

In addition, it is an oracle subgroup as $\boldsymbol{b}'\boldsymbol{G}\boldsymbol{b} = \boldsymbol{b}'\boldsymbol{Q}\boldsymbol{S}\boldsymbol{Q}'\boldsymbol{b} = \boldsymbol{a}'\boldsymbol{S}\boldsymbol{a} = 0$, for any $\boldsymbol{G} \in \mathcal{G}$ and some $\boldsymbol{S} \in \mathcal{S}$. Finally, $\mathcal{G}$ and $\mathcal{S}$ are isomorphic as the map $\boldsymbol{G} = \boldsymbol{Q}\boldsymbol{S}\boldsymbol{Q}'$, $\boldsymbol{S} \in \mathcal{S}$, $\boldsymbol{G} \in \mathcal{G}$, is a bijection from $\mathcal{S}$ to $\mathcal{G}$. □

PROOF OF THEOREM 5. From Lemma 5 we can obtain an oracle subgroup $\mathcal{S}$ of $\mathcal{H}$ with respect to $\boldsymbol{e}_1$ of any desired order p, $1 < p \leq n$. Using Lemma 6, we can transform $\mathcal{S}$ into an isomorphic oracle subgroup of $\mathcal{H}$ with respect to any unit vector. □

B.9. Proof of Proposition 5.

PROOF. This proof is analogous to the proof of Theorem 1 after noticing that one can use Lemma 3 to transform a sample of independent uniform draws from a sphere in n dimensions into a sample of independent draws from the t_{n-1}-distribution and use the sample quantile as a critical value. □

B.10. Proof of Proposition 6. In order to prove Proposition 6, we first prove the following lemma.

LEMMA 7. *If $\mathcal{S}$ is a finite subgroup of $\mathcal{H}$ with matrix representation $\mathfrak{S}$, then the columns of $\mathfrak{S}'\,\mathfrak{S}$ are identical up to permutation.*

PROOF. Without loss of generality, we fix the ordering of the elements of $\mathcal{S} = (\boldsymbol{I}, \boldsymbol{S}_1, \boldsymbol{S}_2, \dots)$. Similarly, let the ordering of $\mathfrak{S}$ be induced by this ordering of $\mathcal{S}$, as

$$\mathfrak{S} := \mathcal{S}\boldsymbol{\iota} := (\boldsymbol{\iota}, \boldsymbol{S}_1\boldsymbol{\iota}, \boldsymbol{S}_2\boldsymbol{\iota}, \dots) =: (\boldsymbol{\iota}, \boldsymbol{s}_1, \boldsymbol{s}_2, \dots),$$

where $\boldsymbol{S}_1, \boldsymbol{S}_2, \dots \in \mathcal{S}$. We will show $\mathfrak{S}'\,\mathfrak{S}\,\boldsymbol{e}_1 = \boldsymbol{P}\,\mathfrak{S}'\,\mathfrak{S}\,\boldsymbol{e}_2$, for some permutation matrix $\boldsymbol{P}$. The same reasoning can be used to show equality of all columns up to permutations. The first column of $\mathfrak{S}'\,\mathfrak{S}$ equals

$$\mathfrak{S}'\,\mathfrak{S}\,\boldsymbol{e}_1 = \mathfrak{S}'\,\boldsymbol{\iota} = \mathfrak{S}'\,\boldsymbol{S}_1'\boldsymbol{S}_1\boldsymbol{\iota} = \mathfrak{S}'\,\boldsymbol{S}_1'\boldsymbol{s}_1 = \mathfrak{S}'\,\boldsymbol{S}_1'\,\mathfrak{S}\,\boldsymbol{e}_2 = (\boldsymbol{S}_1\,\mathfrak{S})'\,\mathfrak{S}\,\boldsymbol{e}_2. \tag{6}$$

Next, notice that

$$\boldsymbol{S}_1\,\mathfrak{S} = \boldsymbol{S}_1\mathcal{S}\boldsymbol{\iota} := (\boldsymbol{S}_1, \boldsymbol{S}_1\boldsymbol{S}_1, \boldsymbol{S}_1\boldsymbol{S}_2, \dots)\boldsymbol{\iota} = (\boldsymbol{S}_1\boldsymbol{\iota}, \boldsymbol{S}_1\boldsymbol{S}_1\boldsymbol{\iota}, \boldsymbol{S}_1\boldsymbol{S}_2\boldsymbol{\iota}, \dots)$$

