

Simultaneous Frequentist Calibration of Confidence Regions for Multiple Functionals in Constrained Inverse Problems

Pau Batlle[†]

pbatllef@caltech.edu

Pratik Patil[‡]

pratikpatil@utexas.edu

Michael Stanley[§]

michael.c.stanley@ama-inc.com

Javier Ruiz Lupon[†]

jruizlup@caltech.edu

Houman Owhadi[†]

owhadi@caltech.edu

Mikael Kuusela[¶]

mkuusela@andrew.cmu.edu

Abstract

Many scientific analyses require simultaneous comparison of multiple functionals of an unknown signal at once, calling for multidimensional confidence regions with guaranteed simultaneous frequentist under structural constraints (e.g., non-negativity, shape, or physics-based). This paper unifies and extends many previous optimization-based approaches to constrained confidence region construction in linear inverse problems through the lens of statistical test inversion. We begin by reviewing the historical development of optimization-based confidence intervals for the single-functional setting, from “strict bounds” to the Burrus conjecture and its recent refutation via the aforementioned test inversion framework. We then extend this framework to the multiple-functional setting. This framework can be used to: (i) improve the calibration constants of previous methods, yielding smaller confidence regions that still preserve frequentist coverage, (ii) obtain tractable multidimensional confidence regions that need not be hyper-rectangles to better capture functional dependence structure, and (iii) generalize beyond Gaussian error distributions to generic log-concave error distributions. We provide theory establishing nominal simultaneous coverage of our methods and show quantitative volume improvements relative to prior approaches using numerical experiments.

[†]Department of Computing and Mathematical Sciences, California Institute of Technology, CA 91125, USA.

[‡]Department of Statistics and Data Sciences, University of Texas, Austin, TX 78712, USA.

[§]Analytical Mechanics Associates, Hampton, VA 23666, USA.

[¶]Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA.

1 Introduction

We study linear inverse problems of the form

$$y = Kx^* + \varepsilon, \quad (1.1)$$

where $y \in \mathbb{R}^n$ denotes the observations, $x^* \in \mathbb{R}^p$ is a fixed unknown parameter vector, $K \in \mathbb{R}^{n \times p}$ is a known arbitrary forward operator, and the random noise vector $\varepsilon \in \mathbb{R}^n$ is drawn from a known log-concave density (with respect to the Lebesgue measure). To connect with and extend previous work, we begin with the Gaussian setting $\varepsilon \sim \mathcal{N}(0, I)$ ¹ and later generalize to generic log-concave settings. The true parameter x^* is known to satisfy certain constraints based on prior physical knowledge. In this work, we focus on the historically important example of non-negative constraints $x^* \geq 0$, and then extend to two broader classes: (i) general linear constraints $Ax^* \leq b$ (including box-constraints $x_{\text{lo}} \leq x^* \leq x_{\text{up}}$), and (ii) cone constraints $x^* \in \mathcal{C}$ for a closed convex cone \mathcal{C} . Let \mathcal{X} denote the set of parameters satisfying our constraints, and for every $x^* \in \mathcal{X}$, write P_{x^*} for the distribution of y under the model (1.1) with true parameter x^* (e.g., $P_{x^*} = \mathcal{N}(Kx^*, I)$ in the Gaussian setting).

We aim not to estimate the (possibly) high-dimensional vector x^* itself, but rather to make inferences about a few key linear functionals of it, such as localized averages or weighted sums over specific components. Formally, given a matrix $H \in \mathbb{R}^{k \times p}$, we are interested in determining a finite-sample $1 - \alpha$ frequentist confidence set for the unknown vector $Hx^* \in \mathbb{R}^k$, i.e., a region $\mathcal{R}_\alpha(y) \subseteq \mathbb{R}^k$ such that, for any $x \in \mathcal{X}$,

$$\mathbb{P}_{y \sim P_x}(Hx \in \mathcal{R}_\alpha(y)) \geq 1 - \alpha. \quad (1.2)$$

