THE GEOMETRY OF HYPOTHESIS TESTING OVER CONVEX CONES: GENERALIZED LIKELIHOOD RATIO TESTS AND MINIMAX RADII

BY YUTING WEI, MARTIN J. WAINWRIGHT[1] AND
ADITYANAND GUNTUBOYINA[2]

University of California, Berkeley

We consider a compound testing problem within the Gaussian sequence model in which the null and alternative are specified by a pair of closed, convex cones. Such cone testing problem arises in various applications, including detection of treatment effects, trend detection in econometrics, signal detection in radar processing and shape-constrained inference in nonparametric statistics. We provide a sharp characterization of the GLRT testing radius up to a universal multiplicative constant in terms of the geometric structure of the underlying convex cones. When applied to concrete examples, this result reveals some interesting phenomena that do not arise in the analogous problems of estimation under convex constraints. In particular, in contrast to estimation error, the testing error no longer depends purely on the problem complexity via a volume-based measure (such as metric entropy or Gaussian complexity); other geometric properties of the cones also play an important role. In order to address the issue of optimality, we prove information-theoretic lower bounds for the minimax testing radius again in terms of geometric quantities. Our general theorems are illustrated by examples including the cases of monotone and orthant cones, and involve some results of independent interest.

1. Introduction. Composite testing problems arise in a wide variety of applications and the generalized likelihood ratio test (GLRT) is a general purpose approach to such problems. The basic idea of the likelihood ratio test dates back to the early works of Fisher, Neyman and Pearson; it attracted further attention following the work of Edwards [14], who emphasized likelihood as a general principle of inference. Recent years have witnessed a great amount of work on the GLRT in various contexts, including the papers [16, 17, 24, 25, 35]. However, despite the wide-spread use of the GLRT, its optimality properties have yet to be fully understood. For suitably regular problems, there is a great deal of asymptotic

theory on the GLRT, and in particular when its distribution under the null is independent of nuisance parameters (e.g., [3, 37, 40]). On the other hand, there are some isolated cases in which the GLRT can be shown to be dominated by other tests (e.g., [24, 30, 32, 50]).

In this paper, we undertake an in-depth study of the GLRT in application to a particular class of composite testing problems of a geometric flavor. In this class of testing problems, the null and alternative hypotheses are specified by a pair of closed convex cones C_1 and C_2, taken to be nested as $C_1 \subset C_2$. Suppose that we are given an observation of the form $y = \theta + w$, where w is a zero-mean Gaussian noise vector. Based on observing y, our goal is to test whether a given parameter θ belongs to the smaller cone C_1—corresponding to the null hypothesis—or belongs to the larger cone C_2. Cone testing problems of this type arise in many different settings, and there is a fairly substantial literature on the behavior of the GLRT in application to such problems (e.g., see the papers and books [8, 13, 23, 31–33, 37–39, 41, 45, 50], as well as references therein).

1.1. *Some motivating examples.*   Before proceeding, let us consider some concrete examples so as to motivate our study.

EXAMPLE 1 (Testing nonnegativity and monotonicity in treatment effects). Suppose that we have a collection of d treatments, say different drugs for a particular medical condition. Letting $\theta_j \in \mathbb{R}$ denote the mean of treatment j, one null hypothesis could be that none of treatments has any effect, that is, $\theta_j = 0$ for all $j = 1, \ldots, d$. Assuming that none of the treatments are directly harmful, a reasonable alternative would be that θ belongs to the *nonnegative orthant cone*

$$(1) \qquad K_+ := \{\theta \in \mathbb{R}^d \mid \theta_j \geq 0 \text{ for all } j = 1, \ldots, d\}.$$

This set-up leads to a particular instance of our general set-up with $C_1 = \{0\}$ and $C_2 = K_+$. Such orthant testing problems have been studied by Kudo [23] and Raubertas et al. [37], among other people.

In other applications, our treatments might consist of an ordered set of dosages of the same drug. In this case, we might have reason to believe that if the drug has any effect, then the treatment means would obey a monotonicity constraint, that is, with higher dosages leading to greater treatment effects. One would then want to detect the presence or absence of such a dose response effect. Monotonicity constraints also arise in various types of econometric models, in which the effects of strategic interventions should be monotone with respect to parameters such as market size (e.g., [12]). For applications of this flavor, a reasonable alternative would be specified by the *monotone cone*

$$(2) \qquad M := \{\theta \in \mathbb{R}^d \mid \theta_1 \leq \theta_2 \leq \cdots \leq \theta_d\}.$$

This set-up leads to another instance of our general problem with $C_1 = \{0\}$ and $C_2 = M$. The behavior of the GLRT for this particular testing problem has also

been studied in past works, including papers by Barlow et al. [3], and Raubertas et al. [37].

As a third instance of the treatment effects problem, we might like to include in our null hypothesis the possibility that the treatments have some (potentially) nonzero effect but one that remains constant across levels, that is, $\theta_1 = \theta_2 = \cdots = \theta_d$. In this case, our null hypothesis is specified by the *ray cone*

$$(3) \qquad R := \{\theta \in \mathbb{R}^d \mid \theta = c\mathbf{1} \text{ for some } c \in \mathbb{R}\}.$$

Supposing that we are interested in testing the alternative that the treatments lead to a monotone effect, we arrive at another instance of our general set-up with $C_1 = R$ and $C_2 = M$. This testing problem has also been studied by Bartholomew [4, 5] and Robertson et al. [42] among other researchers.

In the preceding three examples, the cone C_1 was linear subspace. Let us now consider two more examples, adapted from Menendnez et al. [31], in which C_1 is not a subspace. As before, suppose that component θ_i of the vector $\theta \in \mathbb{R}^d$ denotes the expected response of treatment i. In many applications, it is of interest to test equality of the expected responses of a subset S of the full treatment set $[d] = \{1, \ldots, d\}$. More precisely, for a given subset S containing the index 1, let us consider the problem of testing the the null hypothesis

$$(4) \qquad C_1 \equiv E(S) := \{\theta \in \mathbb{R}^d \mid \theta_i = \theta_1 \ \forall i \in S, \text{ and } \theta_j \geq \theta_1 \ \forall j \notin S\}$$

versus the alternative $C_2 \equiv G(S) = \{\theta \in \mathbb{R}^d \mid \theta_j \geq \theta_1 \ \forall j \in [d]\}$. Note that C_1 here is not a linear subspace.

As a final example, suppose that we have a factorial design consisting of two treatments, each of which can be applied at two different dosages (high and level). Let (θ_1, θ_2) denote the expected responses of the first treatment at the low and high dosages, respectively, with the pair (θ_3, θ_4) defined similarly for the second treatment. Suppose that we are interesting in testing whether the first treatment at the lowest level is more effective than the second treatment at the highest level. This problem can be formulated as testing the null cone

$$(5) \qquad \begin{aligned} &C_1 := \{\theta \in \mathbb{R}^4 \mid \theta_1 \leq \theta_2 \leq \theta_3 \leq \theta_4\} \quad \text{versus the alternative} \\ &C_2 := \{\theta \in \mathbb{R}^4 \mid \theta_1 \leq \theta_2, \text{ and } \theta_3 \leq \theta_4\}. \end{aligned}$$

As before, the null cone C_1 is not a linear subspace.

EXAMPLE 2 (Robust matched filtering in signal processing). In radar detection problems [43], a standard goal is to detect the presence of a known signal of unknown amplitude in the presence of noise. After a matched filtering step, this problem can be reduced to a vector testing problem, where the known signal direction is defined by a vector $\gamma \in \mathbb{R}^d$, whereas the unknown amplitude corresponds to a scalar pre-factor $c \geq 0$. We thus arrive at a ray cone testing problem: the null hypothesis (corresponding to the absence of signal) is given $C_1 = \{0\}$, whereas the alternative is given by the positive ray cone $R_+ = \{\theta \in \mathbb{R}^d \mid \theta = c\gamma \text{ for some } c \geq 0\}$.

In many cases, there may be uncertainty about the target signal, or jamming by adversaries, who introduce additional signals that can be potentially confused with the target signal γ. Signal uncertainties of this type are often modeled by various forms of cones, with the most classical choice being a subspace cone [43]. In more recent work (e.g., [8, 18]), signal uncertainty has been modeled using the *circular cone* defined by the target signal direction, namely

$$(6) \qquad C(\gamma; \alpha) := \{\theta \in \mathbb{R}^d \mid \langle \gamma, \theta \rangle \geq \cos(\alpha) \|\gamma\|_2 \|\theta\|_2\},$$

corresponding to the set of all vectors θ that have angle at least α with the target signal. Thus, we are led to another instance of a cone testing problem involving a circular cone.

EXAMPLE 3 (Cone-constrained testing in linear regression).   Consider the standard linear regression model

$$(7) \qquad\qquad y = X\beta + \sigma Z \qquad \text{where } Z \sim N(0, I_n),$$

where $X \in \mathbb{R}^{n \times p}$ is a fixed and known design matrix. In many applications, we are interested in testing certain properties of the unknown regression vector β, and these can often be encoded in terms of cone-constraints on the vector $\theta := X\beta$. As a very simple example, the problem of testing whether or not $\beta = 0$ corresponds to testing whether $\theta \in C_1 := \{0\}$ versus the alternative that $\theta \in C_2 := \text{range}(X)$. Thus, we arrive at a *subspace testing problem*. We note this problem is known as testing the global null in the linear regression literature (e.g., [9]). If instead we consider the case when the p-dimensional vector β lies in the nonnegative orthant cone (1), then our alternative for the n-dimensional vector θ becomes the *polyhedral cone*

$$(8) \qquad\qquad P := \{\theta \in \mathbb{R}^n \mid \theta = X\beta \text{ for some } \beta \geq 0\}.$$

The corresponding estimation problem with nonnegative constraints on the coefficient vector β has been studied by Slawski et al. [46] and Meinshausen [29]; see also Chen et al. [11] for a survey of this line of work. In addition to these preceding two cases, we can also test various other types of cone alternatives for β, and these are transformed via the design matrix X into other types of cones for the parameter $\theta \in \mathbb{R}^n$.

EXAMPLE 4 (Testing shape-constrained departures from parametric models). Our third example is nonparametric in flavor. Consider the class of functions f that can be decomposed as

$$(9) \qquad\qquad f = \sum_{j=1}^{k} a_j \phi_j + \psi.$$

Here the known functions $\{\phi_j\}_{j=1}^k$ define a linear space, parameterized by the coefficient vector $a \in \mathbb{R}^k$, whereas the unknown function ψ models a structured departure from this linear parametric class. For instance, we might assume that ψ belongs to the class of monotone functions, or the class of convex functions. Given a fixed collection of design points $\{t_i\}_{i=1}^n$, suppose that we make observations of the form $y_i = f(t_i) + \sigma g_i$ for $i = 1, \ldots, n$, where each g_i is a standard normal variable. Defining the shorthand notation $\theta := (f(t_1), \ldots, f(t_n))$ and $g = (g_1, \ldots, g_n)$, our observations can be expressed in the standard form $y = \theta + \sigma g$. If, under the null hypothesis, the function f satisfies the decomposition (9) with $\psi = 0$, then the vector θ must belong to the subspace $\{\Phi a \mid a \in \mathbb{R}^k\}$, where the matrix $\Phi \in \mathbb{R}^{n \times k}$ has entries $\Phi_{ij} = \phi_j(x_i)$.

Now suppose that the alternative is that f satisfies the decomposition (9) with some ψ that is convex. A convexity constraint on ψ implies that we can write $\theta = \Phi a + \gamma$, for some coefficients $a \in \mathbb{R}^k$ and a vector $\gamma \in \mathbb{R}^n$ belonging to the *convex cone*

$$(10) \qquad V\left(\{t_i\}_{i=1}^n\right) := \left\{ \gamma \in \mathbb{R}^n \;\middle|\; \frac{\gamma_2 - \gamma_1}{t_2 - t_1} \leq \frac{\gamma_3 - \gamma_2}{t_3 - t_2} \leq \cdots \leq \frac{\gamma_n - \gamma_{n-1}}{t_n - t_{n-1}} \right\}.$$

This particular cone testing problem and other forms of shape constraints have been studied by Meyer [33], as well as by Sen and Meyer [44].

1.2. *Problem formulation.* Having understood the range of motivations for our problem, let us now set up the problem more precisely. Suppose that we are given observations of the form $y = \theta + \sigma g$, where $\theta \in \mathbb{R}^d$ is a fixed but unknown vector, whereas $g \sim N(0, I_d)$ is a d-dimensional vector of i.i.d. Gaussian entries and σ^2 is a known noise level. Our goal is to distinguish the null hypothesis that $\theta \in C_1$ versus the alternative that $\theta \in C_2 \setminus C_1$, where $C_1 \subset C_2$ are a nested pair of closed, convex cones in $\mathbb{R}^d$.