As $\mathcal{S}$ is a group and $\boldsymbol{S}_1 \in \mathcal{S}$, $\boldsymbol{S}_1\mathcal{S}$ must equal $\mathcal{S}$ up to permutations of its elements. Hence, $\boldsymbol{S}_1\, \mathfrak{S} = \mathfrak{S}\, \boldsymbol{P}'$, for some permutation matrix $\boldsymbol{P}'$. Substituting this identity into the right hand side of the final equality in (6) yields $\mathfrak{S}'\, \mathfrak{S}\, \boldsymbol{e}_1 = \boldsymbol{P}\, \mathfrak{S}'\, \mathfrak{S}\, \boldsymbol{e}_2$. The result follows from the fact that the transpose of a permutation matrix is also a permutation matrix. □

Proof of Proposition 6. As $\boldsymbol{\iota}$ is a column of $\mathfrak{S}$, we can write $\mathfrak{S}'\, \boldsymbol{\iota} = \mathfrak{S}'\, \mathfrak{S}\, \boldsymbol{e}_j$, for some j, so that $\mathfrak{S}'\, \boldsymbol{\iota}$ coincides with column j of the matrix $\mathfrak{S}'\, \mathfrak{S}$. As a consequence,

$$\delta_{\mathcal{S}} = \max_{i, i \neq j} \boldsymbol{e}_i'\, \mathfrak{S}'\, \mathfrak{S}\, \boldsymbol{e}_j.$$

From Lemma 7 we know that $\mathfrak{S}'\, \boldsymbol{\iota}$ coincides with *every* column of $\mathfrak{S}'\, \mathfrak{S}$ up to permutation. Hence,

$$\delta_{\mathcal{S}} = \max_{i, j, i \neq j} \boldsymbol{e}_i'\, \mathfrak{S}'\, \mathfrak{S}\, \boldsymbol{e}_j.$$

□

B.11. Proof of Theorem 8.

Proof. Let $\boldsymbol{S}_1 \in \mathcal{S}$, $\boldsymbol{S}_1 \neq \boldsymbol{I}$ and define $\boldsymbol{s}_1 = \boldsymbol{S}_1\boldsymbol{\iota}$. For $\mathcal{S}$ to be an oracle subgroup of $\mathcal{R}$, we require $\boldsymbol{\iota}'\boldsymbol{s}_1 = 0$. Therefore, n must be even so that the number of positive and negative elements of $\boldsymbol{s}_1$ cancel out in $\boldsymbol{\iota}'\boldsymbol{s}_1$. So, without loss of generality, we write $\boldsymbol{s}_1 = (\boldsymbol{\iota}_{n/2}, -\boldsymbol{\iota}_{n/2})'$. Now suppose we take another element $\boldsymbol{S}_2$ from $\mathcal{S}$, with $\boldsymbol{s}_2 = \boldsymbol{S}_2\boldsymbol{\iota}$. Then $\boldsymbol{s}_2$ must also have an equal number of positive and negative elements. In addition, after multiplying with $\boldsymbol{S}_1$, the resulting $\boldsymbol{s}_3 = \boldsymbol{S}_1\boldsymbol{S}_2\boldsymbol{\iota}$ must also have an equal number of positive and negative elements. Notice that this is possible if and only if exactly half of the positive and negative elements in the $\boldsymbol{s}_1$ correspond to a negative and positive elements in the $\boldsymbol{s}_2$, respectively. So, the second element may be represented as $(\boldsymbol{\iota}_{n/4}, -\boldsymbol{\iota}_{n/4}, \boldsymbol{\iota}_{n/4}, -\boldsymbol{\iota}_{n/4})$. This is possible if and only if n is divisible by 2^2. The resulting set $\{\boldsymbol{I}, \boldsymbol{S}_1, \boldsymbol{S}_2, \boldsymbol{S}_3\}$ constitutes an oracle subgroup of $\mathcal{R}$. This process can be continued exactly until all 2s in the prime factorization of n are exhausted. □

B.12. Proof of Proposition 7.

Proof. We start by proving the first part. Notice that $\mathcal{B} \subset \mathcal{R}$, as $\boldsymbol{R} \in \mathcal{R}$, $\mathcal{A} \subset \mathcal{R}$ and $\mathcal{R}$ is a group. To establish that $\mathcal{B}$ is a subgroup of $\mathcal{R}$, it remains to verify that $\mathcal{B}$ is a group.