We stress that the guarantee (1.2) needs to hold over all parameters x in the constrained parameter space \mathcal{X} , so that in particular holds for the true unknown parameter x^* . In the literature, this problem is sometimes formulated by identifying $Hx^* = (h_1^\top x^*, \dots, h_k^\top x^*) \in \mathbb{R}^k$ as the image of k functionals of interest and then finding k intervals (or one-dimensional regions) $\mathcal{R}_\alpha^i(y)$, $i = 1, \dots, k$. Under this framework, we emphasize that (1.2) specifies *simultaneous* coverage, also named *k-at-a-time*:

$$\mathbb{P}_{y \sim P_{x^*}}(h_i^\top x^* \in \mathcal{R}_\alpha^i(y) \text{ for all } i = 1, \dots, k) \geq 1 - \alpha, \quad (1.3)$$

as opposed to *marginal* or *one-at-a-time* coverage:

$$\mathbb{P}_{y \sim P_{x^*}}(h_i^\top x^* \in \mathcal{R}_\alpha^i(y)) \geq 1 - \alpha \text{ for all } i = 1, \dots, k. \quad (1.4)$$

In this work, we do not restrict the region $\mathcal{R}_\alpha(y)$ to be necessarily a product of one-dimensional intervals (or regions). We allow geometry that captures dependencies among functionals (e.g., polytopes or ellipsoids), provided the region is computationally tractable. We define this for this work as being able to test membership $\mu \in \mathcal{R}_\alpha(y)$ by solving a convex optimization problem whose size scales at most linearly with the problem dimensions.

We generally compare and evaluate competing methods by the size of their regions $\mathcal{R}_\alpha(y)$ (e.g., the k -dimensional Lebesgue measure) subject to the simultaneous coverage guarantee (1.2), since arbitrarily large regions always satisfy such a guarantee. However, this is a challenging problem as the size of $\mathcal{R}_\alpha(y)$ is a random quantity whose distribution depends on the unknown x^* . Consequently, different optimality notions (such as minimax or average expected size or tail quantiles) lead to distinct and generally challenging problems, and no uniformly best method exists across all x^* . Instead, in this work, we focus on establishing provable *uniform* improvements over popular

¹Assuming zero mean and identity covariance is without loss of generality after shifting and whitening.

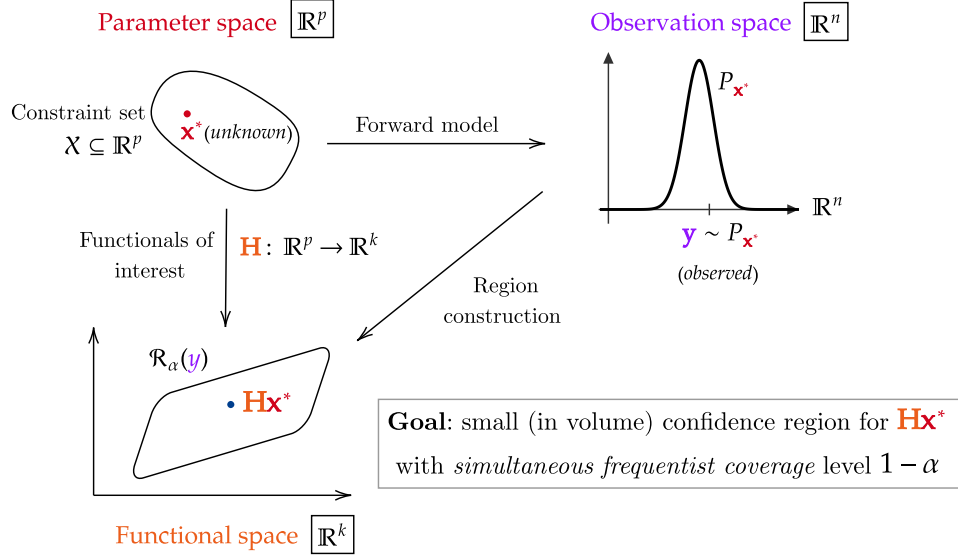


Figure 1.1: Illustration of the problem setup. We seek to construct confidence region $\mathcal{R}_\alpha(\mathbf{y}) \subseteq \mathbb{R}^k$ for $\mathbf{H}\mathbf{x}^* \in \mathbb{R}^k$ from an observation $\mathbf{y} \in \mathbb{R}^n$ sampled from $P_{\mathbf{x}^*}$ that satisfies a frequentist coverage guarantee in finite sample while being as small (in volume) as possible.