In this paper, we study both the fundamental limits of solving this composite testing problem, as well as the performance of a specific procedure, namely the *generalized likelihood ratio test*, or GLRT for short. By definition, the GLRT for the problem of distinguishing between cones C_1 and C_2 is based on the statistic

$$(11\text{a}) \qquad T(y) := -2\log\left(\frac{\sup_{\theta \in C_1} \mathbb{P}_\theta(y)}{\sup_{\theta \in C_2} \mathbb{P}_\theta(y)}\right).$$

It defines a family of tests, parameterized by a threshold parameter $\beta \in [0, \infty)$, of the form

$$(11\text{b}) \qquad \phi_\beta(y) := \mathbb{I}\big(T(y) \geq \beta\big) = \begin{cases} 1 & \text{if } T(y) \geq \beta, \\ 0 & \text{otherwise.} \end{cases}$$

Thus far, our formulation of the testing problem allows for the possibility that θ lies in the set $C_2 \setminus C_1$, but is arbitrarily close to some element of C_1. Thus, under this formulation, it is not possible to make any nontrivial assertions about the

power of the GLRT nor any other test in a uniform sense. Accordingly, so as to be able to make quantitative statements about the performance of different statements, we exclude a certain ε-ball from the alternative. This procedure leads to the notion of the *minimax testing radius* associated this composite decision problem. This minimax formulation was introduced in the seminal work of Ingster and coauthors [20, 21]; since then, it has been studied by many authors (e.g., [2, 15, 26, 27, 47]).

For a given $\varepsilon > 0$, we define the ε-*fattening* of the cone C_1 as

$$(12) \qquad B_2(C_1; \varepsilon) := \left\{ \theta \in \mathbb{R}^d \mid \min_{u \in C_1} \|\theta - u\|_2 \leq \varepsilon \right\},$$

corresponding to the set of vectors in $\mathbb{R}^d$ that are at most Euclidean distance ε from some element of C_1. We then consider the testing problem of distinguishing between the two hypotheses

$$(13) \qquad \mathcal{H}_0 : \theta \in C_1 \quad \text{and} \quad \mathcal{H}_1 : \theta \in C_2 \setminus B_2(C_1; \varepsilon).$$

To be clear, the parameter $\varepsilon > 0$ is a quantity that is used during the course of our analysis in order to titrate the difficulty of the testing problem. All of the tests that we consider, including the GLRT, are not given knowledge of ε. Let us introduce shorthand $\mathcal{T}(C_1, C_2; \varepsilon)$ to denote this testing problem (13).

Obviously, the testing problem (13) becomes more difficult as ε approaches zero, and so it is natural to study this increase in quantitative terms. Letting $\psi : \mathbb{R}^d \to \{0, 1\}$ be any (measurable) test function, we measure its performance in terms of its *uniform error*

$$(14) \qquad \mathcal{E}(\psi; C_1, C_2, \varepsilon) := \sup_{\theta \in C_1} \mathbb{E}_\theta \big[\psi(y) \big] + \sup_{\theta \in C_2 \setminus B_2(\varepsilon; C_1)} \mathbb{E}_\theta \big[1 - \psi(y) \big],$$

which controls the worst-case error over both null and alternative.

For a given error level $\rho \in (0, 1)$, we are interested in the smallest setting of ε for which either the GLRT, or some other test ψ has uniform error at most ρ. More precisely, we define

$$(15a) \qquad \varepsilon_{\text{OPT}}(C_1, C_2; \rho) := \inf \left\{ \varepsilon \mid \inf_\psi \mathcal{E}(\psi; C_1, C_2, \varepsilon) \leq \rho \right\} \quad \text{and}$$

$$(15b) \qquad \varepsilon_{\text{GLR}}(C_1, C_2; \rho) := \inf \left\{ \varepsilon \mid \inf_{\beta \in \mathbb{R}} \mathcal{E}(\phi_\beta; C_1, C_2, \varepsilon) \leq \rho \right\}.$$

When the subspace-cone pair (C_1, C_2) are clear from the context, we occasionally omit this dependence, and write $\varepsilon_{\text{OPT}}(\rho)$ and $\varepsilon_{\text{GLR}}(\rho)$ instead. We refer to these two quantities as the *minimax testing radius* and the *GLRT testing radius*, respectively.

By definition, the minimax testing radius ε_{OPT} corresponds to the smallest separation ε at which there exists *some test* that distinguishes between the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ in equation (13) with uniform error at most ρ. Thus, it provides a

fundamental characterization of the statistical difficulty of the hypothesis testing. On the other hand, the GLRT testing radius $\varepsilon_{\text{GLR}}(\rho)$ provides us with the smallest radius ε for which there exists *some threshold*—say β^*—for which the associated generalized likelihood ratio test ϕ_{β^*} distinguishes between the hypotheses with error at most ρ. Thus, it characterizes the performance limits of the GLRT when an optimal threshold β^* is chosen. Of course, by definition, we always have $\varepsilon_{\text{OPT}}(\rho) \leq \varepsilon_{\text{GLR}}(\rho)$. We write $\varepsilon_{\text{OPT}}(\rho) \asymp \varepsilon_{\text{GLR}}(\rho)$ to mean that—in addition to the previous upper bound—there is also a lower bound $\varepsilon_{\text{OPT}}(\rho) \geq c_\rho \varepsilon_{\text{GLR}}(\rho)$ that matches up to a constant $c_\rho > 0$ depending only on ρ.

1.3. *Overview of our results.* Having set up the problem, let us now provide a high-level overview of the main results of this paper.

• Our first main result, stated as Theorem 1 in Section 3.1, gives a sharp characterization—meaning upper and lower bounds that match up to universal constants—of the GLRT testing radius ε_{GLR} for cone pairs (C_1, C_2) that are nonoblique (we discuss the nonobliqueness property and its significance at length in Section 2.2). We illustrate the consequences of this theorem for a number of concrete cones, include the subspace cone, orthant cone, monotone cone, circular cone and a Cartesian product cone.

• In our second main result, stated as Theorem 2 in Section 3.2, we derive a lower bound that applies to any testing function. It leads to a corollary that provides sufficient conditions for the GLRT to be an optimal test, and we use it to establish optimality for the subspace cone and circular cone, among other examples. We then revisit the Cartesian product cone, first analyzed in the context of Theorem 1, and use Theorem 2 to show that the GLRT is suboptimal for this particular cone, even though it is in no sense a pathological example.

• For the monotone and orthant cones, we find that the lower bound established in Theorem 2 is not sharp, but that the GLRT turns out to be an optimal test. Thus, Section 3.3 is devoted to a detailed analysis of these two cases, in particular using a more refined argument to obtain sharp lower bounds.

The remainder of this paper is organized as follows: Section 2 provides background on conic geometry, including conic projections, the Moreau decomposition and the notion of Gaussian width. It also introduces the notion of a nonoblique pair of cones, which have been studied in the context of the GLRT. In Section 3, we state our main results and illustrate their consequences via a series of examples. Sections 3.1 and 3.2 are devoted, respectively, to our sharp characterization of the GLRT and a general lower bound on the minimax testing radius. Section 3.3 explores the monotone and orthant cones in more detail. In Section 4, we provide the proofs of our main results, with certain more technical aspects deferred to the Appendices in Supplementary Material [52].

Notation. Here in this paper, for functions $f(\sigma, d)$ and $g(\sigma, d)$, we write $f(\sigma, d) \lesssim g(\sigma, d)$ to indicate that $f(\sigma, d) \leq cg(\sigma, d)$ for some constant $c \in (0, \infty)$ that may only depend on ρ but independent of (σ, d), and similarly for $f(\sigma, d) \gtrsim g(\sigma, d)$. We write $f(\sigma, d) \asymp g(\sigma, d)$ if both $f(\sigma, d) \lesssim g(\sigma, d)$ and $f(\sigma, d) \gtrsim g(\sigma, d)$ are satisfied.

2. Background on conic geometry and the GLRT. In this section, we provide some necessary background on cones and their geometry, including the notion of a polar cone and the Moreau decomposition. We also define the notion of a nonoblique pair of cones, and summarize some known results about properties of the GLRT for such cone testing problems.

2.1. *Convex cones and Gaussian widths.* For a given closed convex cone $C \subset \mathbb{R}^d$, we define the Euclidean projection operator $\Pi_C : \mathbb{R}^d \to C$ via

$$(16) \qquad \Pi_C(v) := \arg\min_{u \in C} \|v - u\|_2.$$

By standard properties of projection onto closed convex sets, we are guaranteed that this mapping is well defined. We also define the polar cone

$$(17) \qquad C^* := \{v \in \mathbb{R}^d \mid \langle v, u \rangle \leq 0 \text{ for all } u \in C\}.$$

Figure 1(b) provides an illustration of a cone in comparison to its polar cone. Using Π_{C^*} to denote the projection operator onto this cone, Moreau's theorem [34] ensures that every vector $v \in \mathbb{R}^d$ can be decomposed as

$$(18) \qquad v = \Pi_C(v) + \Pi_{C^*}(v) \quad \text{and such that} \quad \langle \Pi_C(v), \Pi_{C^*}(v) \rangle = 0.$$

We make frequent use of this decomposition in our analysis.

Let $S^{-1} := \{u \in \mathbb{R}^d \mid \|u\|_2 = 1\}$ denotes the Euclidean sphere of unit radius. For every set $A \subseteq S^{-1}$, we define its *Gaussian width* as

$$(19) \qquad \mathbb{W}(A) := \mathbb{E}\Big[\sup_{u \in A} \langle u, g \rangle\Big] \qquad \text{where } g \sim N(0, I_d).$$

This quantity provides a measure of the size of the set A; indeed, it can be related to the volume of A viewed as a subset of the Euclidean sphere. The notion of Gaussian width arises in many different areas, notably in early work on probabilistic methods in Banach spaces [36]; the Gaussian complexity, along with its close relative the Rademacher complexity, plays a central role in empirical process theory [6, 22, 48].

Of interest in this paper are the Gaussian widths of sets of the form $A = C \cap S^{-1}$, where C is a closed convex cone. For a set of this form, using the Moreau decomposition (18), we have the useful equivalence

$$(20) \qquad \mathbb{W}(C \cap S^{-1}) = \mathbb{E}\Big[\sup_{u \in C \cap S^{-1}} \langle u, \Pi_C(g) + \Pi_{C^*}(g) \rangle\Big] = \mathbb{E}\|\Pi_C(g)\|_2,$$

where the final equality uses the fact that $\langle u, \Pi_{C^*}(g)\rangle \le 0$ for all vectors $u \in C$, with equality holding when u is a nonnegative scalar multiple of $\Pi_C(g)$.

For future reference, let us derive a lower bound on $\mathbb{E}\|\Pi_C g\|_2$ that holds for every cone C strictly larger than $\{0\}$. Take some nonzero vector $u \in C$ and let $R_+ = \{cu \mid c \ge 0\}$ be the ray that it defines. Since $R_+ \subseteq C$, we have $\|\Pi_C g\|_2 \ge \|\Pi_{R_+} g\|_2$. But since R_+ is just a ray, the projection $\Pi_{R_+}(g)$ is a standard normal variable truncated to be positive, and hence

$$(21) \qquad \mathbb{E}\|\Pi_C g\|_2 \ge \mathbb{E}\|\Pi_{R_+} g\|_2 = \sqrt{\frac{1}{2\pi}}.$$

This lower bound is useful in parts of our development.

2.2. *Cone-based GLRTs and nonoblique pairs.* In this section, we provide some background on the notion of nonoblique pairs of cones, and their significance for the GLRT. First, let us exploit some properties of closed convex cones in order to derive a simpler expression for the GLRT test statistic (11a). Using the form of the multivariate Gaussian density, we have

$$(22) \quad T(y) = \min_{\theta \in C_1}\|y-\theta\|_2^2 - \min_{\theta \in C_2}\|y-\theta\|_2^2 = \|y - \Pi_{C_1}(y)\|_2^2 - \|y - \Pi_{C_2}(y)\|_2^2$$

$$(23) \qquad = \|\Pi_{C_2}(y)\|_2^2 - \|\Pi_{C_1}(y)\|_2^2,$$

where we have made use of the Moreau decomposition to assert that

$$\|y - \Pi_{C_1}(y)\|_2^2 = \|y\|_2^2 - \|\Pi_{C_1}(y)\|_2^2 \quad \text{and}$$
$$\|y - \Pi_{C_2}(y)\|_2^2 = \|y\|_2^2 - \|\Pi_{C_2}(y)\|_2^2.$$

Thus, we see that a cone-based GLRT has a natural interpretation: it compares the squared amplitude of the projection of y onto the two different cones.