Firstly, notice that the identity element $\boldsymbol{I}$ is in $\mathcal{B}$, as $\boldsymbol{I} \in \mathcal{A}$ and $\mathcal{A} \subseteq \mathcal{B}$. Secondly, we establish that if $\boldsymbol{A}_1, \boldsymbol{A}_2 \in \mathcal{B}$, then $\boldsymbol{A}_1\boldsymbol{A}_2 \in \mathcal{B}$. We distinguish 4 cases:

Case 1: $\boldsymbol{A}_1, \boldsymbol{A}_2 \in \mathcal{A}$.
In this case, $\boldsymbol{A}_1\boldsymbol{A}_2 \in \mathcal{A}$ and $\mathcal{A} \subset \mathcal{B}$, so $\boldsymbol{A}_1\boldsymbol{A}_2 \in \mathcal{B}$.

Case 2: $\boldsymbol{A}_1 \notin \mathcal{A}$, $\boldsymbol{A}_2 \in \mathcal{A}$.
In this case $\boldsymbol{A}_1 = \boldsymbol{R}\boldsymbol{A}$, for some $\boldsymbol{A} \in \mathcal{A}$. Note that, $\boldsymbol{A}\boldsymbol{A}_2 \in \mathcal{A}$. As $\boldsymbol{R}\boldsymbol{B} \in \mathcal{B}$ for all $\boldsymbol{B} \in \mathcal{A}$, we have $\boldsymbol{A}_1\boldsymbol{A}_2 = \boldsymbol{R}\boldsymbol{A}\boldsymbol{A}_2 \in \mathcal{B}$.

Case 3: $\boldsymbol{A}_1 \in \mathcal{A}$, $\boldsymbol{A}_2 \notin \mathcal{A}$.
We have $\boldsymbol{A}_2\boldsymbol{A}_1 = \boldsymbol{A}_1\boldsymbol{A}_2 \in \mathcal{A}$, by commutativity of matrix multiplication for diagonal matrices. Hence, this follows from the preceding case.

Case 4: $\boldsymbol{A}_1, \boldsymbol{A}_2 \notin \mathcal{A}$.
If $\boldsymbol{A}_1, \boldsymbol{A}_2 \notin \mathcal{A}$, then $\boldsymbol{A}_1 = \boldsymbol{R}\boldsymbol{A}$ and $\boldsymbol{A}_2 = \boldsymbol{R}\boldsymbol{B}$ for some $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{A}$. Thus, $\boldsymbol{A}_1\boldsymbol{A}_2 = \boldsymbol{R}\boldsymbol{A}\boldsymbol{R}\boldsymbol{B} =$

$\boldsymbol{RRAB} = \boldsymbol{AB} \in \mathcal{A}$.

Finally, notice that each element of $\mathcal{R}$ is its own inverse. Thus, each element of $\mathcal{B}$ has an inverse in $\mathcal{B}$. Hence, $\mathcal{B}$ is a group, and therefore a subgroup of $\mathcal{R}$.

To prove the second part, notice that $|\mathcal{B}| \leq 2|\mathcal{A}|$ by definition of $\mathcal{B}$. As $\boldsymbol{R} \notin \mathcal{A}$, $\boldsymbol{R} \in \mathcal{B}$ (because $\boldsymbol{I} \in \mathcal{A}$) and $\mathcal{A} \subset \mathcal{B}$, we have $|\mathcal{B}| > |\mathcal{A}|$. We have that $\mathcal{B}$ is a subgroup of $\mathcal{R}$ by Proposition 7. As $\mathcal{R}$ is a Boolean group, all its subgroups are of order 2^k for some $k \in \mathbb{N}$, so $\mathcal{B}$ is of order 2^k for some $k \in \mathbb{N}$. Finally, because $|\mathcal{B}| > |\mathcal{A}|$, $|\mathcal{B}| \leq 2|\mathcal{A}|$ and $|\mathcal{B}| = 2^k$ for some $k \in \mathbb{N}$, we must have $|\mathcal{B}| = 2|\mathcal{A}|$. □

REFERENCES

OEIS FOUNDATION INC. (OEIS) (2021a). The on-line encyclopedia of integer sequences. http://oeis.org/A022166.

OEIS FOUNDATION INC. (OEIS) (2021b). The on-line encyclopedia of integer sequences. http://oeis.org/A006116.

