Submitted to the Annals of Statistics
arXiv: arXiv:1703.06810

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

By Yuting Wei and Martin J. Wainwright * and Adityanand Guntuboyina †

University of California at Berkeley

This supplementary material is organized as follows. In Section A, we first explain the intuition behind the example in Section 3.2.3 [6] where the GLRT is shown to be sub-optimal, and construct a series of other cases where this sub-optimality is observed. We then provide the proofs of Propositions 1 and 2 in Sections B.1 and B.2, respectively. It follows by some background on distance metrics and their properties in Section C. The proofs of Theorem 1 (a) and (b) are completed in Section D and E respectively. The proofs of the lemmas for Theorem 2 are collected in Section F. Finally, the technical lemmas which were crucially used in the proofs of the Proposition 2 and the monotone cone example are proved in Section G.

APPENDIX A: THE GLRT SUB-OPTIMALITY

In this appendix, we first try to understand why the GLRT is sub-optimal for the Cartesian product cone $K_\times = \mathrm{Circ}_{d-1}(\alpha) \times \mathbb{R}$, and use this intuition to construct a more general class of problems for which a similar sub-optimality is witnessed.

A.1. Why is the GLRT sub-optimal?. Let us consider tests with null $C_1 = \{0\}$ and a general product alternative of the form $C_2 = K_\times = K \times \mathbb{R}$, where $K \subseteq \mathbb{R}^{d-1}$ is a base cone. Note that $K = \mathrm{Circ}_{d-1}$ in our previous example.

*Supported in part by Office of Naval Research MURI grant DOD-002888, Air Force Office of Scientific Research Grant AFOSR-FA9550-14-1-001, and National Science Foundation Grants CIF-31712-23800

†Supported by NSF Grant DMS-1309356, NSF CAREER Grant DMS-16-54589

AMS 2000 subject classifications: Primary62F03; secondary 52A05

Keywords and phrases: hypothesis testing, closed convex cone, likelihood ratio test, minimax rate, Gaussian complexity

Now recall the decomposition (22) of the statistic T that underlies the GLRT. By the product nature of the cone, we have

$$T(y) = \|\Pi_{K_\times} y\|_2 = \|(\Pi_K(y_{-d}),\ y_d)\|_2 = \sqrt{\|\Pi_K(y_{-d})\|_2^2 + \|y_d\|_2^2},$$

where $y_{-d} := (y_1, \ldots, y_{d-1}) \in \mathbb{R}^{d-1}$ is formed from the first $d-1$ coordinates of y. Suppose that we are interested in testing between the zero vector and a vector $\theta^* = (0, \ldots, 0, \theta_d^*)$, nonzero only in the last coordinate, which belongs to the alternative. With this particular choice, under the null distribution, we have $y = \sigma g$ whereas under the alternative, we have $y = \theta^* + \sigma g$. Letting $\mathbb{E}_0$ and $\mathbb{E}_1$ denote expectations under these two Gaussian distributions, the performance of the GLRT in this direction is governed by the difference

$$\frac{1}{\sigma}\big\{\mathbb{E}_1[T(y)] - \mathbb{E}_0[T(y)]\big\} = \mathbb{E}_1\sqrt{\|\Pi_K(g_{-d})\|_2^2 + \|\frac{\theta_d^*}{\sigma} + g_d\|_2^2} \\ -\mathbb{E}_0\sqrt{\|\Pi_K(g_{-d})\|_2^2 + \|g_d\|_2^2}.$$

Note both terms in this difference involve a $(d-1)$-dimensional "pure noise" component—namely, the quantity $\|\Pi_K(g_{-d})\|_2^2$ defined by the sub-vector $g_{-d} := (g_1, \ldots, g_{d-1})$—with the only signal lying the last coordinate. For many choices of cone K, the pure noise component acts as a strong mask for the signal component, so that the GLRT is poor at detecting differences in the direction θ^*. Since the vector θ^* belongs to the alternative, this leads to sub-optimality in its overall behavior. Guided by this idea, we can construct a series of other cases where the GLRT is sub-optimal. See Appendix A.2 for details.

A.2. More examples on the GLRT sub-optimality. Now let us construct a larger class of product cones for which the GLRT is sub-optimal. For a given subset $S \subseteq \{1, \ldots, d\}$, define the subvectors $\theta_S = (\theta_i, i \in S)$ and $\theta_{S^c} = (\theta_j, j \in S^c\}$, where S^c denotes the complement of S. For an integer $\ell \geq 1$, consider any cone $K_\ell \subset \mathbb{R}^d$ with the following two properties:

- its Gaussian width scales as $\mathbb{E}\mathbb{W}(K_\ell \cap \mathbb{B}(1)) \asymp \sqrt{d}$, and
- for some fixed subset $\{1, 2 \ldots, d\}$ of cardinality ℓ, there is a scalar $\gamma > 0$ such that

$$\|\theta_S\|_2 \geq \gamma \|\theta_{S^c}\|_2 \quad \text{for all } \theta \in K_\ell.$$

As one concrete example, it is easy to check that the circular cone is a special example with $\ell = 1$ and $\gamma = 1/\tan(\alpha)$. The following result applies to the GLRT when applied to testing the null $C_1 = \{0\}$ versus the alternative $C_2 = K_\times^s = K \times \mathbb{R}$.

Proposition A.1. *For the previously described cone testing problem, the GLRT testing radius is sandwiched as*

$$\epsilon^2_{GLR} \asymp \sqrt{d}\sigma^2,$$

whereas a truncation test can succeed at radius $\epsilon^2 \asymp \sqrt{\ell}\sigma^2$.

Proof. The claimed scaling of the GLRT testing radius follows as a corollary of Theorem 1 after a direct evaluation of $\delta^2_{\mathrm{LR}}(C_1, C_2)$. In order to do so, we begin by observing that

$$\inf_{\eta \in C_2 \times S^{-1}} \langle \eta,\, \mathbb{E}\Pi_{C_2} g\rangle \leq \langle e_d,\, \mathbb{E}\Pi_{c_2} g\rangle \;=\; 0, \quad \text{and}$$
$$\mathbb{E}\mathbb{W}(C_2 \cap \mathbb{B}(1)) = \mathbb{E}\|\Pi_{C_2} g\|_2 \asymp \sqrt{d}$$

which implies that $\delta^2_{\mathrm{LR}}(C_1, C_2) \asymp \sqrt{d}$, and hence implies the sandwich claim on the GLRT via Theorem 1.

On the other hand, for some pre-selected $\beta > 0$, consider the truncation test

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

This test can be viewed as a GLRT for testing the zero null against the alternative $\mathbb{R}^\ell$, and hence it will succeed with separation $\epsilon^2 \asymp \sigma^2\sqrt{\ell}$. Putting these pieces together, we conclude that the GLRT is sub-optimal whenever ℓ is of lower order than d.

□

APPENDIX B: PROOFS FOR PROPOSITION 1 AND 2

In this section, we complete the proofs of Propositions 1 and 2 in Sections B.1 and B.2, respectively.

B.1. Proof of Proposition 1. As in the proof of Theorem 1 and Theorem 2, we can assume without loss of generality that $\sigma = 1$ since K_+ is invariant under rescaling by positive numbers. We split our proof into two cases, depending on whether or not the dimension d is less than 81.

Case 1:. First suppose that $d < 81$. If the separation is upper bounded as $\epsilon^2 \leq \kappa_\rho\sqrt{d}$, then setting $\kappa_\rho = 1/18$ yields

$$\epsilon^2 \leq \kappa_\rho\sqrt{d} < 1/2.$$

Similar to our proof for Theorem 1(b) Case 1, if $\epsilon^2 < 1/2$, every test yields testing error no smaller than 1/2. It is seen by considering a simple verses simple testing problem (84a). So our lower bound directly holds for the case when $d < 81$ satisfies.

Case 2:. Let us consider the case when dimension $d \geq 81$. The idea is to make use of our Lemma 2 to show that the testing error is at least ρ whenever $\epsilon^2 \leq \kappa_\rho \sqrt{d}$. In order to apply Lemma 2, the key is to construct a probability measure $\mathbb{Q}$ supported on set $K \cap B^c(1)$ such that for i.i.d. pair η, η' drawn from $\mathbb{Q}$, quantity $\mathbb{E} e^{\lambda \langle \eta, \eta' \rangle}$ can be well controlled. We claim that there exists such a probability measure $\mathbb{Q}$ that

(59)
$$\mathbb{E}_{\eta,\eta'} e^{\lambda \langle \eta, \eta' \rangle} \leq \exp\left(\exp\left(\frac{2+\lambda}{\sqrt{d}-1} \right) - \left(1 - \frac{1}{\sqrt{d}}\right)^2 \right) \qquad \text{where } \lambda := \epsilon^2.$$

Taking inequality (59) as given for now, letting $\kappa_\rho = 1/8$, we have $\lambda = \epsilon^2 \leq \sqrt{d}/8$. So the right hand side in expression (59) can be further upper bounded as

$$\begin{aligned} \exp\left(\exp\left(\frac{2}{\sqrt{d}-1} + \frac{\sqrt{d}}{\sqrt{d}-1} \frac{\lambda}{\sqrt{d}} \right) - \left(1 - \frac{1}{\sqrt{d}}\right)^2 \right) &\leq \exp\left(\exp\left(\frac{1}{4} + \frac{9}{8} \cdot \frac{1}{8} \right) - \left(1 - \frac{1}{9}\right)^2 \right) \\ &< 2, \end{aligned}$$

where we use the fact that $d \geq 81$. As a consequence of Lemma 2, the testing error of every test satisfies

$$\inf_{\psi} \mathcal{E}(\psi; \{0\}, K_+, \epsilon) \geq 1 - \frac{1}{2} \sqrt{\mathbb{E}_{\eta,\eta'} \exp(\epsilon^2 \langle \eta, \eta' \rangle) - 1} > \frac{1}{2} \geq \rho.$$

Putting these two cases together, our lower bound holds for any dimension thus we complete the proof of Proposition 1.

So it only remains to construct a probability measure $\mathbb{Q}$ such that the inequality (59) holds. We begin by introducing some helpful notation. For an integer s to be specified, consider a collection of vectors $\mathcal{S}$ containing all d-dimensional vectors with exactly s nonzero entries and each nonzero entry equals to $1/\sqrt{s}$. Note that there are in total $M := \binom{d}{s}$ vectors of this type. Letting $\mathbb{Q}$ be the uniform distribution over this set of vectors namely

$$\mathbb{Q}(\{\eta\}) := \frac{1}{M}, \qquad \eta \in \mathcal{S}. \tag{60}$$

Then we can write the expectation as

$$\mathbb{E} e^{\lambda \langle \eta, \eta' \rangle} = \frac{1}{M^2} \sum_{\eta, \eta' \in \mathcal{S}} e^{\lambda \langle \eta, \eta' \rangle}.$$

Note that the inner product $\langle \eta, \eta' \rangle$ takes values i/s, for integer $i \in \{0, 1, \ldots, s\}$ and given every vector η and integer $i \in \{0, 1, \ldots, s\}$, the number of η' such that $\langle \eta, \eta' \rangle = i/s$ equals to $\binom{s}{i}\binom{d-s}{s-i}$. Consequently, we obtain

$$\mathbb{E}e^{\lambda\langle \eta, \eta' \rangle} = \binom{d}{s}^{-1} \sum_{i=0}^{s} \binom{s}{i}\binom{d-s}{s-i} e^{\lambda i/s} = \sum_{i=0}^{s} \frac{A_i z^i}{i!}, \tag{61}$$

where

$$z := e^{\lambda/s} \text{ and } A_i := \frac{(s!(d-s)!)^2}{((s-i)!)^2 d!(d-2s+i)!}.$$

Let us set integer $s := \lfloor \sqrt{d} \rfloor$. We claim quantity A_i satisfies the following bound

$$A_i \leq \exp\Big(- (1 - \frac{1}{\sqrt{d}})^2 + \frac{2i}{\sqrt{d} - 1}\Big) \qquad \text{for all } i \in \{0, 1, \ldots, s\}. \tag{62}$$

Taking expression (62) as given for now and plugging into inequality (61), we have

$$\begin{aligned}
\mathbb{E}e^{\lambda\langle \eta, \eta' \rangle} &\leq \exp\Big(- (1 - \frac{1}{\sqrt{d}})^2\Big) \sum_{i=0}^{s} \frac{(z \exp(\frac{2}{\sqrt{d}-1}))^i}{i!} \\
&\overset{(a)}{\leq} \exp\Big(- (1 - \frac{1}{\sqrt{d}})^2\Big) \exp\Big(z \exp(\frac{2}{\sqrt{d} - 1})\Big) \\
&\overset{(b)}{\leq} \exp\left(- \left(1 - \frac{1}{\sqrt{d}}\right)^2 + \exp\left(\frac{2 + \lambda}{\sqrt{d} - 1}\right)\right),
\end{aligned}$$

where step (a) follows from the standard power series expansion $e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$ and step (b) follows by $z = e^{\lambda/s}$ and $s = \lfloor \sqrt{d} \rfloor > \sqrt{d} - 1$. Therefore it verifies inequality (59) and complete our argument.