When $C_1 = \{0\}$, then it can be shown that under the null hypothesis [i.e., $y \sim N(0, \sigma^2 I_d)$], the statistic $T(y)$ (after rescaling by σ^2) is a mixture of χ^2-distributions (see, e.g., [37]). On the other hand, for a general cone pair (C_1, C_2), it is not straightforward to characterize the distribution of $T(y)$ under the null hypothesis. Thus, past work has studied conditions on the cone pair under which the null distribution has a simple characterization. One such condition is a certain nonobliqueness property that is common to much past work on the GLRT (e.g., [19, 31, 32, 50]). The nonobliqueness condition, first introduced by Warrack et al. [50], is also motivated by the fact that are many instances of oblique cone pairs for which the GLRT is known to dominated by other tests. Menendez et al. [30] provide an explanation for this dominance in a very general context; see also the papers [19, 31] for further studies of nonoblique cone pairs.

A nested pair of closed convex cones $C_1 \subset C_2$ is said to be *nonoblique* if we have the successive projection property

$$(24) \qquad \Pi_{C_1}(x) = \Pi_{C_1}(\Pi_{C_2}(x)) \qquad \text{for all } x \in \mathbb{R}^d.$$

For instance, this condition holds whenever one of the two cones is a subspace, or more generally, whenever there is a subspace L such that $C_1 \subseteq L \subseteq C_2$; see Hu and Wright [19] for details of this latter property. To be clear, these conditions are sufficient—but not necessary—for nonobliqueness to hold. There are many nonoblique cone pairs in which neither cone is a subspace; the cone pairs (4) and (5), as discussed in Example 1 on treatment testing, are two such examples. (We refer the reader to Section 5 of the paper [31] for verification of these properties.) More generally, there are various nonoblique cone pairs that do not sandwich a subspace L.

The significance of the nonobliqueness condition lies in the following decomposition result. For any nested pair of closed convex cones $C_1 \subset C_2$ that are nonoblique, for all $x \in \mathbb{R}^d$ we have

(25) $\Pi_{C_2}(x) = \Pi_{C_1}(x) + \Pi_{C_2 \cap C_1^*}(x) \quad \text{and} \quad \langle \Pi_{C_1}(x), \Pi_{C_2 \cap C_1^*}(x) \rangle = 0.$

This decomposition follows from general theory due to Zarantonello [53], who proves that for nonoblique cones, we have $\Pi_{C_2 \cap C_1^*} = \Pi_{C_1^*} \Pi_{C_2}$; in particular, see Theorem 5.2 in his paper.

An immediate consequence of the decomposition (25) is that the GLRT for any nonoblique cone pair (C_1, C_2) can be written as

$$T(y) = \|\Pi_{C_2}(y)\|_2^2 - \|\Pi_{C_1}(y)\|_2^2 = \|\Pi_{C_2 \cap C_1^*}(y)\|_2^2$$
$$= \|y\|_2^2 - \min_{\theta \in C_2 \cap C_1^*} \|y - \theta\|_2^2.$$

Consequently, we see that the GLRT for the pair (C_1, C_2) is equivalent to—that is, determined by the same statistic as—the GLRT for testing the *reduced hypothesis*

(26) $\widetilde{\mathcal{H}}_0 : \theta = 0 \quad \text{versus} \quad \widetilde{\mathcal{H}}_1 : \theta \in (C_2 \cap C_1^*) \setminus B_2(\varepsilon).$

Following the previous notation, write it as $\mathcal{T}(\{0\}, C_2 \cap C_1^*; \varepsilon)$ and we make frequent use of this convenient reduction in the sequel.

3. Main results and their consequences.

We now turn to the statement of our main results, along with a discussion of some of their consequences. Section 3.1 provides a sharp characterization of the minimax radius for the generalized likelihood ratio test up to a universal constant, along with a number of concrete examples. In Section 3.2, we state and prove a general lower bound on the performance of any test, and use it to establish the optimality of the GLRT in certain settings, as well as its suboptimality in other settings. In Section 3.3, we revisit and study in details two cones of particular interest, namely the orthant and monotone cones.

3.1. *Analysis of the generalized likelihood ratio test*.   Let (C_1, C_2) be a nested pair of closed cones $C_1 \subseteq C_2$ that are nonoblique (24). Consider the polar cone C_1^*

as well as the intersection cone $K = C_2 \cap C_1^*$. Letting $g \in \mathbb{R}^d$ denote a standard Gaussian random vector, we then define the quantity

$$(27) \quad \delta_{\mathrm{LR}}^2(C_1, C_2) := \min\left\{\mathbb{E}\|\Pi_K g\|_2, \left(\frac{\mathbb{E}\|\Pi_K g\|_2}{\max\{0, \inf_{\eta \in K \cap S^{-1}}\langle \eta, \mathbb{E}\Pi_K g\rangle\}}\right)^2\right\}.$$

Note that $\delta_{\mathrm{LR}}^2(C_1, C_2)$ is a purely geometric object, depending on the pair (C_1, C_2) via the new cone $K = C_2 \cap C_1^*$, which arises due to the GLRT equivalence (26) discussed previously.

Recall that the GLRT is based on applying a threshold, at some level $\beta \in [0, \infty)$, to the likelihood ratio statistic $T(y)$; in particular, see equations (11a) and (11b). In the following theorem, we study the performance of the GLRT in terms of the the uniform testing error $\mathcal{E}(\phi_\beta; C_1, C_2, \varepsilon)$ from equation (14). In particular, we show that the critical testing radius for the GLRT is governed by the geometric parameter $\delta_{\mathrm{LR}}^2(C_1, C_2)$.

THEOREM 1. *There are numbers $\{(b_\rho, B_\rho), \rho \in (0, 1/2)\}$ such that for every pair of nonoblique closed convex cones (C_1, C_2) with C_1 strictly contained within C_2:*

(a) *For every error probability $\rho \in (0, 0.5)$, we have*

$$(28a) \quad \inf_{\beta \in [0, \infty)} \mathcal{E}(\phi_\beta; C_1, C_2, \varepsilon) \le \rho \quad \text{for all } \varepsilon^2 \ge B_\rho \sigma^2 \delta_{LR}^2(C_1, C_2).$$

(b) *Conversely, for every error probability $\rho \in (0, 0.11]$, we have*

$$(28b) \quad \inf_{\beta \in [0, \infty)} \mathcal{E}(\phi_\beta; C_1, C_2, \varepsilon) \ge \rho \quad \text{for all } \varepsilon^2 \le b_\rho \sigma^2 \delta_{LR}^2(C_1, C_2).$$

REMARKS. While our proof leads to universal values for the constants B_ρ and b_ρ, we have made no efforts to obtain the sharpest possible ones, so do not state them here. In any case, our main interest is to understand the scaling of the testing radius with respect to σ and the geometric parameters of the problem. In terms of the GLRT testing radius $\varepsilon_{\mathrm{GLR}}$ previously defined (15b), Theorem 1 establishes that

$$(29) \quad \varepsilon_{\mathrm{GLR}}(C_1, C_2; \rho) \asymp \sigma \delta_{\mathrm{LR}}(C_1, C_2),$$

where $\asymp$ denotes equality up to constants depending on ρ, but independent of all other problem parameters. Since $\varepsilon_{\mathrm{GLR}}$ always upper bounds $\varepsilon_{\mathrm{OPT}}$ for every fixed level ρ, we can also conclude from Theorem 1 that

$$\varepsilon_{\mathrm{OPT}}(C_1, C_2; \rho) \lesssim \sigma \delta_{\mathrm{LR}}(C_1, C_2).$$

It is worthwhile noting that the quantity $\delta_{\mathrm{LR}}^2(C_1, C_2)$ depends on the pair (C_1, C_2) only via the new cone $K = C_2 \cap C_1^*$. Indeed, as discussed in Section 2.2, for any

pair of nonoblique closed convex cones, the GLRT for the original testing problem (13) is equivalent to the GLRT for the modified testing problem $\mathcal{T}(\{0\}, K; \varepsilon)$.

Observe that the quantity $\delta_{\mathrm{LR}}^2(C_1, C_2)$ from equation (27) is defined via the minima of two terms. The first term $\mathbb{E}\|\Pi_K g\|_2$ is the (square root of the) Gaussian width of the cone K, and is a familiar quantity from past work on least-squares estimation involving convex sets [10, 49]. The Gaussian width measure the size of the cone K, and it is to be expected that the minimax testing radius should grow with this size, since K characterizes the set of possible alternatives. The second term involving the inner product $\langle \eta, \mathbb{E}\Pi_K g \rangle$ is less immediately intuitive, partly because no such term arises in estimation over convex sets. The second term becomes dominant in cones for which the expectation $v^* := \mathbb{E}[\Pi_K g]$ is relatively large; for such cones, we can test between the null and alternative by performing a univariate test after projecting the data onto the direction v^*. This possibility only arises for cones that are more complicated than subspaces, since $\mathbb{E}[\Pi_K g] = 0$ for any subspace K.

Finally, we note that Theorem 1 gives a sharp characterization of the behavior of the GLRT up to a constant. It is different from the usual minimax guarantee. To the best of our knowledge, it is the first result to provide tight upper and lower control on the uniform performance of a specific test.

3.1.1. *Consequences for convex set alternatives.*

Although Theorem 1 applies to cone-based testing problems, it also has some implications for a more general class of problems based on convex set alternatives. In particular, suppose that we are interested in the testing problem of distinguishing between

$$
(30) \qquad \mathcal{H}_0 : \theta = \theta_0 \quad \text{versus} \quad \mathcal{H}_1 : \theta \in \mathcal{S},
$$

where $\mathcal{S}$ is a not necessarily a cone, but rather an arbitrary closed convex set, and θ_0 is some vector such that $\theta_0 \in \mathcal{S}$. Consider the tangent cone of $\mathcal{S}$ at θ_0, which is given by

$$
(31) \qquad \mathcal{T}_{\mathcal{S}}(\theta) := \{u \in \mathbb{R}^d \mid \text{there exists some } t > 0 \text{ such that } \theta + tu \in \mathcal{S}\}.
$$

Note that $\mathcal{T}_{\mathcal{S}}(\theta_0)$ contains the shifted set $\mathcal{S} - \theta_0$. Consequently, we have

$$
\mathcal{E}(\psi; \{0\}, \mathcal{S} - \theta_0, \varepsilon) \leq \mathbb{E}_{\theta=0}[\psi(y)] + \sup_{\theta \in \mathcal{T}_{\mathcal{S}}(\theta_0) \setminus B_2(0; \varepsilon)} \mathbb{E}_\theta[1 - \psi(y)]
$$

$$
= \mathcal{E}(\psi; \{0\}, \mathcal{T}_{\mathcal{S}}(\theta_0), \varepsilon),
$$

which shows that the tangent cone testing problem

$$
(32) \qquad \mathcal{H}_0 : \theta = 0 \quad \text{versus} \quad \mathcal{H}_1 : \theta \in \mathcal{T}_{\mathcal{S}}(\theta_0),
$$

is more challenging than the original problem (30). Thus, applying Theorem 1 to this cone-testing problem (32), we obtain the following.

COROLLARY 1. *For the convex set testing problem* (30), *we have*

$$
\begin{aligned}
(33) \quad & \varepsilon^2_{\mathrm{OPT}}(\theta_0, \mathcal{S}; \rho) \\
& \lesssim \sigma^2 \min\left\{ \mathbb{E}\|\Pi_{\mathcal{T}_{\mathcal{S}}(\theta_0)} g\|_2, \left(\frac{\mathbb{E}\|\Pi_{\mathcal{T}_{\mathcal{S}}(\theta_0)} g\|_2}{\max\{0, \inf_{\eta \in \mathcal{T}_{\mathcal{S}}(\theta_0) \cap S^{-1}} \langle \eta, \mathbb{E}\Pi_{\mathcal{T}_{\mathcal{S}}(\theta_0)} g \rangle\}} \right)^2 \right\}.
\end{aligned}
$$

This upper bound can be achieved by applying the GLRT to the tangent cone testing problem (32).