OEIS FOUNDATION INC. (OEIS) (2021c). The on-line encyclopedia of integer sequences. http://oeis.org/A076831.

OEIS FOUNDATION INC. (OEIS) (2021d). The on-line encyclopedia of integer sequences. http://oeis.org/A076766.

ANDERSON, M. J. and ROBINSON, J. (2001). Permutation tests for linear models. *Australian & New Zealand Journal of Statistics* **43** 75–88.

ANDREELLA, A. (2021). fMRIdata. https://github.com/angeella/fMRIdata.

ANDREELLA, A., HEMERIK, J., WEEDA, W., FINOS, L. and GOEMAN, J. (2020). Permutation-based true discovery proportions for fMRI cluster analysis. *arXiv preprint arXiv:2012.00368*.

BEKKER, P. A. and LAWFORD, S. (2008). Symmetry-based inference in an instrumental variable setting. *Journal of econometrics* **142** 28–49.

BLANCHARD, G., NEUVIAL, P. and ROQUAIN, E. (2020). Post hoc confidence bounds on false positives using reference families. *The Annals of Statistics* **48** 1281–1303.

CHMIELEWSKI, M. (1981). Elliptically symmetric distributions: A review and bibliography. *International Statistical Review/Revue Internationale de Statistique* 67–74.

CONWAY, J. H. and SLOANE, N. J. A. (1998). *Sphere packings, lattices and groups (Third Edition)*. Springer-Verlag, New York.

DARMOIS, G. (1953). Analyse générale des liaisons stochastiques: etude particulière de l'analyse factorielle linéaire. *Revue de l'Institut International de Statistique* 2–8.

DAVIDSON, R. and FLACHAIRE, E. (2008). The wild bootstrap, tamed at last. *Journal of Econometrics* **146** 162–169.

DEBEER, D. and STROBL, C. (2020). Conditional permutation importance revisited. *BMC bioinformatics* **21** 1–30.

DOBRIBAN, E. (2021). Consistency of invariance-based randomization tests.

DWASS, M. (1957). Modified Randomization Tests for Nonparametric Hypotheses. *The Annals of Mathematical Statistics* **28** 181-187.

EATON, M. L. (1989). Group invariance applications in statistics. In *Regional conference series in Probability and Statistics* i–133. JSTOR.

EDEN, T. and YATES, F. (1933). On the validity of Fisher's z test when applied to an actual example of non-normal data.(With five text-figures.). *The Journal of Agricultural Science* **23** 6–17.

EFRON, B. (1969). Student's t-Test Under Symmetry Conditions. *Journal of the American Statistical Association* **64** 1278–1302.

EKLUND, A., NICHOLS, T. E. and KNUTSSON, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the national academy of sciences* **113** 7900–7905.

FISHER, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* **10** 507–521.

FISHER, R. A. (1925). Applications of "Student's" Distribution. *Metron* **5** 90–104.

FISHER, R. A. (1935). *The design of experiments*. Oliver and Boyd.

GAO, X., BECKER, L. C., BECKER, D. M., STARMER, J. D. and PROVINCE, M. A. (2010). Avoiding the high Bonferroni penalty in genome-wide association studies. *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society* **34** 100–105.

GOOD, P. (2005). *Permutation, Parametric, and Bootstrap Tests of Hypotheses (3rd ed.)*. Springer-Verlag, New York.

HEMERIK, J. and GOEMAN, J. J. (2018a). False discovery proportion estimation by permutations: confidence for significance analysis of microarrays. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80** 137–155.

HEMERIK, J. and GOEMAN, J. J. (2018b). Exact testing with random permutations. *TEST* **27** 811–825.

HEMERIK, J. and GOEMAN, J. (2018c). confSAM. https://cran.r-project.org/web/packages/confSAM/index.html.

HEMERIK, J., GOEMAN, J. J. and FINOS, L. (2020). Robust testing in generalized linear models by sign flipping score contributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **82** 841–864.

HEMERIK, J. and GOEMAN, J. J. (2021). Another look at the lady tasting tea and differences between permutation tests and randomisation tests. *International Statistical Review* **89** 367–381.

HEMERIK, J., SOLARI, A. and GOEMAN, J. (2019). Permutation-based simultaneous confidence bounds for the false discovery proportion. *Biometrika* **106** 635–649.

HEMERIK, J., THORESEN, M. and FINOS, L. (2021). Permutation testing in high-dimensional linear models: an empirical investigation. *Journal of Statistical Computation and Simulation* **91** 897–914.