It is only left for us to check inequality (62) for A_i. Using the fact that $1 - x \leq e^{-x}$, it is guaranteed that

$$A_0 = \frac{((d-s)!)^2}{d!(d-2s)!} = (1 - \frac{s}{d})(1 - \frac{s}{d-1}) \cdots (1 - \frac{s}{d-s+1}) \leq \exp(-s \sum_{i=1}^{s} \frac{1}{d-s+i}). \tag{63a}$$

Recall that integer $s = \lfloor \sqrt{d} \rfloor$, then we can bound the sum in expression (63a) as

$$s \sum_{i=1}^{s} \frac{1}{d-s+i} \geq s \sum_{i=1}^{s} \frac{1}{d} = \frac{s^2}{d} \geq (1 - \frac{1}{\sqrt{d}})^2,$$

which, when combined with inequality (63a), implies that $A_0 \leq \exp(-(1 - \frac{1}{\sqrt{d}})^2)$.

Moreover, direct calculations yield

$$\frac{A_i}{A_{i-1}} = \frac{(s-i+1)^2}{d-2s+i}, \qquad 1 \leq i \leq s. \tag{63b}$$

This ratio is decreasing with index i as $1 \leq i \leq s$, thus is upper bounded by A_1/A_0, which implies that

$$\frac{A_i}{A_{i-1}} \leq \frac{d}{d - 2\sqrt{d} + 1} = (1 + \frac{1}{\sqrt{d}-1})^2 \leq \exp(\frac{2}{\sqrt{d}-1}),$$

where the last inequality follows from $1 + x \leq e^x$. Putting pieces together validates bound (62) thus finishing the proof of Proposition 1.

B.2. Proof of Proposition 2. As in the proof of Theorem 1 and Theorem 2, we can assume without loss of generality that $\sigma = 1$ since L and M are both invariant under rescaling by positive numbers.

We split our proof into two cases, depending on whether or not $\sqrt{\log(ed)} < 14$.

Case 1:. First suppose $\sqrt{\log(ed)} < 14$, so that the choice $\kappa_\rho = 1/28$ yields the upper bound

$$\epsilon^2 \leq \kappa_\rho \sqrt{\log(ed)} < 1/2.$$

Similar to our proof of the lower bound in Theorem 1, by reducing to a simple testing problem (84a), any test yields testing error no smaller than $1/2$ if $\epsilon^2 < 1/2$. Thus, we conclude that the stated lower bound holds when $\sqrt{\log(ed)} < 14$.

Case 2:. Otherwise, we may assume that $\sqrt{\log(ed)} \geq 14$. In this case, we exploit Lemma 2 in order to show that the testing error is at least ρ whenever $\epsilon^2 \leq \kappa_\rho \sqrt{\log(ed)}$. Doing so requires constructing a probability measure $\mathbb{Q}_L$ supported on $M \cap L^\perp \cap B^c(1)$ such that the expectation $\mathbb{E} e^{\epsilon^2 \langle \eta, \eta' \rangle}$ can be well controlled, where (η, η') are drawn i.i.d according to $\mathbb{Q}_L$. Note that L can be either $\{0\}$ or $\mathrm{span}(\mathbf{1})$.

Before doing that, let us first introduce some notation. Let $\delta := 9$ and $r := 1/3$ (note that $\delta = r^{-2}$). Let

$$m := \max \left\{ n \;\Big|\; \sum_{i=1}^{n} \lfloor \frac{\delta - 1}{\delta^i} (d + \log_\delta d + 3) \rfloor \; < \; d \right\}. \tag{64}$$

We claim that the integer m defined above satisfies:

$$\lceil \frac{3}{4}\log_\delta(d)\rceil + 1 \le m \le \lceil \log_\delta d\rceil, \tag{65}$$

where $\lceil x\rceil$ denotes the smallest integer that is greater than or equal to x. To see this, notice that for $t = \lceil \frac{3}{4}\log_\delta(d)\rceil + 1$, we have

$$\begin{aligned}\sum_{i=1}^{t}\lfloor \frac{\delta-1}{\delta^i}(d+\log_\delta d+3)\rfloor \le \sum_{i=1}^{t}\frac{\delta-1}{\delta^i}(d+\log_\delta d+3) &= (1-\frac{1}{\delta^t})(d+\alpha)\\ &\overset{(i)}{\le} d+\alpha-\frac{d+\alpha}{\delta^2 d^{3/4}} \overset{(ii)}{<} d,\end{aligned}$$

where we denote $\alpha := \log_\delta d + 3$. The step (i) follows by definition that $t = \lceil \frac{3}{4}\log_\delta(d)\rceil + 1$ while step (ii) holds because as $\sqrt{\log(ed)} \ge 14$, we have $\alpha = \log_\delta d + 3 < d^{1/4}/\delta^2$. On the other hand, for $t = \lceil \log_\delta d\rceil$, we have

$$\begin{aligned}\sum_{i=1}^{t}\lfloor \frac{\delta-1}{\delta^i}(d+\log_\delta d+3)\rfloor &\ge \sum_{i=1}^{t}\frac{\delta-1}{\delta^i}(d+\alpha)-t\\ &= (1-\frac{1}{\delta^t})(d+\alpha)-t\\ &> d+\alpha-\frac{d+\alpha}{d}-(\log_\delta d+1),\end{aligned}$$

where the last step uses fact $t = \lceil \log_\delta d\rceil$. Since when $\sqrt{\log(ed)} \ge 14$, we have $\alpha = \log_\delta d + 3 < d$, therefore $(d+\alpha)/d + \log_\delta d + 1 \le 2 + \log_\delta d + 1 = \alpha$, which guarantees that

$$\sum_{i=1}^{t}\lfloor \frac{\delta-1}{\delta^i}(d+\log_\delta d+3)\rfloor > d.$$

We thereby established inequality (65).

We now claim that there exists a probability measure $\mathbb{Q}_L$ supported on $M \cap L^\perp \cap B^c(1)$ such that

$$\mathbb{E}_{\eta,\eta'\sim\mathbb{Q}_L} e^{\lambda\langle \eta,\eta'\rangle} \le \exp\left(\exp\left(\frac{9\lambda/4+2}{\sqrt{m}-1}\right) - \left(1-\frac{1}{\sqrt{m}}\right)^2 + \frac{27\lambda}{32(\sqrt{m}-1)}\right), \quad \text{where } \lambda := \epsilon^2. \tag{66}$$

Recall that we showed in inequality (65) that $m \ge \lceil \frac{3}{4}\log_\delta(d)\rceil + 1$. Setting $\kappa_\rho = 1/62$ implies that whenever $\epsilon^2 \le \kappa_\rho\sqrt{\log(ed)}$, we have

$$\epsilon^2 \le \frac{1}{62}\sqrt{\log(ed)} = \frac{1}{62}\sqrt{1+\frac{4}{3}\log\delta\cdot\frac{3}{4}\log_\delta d} \le \frac{1}{62}\sqrt{\frac{4}{3}\log\delta\left(1+\frac{3}{4}\log_\delta d\right)} \le \frac{1}{36}\sqrt{m}. \tag{67}$$

So the right hand side in expression (66) can be made less than 2 by

$$\begin{aligned}
&\exp\left(\frac{9\lambda/4+2}{\sqrt{m}-1}\right) - \left(1-\frac{1}{\sqrt{m}}\right)^2 + \frac{27\lambda}{32(\sqrt{m}-1)} \\
\leq &\exp\left(\frac{9\lambda}{4\sqrt{m}}\frac{\sqrt{m}}{\sqrt{m}-1} + \frac{2}{7}\right) - \left(1-\frac{1}{8}\right)^2 + \frac{27\lambda}{32\sqrt{m}}\frac{\sqrt{m}}{\sqrt{m}-1} \\
\leq &\exp\left(\frac{9}{4\cdot 36}\frac{8}{7} + \frac{2}{7}\right) - \left(1-\frac{1}{8}\right)^2 + \frac{27}{32\cdot 36}\frac{8}{7} < \log 2,
\end{aligned}$$

where we use the fact that $\sqrt{m} \geq \sqrt{1+\frac{3}{4}\log_\delta d} \geq 8$. Lemma 2 thus guarantees the testing error to be no less than

$$\inf_\psi \mathcal{E}(\psi; L, M, \epsilon) \geq 1 - \frac{1}{2}\sqrt{\mathbb{E}_{\eta,\eta'}\, \exp(\epsilon^2\langle\eta,\, \eta'\rangle) - 1} > \frac{1}{2} \geq \rho,$$

which leads to our result in Proposition 2.

Now it only remains to construct a probability measure $\mathbb{Q}_L$ with the right support such that inequality (66) holds. To do this, we make use of a fact from the proof of Proposition 1 for the orthant cone $K_+ \subset \mathbb{R}^m$. Recall that to establish Proposition 1, we constructed a probability measure $\mathbb{D}$ supported on $K_+ \cap S^{m-1} \subset \mathbb{R}^m$ such that if b, b' are an i.i.d pair drawn from $\mathbb{D}$, we have

$$\mathbb{E}_{b,b'\sim\mathbb{D}} e^{\lambda\langle b,\, b'\rangle} \leq \exp\left(\exp\left(\frac{2+\lambda}{\sqrt{m}-1}\right) - \left(1-\frac{1}{\sqrt{m}}\right)^2\right). \tag{68}$$

By construction, $\mathbb{D}$ is a uniform probability measure on the finite set $\mathcal{S}$ which consists of all vectors in $\mathbb{R}^m$ which have s nonzero entries which are all equal to $1/\sqrt{s}$ where $s = \lfloor\sqrt{m}\rfloor$.

Based on this measure $\mathbb{D}$, let us define $\mathbb{Q}_L$ as in the following lemma and establish some of its properties under the assumption that $\sqrt{\log(ed)} \geq 14$.

Lemma B.1. *Let G be the $m \times m$ lower triangular matrix given by*

$$G := \begin{pmatrix} 1 & & & & \\ r & 1 & & & \\ r^2 & r & 1 & & \\ \vdots & \vdots & & \ddots & \\ r^{m-1} & r^{m-2} & \dots & & 1 \end{pmatrix}. \tag{69a}$$

There exists an $d \times m$ matrix F such that

$$F^T F = \mathbb{I}_m \tag{69b}$$

and such that for every $b \in \mathcal{S}$ and $\eta := FGb$, we have

1. *$\eta \in M \cap L^{\perp} \cap B^c(1)$ if $L = \{0\}$, and*
2. *$\eta - \bar{\eta}\mathbf{1} \in M \cap L^{\perp} \cap B^c(1)$ if $L = \text{span}(\mathbf{1})$, where $\bar{\eta} = \sum_{i=1}^{d} \eta_i / d$ denotes the mean of the vector η.*

See Appendix G.2 for the proof of this claim.

If $L = \{0\}$, let probability measure $\mathbb{Q}_L$ be defined as the distribution of $\eta := FGb$ where $b \sim \mathbb{D}$. Otherwise if $L = \text{span}(\mathbf{1})$, let $\mathbb{Q}_L$ be the distribution of $\eta - \bar{\eta}\mathbf{1}$ where again $\eta := FGb$ and $b \sim \mathbb{D}$. From Lemma B.1 we know that $\mathbb{Q}_L$ is supported on $M \cap L^{\perp} \cap B^c(1)$. It only remains to verify the critical inequality (66) to complete the proof of Proposition 2. Let $\eta = FGb$ and $\eta' = FGb'$ with b, b' being i.i.d having distribution $\mathbb{D}$. Using the fact that $F^T F = \mathbb{I}_m$, we can write the inner product of η, η' as

$$\langle \eta,\, \eta' \rangle = b^T G^T F^T F G b' = \langle Gb,\, Gb' \rangle.$$

The following lemma relates inner product $\langle \eta,\, \eta' \rangle$ to $\langle b,\, b' \rangle$, and thereby allows us to derive inequality (66) based on inequality (68). Recall that $\mathcal{S}$ consists of all vectors in $\mathbb{R}^m$ which have s nonzero entries which are all equal to $1/\sqrt{s}$ where $s = \lfloor \sqrt{m} \rfloor$.

LEMMA B.2. *For every $b, b' \in \mathcal{S}$, we have*

$$\langle Gb,\, Gb' \rangle \;\leq\; \frac{\langle b,\, b' \rangle}{(1-r)^2} + \frac{r}{s(1-r)^2(1-r^2)}, \tag{70a}$$

$$\|Gb\|_2^2 \;\geq\; \frac{1}{(1-r)^2} - \frac{2r + r^2}{s(1-r^2)(1-r)^2}. \tag{70b}$$

See Appendix G.3 for the proof of this claim.

We are now ready to prove inequality (66). We consider the two cases $L = \{0\}$ and $L = \text{span}(\mathbf{1})$ separately.

For $L = \{0\}$, recall that $r = 1/3$ and $s = \lfloor \sqrt{m} \rfloor \geq \sqrt{m} - 1$. Therefore as a direct consequence of inequality (70a), we have

$$\mathbb{E}_{\eta,\eta \sim \mathbb{Q}} e^{\lambda \langle \eta,\, \eta' \rangle} \leq \mathbb{E}_{b,b' \sim \mathbb{D}} \exp\left(\frac{9\lambda}{4} \langle b,\, b' \rangle + \frac{27\lambda}{32(\sqrt{m}-1)} \right). \tag{71}$$

Combining inequality (71) with (68) completes the proof of inequality (66).