existing methods: we construct procedures producing improved regions $\tilde{\mathcal{R}}_\alpha(\mathbf{y})$ such that: (i) we retain the same finite-sample simultaneous coverage as in (1.2), yet (ii) we satisfy $\tilde{\mathcal{R}}_\alpha(\mathbf{y}) \subseteq \mathcal{R}_\alpha(\mathbf{y})$ for every data realization \mathbf{y} , where $\mathcal{R}_\alpha(\mathbf{y})$ is a region given by a previous method with guarantees. We make no claim of global optimality among all conceivable methods; rather, we demonstrate strict dominance (by set inclusion) over commonly used prior methods.

Our main baseline throughout this article is the classical Simultaneous Strict Bounds (SSB) method, which achieves the guarantee (1.2) by first constructing a $1 - \alpha$ confidence set for the full true parameter \mathbf{x}^* , intersecting it with the constraint set \mathcal{X} , and then mapping the intersection through the functional matrix \mathbf{H} into the functional space \mathbb{R}^k [30]. While valid, this approach is often conservative and yields regions that are larger than necessary because the initial confidence set for \mathbf{x}^* must be valid for *any* possible functional (including nonlinear ones), not just the specific linear functionals in \mathbf{H} .

The key tool we use to improve upon this method is a link between two classical ways for constructing confidence regions: statistical test inversion and optimization-based methods, which we leverage to shrink regions while preserving coverage.

We briefly situate our contributions within prior work next. For a broader historical context, including the existing connections between test-inversion and optimization-based viewpoints for inference on a single functional, see Section 3.

1.1 Related work

The problem of constructing confidence regions with frequentist guarantees has a long history (with additional context provided in Section 3).

Under weaker model assumptions than the aforementioned SSB method but without constraints, classical statistics has similarly extensively developed simultaneous confidence regions [13, 25]. This

in turn has motivated a long line of research aimed at finding tighter “one-at-a-time” intervals, most famously captured by the Burrus conjecture [3, 23, 24, 32]. This conjecture, which proposed a less conservative calibration for single-functional intervals under non-negativity constraints, remained unresolved for decades before being disproven recently through a modern framework based on the inversion of a likelihood ratio test (LRT) [2]. This approach led to a large variety of one-at-a-time methods with one computational approach developed in [28].

The literature on constrained inference similarly develops theory and methods to incorporate known parameter constraints into hypothesis testing and estimation and is thoroughly summarized in [21, 26]. These approaches heavily rely upon particular hypothesis test frameworks allowing for the use of the chi-bar-squared distribution. The hypothesis test forms explored in this work and [2, 29] fall outside those classical frameworks, though we discuss connections in Section 5.2.

Other statistical inference approaches that rely on splitting or resampling data can also be used for constrained inference. Universal inference [34] is one such method relying on data splitting. Bootstrapping approaches can be used to resample constrained estimators and achieve confidence sets [12] but the validity of these approaches is typically asymptotic. Additionally, we emphasize that the scientific scenarios motivating this work often prohibit the use of data splitting or resampling, since the forward model defining K is often only computationally available [27].

The proposed framework shares some parallels with conformal prediction (CP) (see, e.g., [33, 11, 10] for some early work and [1] for a recent overview of some recent trends), which also typically constructs finite-sample valid regions via test inversion and quantile calibration. The key differences are in goals, guarantees, and assumptions. We target confidence regions for a collection of unknown parameter functionals, whereas conformal (and in general predictive) inference methods typically target prediction intervals for future responses. CP assumes exchangeability of the data and provides marginal coverage with respect to the underlying distribution. By contrast, we assume a (possibly nonlinear) forward model with Gaussian or log-concave noise and structural constraints on the parameter. Under these assumptions, we obtain stronger (uniform worst-case frequentist) coverage guarantees that hold for all feasible parameters in the constraint set, here extended to simultaneous coverage across multiple functionals.