This corollary offers a general recipe of upper bounding the optimal testing radius. In Section 3.1.6, we provide an application of Corollary 1 to the problem of testing

$$
\mathcal{H}_0 : \theta = \theta_0 \quad \text{versus} \quad \mathcal{H}_1 : \theta \in M,
$$

where M is the monotone cone [defined in expression (2)]. When $\theta_0 \neq 0$, this is not a cone testing problem, since the set $\{\theta_0\}$ is not a cone. Using Corollary 1, we prove an upper bound on the optimal testing radius for this problem in terms of the number of constant pieces of θ_0.

In the remainder of this section, we consider some special cases of testing a cone K versus $\{0\}$ in order to illustrate the consequences of Theorem 1. In all cases, we compute the GLRT testing radius for a constant error probability, and so ignore the dependencies on ρ. For this reason, we adopt the more streamlined notation $\varepsilon_{\mathrm{GLR}}(K)$ for the radius $\varepsilon_{\mathrm{GLR}}(\{0\}, K; \rho)$.

3.1.2. *Subspace of dimension k.* Let us begin with an especially simple case—namely, when K is equal to a subspace S_k of dimension $k \leq d$. In this case, the projection Π_K is a linear operator, which can be represented by matrix multiplication using a rank k projection matrix. By symmetry of the Gaussian distribution, we have $\mathbb{E}[\Pi_K g] = 0$. Moreover, by rotation invariance of the Gaussian distribution, the random vector $\|\Pi_K g\|_2^2$ follows a χ^2-distribution with k degrees of freedom, where

$$
\frac{\sqrt{k}}{2} \leq \mathbb{E}\|\Pi_K g\|_2 \leq \sqrt{\mathbb{E}\|\Pi_K g\|_2^2} = \sqrt{k}.
$$

Applying Theorem 1 then yields that the testing radius of the GLRT scales as

$$
(34) \qquad \varepsilon^2_{\mathrm{GLR}}(S_k) \asymp \sigma^2 \sqrt{k}.
$$

Here our notation $\asymp$ denotes equality up to constants independent of (σ, k); we have omitted dependence on the testing error ρ so as to simplify notation, and will do so throughout our discussion.
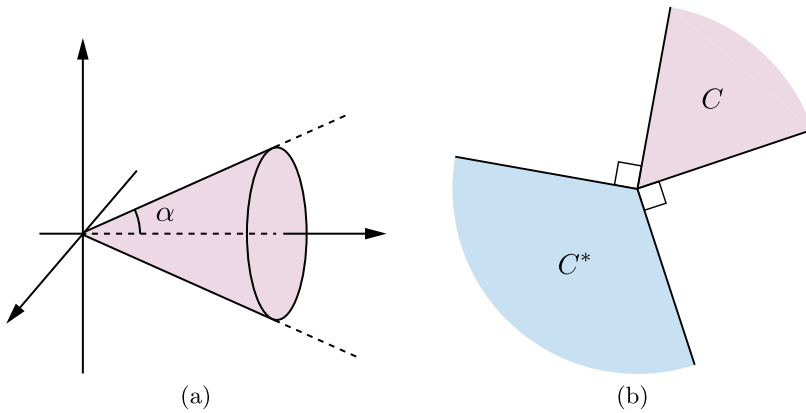
FIG. 1.   (a) *A 3-dimensional circular cone with angle* α. (b) *Illustration of a cone versus its polar cone.*

3.1.3. *Circular cone.*   A circular cone in $\mathbb{R}^d$ with constant angle $\alpha \in (0, \pi/2)$ is given by $\mathrm{Circ}_d(\alpha) := \{\theta \in \mathbb{R}^d \mid \theta_1 \geq \|\theta\|_2 \cos(\alpha)\}$. In geometric terms, it corresponds to the set of all vectors whose angle with the standard basis vector $e_1 = (1, 0, \dots, 0)$ is at most α radians. Figure 1(a) gives an illustration of a circular cone.

Suppose that we want to test the null hypothesis $\theta = 0$ versus the cone alternative $K = \mathrm{Circ}_d(\alpha)$. We claim that, in application to this particular cone, Theorem 1 implies that

$$(35) \qquad \varepsilon_{\mathrm{GLR}}^2(K) \asymp \sigma^2 \min\{\sqrt{d}, 1\} = \sigma^2,$$

where $\asymp$ denotes equality up to constants depending on (ρ, α), but independent of all other problem parameters.

In order to apply Theorem 1, we need to evaluate both terms that define the geometric quantity $\delta_{\mathrm{LR}}^2(C_1, C_2)$. On one hand, by symmetry of the cone $K = \mathrm{Circ}_d(\alpha)$ in its last $(d-1)$-coordinates, we have $\mathbb{E}\Pi_K g = \beta e_1$ for some scalar $\beta > 0$ and e_1 denotes the standard Euclidean basis vector with a 1 in the first coordinate. Moreover, for any $\eta \in K \cap S^{-1}$, we have $\eta_1 \geq \cos(\alpha)$, and hence

$$\inf_{\eta \in K \cap S^{-1}} \langle \eta, \mathbb{E}\Pi_K g \rangle = \eta_1 \beta \geq \cos(\alpha)\beta = \cos(\alpha)\|\mathbb{E}\Pi_K g\|_2.$$

Next, we claim that $\|\mathbb{E}\Pi_K g\|_2 \asymp \mathbb{E}\|\Pi_K g\|_2$. In order to prove this claim, note that Jensen's inequality yields

$$(36) \qquad \begin{aligned} \mathbb{E}\|\Pi_K g\|_2 &\geq \|\mathbb{E}\Pi_K g\|_2 \overset{(a)}{\geq} (\mathbb{E}\Pi_{\mathrm{Circ}_d(\alpha)} g)_1 \\ &= \mathbb{E}(\Pi_{\mathrm{Circ}_d(\alpha)} g)_1 \overset{(b)}{\geq} \mathbb{E}\|\Pi_{\mathrm{Circ}_d(\alpha)} g\|_2 \cos(\alpha), \end{aligned}$$

where in this argument, inequality (a) follows from simply fact that $\|x\|_2 \geq |x_1|$ whereas inequality (b) follows from the definition of circular cone. Plugging into definition $\delta^2_{LR}(C_1, C_2)$, the corresponding second term equals a constant. Therefore, the second term in the definition (27) of $\delta^2_{LR}(C_1, C_2)$ is upper bounded by a constant, independent of the dimension d.

On the other hand, from known results on circular cones (see [28], Section 6.3), there are constants $\kappa_j = \kappa_j(\alpha)$ for $j = 1, 2$ such that $\kappa_1 d \leq \mathbb{E}\|\Pi_K g\|_2^2 \leq \kappa_2 d$. Moreover, we have

$$\mathbb{E}\|\Pi_K g\|_2^2 - 4 \overset{(a)}{\leq} \left(\mathbb{E}\|\Pi_K g\|_2\right)^2 \overset{(b)}{\leq} \mathbb{E}\|\Pi_K g\|_2^2.$$

Here inequality (b) is an immediate consequence of Jensen's inequality, whereas inequality (a) follows from the fact that $\text{var}(\|\Pi_K g\|_2) \leq 4$; see Lemma 4.1 in Section 4.1 and the surrounding discussion for details. Putting together the pieces, we see that $\mathbb{E}\|\Pi_K g\|_2 \asymp \sqrt{d}$ for the circular cone. Combining different elements of our argument leads to the stated claim (35).

3.1.4. *A Cartesian product cone.* We now consider a simple extension of the previous two examples; namely, a convex cone formed by taking the Cartesian product of the real line $\mathbb{R}$ with the circular cone $\text{Circ}_{d-1}(\alpha)$, that is,

$$(37) \qquad K_\times := \text{Circ}_{d-1}(\alpha) \times \mathbb{R}.$$

Please refer to Figure 2 as an illustration of this cone in three dimensions.

This example turns out to be rather interesting because—as will be demonstrated in Section 3.2.3—the GLRT is suboptimal by a factor $\sqrt{d}$ for this cone. In order to set up this later analysis, here we use Theorem 1 to prove that

$$(38) \qquad \varepsilon^2_{GLR}(K_\times) \asymp \sigma^2 \sqrt{d}.$$

Note that this result is strongly suggestive of suboptimality on the part of the GLRT. More concretely, the two cones that form $K_\times$ are both "easy," in that the
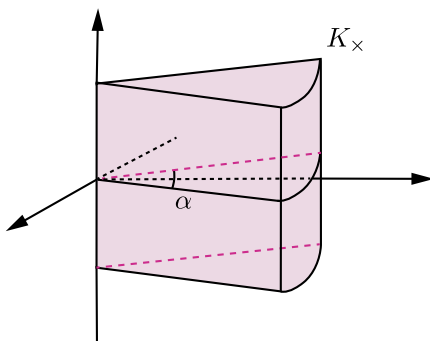

FIG. 2.   *Illustration of the product cone defined in equation* (37).

GLRT radius scales as σ^2 for each. For this reason, one would expect that the squared radius of an optimal test would scale as σ^2—as opposed to the $\sigma^2\sqrt{d}$ of the GLRT—and our later calculations will show that this is indeed the case.

We now prove claim (38) as a consequence of Theorem 1. First, notice that projecting to the product cone $K_\times$ can be viewed as projecting the first $d-1$ dimension to circular cone $\mathrm{Circ}_{d-1}(\alpha)$ and the last coordinate to $\mathbb{R}$. Consequently, we have the following inequality:

$$\mathbb{E}\|\Pi_{\mathrm{Circ}_{d-1}(\alpha)}g\|_2 \leq \mathbb{E}\|\Pi_{K_\times}g\|_2 \overset{(a)}{\leq} \sqrt{\mathbb{E}\|\Pi_{K_\times}g\|_2^2}$$

$$= \sqrt{\mathbb{E}\|\Pi_{\mathrm{Circ}_{d-1}(\alpha)}g\|_2^2 + \mathbb{E}[g_d^2]},$$

where inequality (a) follows by Jensen's inequality. Making use of our previous calculations for the circular cone, we have $\mathbb{E}\|\Pi_{K_\times}g\|_2 \asymp \sqrt{d}$. Moreover, note that the last coordinate of $\mathbb{E}[\Pi_{K_\times}g]$ is equal to 0 by symmetry and the standard basis vector $e_d \in \mathbb{R}^d$, with a single one in its last coordinate, belongs to $K_\times \cap S^{-1}$, we have

$$\inf_{\eta \in K_\times \cap S^{-1}} \langle \eta, \mathbb{E}\Pi_{K_\times}(g)\rangle \leq \langle e_d, \mathbb{E}\Pi_{K_\times}(g)\rangle = 0.$$

Plugging into definition $\delta_{\mathrm{LR}}^2(C_1, C_2)$, the corresponding second term equals infinity. Therefore, the minimum that defines $\delta_{\mathrm{LR}}^2(C_1, C_2)$ is achieved in the first term, and so is proportional to $\sqrt{d}$. Putting together the pieces yields the claim (38).

3.1.5. *Nonnegative orthant cone.* Next, let us consider the (nonnegative) orthant cone given by $K_+ := \{\theta \in \mathbb{R}^d \mid \theta_j \geq 0 \text{ for } j = 1, \ldots, d\}$. Here we use Theorem 1 to show that

(39) $$\varepsilon_{\mathrm{GLR}}^2(K_+) \asymp \sigma^2\sqrt{d}.$$

Turning to the evaluation of the quantity $\delta_{\mathrm{LR}}^2(C_1, C_2)$, it is straightforward to see that $[\Pi_{K_+}(\theta)]_j = \max\{0, \theta_j\}$, and hence $\mathbb{E}\Pi_{K_+}(g) = \frac{1}{2}\mathbb{E}|g_1|\mathbf{1} = \frac{1}{\sqrt{2\pi}}\mathbf{1}$, where $\mathbf{1} \in \mathbb{R}^d$ is a vector of all ones. Thus, we have

$$\|\mathbb{E}\Pi_{K_+}(g)\|_2 = \sqrt{\frac{d}{2\pi}} \quad \text{and}$$

$$\|\mathbb{E}\Pi_{K_+}(g)\|_2 \leq \mathbb{E}\|\Pi_{K_+}(g)\|_2 \leq \sqrt{\mathbb{E}\|\Pi_{K_+}(g)\|_2^2} = \sqrt{\frac{d}{2}},$$

where the second inequality follows from Jensen's inequality. So the first term in the definition of quantity $\delta_{\mathrm{LR}}^2(C_1, C_2)$ is proportional to $\sqrt{d}$. As for the second term, since the standard basis vector $e_1 \in K_+ \cap S^{-1}$, we have

$$\inf_{\eta \in K_+ \cap S^{-1}} \langle \eta, \mathbb{E}\Pi_K g\rangle \leq \left\langle e_1, \frac{1}{\sqrt{2\pi}}\mathbf{1}\right\rangle = \frac{1}{\sqrt{2\pi}}.$$

Consequently, the second term in the definition of quantity $\delta^2_{\mathrm{LR}}(C_1, C_2)$ lower bounded by a universal constant times d. Combining these derivations yields the stated claim (39).