Let us now turn to the case when $L = \text{span}(\mathbf{1})$. The proof is essentially the same as for $L = \{0\}$ with only some minor changes. Again our goal is to check inequality (66). For this, we write

$$\mathbb{E}_{\eta,\eta' \sim \mathbb{Q}_L} e^{\lambda \langle \eta,\, \eta' \rangle} = \mathbb{E}_{\eta,\eta' \sim \mathbb{Q}_{\{0\}}} e^{\lambda \langle \eta - \bar{\eta}\mathbf{1},\, \eta' - \bar{\eta}'\mathbf{1} \rangle} \;\leq\; \mathbb{E}_{\eta,\eta' \sim \mathbb{Q}_{\{0\}}} e^{\lambda \langle \eta,\, \eta' \rangle}.$$

Here the last step use the fact that $\langle \eta - \bar{\eta}\mathbf{1},\, \eta' - \bar{\eta'}\mathbf{1}\rangle = \langle \eta,\, \eta'\rangle - d\bar{\eta}\bar{\eta'} \leq \langle \eta,\, \eta'\rangle$ where the last inequality follows from the nonnegativity of every entry of vectors η and η' (this nonnegativity is a consequence of the nonnegativity of F and G from Lemma B.1 and nonnegativity of vectors in $\mathcal{S}$).

Thus, we have completed the proof of Proposition 2.

APPENDIX C: DISTANCES AND THEIR PROPERTIES

Here we collect some background on distances between probability measures that are useful in analyzing testing error. Suppose $\mathbb{P}_1$ and $\mathbb{P}_2$ are two probability measures on Euclidean space $(\mathbb{R}^d, \mathcal{B})$ equipped with Lebesgue measure. For the purpose of this paper, we assume $\mathbb{P}_1 \ll \mathbb{P}_2$. The *total variation* (TV) distance between $\mathbb{P}_1$ and $\mathbb{P}_2$ is defined as

$$\|\mathbb{P}_1 - \mathbb{P}_2\|_{\mathrm{TV}} := \sup_{B \in \mathcal{B}} |\mathbb{P}_1(B) - \mathbb{P}_2(B)| = \frac{1}{2}\int |d\mathbb{P}_1 - d\mathbb{P}_2|. \tag{72a}$$

A closely related measure of distance is the χ^2 *distance* given by

$$\chi^2(\mathbb{P}_1, \mathbb{P}_2) := \int (\frac{d\mathbb{P}_1}{d\mathbb{P}_2} - 1)^2 d\mathbb{P}_2. \tag{72b}$$

For future reference, we note that the TV distance and χ^2 distance are related via the inequality

$$\|\mathbb{P}_1 - \mathbb{P}_2\|_{\mathrm{TV}} \leq \frac{1}{2}\sqrt{\chi^2(\mathbb{P}_1, \mathbb{P}_2)}. \tag{72c}$$

APPENDIX D: AUXILIARY PROOFS FOR THEOREM 1 (A)

In this appendix, we collect the proofs of lemmas involved in the proof of Theorem 1(a).

D.1. Proof of Lemma 4.1. For future reference, we also note that tail bound (51a) implies that the variance is bounded as

$$\mathrm{var}(Z(\theta)) = \int_0^\infty \mathbb{P}\Big(\big|Z(\theta) - \mathbb{E}[Z(\theta)]\big| \geq \sqrt{u}\Big) du \;\leq\; 2\int_0^\infty e^{-u/2} du \;=\; 4. \tag{73}$$

To prove Lemma 4.1, given every vector θ, we claim that the function $g \mapsto \|\Pi_K(\theta + g)\|_2$ is 1-Lipschitz, whereas the function $g \mapsto \langle \theta,\, \Pi_K g\rangle$ is a $\|\theta\|_2$-Lipschitz function. From these claims, the concentration results then follow from Borell's theorem [2].

In order to establish the Lipschitz property, consider two vectors $g, g' \in \mathbb{R}^d$. By the triangle inequaliuty non-expansiveness of Euclidean projection, we have

$$\Big| \|\Pi_K(\theta + g)\|_2 - \|\Pi_K(\theta + g')\|_2 \Big| \leq \|\Pi_K(\theta + g) - \Pi_K(\theta + g')\|_2 \;\leq\; \|g - g'\|_2.$$

Combined with the Cauchy-Schwarz inequality, we conclude that

$$\big|\langle \theta,\, \Pi_K g\rangle - \langle \theta,\, \Pi_K g'\rangle\big| \leq \|\theta\|_2 \, \|\Pi_K g - \Pi_K g'\|_2 \leq \|\theta\|_2 \, \|g - g'\|_2,$$

which completes the proof of Lemma 4.1.

D.2. Proof of inequality (53). To prove inequality (53), we make use of the following auxiliary Lemma D.1.

LEMMA D.1. *For every closed convex cone K and vector $\theta \in K$, we have the lower bounds*

$$\Gamma(\theta) \geq \frac{\|\theta\|_2^2}{2\|\theta\|_2 + 8\mathbb{E}\|\Pi_K g\|_2} - \frac{2}{\sqrt{e}}. \tag{74a}$$

Moreover, for any vector θ that also satisfies the inequality $\langle \theta,\, \mathbb{E}\Pi_K g\rangle \geq \|\theta\|_2^2$, we have

$$\Gamma(\theta) \geq \alpha^2(\theta) \frac{\langle \theta,\, \mathbb{E}\Pi_K g\rangle - \|\theta\|_2^2}{\alpha(\theta)\|\theta\|_2 + 2\mathbb{E}\|\Pi_K g\|_2} - \frac{2}{\sqrt{e}}, \tag{74b}$$

where $\alpha(\theta) := 1 - \exp\left(\frac{-\langle \theta,\, \mathbb{E}\Pi_K g\rangle^2}{8\|\theta\|_2^2}\right)$.

We now use Lemma D.1 to prove the lower bound (53). Note that the inequality $\|\theta\|_2^2 \geq B_\rho \delta^2_{\mathrm{LR}}(\{0\}, K)$ implies that one of the following two lower bounds must hold:

$$\|\theta\|_2^2 \geq B_\rho \mathbb{E}\|\Pi_K g\|_2, \tag{75a}$$

$$\text{or} \quad \langle \theta,\, \mathbb{E}\Pi_K g\rangle \geq \sqrt{B_\rho}\mathbb{E}\|\Pi_K g\|_2. \tag{75b}$$

We will analyze these two cases separately.

Case 1. In order to show that the lower bound (75a) implies inequality (53), we will prove a stronger result—namely, that the inequality $\|\theta\|_2^2 \geq \sqrt{B_\rho}\mathbb{E}\|\Pi_K g\|_2/2$ implies that inequality (53) holds.

From the lower bound (74a) and the fact that, for each fixed $a > 0$, the function $x \mapsto x^2/(2x+a)$ is increasing on the interval $[0, \infty)$, we find that

$$\Gamma(\theta) \geq \frac{\sqrt{B_\rho \mathbb{E}\|\Pi_K g\|_2}/2}{\sqrt{2}B_\rho^{1/4} + 8\sqrt{\mathbb{E}\|\Pi_K g\|_2}} - \frac{2}{\sqrt{e}}.$$

Further, because of general bound (21) that $\mathbb{E}\|\Pi_K g\|_2 \geq 1/\sqrt{2\pi}$ and the fact that the function $x \mapsto x/(a+x)$ is increasing in x, we obtain

$$\Gamma(\theta) \geq \frac{\sqrt{B_\rho}}{2(8\pi B_\rho)^{1/4} + 16} - \frac{2}{\sqrt{e}},$$

which ensures inequality (53).

Case 2. We now turn to the case when inequality (75b) is satisfied. We may assume the inequality $\|\theta\|_2^2 \geq \sqrt{B_\rho}\mathbb{E}\|\Pi_K g\|_2/2$ is violated because otherwise, inequality (53) follows immediately. When this inequality is violated, we have

$$\langle \theta, \, \mathbb{E}\Pi_K g\rangle \geq \sqrt{B_\rho}\mathbb{E}\|\Pi_K g\|_2 \quad \text{and} \quad \|\theta\|_2^2 < \sqrt{B_\rho}\mathbb{E}\|\Pi_K g\|_2/2. \tag{76}$$

Our strategy is to make use of inequality (74b), and we begin by bounding the quantity α appearing therein. By combining inequality (76) and inequality (21)—namely, $\mathbb{E}\|\Pi_K g\|_2 \geq 1/\sqrt{2\pi}$, we find that

$$\alpha \geq 1 - \exp\left(-\frac{\sqrt{B_\rho}\mathbb{E}\|\Pi_K g\|_2}{4}\right) \geq 1 - \exp\left(-\frac{\sqrt{B_\rho}}{4\sqrt{2\pi}}\right) \geq 1/2, \quad \text{whenever } B_\rho \geq 32\pi.$$

Using expression (76), we deduce that

$$\Gamma(\theta) \;\geq\; \frac{\alpha^2\sqrt{B_\rho \mathbb{E}\|\Pi_K g\|_2}}{\alpha(4B_\rho)^{1/4} + 4\sqrt{\mathbb{E}\|\Pi_K g\|_2}} - \sqrt{\frac{2}{e}} \;\geq\; \frac{\sqrt{B_\rho \mathbb{E}\|\Pi_K g\|_2}}{(2^6 B_\rho)^{1/4} + 16\sqrt{\mathbb{E}\|\Pi_K g\|_2}} - \sqrt{\frac{2}{e}}.$$

where the second inequality uses the previously obtained lower bound $\alpha > 1/2$, and the fact that the function $x \mapsto x^2/(x+b)$ is increasing in x.

This completes the proof of inequality (53).

Proof of Lemma D.1. Now it is only left for us to prove Lemma D.1. We define the random variable $Z(\theta) := \|\Pi_K(\theta+g)\|_2 - \|\Pi_K g\|_2$, as well as its positive and negative parts $Z^+(\theta) = \max\{0, Z(\theta)\}$ and $Z^-(\theta) = \max\{0, -Z(\theta)\}$, so that $\Gamma(\theta) = \mathbb{E}Z(\theta) = \mathbb{E}Z^+(\theta) - \mathbb{E}Z^-(\theta)$. Our strategy is to bound $\mathbb{E}Z^-(\theta)$ from above and then bound $\mathbb{E}Z^+(\theta)$ from below. The following auxiliary lemma is useful for these purposes:

LEMMA D.2. *For every closed convex cone $K \subset \mathbb{R}^d$ and vectors $x \in K$ and $y \in \mathbb{R}^d$, we have:*

(77)
$$\Big| \|\Pi_K(x+y)\|_2 - \|\Pi_K(y)\|_2 \Big| \leq \|x\|_2, \qquad \textit{and}$$

(78)
$$\max\left\{2\langle x,\, y\rangle + \|x\|_2^2,\, 2\langle x,\, \Pi_K y\rangle - \|x\|_2^2\right\} \overset{(i)}{\leq} \|\Pi_K(x+y)\|_2^2 - \|\Pi_K(y)\|_2^2 \overset{(ii)}{\leq} 2\langle x,\, \Pi_K y\rangle + \|x\|_2^2.$$

We return to prove this claim in Appendix D.3.

Inequality (77) implies that $Z(\theta) \geq -\|\theta\|_2$ and thus $\mathbb{E}Z^-(\theta) \leq \|\theta\|_2 \mathbb{P}\{Z(\theta) \leq 0\}$. The lower bound in inequality (78) then implies that $\mathbb{P}\{Z(\theta) \leq 0\} \leq \mathbb{P}\{\langle\theta,\, g\rangle \leq -\|\theta\|_2^2/2\} \leq \exp\big(-\frac{\|\theta\|_2^2}{8}\big)$, whence

$$\mathbb{E}Z^-(\theta) \leq \|\theta\|_2 \exp\left(\frac{-\|\theta\|_2^2}{8}\right) \leq \sup_{u>0}\left(ue^{-u^2/8}\right) = \frac{2}{\sqrt{e}}.$$

Putting together the pieces, we have established the lower bound

$$\mathbb{E}Z(\theta) \;=\; \mathbb{E}Z^+(\theta) - \mathbb{E}Z^-(\theta) \;\geq\; \mathbb{E}Z^+(\theta) - \frac{2}{\sqrt{e}}. \tag{79}$$

The next task is to lower bound the expectation $\mathbb{E}Z^+(\theta)$. By the triangle inequality, we have

$$\begin{aligned}\|\Pi_K(\theta+g)\|_2 &\leq \|\Pi_K(\theta+g) - \Pi_K(g)\|_2 + \|\Pi_K(g)\|_2 \\ &\leq \|\theta\|_2 + \|\Pi_K(g)\|_2,\end{aligned}$$

where the second inequality uses non-expansiveness of the projection. Consequently, we have the lower bound

(80)
$$\mathbb{E}Z^+(\theta) = \mathbb{E}\frac{\left(\|\Pi_K(\theta+g)\|_2^2 - \|\Pi_K g\|_2^2\right)^+}{\|\Pi_K(\theta+g)\|_2 + \|\Pi_K g\|_2} \geq \mathbb{E}\frac{\left(\|\Pi_K(\theta+g)\|_2^2 - \|\Pi_K g\|_2^2\right)^+}{\|\theta\|_2 + 2\|\Pi_K g\|_2}.$$

Note that inequality (78)(i) implies two lower bounds on the difference $\|\Pi_K(\theta+g)\|_2^2 - \|\Pi_K g\|_2^2$. We treat each of these lower bounds in turn, and show how they lead to inequalities (74a) and (74b).