1.2 Summary of contributions

Broadly speaking, we construct and analyze (via test inversion) simultaneous confidence regions that retain finite-sample frequentist coverage while being uniformly less conservative than classical SSB. More specifically:

- We synthesize the extensive literature on single-functional methods, from classical strict bounds to modern optimization-based procedures, within a unified test-inversion language (Sections 2 and 3).
- We extend this framework from single to multiple functionals, introducing joint test statistics and their associated calibration problems; for several choices of test statistics, we derive optimal calibration constants (Sections 4 and 5).
- We propose computationally practical procedures for constructing these resulting regions, including algorithms to compute the aforementioned optimal constants and reduction and splitting techniques for high-dimensional settings (Section 6).

The paper proceeds as follows. Section 2 reviews the classical unconstrained problem as a warm-up

and shows the equivalence between parameter-space and functional-space formulations and casts the standard Wald ellipsoid via test inversion. Section 3 gives historical context for the single-functional, non-negatively constrained case, tracing the evolution of the Burrus conjecture and its eventual resolution, which motivates the modern test-inversion framework. In Section 4, we formally extend this framework to the multiple-functional case, formulating joint hypothesis tests and the construction of simultaneous confidence regions. Section 5 analyses the calibration problem in details and compares three key test statistics identifying their theoretical properties, which are crucial for practical implementation. Section 6 presents computationally practical algorithms for high-dimensional problems, including generalized reductions for multiple functionals or and/box-constrained problems and a new row-null splitting technique for rank-deficient models. Section 7 provides numerical experiments demonstrating nominal coverage for our methods with substantially smaller (in volume) confidence regions compared to classical approaches. We conclude with a summary and discussion of some open problems in Section 8.

2 Warming up with the unconstrained case

We begin by reviewing the classical problem of constructing a confidence region for Hx^* given $y = Kx^* + \varepsilon$ with $\varepsilon \sim \mathcal{N}(0, I)$ in the unconstrained setting, where the parameter space is $\mathcal{X} = \mathbb{R}^p$. This well-understood case serves as an essential baseline and allows us to introduce the key recurring concepts of observability of the functionals, parameter description, and functional description of confidence sets, and the test-inversion viewpoint that we will generalize to the constrained setting. We then preview how adding constraints can (in a sense) “create” observability that can result in finite regions even when unconstrained regions would be unbounded.

2.1 Classical unconstrained framework

In the unconstrained setting, we first consider an idealized “oracle” case with noiseless observations, which leads to a fundamental condition for obtaining bounded confidence regions. We then introduce noise and review the classical Wald ellipsoid, its relationship to parameter-space representations, and its modern interpretation through test inversion.

Define the oracle compatibility region \mathcal{R}^O as the set of functionals $\mu \in \mathbb{R}^k$ (i.e., values of $Hx^* = (h_1^\top x^*, \dots, h_k^\top x^*)$) that would be compatible with our data y if that data was observed without noise (i.e., $y = Kx^*$):

$$\mathcal{R}^O = \{\mu \in \mathbb{R}^k : \exists x \in \mathbb{R}^p \text{ such that } Kx = y, Hx = \mu\}. \quad (2.1)$$

A simple argument shows that

$$\mathcal{R}^O = \{Hx : x \in \mathbb{R}^p, Kx = y\} = \{HK^\dagger y + H(I - K^\dagger K)z : z \in \mathbb{R}^p\}. \quad (2.2)$$

This simple equality between a set in \mathbb{R}^k (functional space) and a linear image, under H , of a set in \mathbb{R}^p (parameter space) will have strong consequences in noisy constrained settings. We will henceforth refer to descriptions such as (2.1), directly in terms of elements of the functional space, as the μ -description of the set, and to descriptions such as (2.2), in terms of the linear image of elements in the parameter space, as the equivalent x -description of the set, and their equivalence will be an important tool for the analysis of noisy and/or constrained settings.

In this simple noiseless unconstrained setting, any functional of interest h_k can be either exactly recovered by $h_k^\top K^\dagger y$ if $h_k \in \text{row}(K)$ or cannot be estimated from the data, so that \mathcal{R}^O is the

coordinate-wise product of singletons and infinite intervals. Overall, a finite volume region for Hx^* is possible if and only if all of the functionals of interest (rows of H) lie in $\text{row}(K)$ (the same observability condition holds with noise as well). A general, not necessarily bounded, confidence region can be constructed by taking the product of the bounded confidence region of the functionals of the row space and copies of \mathbb{R} for those unobservable functionals (i.e., not in the row space). We henceforth assume that H lies in $\text{row}(K)$ to build the confidence region for observable functionals. We furthermore assume that the rows of H are linearly independent without loss of generality, since if a functional lies in the span of the other rows of H , its inference follows directly by projection of the confidence region of the other functionals.