HOEFFDING, W. (1952). The large-sample power of tests based on permutations of observations. *The Annals of Mathematical Statistics* **23** 169–192.

HOTELLING, H. (1953). New light on the correlation coefficient and its transforms. *Journal of the Royal Statistical Society. Series B (Methodological)* **15** 193–232.

KOFLER, R. and SCHLÖTTERER, C. (2012). Gowinda: unbiased analysis of gene set enrichment for genome-wide association studies. *Bioinformatics* **28** 2084–2085.

KONING, N. (2019). Directing Power Towards Conic Parameter Subspaces. *arXiv preprint arXiv:1907.05077*.

KONING, N. W. (2022a). Database of Near Oracle Subgroups. https://github.com/nickwkoning/NOSdata.

KONING, N. W. (2022b). Near Oracle Subgroups. https://github.com/nickwkoning/NOS.

KONING, N. W. (2022c). fast confSAM. https://github.com/nickwkoning/fastconfSAM.

LANGSRUD, Ø. (2005). Rotation tests. *Statistics and computing* **15** 53–60.

LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing statistical hypotheses*. Springer Science & Business Media.

LI, J. and TIBSHIRANI, R. (2013). Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. *Statistical methods in medical research* **22** 519–536.

MEINSHAUSEN, N. (2006). False discovery control for multiple tests of association under general dependence. *Scandinavian Journal of Statistics* **33** 227–237.

MEINSHAUSEN, N., MAATHUIS, M. H., BÜHLMANN, P. et al. (2011). Asymptotic optimality of the Westfall–Young permutation procedure for multiple testing under dependence. *The Annals of Statistics* **39** 3369–3391.

ONGHENA, P. (2018). Randomization tests or permutation tests? A historical and terminological clarification. *Randomization, masking, and allocation concealment* 209–227.

PESARIN, F. and SALMASO, L. (2010). *Permutation tests for complex data: theory, applications and software*. John Wiley & Sons.

PHIPSON, B. and SMYTH, G. K. (2010). Permutation P-values should never be zero: calculating exact P-values when permutations are randomly drawn. *Statistical applications in genetics and molecular biology* **9** 39.

SKITOVITCH, V. P. (1953). On a property of the normal distribution. *Doklady Akad. Nauk SSSR (N.S)* **89** 217–219.

SLEPIAN, D. (1968). Group codes for the Gaussian channel. *Bell System Technical Journal* **47** 575–602.

SLOANE, N. J. A., HARDIN, R. H., SMITH, W. D. et al. (1996). Tables of Spherical Codes. http://neilsloane.com/packings/. Accessed: 2021-11-19.

SMEETS, P. A., KROESE, F. M., EVERS, C. and DE RIDDER, D. T. (2013). Allured or alarmed: counteractive control responses to food temptations in the brain. *Behavioural brain research* **248** 41–45.

SOLARI, A., FINOS, L. and GOEMAN, J. J. (2014). Rotation-based multiple testing in the multivariate linear model. *Biometrics* **70** 954–961.

SOUTHWORTH, L. K., KIM, S. K. and OWEN, A. B. (2009). Properties of balanced permutations. *Journal of Computational Biology* **16** 625–638.

TUSHER, V. G., TIBSHIRANI, R. and CHU, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences* **98** 5116–5121.

VESELY, A., FINOS, L. and GOEMAN, J. J. (2021). Permutation-based true discovery guarantee by sum tests. *arXiv preprint arXiv:2102.11759*.

WESTFALL, P. H. and TROENDLE, J. F. (2008). Multiple testing with minimal assumptions. *Biometrical Journal: Journal of Mathematical Methods in Biosciences* **50** 745–755.

WESTFALL, P. H. and YOUNG, S. S. (1993). *Resampling-based multiple testing: Examples and methods for p-value adjustment* **279**. John Wiley & Sons.

WINKLER, A. M., RIDGWAY, G. R., WEBSTER, M. A., SMITH, S. M. and NICHOLS, T. E. (2014). Permutation inference for the general linear model. *Neuroimage* **92** 381–397.

WINKLER, A. M., RIDGWAY, G. R., DOUAUD, G., NICHOLS, T. E. and SMITH, S. M. (2016). Faster permutation inference in brain imaging. *NeuroImage* **141** 502–516.

YOUNG, A. (2019). Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *The Quarterly Journal of Economics* **134** 557–598.