Proof of inequality (74a)*:.* Inequality (80) and the first lower bound term from inequality (78)(i) imply that

$$\mathbb{E}Z^+(\theta) \geq \mathbb{E}\frac{\left(2\langle\theta,\, g\rangle + \|\theta\|_2^2\right)^+}{\|\theta\|_2 + 2\|\Pi_K g\|_2} \geq \mathbb{E}\frac{\|\theta\|_2^2}{\|\theta\|_2 + 2\|\Pi_K g\|_2}\mathbb{I}\{\langle\theta,\, g\rangle \geq 0\}.$$

Jensen's inequality (and the fact that $\mathbb{P}\{\langle\theta,\, g\rangle \geq 0\} = 1/2$) now allow us to deduce

$$\mathbb{E}Z^+(\theta) \geq \mathbb{P}\left\{\langle\theta,\, g\rangle \geq 0\right\}\|\theta\|_2^2\left(\|\theta\|_2 + \frac{2\mathbb{E}\|\Pi_K g\|_2}{P\left\{\langle\theta,\, g\rangle \geq 0\right\}}\right)^{-1} = \frac{\|\theta\|_2^2}{2\|\theta\|_2 + 8\mathbb{E}\|\Pi_K g\|_2}$$

and this gives inequality (74a).

Proof of inequality (74b)*:.* Putting inequality (80), the second term on the left hand side of inequality (78)(i), along with the fact that $\langle\theta,\, \mathbb{E}\Pi_K g\rangle \geq \|\theta\|_2^2$ together guarantees that

$$\mathbb{E}Z^+(\theta) \geq \mathbb{E}\frac{\left(2\langle\theta,\, \Pi_K g\rangle - \|\theta\|_2^2\right)^+}{\|\theta\|_2 + 2\|\Pi_K g\|_2} \geq \mathbb{E}\frac{\langle\theta,\, \mathbb{E}\Pi_K g\rangle - \|\theta\|_2^2}{\|\theta\|_2 + 2\|\Pi_K g\|_2}\;\mathbb{I}\left\{\langle\theta,\, \Pi_K g\rangle > \frac{1}{2}\langle\theta,\, \mathbb{E}\Pi_K g\rangle\right\}.$$

Now introducing the event $\mathcal{D} := \left\{\langle\theta,\, \Pi_K g\rangle > \langle\theta,\, \mathbb{E}\Pi_K g\rangle/2\right\}$, Jensen's inequality implies that

$$\mathbb{E}Z^+(\theta) \geq \mathbb{P}(\mathcal{D})\,\mathbb{E}\frac{\langle\theta,\, \mathbb{E}\Pi_K g\rangle - \|\theta\|_2^2}{\|\theta\|_2 + 2\frac{\mathbb{E}\|\Pi_K g\|_2}{\mathbb{P}(\mathcal{D})}}. \tag{81}$$

The concentration inequality (51b) from Lemma 4.1 gives us that

$$\mathbb{P}(\mathcal{D}) \geq \mathbb{P}\left\{\langle\theta,\, \Pi_K g\rangle > \frac{1}{2}\langle\theta,\, \mathbb{E}\Pi_K g\rangle\right\} \geq 1 - \exp\left(-\frac{\langle\theta,\, \mathbb{E}\Pi_K g\rangle^2}{8\|\theta\|_2^2}\right). \tag{82}$$

Inequality (74b) now follows by combining inequalities (79), (81) and (82).

D.3. Proof of Lemma D.2. It remains to prove Lemma D.2. Inequality (77) is a standard Lipschitz property of projection onto a closed convex cone. Turning to inequality (78), recall the polar cone $K^* := \{z \mid \langle z,\, \theta\rangle \leq 0,\ \forall\ \theta \in K\}$, as well as the Moreau decomposition (18)—namely, $z = \Pi_K(z) + \Pi_{K^*}(z)$. Using this notation, we have

$$\begin{aligned}\|\Pi_K(x+y)\|_2^2 - \|\Pi_K y\|_2^2 &= \|x + y - \Pi_{K^*}(x+y)\|_2^2 - \|y - \Pi_{K^*}y\|_2^2\\ &= \|x\|_2^2 + 2\langle x,\, y - \Pi_{K^*}(x+y)\rangle + \|y - \Pi_{K^*}(x+y)\|_2^2 - \|y - \Pi_{K^*}y\|_2^2.\end{aligned}$$

Since $\Pi_{K^*}(y)$ is the closest point in K^* to y, we have $\|y - \Pi_{K^*}(x+y)\|_2 \geq \|y - \Pi_{K^*}(y)\|_2$, and hence

$$\|\Pi_K(x+y)\|_2^2 - \|\Pi_K y\|_2^2 \geq \|x\|_2^2 + 2\langle x,\, y - \Pi_{K^*}(x+y)\rangle. \tag{83}$$

Since $x \in K$ and $\Pi_{K^*}(x+y) \in K^*$, we have $\langle x,\, \Pi_{K^*}(x+y)\rangle \leq 0$, and hence, inequality (83) leads to the bound (i) in equation (78). In order to establish inequality (ii) in equation (78), we begin by rewriting expression (83) as

$$\|\Pi_K(x+y)\|_2^2 - \|\Pi_K y\|_2^2 \geq \|x\|_2^2 + 2\langle x,\, y - \Pi_{K^*} y\rangle + 2\langle x,\, \Pi_{K^*} y - \Pi_{K^*}(x+y)\rangle.$$

Applying the Cauchy-Schwarz inequality to the final term above and using the 1-Lipschitz property of $z \mapsto \Pi_{K^*} z$, we obtain:

$$\langle x,\, \Pi_{K^*} y - \Pi_{K^*}(x+y)\rangle \geq -\|x\|_2 \|\Pi_{K^*} y - \Pi_{K^*}(x+y)\|_2 \geq -\|x\|_2^2,$$

which establishes the upper bound of inequality (78).

Finally, in order to prove the lower bound in inequality (78), we write

$$\begin{aligned}
&\|\Pi_K(x+y)\|_2^2 - \|\Pi_K y\|_2^2 \\
=&\|x + y - \Pi_{K^*}(x+y)\|_2^2 - \|x + y - \Pi_{K^*} y - x\|_2^2 \\
=&\|x + y - \Pi_{K^*}(x+y)\|_2^2 - \|x + y - \Pi_{K^*} y\|_2^2 + 2\langle x,\, x + y - \Pi_{K^*} y\rangle - \|x\|_2^2.
\end{aligned}$$

Since the vector $\Pi_{K^*}(x+y)$ corresponds to the projection of $x+y$ onto K^*, we have $\|x + y - \Pi_{K^*}(x+y)\|_2 \leq \|x + y - \Pi_{K^*} y\|_2$ and thus

$$\|\Pi_K(x+y)\|_2^2 - \|\Pi_K y\|_2^2 \leq \|x\|_2^2 + 2\langle x,\, \Pi_K y\rangle,$$

which completes the proof of inequality (78).

APPENDIX E: AUXILIARY PROOFS FOR THEOREM 1 (B)

In this appendix, we collect the proofs of lemmas involved in the proof of Theorem 1(b), corresponding to the lower bound on the GLRT performance.

E.1. Proof for scenario $\mathbb{E}\|\Pi_K g\|_2 < 128$. When $\mathbb{E}\|\Pi_K g\|_2 < 128$, we begin by setting $b_\rho = \frac{1}{256}$. The assumed bound $\epsilon^2 \leq \frac{1}{256}\delta^2_{\mathrm{LR}}(\{0\}, K)$ then implies that

$$\epsilon^2 \leq \frac{1}{256}\delta^2_{\mathrm{LR}}(\{0\}, K) \leq \frac{\mathbb{E}\|\Pi_K g\|_2}{256} < \frac{1}{2}.$$

For every $\epsilon^2 \leq \frac{1}{2}$, we claim that $\mathcal{E}(\phi; \{0\}, K, \epsilon) \geq 1/2$. Note that the uniform error $\mathcal{E}(\phi; \{0\}, K, \epsilon)$ is at least as large as the error in the simple binary test

$$\mathcal{H}_0 : y \sim N(0, I_d) \quad \text{versus} \quad \mathcal{H}_1 : y \sim N(\theta, I_d), \tag{84a}$$

where $\theta \in K$ is any vector such that $\|\theta\|_2 = \epsilon$. We claim that the error for the simple binary test (84a) is lower bounded as

$$\inf_{\psi} \mathcal{E}(\psi; \{0\}, \{\theta\}, \epsilon) \geq 1/2 \qquad \text{whenever } \epsilon^2 \leq 1/2. \tag{84b}$$

The proof of this claim is straightforward: introducing the shorthand $\mathbb{P}_\theta = N(\theta, I_d)$ and $\mathbb{P}_0 = N(0, I_d)$, we have

$$\inf_{\psi} \mathcal{E}(\psi; \{0\}, \{\theta\}, \epsilon) = 1 - \|\mathbb{P}_\theta - \mathbb{P}_0\|_{\mathrm{TV}}.$$

Using the relation between χ^2 distance and TV-distance in expression (72c) and the fact that $\chi^2(\mathbb{P}_\theta, \mathbb{P}_0) = \exp(\epsilon^2) - 1$, we find that the testing error satisfies

$$\inf_{\psi} \mathcal{E}(\psi; \{0\}, \{\theta\}, \epsilon) \geq 1 - \frac{1}{2}\sqrt{\exp(\epsilon^2) - 1} \geq 1/2, \qquad \text{whenever } \epsilon^2 \leq 1/2.$$

(See Section C for more details on the relation between the TV and χ^2-distances.) This completes the proof under the condition $\mathbb{E}\|\Pi_K g\|_2 < 128$.

E.2. Proof of Lemma E.1. Let us first state Lemma E.1 and give a proof of it.

LEMMA E.1. *For any constant $a \geq 1$ and for every closed convex cone $K \neq \{0\}$, we have*

$$0 \leq \Gamma(\theta) \leq \frac{2a\|\theta\|_2^2 + 4\langle \theta,\, \mathbb{E}\Pi_K g\rangle}{\mathbb{E}\|\Pi_K g\|_2} + b\|\theta\|_2 \qquad \textit{for all } \theta \in K, \tag{85a}$$

where

$$b := 3\exp\Big(-\frac{(\mathbb{E}\|\Pi_K g\|_2)^2}{8}\Big) + 24\exp\Big(-\frac{a^2\|\theta\|_2^2}{16}\Big). \tag{85b}$$

In order to prove that $\Gamma(\theta) \geq 0$, we first introduce the convenient shorthand notation $v_1 := \Pi_{K^*}(\theta + g)$ and $v_2 := \Pi_{K^*} g$. Recall that K^* denotes the polar cone of K defined in expression (17). With this notation, the the Moreau decomposition (18) then implies that

$$\begin{aligned}\|\Pi_K(\theta + g)\|_2^2 - \|\Pi_K g\|_2^2 &= \|\theta + g - v_1\|_2^2 - \|g - v_2\|_2^2 \\ &= \|\theta\|_2^2 + 2\langle \theta,\, g - v_1\rangle + \|g - v_1\|_2^2 - \|g - v_2\|_2^2.\end{aligned}$$

The right hand side above is greater than $\|\theta\|_2^2 + 2\langle \theta,\, g - v_1\rangle$ because $\|g - v_1\|_2^2 \geq \min_{v \in K^*} \|g - v\|_2^2 = \|g - v_2\|_2^2$. From the fact that $\mathbb{E}\langle \theta,\, g\rangle = 0$ and $\langle \theta,\, v\rangle \leq 0$ for all $v \in K^*$, we have $\Gamma(\theta) \geq 0$.

Now let us prove the upper bound for expected difference $\Gamma(\theta)$. Using the convenient shorthand notation $Z(\theta) := \|\Pi_K(\theta+g)\|_2 - \|\Pi_K g\|_2$, we define the event

$$\mathcal{B} := \{\|\Pi_K g\|_2 \geq \frac{1}{2}\mathbb{E}\|\Pi_K g\|_2\}, \qquad \text{where } g \sim N(0, I_d).$$

Our proof is then based on the decomposition $\Gamma(\theta) = \mathbb{E}Z(\theta) = \mathbb{E}Z(\theta)\mathbb{I}(\mathcal{B}^c) + \mathbb{E}Z(\theta)\mathbb{I}(\mathcal{B})$. In particular, we upper bound each of these two terms separately.