Under this assumption, there exists a matrix B such that $B^\top K = H$. With Gaussian noise $\varepsilon \sim \mathcal{N}(0, I)$, we have $B^\top y = B^\top Kx^* + B^\top \varepsilon \sim \mathcal{N}(Hx^*, B^\top B)$. The matrix satisfying this that minimizes the variances in each column is $B = K(K^\top K)^\dagger H^\top$. Since $B^\top y \sim \mathcal{N}(Hx^*, B^\top B)$, one obtains the following $1 - \alpha$ ellipsoidal region:

$$\mathcal{R}_y^\mu = \{\mu \in \mathbb{R}^k : \|\mu - B^\top y\|_{B^\top B}^2 \leq Q_{\chi_k^2, 1-\alpha}\}, \quad (2.3)$$

where we adopt the convention $\|v\|_{B^\top B}^2 := v^\top (B^\top B)^{-1} v$ ² and $Q_{X, 1-\alpha}$ denotes the $(1 - \alpha)$ -quantile of a random variable X (satisfying $\mathbb{P}_{x \sim X}(x \leq Q_{X, 1-\alpha}) = 1 - \alpha$), with a slight abuse of notation, we will also use $Q_{P, 1-\alpha}$ to denote the $(1 - \alpha)$ -quantile of a distribution P .

Since $B^\top y = H(K^\top K)^\dagger K^\top y = HK^\dagger y$, we can view (2.3) as the ellipsoid around the (minimum ℓ_2 -norm) least squares estimator of Hx^* , which is known as the Wald form.

As in the noiseless case, we can identify (2.3) as the linear image under H of a set of “compatible parameters” in the original parameter space. Defining

$$\mathcal{R}_y^x = \{Hx : x \in \mathbb{R}^p, \|Kx - y\|_2^2 \leq \min_{x'} \|Kx' - y\|_2^2 + Q_{\chi_k^2, 1-\alpha}\}, \quad (2.4)$$

we have $\mathcal{R}_y^\mu = \mathcal{R}_y^x$ (equivalence between the μ -description (2.3) and the x -description (2.4)).

We prove a useful generalization of this result in the following proposition, in terms of a function of the data $f(y)$ and a real-valued constant M , which we separate from the function f to facilitate conceptual understanding of the results to follow.

Proposition 2.1 (Equivalence between x -description and μ -description). *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $M \in \mathbb{R}$, and define the set*

$$\mathcal{S}_y = \{x \in \mathbb{R}^p : \|Kx - y\|_2^2 \leq f(y) + M\}. \quad (2.5)$$

Then, we have

$$\{Hx : x \in \mathcal{S}_y\} = \{\mu \in \mathbb{R}^k : \inf_{Hx=\mu} \|Kx - y\|_2^2 \leq f(y) + M\} \quad (2.6)$$

$$= \{\mu \in \mathbb{R}^k : \|\mu - B^\top y\|_{B^\top B}^2 \leq f(y) + M - \inf_{x'} \|Kx' - y\|_2^2\}. \quad (2.7)$$

Proof. Let $c(y) := \min_{x'} \|Kx' - y\|_2^2 = \|(I - KK^\dagger)y\|_2^2$ and $\phi(\mu) := \min_{Hx=\mu} \|Kx - y\|_2^2$. Since $H \subseteq \text{row}(K)$, there is $B = K(K^\top K)^\dagger H^\top$ with $B^\top K = H$. A standard constrained LS argument gives $\phi(\mu) = c(y) + \|\mu - B^\top y\|_{B^\top B}^2$.

²We will use this notation as well as $\|v\|_M^2 = v^\top M^\dagger v$ for the case in which M is not invertible

(\subseteq) If $x \in \mathcal{S}_y = \{x : \|Kx - y\|_2^2 \leq f(y) + M\}$ and $\mu = Hx$, then $\phi(\mu) \leq \|Kx - y\|_2^2 \leq f(y) + M$, hence μ belongs to the RHS.