3.1.6. *Monotone cone.* As our final example, consider testing in the monotone cone given by $M := \{\theta \in \mathbb{R}^d \mid \theta_1 \leq \theta_2 \leq \cdots \leq \theta_d\}$. Testing with monotone cone constraint has also been studied in different settings before, where it is known in some cases that restricting to monotone cone helps reduce the hardness of the problem to be logarithmically dependent on the dimension (e.g., [7, 51]).

Here we use Theorem 1 to show that

$$
\varepsilon^2_{\mathrm{GLR}}(M) \asymp \sigma^2 \sqrt{\log d}. \tag{40}
$$

From known results on monotone cone (see Section 3.5, [1]), we know that $\mathbb{E}\|\Pi_M g\|_2 \asymp \sqrt{\log d}$. So the only remaining detail is to control the second term defining $\delta^2_{\mathrm{LR}}(C_1, C_2)$. We claim that the second term is actually infinity since

$$
\max\left\{0, \inf_{\eta \in M \cap S^{-1}} \langle \eta, \mathbb{E}\Pi_M g \rangle\right\} = 0, \tag{41}
$$

which can be seen by simply noticing vectors $\frac{1}{\sqrt{d}}\mathbf{1}, -\frac{1}{\sqrt{d}}\mathbf{1} \in M \cap S^{-1}$ and

$$
\min\left\{\left\langle \frac{\mathbf{1}}{\sqrt{d}}, \mathbb{E}\Pi_M g \right\rangle, \left\langle -\frac{\mathbf{1}}{\sqrt{d}}, \mathbb{E}\Pi_M g \right\rangle\right\} \leq 0.
$$

Here $\mathbf{1} \in \mathbb{R}^d$ denotes the vector of all ones. Combining the pieces yields the claim (40).

Testing constant versus monotone. It is worth noting that the same GLRT bound also holds for the more general problem of testing the monotone cone M versus the linear subspace $L = \mathrm{span}(\mathbf{1})$ of constant vectors, namely,

$$
\varepsilon^2_{\mathrm{GLR}}(L, M) \asymp \sigma^2 \sqrt{\log d}. \tag{42}
$$

In particular, the following lemma provides the control that we need.

LEMMA 1. *For the monotone cone M and the linear space $L = \mathrm{span}(\mathbf{1})$, there is a universal constant c such that*

$$
\inf_{\eta \in K \cap S^{-1}} \langle \eta, \mathbb{E}\Pi_K g \rangle \leq c, \qquad K := M \cap L^\perp.
$$

See the Supplementary Material ([52], Appendix G.1) for the proof of this lemma.

Testing an arbitrary vector θ_0 versus the monotone cone.   Finally, let us consider an important implication of Corollary 1 in the context of testing departures in monotone cone. More precisely, for a fixed vector $\theta_0 \in M$, consider the testing problem

$$\text{(43)} \qquad \mathcal{H}_0 : \theta = \theta_0 \quad \text{versus} \quad \mathcal{H}_1 : \theta \in M.$$

Let us define $k(\theta_0)$ as the number of constant *pieces* of θ_0, by which we mean there exist integers $d_1, \ldots, d_{k(\theta_0)}$ with $d_i \geq 1$ and $d_1 + \cdots + d_{k(\theta_0)} = d$ such that θ_0 is a constant on each set $S_i := \{j : \sum_{t=1}^{i-1} d_t + 1 \leq j \leq \sum_{t=1}^{i} d_i\}$, for $1 \leq i \leq k(\theta_0)$.

We claim that Corollary 1 guarantees that the optimal testing radius satisfies

$$\text{(44)} \qquad \varepsilon_{\mathrm{OPT}}^2(\theta_0, M; \rho) \lesssim \sigma^2 \sqrt{k(\theta_0) \log\left(\frac{d}{k(\theta_0)}\right)}.$$

Note that this upper bound depends on the structure of θ_0 through how many pieces θ_0 possesses, which reveals the adaptive nature of Corollary 1.

In order to prove inequality (44), let us use shorthand k to denote $k(\theta_0)$. First, notice that both $\mathbf{1}/\sqrt{d}, -\mathbf{1}/\sqrt{d} \in \mathcal{T}_M(\theta_0)$, then

$$\max\left\{0, \inf_{\eta \in \mathcal{T}_M(\theta_0) \cap S^{-1}} \langle \eta, \mathbb{E}\Pi_{\mathcal{T}_M(\theta_0)} g \rangle\right\} \leq 0,$$

which implies the second term for $\delta_{\mathrm{LR}}^2(C_1, C_2)$ goes to infinity. It only remains to calculate $\mathbb{E}\|\Pi_{\mathcal{T}_M(\theta_0)} g\|_2$. Since the tangent cone $\mathcal{T}_M(\theta_0)$ equals the Cartesian product of k monotone cones, namely $\mathcal{T}_M(\theta_0) = M_{d_1} \times \cdots \times M_{d_k}$, we have

$$\mathbb{E}\|\Pi_{\mathcal{T}_M(\theta_0)} g\|_2^2 = \mathbb{E}\|\Pi_{M_{d_1}} g\|_2^2 + \cdots + \mathbb{E}\|\Pi_{M_{d_k}} g\|_2^2 = \log(d_1) + \cdots + \log(d_k)$$

$$\leq k \log\left(\frac{d}{k}\right),$$

where the last step follows from convexity of the logarithm function. Therefore, Jensen's inequality guarantees that

$$\mathbb{E}\|\Pi_{\mathcal{T}_M(\theta_0)} g\|_2 \leq \sqrt{\mathbb{E}\|\Pi_{\mathcal{T}_M(\theta_0)} g\|_2^2} \leq \sqrt{k \log\left(\frac{d}{k}\right)}.$$

Putting the pieces together, Corollary 1 guarantees that the claimed inequality (44) holds for the testing problem (43).

3.2. *Lower bounds on the testing radius.*   Thus far, we have derived sharp bounds for a particular procedure—namely, the GLRT. Of course, it is of interest to understand when the GLRT is actually an optimal test, meaning that there is no other test that can discriminate between the null and alternative for smaller separations. In this section, we use information-theoretic methods to derive a lower bound on the optimal testing radius $\varepsilon_{\mathrm{OPT}}$ for every pair of nonoblique and nested

closed convex cones (C_1, C_2). Similar to Theorem 1, this bound depends on the geometric structure of intersection cone $K := C_2 \cap C_1^*$, where C_1^* is the polar cone to C_1.

In particular, let us define the quantity

$$(45) \qquad \delta_{\mathrm{OPT}}^2(C_1, C_2) := \min\left\{ \mathbb{E}\|\Pi_K g\|_2, \left(\frac{\mathbb{E}\|\Pi_K g\|_2}{\sup_{\eta \in K \cap S^{-1}} \langle \eta, \mathbb{E}\Pi_K g \rangle} \right)^2 \right\}.$$

Note that the only difference from $\delta_{\mathrm{LR}}^2(C_1, C_2)$ is the replacement of the infimum over $K \cap S^{-1}$ with a supremum, in the denominator of the second term. Moreover, since the supremum is achieved at $\frac{\mathbb{E}\Pi_K g}{\|\mathbb{E}\Pi_K g\|_2}$, we have $\sup_{\eta \in K \cap S^{-1}} \langle \eta, \mathbb{E}\Pi_K g \rangle = \|\mathbb{E}\Pi_K g\|_2$. Consequently, the second term on the right-hand side of equation (45) can be also written in the equivalent form $(\frac{\mathbb{E}\|\Pi_K g\|_2}{\|\mathbb{E}\Pi_K g\|_2})^2$.

With this notation in hand, we are now ready to state a general lower bound for minimax optimal testing radius.

THEOREM 2.   *There are numbers $\{\kappa_\rho, \rho \in (0, 1/2]\}$ such that for every nested pair of nonoblique closed convex cones $C_1 \subset C_2$, we have*

$$(46) \qquad \inf_\psi \mathcal{E}(\psi; C_1, C_2, \varepsilon) \geq \rho \qquad \text{whenever } \varepsilon^2 \leq \kappa_\rho \sigma^2 \delta_{\mathrm{OPT}}^2(C_1, C_2).$$

In particular, we can take $\kappa_\rho = 1/14$ for all $\rho \in (0, 1/2]$.

REMARKS.   In more compact terms, Theorem 2 can be understood as guaranteeing

$$\varepsilon_{\mathrm{OPT}}(C_1, C_2; \rho) \gtrsim \sigma \delta_{\mathrm{OPT}}(C_1, C_2),$$

where $\gtrsim$ denotes an inequality up to constants (with ρ viewed as fixed).

Theorem 2 is proved by constructing a distribution over the alternative $\mathcal{H}_1$ supported only on those points in $\mathcal{H}_1$ that are hard to distinguish from $\mathcal{H}_0$. Based on this construction, the testing error can be lower bound by controlling the total variation distance between two marginal likelihood functions. We refer our readers to our Section 4.2 for more details on this proof.

One useful consequence of Theorem 2 is in providing a sufficient condition for optimality of the GLRT, which we summarize here.

COROLLARY 2 (Sufficient condition for optimality of GLRT).   *Given the cone $K = C_2 \cap C_1^*$, suppose that there is a numerical constant $b > 1$, independent of K and all other problem parameters, such that*

$$(47) \qquad \sup_{\eta \in K \cap S^{-1}} \langle \eta, \mathbb{E}\Pi_K g \rangle = \|\mathbb{E}\Pi_K g\|_2 \leq b \inf_{\eta \in K \cap S^{-1}} \langle \eta, \mathbb{E}\Pi_K g \rangle.$$

Then the GLRT is a minimax optimal test, that is,

$$\varepsilon_{\mathrm{GLR}}(C_1, C_2; \rho) \asymp \varepsilon_{\mathrm{OPT}}(C_1, C_2; \rho).$$

It is natural to wonder whether the condition (47) is also necessary for optimality of the GLRT. This turns out not to be the case. The monotone cone, to be revisited in Section 3.3.2, provides an instance of a cone testing problem for which the GLRT is optimal while condition (47) is violated. Let us now return to these concrete examples.

3.2.1. *Revisiting the k-dimensional subspace.* Let S_k be a subspace of dimension $k \leq d$. In our earlier discussion in Section 3.1.2, we established that $\varepsilon_{\mathrm{GLR}}^2(S_k) \asymp \sigma^2 \sqrt{k}$. Let us use Corollary 2 to verify that the GLRT is optimal for this problem. For a k-dimensional subspace $K = S_k$, we have $\mathbb{E}\Pi_K g = 0$ by symmetry; consequently, condition (47) holds in a trivial manner. Thus, we conclude that $\varepsilon_{\mathrm{OPT}}^2(S_k) \asymp \varepsilon_{\mathrm{GLR}}^2(S_k)$, showing that the GLRT is optimal over all tests.

3.2.2. *Revisiting the circular cone.* Recall the circular cone $K = \{\theta \in \mathbb{R}^d \mid \theta_1 \geq \|\theta\|_2 \cos(\alpha)\}$ for fixed $0 < \alpha < \pi/2$. In our earlier discussion, we proved that $\varepsilon_{\mathrm{GLR}}^2(K) \asymp \sigma^2$. Here let us verify that this scaling is optimal over all tests. By symmetry, we find that $\mathbb{E}\Pi_K g = \beta e_1 \in \mathbb{R}^d$, where e_1 denotes the standard Euclidean basis vector with a 1 in the first coordinate, and $\beta > 0$ is some scalar. For any vector $\eta \in K \cap S^{-1}$, we have $\eta_1 \geq \cos(\alpha)$, and hence

$$\inf_{\eta \in K \cap S^{-1}} \langle \eta, \mathbb{E}\Pi_K g \rangle \geq \cos(\alpha)\beta = \cos(\alpha)\|\mathbb{E}\Pi_K g\|_2.$$

Consequently, we see that condition (47) is satisfied with $b = \frac{1}{\cos(\alpha)} > 0$, so that the GLRT is optimal over all tests for each fixed α. (To be clear, in this example, our theory does not provide a sharp bound uniformly over varying α.)

3.2.3. *Revisiting the product cone.* Recall from Section 3.1.4 our discussion of the Cartesian product cone $K_\times = \mathrm{Circ}_{d-1}(\alpha) \times \mathbb{R}$. In this section, we establish that the GLRT, when applied to a testing problem based on this case, is suboptimal by a factor of $\sqrt{d}$.