Bounding $\mathbb{E}[Z(\theta)\mathbb{I}(\mathcal{B}^c)]$*:.* The analysis of this term is straightforward: inequality (77) from Lemma D.2 guarantees that $Z(\theta) \leq \|\theta\|_2$, whence

$$\mathbb{E}Z(\theta)\mathbb{I}(\mathcal{B}^c) \leq \|\theta\|_2 \mathbb{P}(\mathcal{B}^c). \tag{86}$$

Bounding $\mathbb{E}[Z(\theta)\mathbb{I}(\mathcal{B})]$*:.* Turning to the second term, we have

$$\begin{aligned}\mathbb{E}Z(\theta)\mathbb{I}(\mathcal{B}) &\leq \mathbb{E}Z^+(\theta)\mathbb{I}(\mathcal{B}) \\ &= \mathbb{E}\frac{\left(\|\Pi_K(\theta+g)\|_2^2 - \|\Pi_K g\|_2^2\right)^+}{\|\Pi_K(\theta+g)\|_2 + \|\Pi_K g\|_2}\mathbb{I}(\mathcal{B}) \leq \mathbb{E}\frac{\left(\|\Pi_K(\theta+g)\|_2^2 - \|\Pi_K g\|_2^2\right)^+}{\|\Pi_K g\|_2}\mathbb{I}(\mathcal{B}).\end{aligned}$$

On event $\mathcal{B}$, we can lower bound quantity $\|\Pi_K g\|_2$ with $\mathbb{E}\|\Pi_K g\|_2/2$ thus

$$\mathbb{E}\frac{\left(\|\Pi_K(\theta+g)\|_2^2 - \|\Pi_K g\|_2^2\right)^+}{\|\Pi_K g\|_2}\mathbb{I}(\mathcal{B}) \leq \underbrace{\mathbb{E}\frac{\left(\|\Pi_K(\theta+g)\|_2^2 - \|\Pi_K g\|_2^2\right)^+\mathbb{I}(\mathcal{B})}{\mathbb{E}\|\Pi_K g\|_2/2}}_{:=T_1}. \tag{87}$$

Next we use inequality (78) to bound the numerator of the quantity T_1, namely

$$\begin{aligned}\mathbb{E}\left(\|\Pi_K(\theta+g)\|_2^2 - \|\Pi_K g\|_2^2\right)^+\mathbb{I}(\mathcal{B}) &\leq \mathbb{E}\left(2\langle\theta,\, \Pi_K g\rangle + \|\theta\|_2^2\right)^+\mathbb{I}(\mathcal{B}) \\ &\leq \mathbb{E}\left(2\langle\theta,\, \Pi_K g\rangle + a\|\theta\|_2^2\right)^+\mathbb{I}(\mathcal{B}),\end{aligned}$$

for every constant $a \geq 1$. To further simplify notation, introduce event $\mathcal{C} := \{\theta^T\Pi_K g \geq -a\|\theta\|_2^2/2\}$ and by definition, we obtain

$$\begin{aligned}\mathbb{E}\left(2\langle\theta,\, \Pi_K g\rangle + a\|\theta\|_2^2\right)^+\mathbb{I}(\mathcal{B}) &= \mathbb{E}\left(2\langle\theta,\, \Pi_K g\rangle + a\|\theta\|_2^2\right)\mathbb{I}(\mathcal{B}\cap\mathcal{C}) \\ &\leq a\|\theta\|_2^2 + 2\mathbb{E}[\langle\theta,\, \Pi_K g\rangle\mathbb{I}(\mathcal{B}\cap\mathcal{C})]. \qquad (88)\end{aligned}$$

The right hand side of inequality (88) consists of two terms. The first term $a\|\theta\|_2^2$ is a constant, so that we only need to further bound the second term $2\mathbb{E}\langle\theta,\, \Pi_K g\rangle\mathbb{I}(\mathcal{B}\cap\mathcal{C})$. We claim that

$$\mathbb{E}[\langle\theta,\, \Pi_K g\rangle\mathbb{I}(\mathcal{B}\cap\mathcal{C})] \leq \mathbb{E}\langle\theta,\, \Pi_K g\rangle + \|\theta\|_2\mathbb{E}\|\Pi_K g\|_2(6\sqrt{\mathbb{P}(\mathcal{C}^c)} + \mathbb{P}(\mathcal{B}^c)/2). \tag{89}$$

Taking inequality (89) as given for the moment, combining inequalities (87), (88) and (89) yields

$$\mathbb{E}Z^+(\theta)\mathbb{I}(\mathcal{B}) \le T_1 \le \frac{2a\|\theta\|_2^2 + 4\mathbb{E}\langle\theta,\, \Pi_K g\rangle}{\mathbb{E}\|\Pi_K g\|_2} + \|\theta\|_2(24\sqrt{\mathbb{P}(\mathcal{C}^c)} + 2\mathbb{P}(\mathcal{B}^c)). \tag{90}$$

As a summary of the above two parts—namely inequalities (86) and (90), if we assume inequality (89), we have

$$\Gamma(\theta) \le \frac{2a\|\theta\|_2^2 + 4\mathbb{E}\langle\theta,\, \Pi_K g\rangle}{\mathbb{E}\|\Pi_K g\|_2} + \|\theta\|_2(24\sqrt{\mathbb{P}(\mathcal{C}^c)} + 3\mathbb{P}(\mathcal{B}^c)). \tag{91}$$

Based on expression (91), the last step in proving Lemma E.1 is to control the probabilities $\mathbb{P}(\mathcal{C}^c)$ and $\mathbb{P}(\mathcal{B}^c)$ respectively. Using the fact that $\langle\theta,\, \Pi_K g\rangle = \langle\theta,\, (g - \Pi_{K^*} g)\rangle \ge \langle\theta,\, g\rangle$ and the concentration of $\langle\theta,\, g\rangle$, we have

$$\mathbb{P}(\mathcal{C}^c) = \mathbb{P}(\langle\theta,\, \Pi_K g\rangle < -\frac{a}{2}\|\theta\|_2^2) \le \mathbb{P}(\langle\theta,\, g\rangle < -\frac{a}{2}\|\theta\|_2^2) \le \exp(-\frac{a^2\|\theta\|_2^2}{8}),$$

$$\text{and} \quad \mathbb{P}(\mathcal{B}^c) = \mathbb{P}(\|\Pi_K g\|_2 < \frac{1}{2}\mathbb{E}\|\Pi_K g\|_2) \le \exp(-\frac{(\mathbb{E}\|\Pi_K g\|_2)^2}{8}).$$

where the second inequality follows directly from concentration result in Lemma 4.1 (51a). Substituting the above two inequalities into expression (91) yields Lemma E.1.

So it is only left for us to show inequality (89). To see this, first notice that

$$\mathbb{E}[\langle\theta,\, \Pi_K g\rangle\mathbb{I}(\mathcal{B} \cap \mathcal{C})] = \mathbb{E}\langle\theta,\, \Pi_K g\rangle - \mathbb{E}\langle\theta,\, \Pi_K g\rangle\mathbb{I}(\mathcal{C}^c \cup \mathcal{B}^c). \tag{92}$$

The Cauchy-Schwarz inequality and triangle inequality allow us to deduce

$$\begin{aligned} -\mathbb{E}\langle\theta,\, \Pi_K g\rangle\mathbb{I}(\mathcal{C}^c \cup \mathcal{B}^c) &= \langle\theta,\, -\mathbb{E}[\Pi_K g\mathbb{I}(\mathcal{C}^c \cup \mathcal{B}^c)]\rangle \\ &\le \|\theta\|_2\|\mathbb{E}[\Pi_K g\mathbb{I}(\mathcal{C}^c \cup \mathcal{B}^c)]\|_2 \\ &\le \|\theta\|_2\Big\{\|\mathbb{E}\Pi_K g\mathbb{I}(\mathcal{C}^c)\|_2 + \|\mathbb{E}\Pi_K g\mathbb{I}(\mathcal{B}^c)\|_2\Big\}. \end{aligned}$$

Jensen's inequality further guarantees that

$$-\mathbb{E}\langle\theta,\, \Pi_K g\rangle\mathbb{I}(\mathcal{C}^c \cup \mathcal{B}^c) \le \|\theta\|_2\Big\{\underbrace{\mathbb{E}[\|\Pi_K g\|_2\mathbb{I}(\mathcal{C}^c)]}_{:=T_2} + \underbrace{\mathbb{E}[\|\Pi_K g\|_2\mathbb{I}(\mathcal{B}^c)]}_{:=T_3}\Big\}, \tag{93}$$

By definition, on event $\mathcal{B}^c$, we have $\|\Pi_K g\|_2 \leq \mathbb{E}\|\Pi_K g\|_2/2$, and consequently

$$T_3 \leq \frac{\mathbb{E}\|\Pi_K g\|_2 \mathbb{P}(\mathcal{B}^c)}{2}. \tag{94}$$

Turning to the quantity T_2, applying Cauchy-Schwartz inequality yields

$$T_2 \leq \sqrt{\mathbb{E}\|\Pi_K g\|_2^2}\sqrt{\mathbb{E}\mathbb{I}(\mathcal{C}^c)} = \sqrt{(\mathbb{E}\|\Pi_K g\|_2)^2 + \operatorname{var}(\|\Pi_K g\|_2)}\sqrt{\mathbb{P}(\mathcal{C}^c)}.$$

The variance term can be bounded as in inequality (73) which says that $\operatorname{var}(\|\Pi_K g\|_2) \leq 4$.

From inequality (21), for every nontrivial cone ($K \neq \{0\}$), we are guaranteed that $\mathbb{E}\|\Pi_K g\|_2 \geq 1/\sqrt{2\pi}$, and hence $\operatorname{var}(\|\Pi_K g\|_2) \leq 8\pi(\mathbb{E}\|\Pi_K g\|_2)^2$. Consequently, the quantity T_2 can be further bounded as

$$T_2 \leq \sqrt{1+8\pi}\,\mathbb{E}\|\Pi_K g\|_2\sqrt{\mathbb{P}(\mathcal{C}^c)} \leq 6\mathbb{E}\|\Pi_K g\|_2\sqrt{\mathbb{P}(\mathcal{C}^c)}. \tag{95}$$

Putting together inequalities (94), (95) and (93) yields

$$-\mathbb{E}[\langle\theta,\, \Pi_K g\rangle \mathbb{I}(\mathcal{C}^c \cup (\mathcal{C} \cap \mathcal{B}^c))] \leq \|\theta\|_2 \mathbb{E}\|\Pi_K g\|_2 (6\sqrt{\mathbb{P}(\mathcal{C}^c)} + \mathbb{P}(\mathcal{B}^c)/2),$$

which validates claim (89) when combined with inequality (92). We finish the proof of Lemma E.1.

E.3. Calculate the testing error. The following lemma allows us to relate $\|\Pi_K g\|_2$ to its expectation:

LEMMA E.2. *Given every closed convex cone K such that $\mathbb{E}\|\Pi_K g\|_2 \geq 128$, we have*

$$\mathbb{P}(\|\Pi_K g\|_2 > \mathbb{E}\|\Pi_K g\|_2) \;>\; 7/16. \tag{96}$$

See supplementary file [Appendix E.5] for the proof of this claim.

For future reference, we note that it is relatively straightforward to show that the random variable $\|\Pi_K g\|_2$ is distributed as a mixture of χ-distributions, and indeed, the Lemma E.2 can be proved via this route. Raubertas et al. [4] proved that the squared quantity $\|\Pi_K g\|_2^2$ is a mixture of χ^2 distributions, and a very similar argument yields the analogous statement for $\|\Pi_K g\|_2$.

We are now ready to calculate the testing error for the GLRT given in equation (11b). Our goal is to lower bound the error $\mathcal{E}(\phi_\beta; \{0\}, K, \epsilon)$ uniformly over the chosen threshold $\beta \in [0, \infty)$. We divide the choice of β into three cases, depending on the relationship between β and $\mathbb{E}\|\Pi_K g\|_2$, $\mathbb{E}\|\Pi_K(\theta + g)\|_2$. Notice this particular θ is chosen to be the one that satisfies inequality (55).

Case 1. First, consider a threshold $\beta \in [0, \; \mathbb{E}\|\Pi_K g\|_2]$. It then follows immediately from inequality (96) that the type I error by its own satisfies

$$\text{type I error} = \mathbb{P}_0(\|\Pi_K y\|_2 \geq \beta) \geq \mathbb{P}(\|\Pi_K g\|_2 \geq \mathbb{E}\|\Pi_K g\|_2) \geq \frac{7}{16}.$$

Case 2. Otherwise, consider a threshold $\beta \in \big(\mathbb{E}\|\Pi_K g\|_2, \; \mathbb{E}\|\Pi_K(\theta+g)\|_2\big]$. In this case, we again use inequality (96) to bound the type I error, namely

$$\begin{aligned} \text{type I error} &= \mathbb{P}_0(\|\Pi_K y\|_2 \geq \beta) \\ &= \mathbb{P}\Big[\|\Pi_K g\|_2 \geq \mathbb{E}\|\Pi_K g\|_2\Big] - \mathbb{P}\Big[\|\Pi_K g\|_2 \in [\mathbb{E}\|\Pi_K g\|_2, \beta)\Big] \\ &\geq \frac{7}{16} - \max_x \{f_{\|\Pi_K g\|_2}(x)(\beta - \mathbb{E}\|\Pi_K g\|_2)\}, \end{aligned}$$

where we use $f_{\|\Pi_K g\|_2}$ to denote the density function of the random variable $\|\Pi_K g\|_2$ As discussed earlier, the random variable $\|\Pi_K g\|_2$ is distributed as a mixture of χ-distributions; in particular, see Lemma E.2 above and the surrounding discussion for details. As can be verified by direct numerical calculation, any χ_k variable has a density that bounded from above by $4/5$. Using this fact, we have

$$\text{type I error} \geq \frac{7}{16} - \frac{4}{5}(\beta - \mathbb{E}\|\Pi_K g\|_2) \overset{(i)}{\geq} \frac{7}{16} - \frac{4}{5}\Gamma(\theta) \overset{(ii)}{>} 3/8,$$

where step (i) follows by the assumption that β belongs to the interval $\big(\mathbb{E}\|\Pi_K g\|_2, \; \mathbb{E}\|\Pi_K(\theta+g)\|_2\big]$, and step (ii) follows since $\Gamma(\theta) \leq 1/16$.