(\supseteq) If $\phi(\mu) \leq f(y) + M$, pick x_μ attaining $\phi(\mu)$. Then, $\|Kx_\mu - y\|_2^2 = \phi(\mu) \leq f(y) + M$ and $Hx_\mu = \mu$, so $\mu \in H(\mathcal{S}_y)$.

Thus, $H(\mathcal{S}_y) = \{\mu : \phi(\mu) \leq f(y) + M\}$, and substituting ϕ gives the quadratic form description. \square

This result lets us translate between μ -descriptions and x -descriptions. This is key for our analysis of coverage: even though the set is constructed in μ -space, coverage has to be enforced in x -space, because it has to hold for every parameter in the original space.

This coverage analysis is trivial to do in the unconstrained case, but becomes challenging in the constrained space, hence the usefulness of this translation tool between the two spaces.

Another consequence is the ability to describe confidence regions in μ -space for SSB procedures, that come from mapping $1 - \alpha$ confidence regions in x -space through H . While this is a conservative assumption that usually leads to overcoverage, the best-known regions in the constrained case prior to our work follow this procedure. For example, in this setting, one can build the $1 - \alpha$ confidence region $\{x \in \mathbb{R}^p : \|Kx - y\|_2^2 \leq Q_{\chi_n^2, 1-\alpha}\}$ and map it through H . Proposition 2.1 shows that in this case one obtains the region

$$\{\mu \in \mathbb{R}^k : \|\mu - B^\top y\|_{B^\top B}^2 \leq Q_{\chi_n^2, 1-\alpha} - \inf_{x'} \|Kx' - y\|_2^2\}, \quad (2.8)$$

which is a valid $1 - \alpha$ ellipsoidal confidence region (coverage is inherited from the x -space coverage). Unlike the Wald ellipsoid (2.3), the length of the different axes of the ellipse in (2.8) is dependent on the data y . Depending on the observation y , the ellipse (2.8) might be larger or smaller than the Wald ellipsoid.

As a summary, we have two different $1 - \alpha$ confidence sets, each allowing for three equivalent descriptions:

$$\mathcal{R}^1 = \{\mu \in \mathbb{R}^k : \|\mu - B^\top y\|_{B^\top B}^2 \leq Q_{\chi_n^2, 1-\alpha} - \inf_{x'} \|Kx' - y\|_2^2\} \quad (2.9)$$

$$= \{\mu \in \mathbb{R}^k : \inf_{Hx=\mu} \|Kx - y\|_2^2 \leq Q_{\chi_n^2, 1-\alpha}\} \quad (2.10)$$

$$= \{Hx : \|Kx - y\|_2^2 \leq Q_{\chi_n^2, 1-\alpha}\}, \quad (2.11)$$

corresponding to $f(y) = 0$, $M = Q_{\chi_n^2, 1-\alpha}$ in Proposition 2.1, and

$$\mathcal{R}^2 = \{\mu \in \mathbb{R}^k : \|\mu - B^\top y\|_{B^\top B}^2 \leq Q_{\chi_k^2, 1-\alpha}\} \quad (2.12)$$

$$= \{\mu \in \mathbb{R}^k : \inf_{Hx=\mu} \|Kx - y\|_2^2 \leq \inf_{x'} \|Kx' - y\|_2^2 + Q_{\chi_k^2, 1-\alpha}\} \quad (2.13)$$

$$= \{Hx : \|Kx - y\|_2^2 \leq \inf_{x'} \|Kx' - y\|_2^2 + Q_{\chi_k^2, 1-\alpha}\}, \quad (2.14)$$

corresponding to $f(y) = \inf_{x'} \|Kx' - y\|_2^2$ and $M = Q_{\chi_k^2, 1-\alpha}$ in Proposition 2.1.

In order to later generalize in the constrained case, it is helpful to interpret both of these sets (and any of the form of Proposition 2.1) as the inversion set of a hypothesis test:

$$H_0 : Hx^* = \mu \quad \text{versus} \quad H_1 : Hx^* \neq \mu. \quad (2.15)$$

with test statistic $\lambda(\mu, y) = \inf_{Hx=\mu} \|Kx - y\|_2^2 - f(y)$, resulting in confidence regions of the form $\{\mu : \lambda(\mu, y) \leq M\}$.