Let us first prove that the sufficient condition (47) is violated, so that Corollary 2 does *not* imply optimality of the GLRT. From our earlier calculations, we know that $\mathbb{E}\|\Pi_{K_\times} g\|_2 \asymp \sqrt{d}$. On the other hand, we also know that $\mathbb{E}\Pi_{K_\times} g$ is equal to zero in its last coordinate. Since the standard basis vector e_d belongs to the set $K_\times \cap S^{-1}$, we have

$$\inf_{\eta \in K_\times \cap S^{-1}} \langle \eta, \mathbb{E}\Pi_{K_\times} g \rangle \leq \langle e_d, \mathbb{E}\Pi_{K_\times} g \rangle = 0,$$

so that condition (47) does not hold.

From this calculation alone, we cannot conclude that the GLRT is suboptimal. So let us now compute the lower bound guaranteed by Theorem 2. From our previous discussion, we know that $\mathbb{E}\Pi_{K_\times} g = \beta e_1$ for some scalar $\beta > 0$. Moreover, we also have $\|\mathbb{E}\Pi_{K_\times} g\|_2 = \beta \asymp \sqrt{d}$; this scaling follows because we have

$\|\mathbb{E}\Pi_{K_\times}g\|_2 = \|\mathbb{E}\Pi_{\mathrm{Circ}_{d-1}(\alpha)}g\|_2 \asymp \sqrt{d-1}$, where we have used the previous inequality (36) for circular cone. Putting together the pieces, we find that Theorem 2 implies that

$$(48) \qquad\qquad \varepsilon_{\mathrm{OPT}}^2(K_\times) \gtrsim \sigma^2,$$

which differs from the GLRT scaling in a factor of $\sqrt{d}$.

Does there exist a test that achieves the lower bound (48)? It turns out that a simple truncation test does so, and hence is optimal. To provide intuition for the test, observe that for any vector $\theta \in K_\times \cap S^{-1}$, we have $\theta_1^2 + \theta_d^2 \geq \cos^2(\alpha)$. To verify this claim, note that

$$\frac{1}{\cos^2(\alpha)}(\theta_1^2 + \theta_d^2) \geq \frac{\theta_1^2}{\cos^2(\alpha)} + \theta_d^2 \geq \sum_{j=1}^{d-1} \theta_j^2 + \theta_d^2 = 1.$$

Consequently, the two coordinates (y_1, y_d) provide sufficient information for constructing a good test. In particular, consider the truncation test

$$\varphi(y) := \mathbb{I}\big[\big\|(y_1, y_d)\big\|_2 \geq \beta\big]$$

for some threshold $\beta > 0$ to be determined. This can be viewed as a GLRT for testing the standard null against the alternative $\mathbb{R}^2$, and hence our general theory guarantees that it will succeed with separation $\varepsilon^2 \gtrsim \sigma^2$. This guarantee matches our lower bound (48), showing that the truncation test is indeed optimal, and moreover, that the GLRT is suboptimal by a factor of $\sqrt{d}$ for this particular problem.

We provide more intuition on why the the GLRT suboptimal and use this intuition to construct a more general class of problems for which a similar suboptimality is witnessed in the Supplementary Material ([52], Appendix A).

3.3. *Detailed analysis of two cases.*   This section is devoted to a detailed analysis of the orthant cone, followed by the monotone cone. Here we find that the GLRT is again optimal for both of these cones, but establishing this optimality requires a more delicate analysis.

3.3.1. *Revisiting the orthant cone.*   Recall from Section 3.1.5 our discussion of the (nonnegative) orthant cone

$$K_+ := \{\theta \in \mathbb{R}^d \mid \theta_j \geq 0 \text{ for } j = 1, \dots, d\},$$

where we proved that $\varepsilon_{\mathrm{GLR}}^2(K_+) \asymp \sigma^2\sqrt{d}$. Let us first show that the sufficient condition (47) does not hold, so that Corollary 2 does *not* imply optimality of the GLRT. As we have computed in our Section 3.1.5, quantity $\mathbb{E}\|\Pi_{K_+}(g)\|_2 \asymp \sqrt{d}$ and

$$\inf_{\eta \in K_+ \cap S^{-1}} \langle \eta, \mathbb{E}\Pi_K g \rangle \leq \Big\langle e_1, \frac{1}{\sqrt{2\pi}}\mathbf{1} \Big\rangle = \frac{1}{\sqrt{2\pi}},$$

where use the fact that $\mathbb{E}\Pi_{K_+}(g) = \frac{1}{\sqrt{2\pi}}\mathbf{1}$. So that condition (47) is violated.

Does this mean the GLRT is suboptimal? It turns out that the GLRT is actually optimal over all tests, as we can demonstrate by proving a lower bound—tighter than the one given in Theorem 2—that matches the performance of the GLRT. We summarize it as follows.

PROPOSITION 1.   *There are numbers* $\{\kappa_\rho, \rho \in (0, 1/2]\}$ *such that for the (non-negative) orthant cone* K_+, *we have*

$$(49) \qquad \inf_\psi \mathcal{E}(\psi; \{0\}, K_+, \varepsilon) \geq \rho \qquad \text{whenever } \varepsilon^2 \leq \kappa_\rho \sigma^2 \sqrt{d}.$$

See the Supplementary Material ([52], Section B.1) for the proof of this proposition.

From Proposition 1, we see that the optimal testing radius satisfies $\varepsilon^2_{\text{OPT}}(K_+) \gtrsim \sigma^2 \sqrt{d}$. Compared to the GLRT radius $\varepsilon^2_{\text{GLR}}(K_+)$ established in expression (39), it implies the optimality of the GLRT.

3.3.2. *Revisiting the monotone cone.*   Recall the monotone cone given by $M := \{\theta \in \mathbb{R}^d \mid \theta_1 \leq \theta_2 \leq \cdots \leq \theta_d\}$. In our previous discussion in Section 3.1.6, we established that $\varepsilon^2_{\text{GLR}}(M) \asymp \sigma^2 \sqrt{\log d}$. We also pointed out that this scaling holds for a more general problem, namely, testing cone M versus linear subspace $L = \text{span}(\mathbf{1})$. In this section, we show that the GLRT is also optimal for both cases.

First, observe that Corollary 2 does not imply optimality of the GLRT. In particular, using symmetry of the inner product, we have shown in expression (41) that

$$\max\left\{0, \inf_{\eta \in M \cap S^{-1}} \langle \eta, \mathbb{E}\Pi_M g \rangle\right\} = 0$$

for cone pair $(C_1, C_2) = (\{0\}, M)$. Also note that from Lemma 1 we know that for cone pair $(C_1, C_2) = (\text{span}(\mathbf{1}), M)$, there is a universal constant c such that

$$\inf_{\eta \in K \cap S^{-1}} \langle \eta, \mathbb{E}\Pi_K g \rangle \leq c, \qquad K := M \cap L^\perp.$$

In both cases, since $\mathbb{E}\|\Pi_K g\|_2 \asymp \sqrt{\log d}$, so that the sufficient condition (47) for GLRT optimality fails to hold.

It turns out that we can demonstrate a matching lower bound for $\varepsilon^2_{\text{OPT}}(M)$ in a more direct way by carefully constructing a prior distribution on the alternatives and control the testing error. Doing so allows us to conclude that the GLRT is optimal, and we summarize our conclusions in the following.

PROPOSITION 2.   *There are numbers* $\{\kappa_\rho, \rho \in (0, 1/2]\}$ *such that for the monotone cone* M *and subspace* $L = \{0\}$ *or* $\text{span}(\mathbf{1})$, *we have*

$$(50) \qquad \inf_\psi \mathcal{E}(\psi; L, M, \varepsilon) \geq \rho \qquad \text{whenever } \varepsilon^2 \leq \kappa_\rho \sigma^2 \sqrt{\log(ed)}.$$

See the Supplementary Material ([52], Section B.2) for the proof of this proposition.

Proposition 2, equipped with previous achievable results by GLRT (40), gives a sharp rate characterization on the testing radius for both problems with regard to monotone cone:

$$\mathcal{H}_0 : \theta = 0 \quad \text{versus} \quad \mathcal{H}_1 : \theta \in M \quad \text{and}$$

$$\mathcal{H}_0 : \theta \in \text{span}(\mathbf{1}) \quad \text{versus} \quad \mathcal{H}_1 : \theta \in M.$$

In both cases, the optimal testing radius satisfies $\varepsilon_{\text{OPT}}^2(L, M, \rho) \asymp \sigma^2 \sqrt{\log(ed)}$. As a consequence, the GLRT is optimal up to an universal constant. As far as we know, the problem of testing a zero or constant vector versus the monotone cone as the alternative has not been fully characterized in any past work.

4. Proofs of main results. We now turn to the proofs of our main results, with the proof of Theorems 1 and 2 given in Sections 4.1 and 4.2 respectively. In all cases, we defer the proofs of certain more technical lemmas to the Appendices in the Supplementary Material [52].

4.1. *Proof of Theorem* 1. Since the cones (C_1, C_2) are both invariant under rescaling by positive numbers, we can first prove the result for noise level $\sigma = 1$, and then recapture the general result by rescaling appropriately. Thus, we fix $\sigma = 1$ throughout the remainder of the proof so as to simplify notation. Moreover, let us recall that the GLRT consists of tests of the form $\phi_\beta(y) := \mathbb{I}(T(y) \geq \beta)$, where the likelihood ratio $T(y)$ is given in equation (11a). Note here the cut-off $\beta \in [0, \infty)$ is a constant that does not depend on the data vector y.

By the previously discussed equivalence (26), we can focus our attention on the simpler problem $\mathcal{T}(\{0\}, K; \varepsilon)$, where $K = C_2 \cap C_1^*$. By the monotonicity of the square function for positive numbers, the GLRT is controlled by the behavior of the statistic $\|\Pi_K(y)\|_2$, and in particular how it varies depending on whether y is drawn according to $\mathcal{H}_0$ or $\mathcal{H}_1$.

Letting $g \in \mathbb{R}^d$ denote a standard Gaussian random vector, let us introduce the random variable $Z(\theta) := \|\Pi_K(\theta + g)\|_2$ for each $\theta \in \mathbb{R}^d$. Observe that the statistic $\|\Pi_K(y)\|_2$ is distributed according to $Z(0)$ under the null $\mathcal{H}_0$, and according to $Z(\theta)$ for some $\theta \in K$ under the alternative $\mathcal{H}_1$. The following lemma which is proved in Appendix D.1 guarantees random variables of the type $Z(\theta)$ and $\langle \theta, \Pi_K g \rangle$ are sharply concentrated around their expectations.

LEMMA 4.1. *For a standard Gaussian random vector* $g \sim N(0, I_d)$, *closed convex cone* $K \in \mathbb{R}^d$ *and vector* $\theta \in \mathbb{R}^d$, *we have*

$$(51a) \qquad \mathbb{P}\big(\pm (Z(\theta) - \mathbb{E}[Z(\theta)]) \geq t\big) \leq \exp\left(-\frac{t^2}{2}\right) \quad \text{and}$$

$$(51b) \qquad \mathbb{P}\big(\pm (\langle \theta, \Pi_K g \rangle - \mathbb{E}\langle \theta, \Pi_K g \rangle) \geq t\big) \leq \exp\left(-\frac{t^2}{2\|\theta\|_2^2}\right),$$

where both inequalities hold for all $t \geq 0$.

As shown in the sequel, using the concentration bound (51a), the study of the GLRT can be reduced to the problem of bounding the mean difference

$$\Gamma(\theta) := \mathbb{E}\big(\|\Pi_K(\theta + g)\|_2 - \|\Pi_K g\|_2\big) \tag{52}$$

for each $\theta \in K$. In particular, in order to prove the achievability result stated in part (a) of Theorem 1, we need to lower bound $\Gamma(\theta)$ uniformly over $\theta \in K$, whereas a uniform upper bound on $\Gamma(\theta)$ is required in order to prove the negative result in part (b).

4.1.1. *Proof of GLRT achievability result* [*Theorem* 1(a)]. By assumption, we can restrict our attention to alternative distributions defined by vectors $\theta \in K$ satisfying the lower bound $\|\theta\|_2^2 \geq B_\rho \delta_{\mathrm{LR}}^2(\{0\}, K)$, where for every target level $\rho \in (0, 1)$, constant B_ρ is chosen such that

$$B_\rho := \max\left\{32\pi, \inf\left(B > 0 \;\Big|\; \frac{B^{1/2}}{(2^7 \pi B)^{1/4} + 16} - \frac{2}{\sqrt{e}} \geq \sqrt{-8\log(\rho/2)}\right)\right\}.$$

Since function $f(x) := \frac{x^{1/2}}{(2^7 \pi x)^{1/4} + 16} - \frac{2}{\sqrt{e}}$ is strictly increasing and goes to infinity, so that the constant B_ρ defined above is always finite.