Case 3. Otherwise, given a threshold $\beta \in \big(\mathbb{E}\|\Pi_K(g+\theta)\|_2, \infty\big)$, we define the scalar $x := \beta - \mathbb{E}\|\Pi_K(g+\theta)\|_2$. From the concentration inequality given in Lemma 4.1, we can deduce that

$$\begin{aligned} \text{type II error} &\geq \mathbb{P}_\theta(\|\Pi_K y\|_2 \leq \beta) \\ &= 1 - \mathbb{P}\Big(\|\Pi_K(\theta+g)\|_2 - \mathbb{E}\|\Pi_K(\theta+g)\|_2 > \beta - \mathbb{E}\|\Pi_K(\theta+g)\|_2\Big) \\ &\geq 1 - \exp(-x^2/2). \end{aligned}$$

At the same time,

$$\begin{aligned} \text{type I error} = \mathbb{P}_0(\|\Pi_K y\|_2 \geq \beta) \; &= \; \mathbb{P}(\|\Pi_K g\|_2 \geq \mathbb{E}\|\Pi_K g\|_2) - \mathbb{P}(\|\Pi_K g\|_2 \in [\mathbb{E}\|\Pi_K g\|_2, \beta)) \\ &\geq \frac{7}{16} - \frac{4}{5}(\beta - \mathbb{E}\|\Pi_K g\|_2), \end{aligned}$$

where we again use inequality (96) and the boundedness of the density of $\|\Pi_K g\|_2$. Recalling that we have defined $x := \beta - \mathbb{E}\|\Pi_K(g+\theta)\|_2$ as well as

$\Gamma(\theta) = \mathbb{E}\big(\|\Pi_K(\theta+g)\|_2 - \|\Pi_K g\|_2\big)$, we have

$$\beta - \mathbb{E}\|\Pi_K g\|_2 = x + \Gamma(\theta) \;\leq\; x + \frac{1}{16},$$

where the last step uses the fact that $\Gamma(\theta) \leq 1/16$. Consequently, the type I error is lower bounded as

$$\text{type I error} \geq \frac{7}{16} - \frac{4}{5}(x + 1/16) \;=\; \frac{31}{80} - \frac{4}{5}x.$$

Combining the two types of error, we find that the testing error is lower bounded as

$$\inf_{x>0}\Big\{(\frac{31}{80} - \frac{4}{5}x)^+ + 1 - \exp(-x^2/2)\Big\} = 1 - \exp(-\frac{31^2}{2\times 64^2}) \geq 0.11.$$

Putting pieces together, the GLRT cannot succeed with error smaller than 0.11 no matter how the cut-off β is chosen.

E.4. Proof of inequality (55). Now let us turn to the proof of inequality (55). First notice that if the radius satisfies $\epsilon^2 \leq b_\rho \delta^2_{\mathrm{LR}}(\{0\}, K)$, then there exists some $\theta \in \mathcal{H}_1$ with $\|\theta\|_2 = \epsilon$ that satisfies

$$\|\theta\|_2^2 \leq b_\rho \mathbb{E}\|\Pi_K g\|_2 \text{ and } \langle\theta,\, \mathbb{E}\Pi_K g\rangle \leq \sqrt{b_\rho}\mathbb{E}\|\Pi_K g\|_2. \tag{97}$$

Setting $a = 4/\sqrt{b_\rho} \geq 1$ in inequality (85a) yields

$$\Gamma(\theta) \leq \frac{8\|\theta\|_2^2/\sqrt{b_\rho} + 4\langle\theta,\, \mathbb{E}\Pi_K g\rangle}{\mathbb{E}\|\Pi_K g\|_2} + b\|\theta\|_2$$

where $b := 3\exp(-\frac{(\mathbb{E}\|\Pi_K g\|_2)^2}{8}) + 24\exp(-\frac{\|\theta\|_2^2}{b_\rho})$. Now we only need to bound the two terms in the upper bound separately. First, note that inequality (97) yields

$$\frac{8\|\theta\|_2^2/\sqrt{b_\rho} + 4\langle\theta,\, \mathbb{E}\Pi_K g\rangle}{\mathbb{E}\|\Pi_K g\|_2} \leq 12\sqrt{b_\rho}. \tag{98}$$

On the other hand, again by applying inequality (97), it is straightforward to verify the following two facts that

$$\begin{aligned}\|\theta\|_2 \exp(-\frac{(\mathbb{E}\|\Pi_K g\|_2)^2}{8}) &\leq \sqrt{b_\rho \mathbb{E}\|\Pi_K g\|_2}\exp(-\frac{(\mathbb{E}\|\Pi_K g\|_2)^2}{8}) \\ &\leq \sqrt{b_\rho}\max_{x\in(0,\infty)}\sqrt{x}\exp(-\frac{x^2}{8}) = \sqrt{b_\rho}\left(\frac{2}{e}\right)^{1/4},\end{aligned}$$

$$\text{and} \quad \|\theta\|_2\exp(-\frac{\|\theta\|_2^2}{b_\rho}) \leq \sup_{x\in(0,\infty)} x\exp(-\frac{x^2}{b_\rho}) = \sqrt{\frac{b_\rho}{2e}}.$$

Combining the above two inequalities ensures an upper bound for product $b\|\theta\|_2$ and directly leads to upper bound of quantity $\Gamma(\theta)$, namely

$$\Gamma(\theta) \leq 12\sqrt{b_\rho} + 3\sqrt{b_\rho}\left(\frac{2}{e}\right)^{1/4} + 24\sqrt{\frac{b_\rho}{2e}},$$

With the choice of b_ρ, we established inequality (55).

E.5. Proof of Lemma E.2. In order to prove this result, we first define random variable $F := \|\Pi_K g\|_2^2 - m$, where $m := \mathbb{E}\|\Pi_K g\|_2^2$ and $\tilde{\sigma}^2 := \mathrm{var}(F)$. We make use of the Theorem 2.1 in Goldstein et al. [3] which shows that the distribution of F and Gaussian distribution $Z \sim N(0, \tilde{\sigma}^2)$ are very close, more specifically, the Theorem says

$$\|F - Z\|_{\mathrm{TV}} \leq \frac{16}{\tilde{\sigma}^2}\sqrt{m} \leq \frac{8}{\mathbb{E}\|\Pi_K g\|_2}. \tag{99}$$

In the last inequality, we use the facts that $\tilde{\sigma}^2 \geq 2m$ and $\sqrt{\mathbb{E}\|\Pi_K g\|_2^2} \geq \mathbb{E}\|\Pi_K g\|_2$.

It is known that the quantity $\|\Pi_K g\|_2^2$ is distributed as a mixture of χ^2 distributions(see e.g., [4, 3])—in particular, we can write

$$\|\Pi_K g\|_2^2 \overset{\text{law}}{=} \sum_{i=1}^{V_K} X_i = W_K + V_K, \qquad W_K = \sum_{i=1}^{V_K}(X_i - 1),$$

where each $\{X_i\}_{i\geq 1}$ is an i.i.d. sequence χ_1^2 variables, independent of V_K. Applying the decomposition of variance yields

$$\tilde{\sigma}^2 = \mathrm{var}(V_K) + 2\mathbb{E}\|\Pi_K g\|_2^2 \geq 2m.$$

We can write the probability $\mathbb{P}(\|\Pi_K g\|_2 > \mathbb{E}\|\Pi_K g\|_2)$ as

$$\mathbb{P}(\|\Pi_K g\|_2 > \mathbb{E}\|\Pi_K g\|_2) = \mathbb{P}(\|\Pi_K g\|_2^2 - \mathbb{E}\|\Pi_K g\|_2^2 > (\mathbb{E}\|\Pi_K g\|_2)^2 - \mathbb{E}\|\Pi_K g\|_2^2) \geq \mathbb{P}(F > 0).$$

So if $\mathbb{E}\|\Pi_K g\|_2 \geq 128$, then inequality (99) ensures that $d_{TV}(F, N) \leq 1/16$, and hence

$$\mathbb{P}(F > 0) \geq \mathbb{P}(Z > 0) - \|F - Z\|_{\mathrm{TV}} \geq \frac{7}{16}.$$

We finish the proof of Lemma E.2.

APPENDIX F: AUXILIARY PROOFS FOR THEOREM 2

In this appendix, we collect the proofs of various lemmas used in the proof of Theorem 2.

F.1. Proof of Lemma 2. For every probability measure $\mathbb{Q}$ supported on $K \cap B^c(1)$, let vector θ be distributed accordingly to measure $\epsilon\mathbb{Q}$ then it is supported on $K \cap B^c(\epsilon)$. Consider a mixture of distributions,

$$\mathbb{P}_1(y) = \mathbb{E}_\theta \ (2\pi)^{-d/2} \exp(-\frac{\|y-\theta\|_2^2}{2}). \tag{100}$$

Let us first control the χ^2 distance between distributions $\mathbb{P}_1$ and $\mathbb{P}_0 := N(0, I_d)$. Direct calculations yield

$$\begin{aligned}\chi^2(\mathbb{P}_1,\mathbb{P}_0) + 1 = \mathbb{E}_{\mathbb{P}_0}\left(\frac{\mathbb{P}_1}{\mathbb{P}_0}\right)^2 &= \mathbb{E}_{\mathbb{P}_0}\left(\mathbb{E}_\theta \exp\{-\frac{\|y-\theta\|_2^2}{2} + \frac{\|y\|_2^2}{2}\}\right)^2 \\ &= \mathbb{E}_{\mathbb{P}_0}\left(\mathbb{E}_\theta \exp\{\langle y,\, \theta\rangle - \frac{\|\theta\|_2^2}{2}\}\right)^2.\end{aligned}$$

Suppose random vector θ' is an independent copy of random vector θ, then

$$\begin{aligned}\chi^2(\mathbb{P}_1,\mathbb{P}_0) + 1 &= \mathbb{E}_{\mathbb{P}_0}\mathbb{E}_{\theta,\theta'} \exp\{\langle y,\, \theta+\theta'\rangle - \frac{\|\theta\|_2^2 + \|\theta'\|_2^2}{2}\} \\ &= \mathbb{E}_{\theta,\theta'} \exp\{\frac{\|\theta+\theta'\|_2^2}{2} - \frac{\|\theta\|_2^2 + \|\theta'\|_2^2}{2}\} \\ &= \mathbb{E}_{\theta,\theta'} \exp(\langle \theta,\, \theta'\rangle) \\ &= \mathbb{E} \exp(\epsilon^2\langle \eta,\, \eta'\rangle),\end{aligned} \tag{101}$$

where the second step uses the fact the moment generating function of multivariate normal distribution. As we know, the testing error is always bounded below by $1 - \|\mathbb{P}_1, \mathbb{P}_0\|_{\mathrm{TV}}$, so by the relation between the χ^2 distance and TV distance, we have:

$$\text{testing error} \ \geq \ 1 - \frac{1}{2}\sqrt{\mathbb{E}\exp\left(\epsilon^2\langle \eta,\, \eta'\rangle\right) - 1},$$

which completes our proof.

F.2. Proof of Lemma F.1. Let us first provide a formal statement of Lemma F.1 and then prove it.

Lemma F.1. *Letting η and η' denote an i.i.d pair of random variables drawn from the distribution $\mathbb{Q}$ defined in equation (57), we have*

$$\mathbb{E}_{\eta,\eta'} \exp(\epsilon^2\langle \eta,\, \eta'\rangle) \leq \frac{1}{a^2}\exp\left(\frac{5\epsilon^2\|\mathbb{E}\Pi_K g\|_2^2}{(\mathbb{E}\|\Pi_K g\|_2)^2} + \frac{40\epsilon^4\mathbb{E}(\|\Pi_K g\|_2^2)}{(\mathbb{E}\|\Pi_K g\|_2)^4}\right), \tag{102}$$

where $a := \mathbb{P}(\|\Pi_K g\|_2 \geq \frac{1}{2}\mathbb{E}\|\Pi_K g\|_2)$ and $\epsilon > 0$ satisfies the inequality $\epsilon^2 \leq (\mathbb{E}\|\Pi_K g\|_2)^2/32$.

To prove this result, we use Borell's lemma [2] which states that for a standard Gaussian vector $Z \sim N(0, I_d)$ and a function $f : \mathbb{R}^d \to \mathbb{R}$ which is L-Lipschitz, we have

$$\mathbb{E} \exp(af(Z)) \leq \exp(a\mathbb{E}f(Z) + a^2L^2/2) \tag{103}$$

for every $a \geq 0$.

Let g, g' be i.i.d standard normal vectors in $\mathbb{R}^d$. Let

$$\mathcal{A}(g) := \{\|\Pi_K g\|_2 > \frac{1}{2}\mathbb{E}\|\Pi_K g\|_2\} \text{ and } \mathcal{A}(g') := \{\|\Pi_K g'\|_2 > \frac{1}{2}\mathbb{E}\|\Pi_K g'\|_2\}$$

By definition of the probability measure $\mathbb{Q}$ in expression (57), we have

$$\begin{aligned}\mathbb{E}_{\eta,\eta'} \exp(\epsilon^2\langle\eta,\, \eta'\rangle) &= \mathbb{E}_{g,g'}\left[\exp\left(\frac{4\epsilon^2\langle\Pi_K g,\, \Pi_K g'\rangle}{\mathbb{E}\|\Pi_K g\|_2\mathbb{E}\|\Pi_K g'\|_2}\right) \Bigm| \mathcal{A}(g)\cap\mathcal{A}(g')\right] \\ &= \frac{1}{\mathbb{P}(\mathcal{A}(g)\cap\mathcal{A}(g'))}\mathbb{E}_{g,g'}\exp\left(\frac{4\epsilon^2\langle\Pi_K g,\, \Pi_K g'\rangle}{\mathbb{E}\|\Pi_K g\|_2\mathbb{E}\|\Pi_K g'\|_2}\right)\mathbb{I}(\mathcal{A}(g)\cap\mathcal{A}(g')).\end{aligned}$$

Using the independence of g, g' and nonnegativity of the exponential function, we have

$$\mathbb{E}_{\eta,\eta'} \exp(\epsilon^2\langle\eta,\, \eta'\rangle) \leq \frac{1}{\mathbb{P}(\mathcal{A}(g))^2}\underbrace{\mathbb{E}_{g,g'}\exp\left(\frac{4\epsilon^2\langle\Pi_K g,\, \Pi_K g'\rangle}{\mathbb{E}\|\Pi_K g\|_2\mathbb{E}\|\Pi_K g'\|_2}\right)}_{:=T_1}. \tag{104}$$

To simplify the notation, we write $\lambda := 4\epsilon^2/(\mathbb{E}\|\Pi_K g\|_2)^2$ so that

$$T_1 = \mathbb{E}_{g,g'}\exp\left(\lambda\langle\Pi_K g,\, \Pi_K g'\rangle\right). \tag{105}$$

Now for every fixed value of g, the function $h \mapsto \langle\Pi_K g,\, \Pi_K h\rangle$ is Lipschitz with Lipschitz constant equal to $\|\Pi_K g\|_2$. This is because

$$|\langle\Pi_K g,\, \Pi_K h\rangle - \langle\Pi_K g,\, \Pi_K h'\rangle| \leq \|\Pi_K g\|_2\|\Pi_K h - \Pi_K h'\|_2 \leq \|\Pi_K g\|_2\|h - h'\|_2,$$

where we used Cauchy-Schwartz inequality and the non-expansive property of convex projection. As a consequence of inequality (103) and Cauchy-Schwartz inequality, the term T_1 can be upper bounded as

$$\begin{aligned}T_1 &\leq \mathbb{E}_g\exp\left(\lambda\langle\Pi_K g,\, \mathbb{E}\Pi_K g'\rangle + \frac{\lambda^2\|\Pi_K g\|_2^2}{2}\right) \\ &\leq \underbrace{\sqrt{\mathbb{E}_g\exp\left(2\lambda\langle\Pi_K g,\, \mathbb{E}\Pi_K g'\rangle\right)}}_{:=T_2}\underbrace{\sqrt{\mathbb{E}_g\exp\left(\lambda^2\|\Pi_K g\|_2^2\right)}}_{:=T_3}. \end{aligned} \tag{106}$$

We now control T_2, T_3 separately. For T_2, note again that $h \mapsto \langle \Pi_K h,\, \mathbb{E}\Pi_K g' \rangle$ is a Lipschitz function with Lipschitz constant equal to $\|\mathbb{E}\Pi_K g'\|_2$. Inequality (103) implies therefore that

$$T_2 \leq \sqrt{\exp\left(2\lambda\langle \mathbb{E}\Pi_K g,\, \mathbb{E}\Pi_K g'\rangle + 2\lambda^2\|\mathbb{E}\Pi_K g'\|_2^2\right)}. \tag{107}$$

To control quantity T_3, we use a result from [1, Sublemma E.3] on the moment generating function of $\|\Pi_K g\|^2$ which gives

$$T_3 \leq \sqrt{\exp\left(\lambda^2\mathbb{E}(\|\Pi_K g\|_2^2) + \frac{2\lambda^4\mathbb{E}(\|\Pi_K g\|_2^2)}{1-4\lambda^2}\right)}, \qquad \text{whenever } \lambda < 1/4. \tag{108}$$

Because of the assumption that $\epsilon^2 \leq (\mathbb{E}\|\Pi_K g\|_2)^2/32$, we have $\lambda \leq 1/8 < 1/4$. Therefore putting all the pieces together as above, we obtain

$$\begin{aligned}
\mathbb{E}_{\eta,\eta'} \exp(\epsilon^2\langle \eta,\, \eta'\rangle) &\leq \frac{1}{\mathbb{P}(\mathcal{A}(g))^2} \exp\left((\lambda+\lambda^2)\|\mathbb{E}\Pi_K g\|_2^2 + \frac{\lambda^2\mathbb{E}(\|\Pi_K g\|_2^2)}{2} + \frac{\lambda^4\mathbb{E}(\|\Pi_K g\|_2^2)}{1-4\lambda^2}\right) \\
&\leq \frac{1}{\mathbb{P}(\mathcal{A}(g))^2} \exp\left(1.25\lambda\|\mathbb{E}\Pi_K g\|_2^2 + 2.5\lambda^2\mathbb{E}(\|\Pi_K g\|_2^2)\right) \\
&= \frac{1}{\mathbb{P}(\mathcal{A}(g))^2} \exp\left(\frac{5\epsilon^2\|\mathbb{E}\Pi_K g\|_2^2}{(\mathbb{E}(\|\Pi_K g\|_2^2)} + \frac{40\epsilon^4\mathbb{E}(\|\Pi_K g\|_2^2)}{(\mathbb{E}\|\Pi_K g\|_2)^4})\right).
\end{aligned}$$

This completes the proof of inequality (102).

APPENDIX G: AUXILIARY PROOFS FOR PROPOSITION 2 AND THE MONOTONE CONE

In this appendix, we collect various results related to the monotone cone, and the proof of Proposition 2.

G.1. Proof of Lemma 1. So as to simplify notation, we define $\xi = \Pi_K g$, with j^{th} coordinate denoted as ξ_j. Moreover, for a given vector $g \in \mathbb{R}^d$ and integers $1 \leq u < v \leq d$, we define the u to v average as

$$\bar{g}_{uv} := \frac{1}{v-u+1}\sum_{j=u}^{v} g_j.$$

To demonstrate an upper bound for the inner product $\inf_{\eta \in K\cap S^{d-1}} \langle \eta,\, \mathbb{E}\Pi_K g\rangle$, it turns out that it is enough to take $\eta = \frac{1}{\sqrt{2}}(-1, 1, 0, \ldots, 0) \in K \cap S^{d-1}$ and uses the fact that

$$\inf_{\eta \in K\cap S^{d-1}} \langle \eta,\, \mathbb{E}\Pi_K g\rangle \leq \frac{1}{\sqrt{2}}\mathbb{E}(\xi_2 - \xi_1). \tag{109}$$

So it is only left for us to analyze $\mathbb{E}(\xi_2 - \xi_1)$ which actually has an explicit form based on the explicit representation of projection to the monotone cone (see Robertson et al. [5], Chapter 1) where

$$\xi_i = \lambda_i - \bar{\lambda}, \qquad \lambda_i = \max_{u \le j} \min_{v \ge j} \bar{g}_{uv}. \tag{110}$$

This is true because projecting to cone $K = M \cap L^{\perp}$ can be written into two steps $\Pi_K g = \Pi_{L^{\perp}}(\Pi_M g)$ and projecting to subspace $L^{\perp}$ only shifts the vector to be mean zero.

We claim that the difference satisfies

$$\xi_2 - \xi_1 \le \max_{v \ge 2} |\bar{g}_{2v}| + \max_{v \ge 1} |\bar{g}_{1v}|. \tag{111}$$

To see this, as a consequence of expression (110), we have

$$\xi_2 - \xi_1 = \max\{\min_{v \ge 2} \bar{g}_{1v},\ \min_{v \ge 2} \bar{g}_{2v}\} - \min_{v \ge 1} \bar{g}_{1v}.$$

The right hand side above only takes value in set $\{\min_{v\ge 2} \bar{g}_{1v} - g_1,\ 0,\ \min_{v \ge 2} \bar{g}_{2v} - \min_{v \ge 1} \bar{g}_{1v}\}$ where the last two values agree with bound (111) obviously while the first value can be written as

$$\min_{v \ge 2} \bar{g}_{1v} - g_1 = \min_{v \ge 2} \left(\frac{1}{v} \sum_{i=2}^{v} g_i - (1 - \frac{1}{v}) g_1 \right) = \min_{v \ge 2} (1 - \frac{1}{v})(\bar{g}_{2v} - g_1) \le |\bar{g}_{2v}| + |g_1|,$$

which also agrees with inequality (111).

Next let us prove that for every $j = 1, 2$, we have

$$\mathbb{E} \max_{v \ge j} |\bar{g}_{jv}| < 20\sqrt{2}, \tag{112}$$

and combine this fact with expressions (111) and (109) gives us $\inf_{\eta \in K \cap S^{d-1}} \langle \eta, \mathbb{E}\Pi_K g \rangle \le 40$ which validates the conclusion in Lemma 1.

It is only left for us to verify inequality (112). First as we can partition the interval $[j, d]$ into k smaller intervals where each smaller interval is of length 2^m except the last one, then

$$\mathbb{E} \max_{j \le v \le d} |\bar{g}_{jv}| = \mathbb{E} \max_{1 \le m \le k} \max_{v \in I_m} |\bar{g}_{jv}| \le \sum_{m=1}^{k} \mathbb{E} \max_{v \in I_k} |\bar{g}_{jv}|, \tag{113}$$

where $I_m = [2^m + j - 2, 2^{m+1} + j - 3],\ 1 \le m < k$, the number of intervals k and length of I_k are chosen to make those intervals sum up to d.

Given index $2^m + j - 2 \leq v \leq 2^{m+1} + j - 3$, random variables $\bar{g}_{jv}$ are Gaussian distributed with mean zero and variance $1/(v - j + 1)$. Suppose we have Gaussian random variable X_v with mean zero and variance $\sigma_m^2 = 1/(2^m - 1)$ and the covariance satisfies $\mathrm{cov}(X_v, X_{v'}) = \mathrm{cov}(\bar{g}_{jv}, \bar{g}_{jv'})$. Since $\sigma_m^2 \geq 1/(v - j + 1)$, the variable $\max_{v \in I_m} |\bar{g}_{jv}|$ is stochastically dominated by the maximum $\max_{2^m \leq v \leq 2^{m+1}-1} |X_v|$, and therefore

$$\sum_{m=1}^{k} \mathbb{E} \max_{v \in I_m} |\bar{g}_{jv}| \leq \sum_{m=1}^{k} \mathbb{E} \max_{2^m \leq v \leq 2^{m+1}-1} |X_v|.$$

Applying the fact that for $t \geq 2$ number of Gaussian random variable $\epsilon_i \sim N(0, \sigma^2)$, we have $\mathbb{E} \max_{1 \leq i \leq t} |\epsilon_i| \leq 4\sigma\sqrt{2 \log t}$ which gives

$$\sum_{m=1}^{k} \mathbb{E} \max_{v \in I_m} |\bar{g}_{jv}| \leq \sum_{m=1}^{k} 4\sigma_m \sqrt{2 \log(2^m)} = 4\sqrt{2 \log 2} \left(\sum_{m=1}^{k} \sqrt{\frac{m}{2^m - 1}} \right). \tag{114}$$

The last step is to control the sum $\sum_{m=1}^{k} \sqrt{\frac{m}{2^m-1}}$. There are many ways to show that it is upper bounded by some constant. One crude way is use the fact that $\frac{\sqrt{m}}{2^m-1} \leq 2^{m/4}$ whenever $m \geq 5$, therefore we have

$$\begin{aligned} \sum_{m=1}^{k} \sqrt{\frac{m}{2^m - 1}} = \sum_{m=1}^{4} \sqrt{\frac{m}{2^m - 1}} + \sum_{m=5}^{k} \sqrt{\frac{m}{2^m - 1}} &< \sum_{m=1}^{4} \sqrt{\frac{m}{2^m - 1}} + \sum_{m=5}^{k} \frac{1}{2^{m/4}} \\ &< \sum_{m=1}^{4} \sqrt{\frac{m}{2^m - 1}} + \frac{2^{-5/4}}{1 - 2^{-1/4}} < 6, \end{aligned}$$

which validates inequality (112) when combined with inequalities (113) and (114). This completes the proof of Lemma 1.

G.2. Proof of Lemma B.1. The proof of Lemma B.1 involves two parts. First, we define the matrices G, F. Then we prove that the distribution of η has the right support where we make use of Lemma B.2.

As stated, matrix G is a lower triangular matrix satisfying (69a). Let us now specify the matrix F. Recall that we denote $\delta := r^{-2}$ and $r := 1/3$. To define matrix F, let us first define a partition of $[d]$ into m consecutive intervals $\{I_1, \ldots, I_m\}$ with m specified in expression (64) and the length of each interval $|I_i| = \ell_i$ where ℓ_i is defined as

$$\ell_i := \lfloor \frac{\delta - 1}{\delta^i} (d + \log_\delta d + 3) \rfloor, \qquad 1 \leq i \leq m - 1, \tag{115}$$

and $\ell_m := d - \sum_{i=1}^{m-1} \ell_i$.