We first claim that it suffices to show that for such vector, the difference (52) is lower bounded as

$$\Gamma(\theta) \geq \frac{B_\rho^{1/2}}{(2^7 \pi B_\rho)^{1/4} + 16} - \frac{2}{\sqrt{e}} = f(B_\rho). \tag{53}$$

Taking inequality (53) as given for the moment, we claim that the test

$$\phi_\tau(y) = \mathbb{I}\big[\|\Pi_K(y)\|_2^2 \geq \tau\big]$$

with threshold

$$\tau := \left(\frac{1}{2} f(B_\rho) + \mathbb{E}\big[\|\Pi_K(g)\|_2\big]\right)^2$$

has uniform error probability controlled as

$$\mathcal{E}(\phi_\tau; \{0\}, K, \varepsilon) := \mathbb{E}_0[\phi_\tau(y)] + \sup_{\theta \in K, \|\theta\|_2^2 \geq \varepsilon^2} \mathbb{E}_\theta\big[1 - \phi_\tau(y)\big]$$

$$\leq 2e^{-f^2(B_\rho)/8} < \rho, \tag{54}$$

where the last inequality follows from the definition of B_ρ.

Establishing the error control (54).   Beginning with errors under the null $\mathcal{H}_0$, we have

$$\mathbb{E}_0[\phi_\tau(y)] = \mathbb{P}_0(\|\Pi_K g\|_2 \geq \sqrt{\tau}) = \mathbb{P}_0[\|\Pi_K g\|_2 - \mathbb{E}[\|\Pi_K g\|_2] \geq f(B_\rho)/2]$$
$$\leq \exp(-f^2(B_\rho)/8),$$

where the final inequality follows from the concentration bound (51a) in Lemma 4.1, as along as $f(B_\rho) > 0$.

On the other hand, we have

$$\sup_{\theta \in K, \|\theta\|_2^2 \geq \varepsilon^2} \mathbb{E}_\theta[1 - \phi_\tau(y)]$$

$$= \mathbb{P}\left[\|\Pi_K(\theta + g)\|_2 \leq \frac{1}{2}f(B_\rho) + \mathbb{E}\|\Pi_K g\|_2\right]$$

$$= \mathbb{P}\left[\|\Pi_K(\theta + g)\|_2 - \mathbb{E}\|\Pi_K(\theta + g)\|_2 \leq \frac{1}{2}f(B_\rho) - \Gamma(\theta)\right],$$

where the last equality follows by substituting $\Gamma(\theta) = \mathbb{E}[\|\Pi_K(\theta + g)\|_2] - \mathbb{E}[\|\Pi_K g\|_2]$. Since the lower bound (53) guarantees that $\frac{1}{2}f(B_\rho) - \Gamma(\theta) \leq -\frac{1}{2}f(B_\rho)$, we find that

$$\sup_{\theta \in K, \|\theta\|_2^2 \geq \varepsilon^2} \mathbb{E}_\theta[1 - \phi_\tau(y)] \leq \mathbb{P}\left[\|\Pi_K(\theta + g)\|_2 - \mathbb{E}\|\Pi_K(\theta + g)\|_2 \leq -\frac{1}{2}f(B_\rho)\right]$$

$$\leq \exp(-f^2(B_\rho)/8),$$

where the final inequality again uses the concentration inequality (51a). Putting the pieces together yields the claim (54).

The only remaining detail is to prove the lower bound (53) on the difference (52). To mainstream our proof, we leave the proof of this detail to Supplementary Material ([52], Appendix D.2).

4.1.2. *Proof of GLRT lower bound* [*Theorem* 1(b)].   We divide our proof into two scenarios, depending on whether or not $\mathbb{E}\|\Pi_K g\|_2$ is less than 128. We focus on the case when $\mathbb{E}\|\Pi_K g\|_2 \geq 128$ and leave the other scenario to Appendix E.1.

In this case, our strategy is to exhibit some $\theta \in \mathcal{H}_1$ for which the expected difference $\Gamma(\theta) = \mathbb{E}(\|\Pi_K(\theta + g)\|_2 - \|\Pi_K g\|_2)$ is small, which then leads to significant error when using the GLRT. In order to do so, we require an auxiliary lemma (Lemma E.1) to suitable control $\Gamma(\theta)$ which is stated and proved in the Supplementary Material ([52], Appendix E.2).

We now proceed to prove our main claim. Based on Lemma E.1, we claim that if $\varepsilon^2 \leq b_\rho \delta_{\mathrm{LR}}^2(\{0\}, K)$ for a suitably small constant b_ρ such that

$$b_\rho := \sup\left\{b_\rho > 0 \,\Big|\, 12\sqrt{b_\rho} + 3\sqrt{b_\rho}\left(\frac{2}{e}\right)^{1/4} + 24\sqrt{\frac{b_\rho}{2e}} \leq \frac{1}{16}\right\},$$

then

(55) $$\Gamma(\theta) \le \frac{1}{16} \qquad \text{for some } \theta \in K, \|\theta\|_2 \ge \varepsilon.$$

We take inequality (55) as given for now, returning to prove it in Appendix E.4 in the Supplementary Material [52]. In summary then, we have exhibited some $\theta \in \mathcal{H}_1$—namely, a vector $\theta \in K$ with $\|\theta\|_2 \ge \varepsilon$—such that $\Gamma(\theta) \le 1/16$. This special vector θ plays a central role in our proof.

We claim that the GLRT cannot succeed with error smaller than 0.11 no matter how the cut-off β is chosen. We leave this calculation to Appendix E.3 in the Supplementary Material [52].

4.2. *Proof of Theorem* 2. We now turn to the proof of Theorem 2. As in the proof of Theorem 1, we can assume without loss of generality that $\sigma = 1$. Since $0 \in C_1$ and $K := C_2 \cap C_1^* \subseteq C_2$, it suffices to prove a lower bound for the reduced problem of testing

$$\mathcal{H}_0 : \theta = 0 \quad \text{versus} \quad \mathcal{H}_1 : \|\theta\|_2 \ge \varepsilon, \qquad \theta \in K.$$

Let $B(1) = \{\theta \in \mathbb{R}^d \mid \|\theta\|_2 < 1\}$ denotes the open Euclidean ball of radius 1, and let $B^c(1) := \mathbb{R}^d \setminus B(1)$ denotes its complement.

We divide our analysis into two cases, depending on whether or not $\mathbb{E}\|\Pi_K g\|_2$ is less than 7. In both cases, let us set $\kappa_\rho = 1/14$.

Case 1. Suppose that $\mathbb{E}\|\Pi_K g\|_2 < 7$. In this case,

$$\varepsilon^2 \le \kappa_\rho \delta_{\mathrm{OPT}}^2(\{0\}, K) \le \kappa_\rho \mathbb{E}\|\Pi_K g\|_2 < 1/2.$$

Similar to our proof of Theorem 1(b), Case 1, by reducing to the simple verses simple testing problem (85a), any test yields testing error no smaller than $1/2$ if $\varepsilon^2 < 1/2$. So our lower bound directly holds for the case when $\mathbb{E}\|\Pi_K g\|_2 < 7$.

Case 2. Otherwise, suppose we have $\mathbb{E}\|\Pi_K g\|_2 \ge 7$. The following lemma provides a generic way to lower bound the testing error.

LEMMA 2. *For every nontrivial closed convex cone K and probability measure $\mathbb{Q}$ supported on $K \cap B^c(1)$, the testing error is lower bounded as*

(56) $$\inf_\psi \mathcal{E}(\psi; \{0\}, K, \varepsilon) \ge 1 - \frac{1}{2}\sqrt{\mathbb{E}_{\eta, \eta'} \exp(\varepsilon^2 \langle \eta, \eta' \rangle) - 1},$$

where $\mathbb{E}_{\eta, \eta'}$ denotes expectation with respect to an i.i.d. pair $\eta, \eta' \sim \mathbb{Q}$.

See the Supplementary Material ([52], Appendix F.1) for the proof of this claim.

We apply Lemma 2 with the probability measure $\mathbb{Q}$ defined as

$$(57) \qquad \mathbb{Q}(A) := \mathbb{P}\left(\frac{\Pi_K g}{\mathbb{E}\|\Pi_K g\|_2/2} \in A \,\Big|\, \|\Pi_K g\|_2 \geq \mathbb{E}\|\Pi_K g\|_2/2 \right)$$

for measurable set $A \subset \mathbb{R}^d$ where g denotes a standard d-dimensional Gaussian random vector, that is, $g \sim N(0, I_d)$. It is easy to check that measure $\mathbb{Q}$ is supported on $K \cap B^c(1)$. We make use of Lemma F.1 in Appendix F.2 to control $\mathbb{E}_{\eta, \eta'} \exp(\varepsilon^2 \langle \eta, \eta' \rangle)$ and thus upper bounding the testing error.

We now lower bound the testing error when $\varepsilon^2 \leq \kappa_\rho \delta_{\mathrm{OPT}}^2(\{0\}, K)$. By definition of $\delta_{\mathrm{OPT}}^2(\{0\}, K)$, the inequality $\varepsilon^2 \leq \kappa_\rho \delta_{\mathrm{OPT}}^2(\{0\}, K)$ implies that

$$\varepsilon^2 \leq \kappa_\rho \mathbb{E}\|\Pi_K g\|_2 \quad \text{and} \quad \varepsilon^2 \leq \kappa_\rho \left(\frac{\mathbb{E}\|\Pi_K g\|_2}{\|\mathbb{E}\Pi_K g\|_2} \right)^2.$$

The first inequality above implies, with $\kappa_\rho = 1/14$, that $\varepsilon^2 \leq \mathbb{E}\|\Pi_K g\|_2/14 \leq (\mathbb{E}\|\Pi_K g\|_2)^2/32$ (note that $\mathbb{E}\|\Pi_K g\|_2 \geq 7$). Therefore, the assumption in Lemma F.1 is satisfied so that inequality (103) gives

$$(58) \qquad \mathbb{E}_{\eta, \eta'} \exp(\varepsilon^2 \langle \eta, \eta' \rangle) \leq \frac{1}{a^2} \exp\left(5\kappa_\rho + \frac{40\kappa_\rho^2 \mathbb{E}(\|\Pi_K g\|_2^2)}{(\mathbb{E}\|\Pi_K g\|_2)^2} \right).$$

So it suffices to control the right-hand side above. From the concentration result in Lemma 4.1, we obtain

$$a = \mathbb{P}\left(\|\Pi_K g\|_2 - \mathbb{E}\|\Pi_K g\|_2 \geq -\frac{1}{2} \mathbb{E}\|\Pi_K g\|_2 \right)$$

$$\geq 1 - \exp\left(-\frac{(\mathbb{E}\|\Pi_K g\|_2)^2}{8} \right) > 1 - \exp(-6),$$

where the last step uses $\mathbb{E}\|\Pi_K g\|_2 \geq 7$, and

$$\mathbb{E}\|\Pi_K g\|_2^2 = (\mathbb{E}\|\Pi_K g\|_2)^2 + \mathrm{var}(\|\Pi_K g\|_2) \leq (\mathbb{E}\|\Pi_K g\|_2)^2 + 4.$$

Here the last inequality follows from the fact that $\mathrm{var}(\|\Pi_K g\|_2) \leq 4$; see Lemma 4.1. Plugging these two inequalities into expression (58) gives

$$\mathbb{E}_{\eta, \eta'} \exp(\varepsilon^2 \langle \eta, \eta' \rangle) \leq \left(\frac{1}{1 - \exp(-6)} \right)^2 \exp\left(5\kappa_\rho + 40\kappa_\rho^2 + \frac{160\kappa_\rho^2}{(\mathbb{E}\|\Pi_K g\|_2)^2} \right),$$

where the right hand side is less than 2 when $\kappa_\rho = 1/14$ and $\mathbb{E}\|\Pi_K g\|_2 \geq 7$. Combining with inequality (56) forces the testing error to be lower bounded as

$$\forall \psi, \qquad \mathcal{E}(\psi; \{0\}, K, \varepsilon) \geq 1 - \frac{1}{2}\sqrt{\mathbb{E}_{\eta, \eta'} \exp(\varepsilon^2 \langle \eta, \eta' \rangle) - 1} \geq \frac{1}{2} > \rho,$$

which completes the proof of Theorem 2.