Following directly from the definition (115), each length $\ell_i \geq 1$ and ℓ_i is a decreasing sequence with regard to i. Also ℓ_i satisfies the following

$$\ell_1 = \lfloor \frac{\delta - 1}{\delta}(d + \log_\delta d + 3) \rfloor < d \qquad \text{and} \quad \ell_i \geq \delta \ell_{i+1}, \text{ for } 1 \leq i \leq m-1, \tag{116}$$

where the first inequality holds since as $\sqrt{\log(ed)} \geq 14$, we have $(\delta - 1)(\log_\delta d + 3) \leq d$ and the last inequality follows from the fact that $\lfloor ab \rfloor \geq a\lfloor b \rfloor$ for positive integer a and $b \geq 0$ (because $a\lfloor b \rfloor$ is an integer that is smaller than ab).

We are now ready to define the $d \times m$ matrix F. We take

$$F(i,j) = \begin{cases} \frac{1}{\sqrt{\ell_j}} & i \in I_j, \\ 0 & \text{otherwise.} \end{cases} \tag{117}$$

It is easy to check that matrix F satisfies $F^T F = \mathbb{I}_m$ which validates inequality (69b).

First we show that both $\eta = FGb$ and $\eta - \bar{\eta}\mathbf{1}$ belong to M. The i-th coordinate of η can be written as

$$\eta_i = \frac{1}{\sqrt{\ell_j}} \sum_{t=1}^{j} r^{j-t} b_t, \qquad \forall\ i \in I_j.$$

Therefore we can denote u_j as the value of η_i for $i \in I_j$. To establish monotonicity, we only need to compare the value in the consecutive blocks. Direct calculation of the consecutive ratio yields

$$\frac{u_{j+1}}{u_j} = \frac{r(\sum_{t=1}^{j} r^{j-t} b_t) + b_{j+1}}{\sqrt{\ell_{j+1}}} \frac{\sqrt{\ell_j}}{\sum_{t=1}^{j} r^{j-t} b_t} \geq r\sqrt{\frac{\ell_j}{\ell_{j+1}}} \geq 1,$$

where we used the nonnegativity of coordinates of vector b and the last inequality follows from inequality (116) and $\delta = r^{-2}$. The monotonicity of $\eta - \bar{\eta}\mathbf{1}$ thus inherits directly from the monotonicity of η.

To complete the proof of Lemma B.1, we only need to prove lower bounds on $\|\eta\|_2$ and $\|\eta - \bar{\eta}\|_2$. For these, we shall use inequality (70b) of Lemma B.2.

Proof of the bound $\|\eta\|_2 \geq 1$:. Recall that $r = 1/3$ and as a direct consequence of inequality (70b) in Lemma B.2, we have

$$\langle \eta,\, \eta \rangle = \|Gb\|_2^2 \ \geq\ \frac{9}{4} - \frac{63}{32s} > 1.96, \tag{118}$$

where the last step follows form the fact that $s = \lfloor \sqrt{m} \rfloor \geq 7$. Therefore, the norm condition holds so η is supported on $M \cap L^T \cap B^c(1)$.

Proof of the bound $\|\eta - \bar{\eta}\mathbf{1}\|_2 \geq 1$*:*. The norm $\|\eta - \bar{\eta}\mathbf{1}\|_2^2$ has the following decomposition where

$$\|\eta - \bar{\eta}\mathbf{1}\|_2^2 = \|\eta\|_2^2 - d(\bar{\eta})^2.$$

We claim that $d(\bar{\eta})^2 \leq 0.2$. If we take this for now, combining with inequality (118) which says $\|\eta\|_2^2$ is greater than 1.96, we can deduce that $\|\eta - \bar{\eta}\mathbf{1}\|_2^2 \geq 1$. So it suffices to verify the claim $d(\bar{\eta})^2 \leq 0.2$. Recall that $\eta = FGb$. Direct calculation yields

$$d\bar{\eta} = \langle \mathbf{1}, \eta \rangle = \mathbf{1}^T \cdot FGb = \sum_{k=1}^m b_k \underbrace{\sum_{i=k}^m \sqrt{\ell_i} r^{i-k}}_{:=a_k}.$$

Plugging into the definitions of r and ℓ_i guarantees that

$$a_k \leq \sum_{i=k}^m \sqrt{\frac{(\delta-1)(d+\log_\delta d+3)}{\delta^i}} \frac{1}{\delta^{(i-k)/2}} = \sqrt{(\delta-1)(d+\log_\delta d+3)\delta^k} \sum_{i=k}^m \delta^{-i}$$
$$\leq \sqrt{\frac{(d+\log_\delta d+3)}{(\delta-1)\delta^{k-2}}},$$

where the last step uses the summability of a geometric sequence—namely $\sum_{i=k}^m \delta^{-i} \leq \delta^{-k+1}/(\delta-1)$. Now for every vector b, our goal is to control $\sum a_k b_k$. Recall that every vector b has s nonzero entries which equal to $1/\sqrt{s}$ where $s = \lfloor\sqrt{m}\rfloor$. Since a_k decreases with k, this inner product $\sum a_k b_k$ is largest when the first s coordinates of b are nonzero, therefore

$$d\bar{\eta} \leq \sum_{k=1}^s a_k \frac{1}{\sqrt{s}} \leq \frac{1}{\sqrt{s}} \sqrt{\frac{\delta^2(d+\log_\delta d+3)}{\delta-1}} \sum_{k=1}^s \frac{1}{\delta^{k/2}} \leq \frac{1}{\sqrt{s}} \sqrt{\frac{\delta^2(d+\log_\delta d+3)}{\delta-1}} \frac{1}{\sqrt{\delta}-1},$$

and thus we have

$$d(\bar{\eta})^2 \leq \frac{1}{\sqrt{m}-1} \frac{(d+\log_\delta d+3)}{d} \frac{\delta^2}{(\delta-1)(\sqrt{\delta}-1)^2} \leq \frac{81(d+\log_\delta d+3)}{32d(\sqrt{m}-1)} < 0.2,$$

where the last step uses $\sqrt{m} \geq 8$. Therefore, the norm condition also holds so $\eta - \bar{\eta}\mathbf{1}$ is supported on $M \cap L^T \cap B^c(1)$.

Thus, we have completed the proof of Lemma B.1.

G.3. Proof of Lemma B.2. By definition of the matrix G, we have

$$\begin{aligned}\langle Gb,\, Gb'\rangle = \sum_{t=1}^{m}(Gb)_t(Gb')_t &= \sum_{t=1}^{m}(b_t + rb_{t-1} + \cdots + r^{t-1}b_1)(b'_t + rb'_{t-1} + \cdots + r^{t-1}b'_1) \\ &= \sum_{t=1}^{m}\sum_{u=1}^{t}\sum_{v=1}^{t} r^{2t-u-v} b_u b'_v.\end{aligned}$$

Switching the order of summation yields

$$\begin{aligned}\langle Gb,\, Gb'\rangle &= \sum_{u=1}^{m}\sum_{v=1}^{m} b_u b'_v \sum_{t=\max\{u,v\}}^{m} r^{2t-u-v} \\ &= \sum_{u=1}^{m}\sum_{v=1}^{m} \frac{b_u b'_v}{r^{u+v}} \frac{r^{2\max\{u,v\}} - r^{2m+2}}{1-r^2} \\ &= \underbrace{\frac{1}{1-r^2}\sum_{u=1}^{m}\sum_{v=1}^{m} b_u b'_v r^{|u-v|}}_{:=\Delta_1} - \underbrace{\frac{1}{1-r^2}\sum_{u=1}^{m}\sum_{v=1}^{m} b_u b'_v r^{2m+2-u-v}}_{:=\Delta_2}. \end{aligned} \tag{119}$$

We bound the two terms Δ_1 and Δ_2 separately.

Recall the fact that b, b' belong to $\mathcal{S}$, so there are exactly $s = \lfloor\sqrt{m}\rfloor$ nonzero entry in both b and b' and these entries equal to $1/\sqrt{s}$. The summation defining Δ_1 is not affected by the permutation of coordinates, so that we can assume without loss of generality that the indices of nonzero entries in b are indexed by $\{1, \ldots, s\}$, and that the indices of nonzero entries in b' are indexed by $\{k, k+1, \ldots, k+s-1\}$ for some $1 \le k \le m+1-s$.

We split our proof into two cases depending on whether $k \le s$ or $k > s$.

Case 1 ($k \le s$):. The summation Δ_1 can be written as

$$s(1-r^2)\Delta_1 = s\sum_{u=1}^{m}\sum_{v=1}^{m} b_u b'_v r^{|u-v|} = \sum_{u=1}^{s}\sum_{v=k}^{k+s-1} r^{|u-v|}.$$

Direct calculation yields

$$\begin{aligned}s(1-r^2)\Delta_1 &= \sum_{u=1}^{k-1}\sum_{v=k}^{k+s-1} r^{v-u} + \sum_{u=k}^{s}\sum_{v=k}^{u} r^{u-v} + \sum_{u=k}^{s}\sum_{v=u+1}^{k+s-1} r^{v-u} \\ &= \frac{(1-r^s)(r-r^k)}{(1-r)^2} + \frac{s-k+1}{1-r} - \frac{r}{(1-r)^2}(1-r^{s-k+1}) + \frac{r(s-k+1)}{1-r} - \frac{r^k - r^{s+1}}{(1-r)^2} \\ &= \frac{1+r}{1-r}(s-k+1) + \frac{r^k(r^s + r^{s+2} - 2)}{(1-r)^2}.\end{aligned}$$

Notice the following two facts that

$$\langle b,\, b'\rangle = \frac{s-k+1}{s} \qquad \text{and} \quad \frac{-2r}{(1-r)^2} \leq \frac{r^k(r^s + r^{s+2} - 2)}{(1-r)^2} < 0,$$

so that

$$\frac{1}{(1-r)^2}\langle b,\, b'\rangle + \frac{-2r}{s(1-r^2)(1-r)^2} \leq \Delta_1 \leq \frac{1}{(1-r)^2}\langle b,\, b'\rangle. \tag{120}$$

Case 2 ($k > s$):. The summation Δ_1 satisfies the bounds

$$s(1-r^2)\Delta_1 = s\sum_{u=1}^{m}\sum_{v=1}^{m} b_u b'_v r^{|u-v|} = \sum_{u=1}^{s}\sum_{v=k}^{k+s-1} r^{v-u} = \frac{r^{k-s}(1-r^s)^2}{(1-r)^2}.$$

Since $k - s \geq 1$, we have $\langle b,\, b'\rangle = 0$ and consequently

$$\Delta_1 \leq \frac{1}{(1-r)^2}\langle b,\, b'\rangle + \frac{r}{s(1-r^2)(1-r)^2}. \tag{121}$$

Combining inequalities (119), (120) and (121), we can deduce that

$$\langle Gb,\, Gb'\rangle \;\leq\; \Delta_1 \;\leq\; \frac{1}{(1-r)^2}\langle b,\, b'\rangle + \frac{r}{s(1-r^2)(1-r)^2},$$

which validates inequality (70a).

On the other hand, when $b = b'$, the summation Δ_2 is the largest when the nonzero entries of b lie on coordinates $m-s+1, \ldots, m$. Thus we have

$$s(1-r^2)\Delta_2 \leq \sum_{u=m-s+1}^{m}\;\sum_{v=m-s+1}^{m} r^{2m+2-u-v} = \frac{r^2(1-r^s)^2}{(1-r)^2} < \frac{r^2}{(1-r)^2}. \tag{122}$$

Combining decomposition (119) with the inequalities (120), we can deduce that

$$\langle Gb,\, Gb\rangle \leq \frac{1}{(1-r)^2} - \frac{2r}{s(1-r^2)(1-r)^2} - \frac{r^2}{s(1-r^2)(1-r)^2},$$

where we use the fact that $\langle b,\, b\rangle = 1$. This completes the proof of inequality (70b).

REFERENCES

[1] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp. Living on the edge: phase transitions in convex programs with random data. *Information and Inference*, 3(3):224–294, 2014.

[2] C. Borell. The Brunn-Minkowski inequality in Gauss space. *Inventiones mathematicae*, 30:207–216, 1975.

[3] L. Goldstein, I. Nourdin, and G. Peccati. Gaussian phase transitions and conic intrinsic volumes: Steining the Steiner formula. *arXiv preprint arXiv:1411.6265*, 2014.

[4] R. F. Raubertas, C.-I. Charles Lee, and E. V. Nordheim. Hypothesis tests for normal means constrained by linear inequalities. *Communications in Statistics - Theory and Methods*, 15(9):2809–2833, 1986.

[5] T. Robertson, F. T. Wright, and R. L. Dykstra. *Order Restricted Statistical Inference*. Wiley Series in Probability and Mathematical Statistics, 1988.

[6] Y. Wei, M. J. Wainwright, and A. Guntuboyina. The geometry of hypothesis testing over convex cones: Generalized likelihood tests and minimax radii. 2017.

Department of Statistics
University of California, Berkeley
Berkeley, California 94720
E-mail: ytwei@berkeley.edu
wainwrig@berkeley.edu
aditya@stat.berkeley.edu