5. Discussion.   In this paper, we have studied the the problem of testing between two hypotheses that are specified by a pair of nonoblique closed convex cones. Our first main result provided a characterization, sharp up to universal multiplicative constants, of the testing radius achieved by the generalized likelihood ratio test. This characterization was geometric in nature, depending on a combination of the Gaussian width of an induced cone, and a second geometric parameter. Due to the combination of these parameters, our analysis shows that the GLRT can have very different behavior even for cones that have the same Gaussian width; for instance, compare our results for the circular and orthant cone in Section 3.1. It is worth noting that this behavior is in sharp contrast to the situation for estimation problems over convex sets, where it is understood that (localized) Gaussian widths completely determine the estimation error associated with the least-squares estimator [10, 49]. In this way, our analysis reveals a fundamental difference between minimax testing and estimation.

Our analysis also highlights some new settings in which the GLRT is nonoptimal. Although past work [32, 35, 50] has exhibited nonoptimality of the GLRT in certain settings, in the context of cones, all of these past examples involve oblique cones. In Section 3.1.4, we gave an example of suboptimality which, to the best of our knowledge, is the first for a nonoblique pair of cones—namely, the cone {0}, and a certain type of Cartesian product cone.

Our work leaves open various questions, and we conclude by highlighting a few here. First, in Section 3.2, we proved a general information-theoretic lower bound for the minimax testing radius. This lower bound provides a sufficient condition for the GLRT to be minimax optimal up to constants. Despite being tight in many nontrivial situations, our information-theoretic lower bound is not tight for all cones; proving such a sharp lower bound is an interesting topic for future research. Second, as with a long line of past work on this topic [30, 31, 37, 50], our analysis is based on assuming that the noise variance σ^2 is known. In practice, this may or may not be a realistic assumption, and so it is interesting to consider the extension of our results to this setting.

We note that our minimax lower bounds are proved by constructing prior distributions on $\mathcal{H}_0$ and $\mathcal{H}_1$ and then control the distance between marginal likelihood functions. Following this idea, we can also consider our testing problem in the Bayesian framework. Without any prior preference on which hypothesis to take, we will let $\Pr(\mathcal{H}_0) = \Pr(\mathcal{H}_1) = 1/2$; thus the Bayesian testing procedure makes decision based on quantity

$$(59) \qquad B_{01} := \frac{m(y \mid \mathcal{H}_0)}{m(y \mid \mathcal{H}_1)} = \frac{\int_{\theta \in C_1} \mathbb{P}_\theta(y) \pi_1(\theta) \, d\theta}{\int_{\theta \in C_2} \mathbb{P}_\theta(y) \pi_2(\theta) \, d\theta},$$

which is often called Bayesian factor in literature. Analyzing the behavior of this statistic is an interesting direction to pursue in the future.

SUPPLEMENTARY MATERIAL

Supplement to "The geometry of hypothesis testing over convex cones: Generalized likelihood ratio tests and minimax radii" (DOI: 10.1214/18-AOS1701SUPP; .pdf). The supplementary material includes the explanation for the GLRT suboptimality and the proofs of more technical aspects.

REFERENCES

[1] AMELUNXEN, D., LOTZ, M., MCCOY, M. B. and TROPP, J. A. (2014). Living on the edge: Phase transitions in convex programs with random data. *Inf. Inference* **3** 224–294. MR3311453

[2] BARAUD, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli* **8** 577–606. MR1935648

[3] BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference Under Order Restrictions. The Theory and Application of Isotonic Regression*. Wiley, New York. MR0326887

[4] BARTHOLOMEW, D. J. (1959). A test of homogeneity for ordered alternatives. *Biometrika* **46** 36–48. MR0104312

[5] BARTHOLOMEW, D. J. (1959). A test of homogeneity for ordered alternatives. II. *Biometrika* **46** 328–335. MR0112204

[6] BARTLETT, P. L., BOUSQUET, O. and MENDELSON, S. (2005). Local Rademacher complexities. *Ann. Statist.* **33** 1497–1537. MR2166554

[7] BATU, T., KUMAR, R. and RUBINFELD, R. (2004). Sublinear algorithms for testing monotone and unimodal distributions. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing* 381–390. ACM, New York. MR2121623

[8] BESSON, O. (2006). Adaptive detection of a signal whose signature belongs to a cone. In *Proceedings SAM Conference*.

[9] BÜHLMANN, P. (2013). Statistical significance in high-dimensional linear models. *Bernoulli* **19** 1212–1242. MR3102549

[10] CHATTERJEE, S. (2014). A new perspective on least squares under convex constraint. *Ann. Statist.* **42** 2340–2381. MR3269982

[11] CHEN, D. and PLEMMONS, R. J. (2010). Nonnegativity constraints in numerical analysis. In *The Birth of Numerical Analysis* 109–139. World Sci. Publ., Hackensack, NJ. MR2604144

[12] CHETVERIKOV, D. (2012). Testing regression monotonicity in econometric models Technical report, UCLA. Available at arXiv:1212.6757.

[13] DYKSTRA, R. L. and ROBERTSON, T. (1983). On testing monotone tendencies. *J. Amer. Statist. Assoc.* **78** 342–350. MR0711109

[14] EDWARDS, A. W. F. (1984). *Likelihood*. Cambridge Univ. Press, Cambridge. Reprint of the 1972 original. MR0788494

[15] ERMAKOV, M. S. (1991). Minimax detection of a signal in Gaussian white noise. *Theory Probab. Appl.* **35** 667–679. MR1090496

[16] FAN, J. and JIANG, J. (2007). Nonparametric inference with generalized likelihood ratio tests. *TEST* **16** 409–444. MR2365172

[17] FAN, J., ZHANG, C. and ZHANG, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.* **29** 153–193. MR1833962

[18] GRECO, M., GINI, F. and FARINA, A. (2008). Radar detection and classification of jamming signals belonging to a cone class. *IEEE Trans. Signal Process.* **56** 1984–1993. MR2518695

[19] HU, X. and WRIGHT, F. T. (1994). Likelihood ratio tests for a class of nonoblique hypotheses. *Ann. Inst. Statist. Math.* **46** 137–145. MR1272754

[20] INGSTER, YU. I. (1987). A minimax test of nonparametric hypotheses on the density of a distribution in L_p metrics. *Theory Probab. Appl.* **31** 333–337. MR0851000

[21] INGSTER, YU. I. and SUSLINA, I. A. (2003). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models. Lecture Notes in Statistics* **169**. Springer, New York. MR1991446

[22] KOLTCHINSKII, V. (2001). Rademacher penalties and structural risk minimization. *IEEE Trans. Inform. Theory* **47** 1902–1914. MR1842526

[23] KUDÔ, A. (1963). A multivariate analogue of the one-sided test. *Biometrika* **50** 403–418. MR0163386

[24] LEHMANN, E. L. (2012). On likelihood ratio tests. In *Selected Works of E. L. Lehmann* (J. Rojo, ed.) 209–216. Springer, New York. MR2920510

[25] LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. Springer, New York. MR2135927

[26] LEPSKI, O. V. and SPOKOINY, V. G. (1999). Minimax nonparametric hypothesis testing: The case of an inhomogeneous alternative. *Bernoulli* **5** 333–358. MR1681702

[27] LEPSKI, O. V. and TSYBAKOV, A. B. (2000). Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probab. Theory Related Fields* **117** 17–48. MR1759508

[28] MCCOY, M. B. and TROPP, J. A. (2014). From Steiner formulas for cones to concentration of intrinsic volumes. *Discrete Comput. Geom.* **51** 926–963. MR3216671

[29] MEINSHAUSEN, N. (2013). Sign-constrained least squares estimation for high-dimensional regression. *Electron. J. Stat.* **7** 1607–1631. MR3066380

[30] MENÉNDEZ, J. A., RUEDA, C. and SALVADOR, B. (1992). Dominance of likelihood ratio tests under cone constraints. *Ann. Statist.* **20** 2087–2099. MR1193327

[31] MENÉNDEZ, J. A., RUEDA, C. and SALVADOR, B. (1992). Testing nonoblique hypotheses. *Comm. Statist. Theory Methods* **21** 471–484. MR1158571

[32] MENÉNDEZ, J. A. and SALVADOR, B. (1991). Anomalies of the likelihood ratio tests for testing restricted hypotheses. *Ann. Statist.* **19** 889–898. MR1105850

[33] MEYER, M. C. (2003). A test for linear versus convex regression function using shape-restricted regression. *Biometrika* **90** 223–232. MR1966562

[34] MOREAU, J.-J. (1962). Décomposition orthogonale d'un espace hilbertien selon deux cônes mutuellement polaires. *C. R. Acad. Sci. Paris* **255** 238–240. MR0139919

[35] PERLMAN, M. D. and WU, L. (1999). The emperor's new tests. *Statist. Sci.* **14** 355–381. MR1765215

[36] PISIER, G. (1986). Probabilistic methods in the geometry of Banach spaces. In *Probability and Analysis* (*Varenna*, 1985). *Lecture Notes in Math.* **1206** 167–241. Springer, Berlin. MR0864714

[37] RAUBERTAS, R. F., LEE, C.-I. C. and NORDHEIM, E. V. (1986). Hypothesis tests for normal means constrained by linear inequalities. *Comm. Statist. Theory Methods* **15** 2809–2833. MR0855765

[38] ROBERTSON, T. (1978). Testing for and against an order restriction on multinomial parameters. *J. Amer. Statist. Assoc.* **73** 197–202. MR0488300

[39] ROBERTSON, T. (1986). On testing symmetry and unimodality. In *Advances in Order Restricted Statistical Inference* (*Iowa City, Iowa*, 1985). *Lecture Notes in Statistics* **37** 231–248. Springer, Berlin. MR0875656

[40] ROBERTSON, T. and WEGMAN, E. J. (1978). Likelihood ratio tests for order restrictions in exponential families. *Ann. Statist.* **6** 485–505. MR0471147

[41] ROBERTSON, T. and WRIGHT, F. T. (1982). On measuring the conformity of a parameter set to a trend, with applications. *Ann. Statist.* **10** 1234–1245. MR0673658

[42] ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. Wiley, Chichester. MR0961262

[43] SCHARF, L. L. (1991). *Statistical Signal Processing*: *Detection*, *Estimation and Time Series Analysis*. Addison-Wesley, Reading, MA.

[44] SEN, B. and MEYER, M. (2017). Testing against a linear regression model using ideas from shape-restricted estimation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 423–448. MR3611753

[45] SHAPIRO, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *Int. Stat. Rev.* **56** 49–62. MR0963140

[46] SLAWSKI, M. and HEIN, M. (2013). Non-negative least squares for high-dimensional linear models: Consistency and sparse recovery without regularization. *Electron. J. Stat.* **7** 3004–3056. MR3151760

[47] SPOKOINY, V. G. (1998). Adaptive and spatially adaptive testing of a nonparametric hypothesis. *Math. Methods Statist.* **7** 245–273. MR1651777

[48] VAN DE GEER, S. A. (2000). *Applications of Empirical Process Theory*. *Cambridge Series in Statistical and Probabilistic Mathematics* **6**. Cambridge Univ. Press, Cambridge. MR1739079

[49] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*: *With Applications to Statistics*. Springer, New York. MR1385671

[50] WARRACK, G. and ROBERTSON, T. (1984). A likelihood ratio test regarding two nested but oblique order-restricted hypotheses. *J. Amer. Statist. Assoc.* **79** 881–886. MR0770282

[51] WEI, Y. and WAINWRIGHT, M. J. (2016). Sharp minimax bounds for testing discrete monotone distributions. In *IEEE International Symposium on Information Theory* (*ISIT*) 2684–2688. IEEE, Los Alamitos, CA.

[52] WEI, Y., WAINWRIGHT, M. J and GUNTUBOYINA, A. (2019). Supplement to "The geometry of hypothesis testing over convex cones: Generalized likelihood ratio tests and minimax radii." DOI:10.1214/18-AOS1701SUPP.

[53] ZARANTONELLO, E. H. (1971). Projections on convex sets in Hilbert space and spectral theory. In *Contributions to Nonlinear Functional Analysis* (*Proc. Sympos.*, *Math. Res. Center*, *Univ. Wisconsin*, *Madison*, *Wis.*, 1971) 237–424. Academic Press, New York. MR0388177

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, BERKELEY
BERKELEY, CALIFORNIA 94720
USA
E-MAIL: ytwei@berkeley.edu
        wainwrig@berkeley.edu
        aditya@stat.berkeley.edu