Therefore, when selecting f in Proposition 2.1, we are equivalently choosing a test statistic, and then choosing the constant M is the *calibration* problem for that test statistic.

We then observe that \mathcal{R}^1 is the inversion region of the test statistic with one term $\lambda^1(\mu, y) = \inf_{Hx=\mu} \|Kx - y\|_2^2$, and that \mathcal{R}^2 is the inversion region of the test statistic with two terms $\lambda^2(\mu, y) = \inf_{Hx=\mu} \|Kx - y\|_2^2 - \inf_{x'} \|Kx' - y\|_2^2$, which corresponds to a scaled negative log-likelihood ratio (LLR) test-statistic.

We re-prove coverage of the regions obtained by taking $M = Q_{\chi_n^2, 1-\alpha}$ for λ^1 and $M = Q_{\chi_k^2, 1-\alpha}$ for λ^2 below through the test inversion lens by analyzing the distributions of the test statistics under the null. This test-inversion argument for coverage will extend cleanly to the constrained setting, in which other direct techniques for coverage validity become more challenging. We do so with full generality over K and H , we then recover the case in which $H \subseteq \text{row}(K)$ and its rows are linearly independent.

Proposition 2.2 (Exact null laws for both test statistics and orthogonal decomposition). *Assume $H \subseteq \text{row}(K)$, and let $r := \text{rank}(H)$ and $R := \text{rank}(K)$. Define $B := K(K^\top K)^\dagger H^\top$ so that $B^\top K = H$, and set $c(y) := \inf_x \|Kx - y\|_2^2 = \|(I - P_K)y\|_2^2$, where P_K projects into the column space of K . For $\mu \in \mathbb{R}^k$, consider the statistics*

$$\lambda^1(\mu, y) = \inf_{Hx=\mu} \|Kx - y\|_2^2, \quad \lambda^2(\mu, y) = \inf_{Hx=\mu} \|Kx - y\|_2^2 - \inf_x \|Kx - y\|_2^2.$$

Then, under $H_0 : Hx^* = \mu$ with $y = Kx^* + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, I)$, the following hold:

- (Exact null distributions) Under H_0 ,

$$c(y) \sim \chi_{n-R}^2, \quad \lambda^1(\mu, y) \sim \chi_{n-R+r}^2, \quad \lambda^2(\mu, y) \sim \chi_r^2.$$

Moreover, $c(y)$ and $\lambda^2(\mu, y)$ are independent. If $H \subseteq \text{row}(K)$ and H has full row rank, then $r = \text{rank}(H) = k \leq R = \text{rank}(K) \leq n$.

- (Inversion sets with exact calibration) For any $0 < \alpha < 1$, the regions

$$\begin{aligned} \mathcal{R}^1 &= \{\mu : \lambda^1(\mu, y) \leq Q_{\chi_{n-R+r}^2, 1-\alpha}\} = \left\{ \mu : \|\mu - B^\top y\|_{B^\top B}^2 \leq Q_{\chi_{n-R+r}^2, 1-\alpha} - c(y) \right\} \\ \mathcal{R}^2 &= \{\mu : \lambda^2(\mu, y) \leq Q_{\chi_r^2, 1-\alpha}\} = \{\mu : \|\mu - B^\top y\|_{B^\top B}^2 \leq Q_{\chi_r^2, 1-\alpha}\} \end{aligned}$$

have $1 - \alpha$ frequentist coverage.

Proof. Recall, from the proof of Proposition 2.1:

$$\inf_{Hx=\mu} \|Kx - y\|_2^2 = c(y) + \|\mu - B^\top y\|_{B^\top B}^2 \quad (2.16)$$

This immediately shows $\lambda^2(\mu, y) = \|\mu - B^\top y\|_{B^\top B}^2$. For the first claim, write $y = Kx^* + \varepsilon$ with $\varepsilon \sim \mathcal{N}(0, I)$ under $H_0 : Hx^* = \mu$. Then $B^\top y = B^\top(Kx^* + \varepsilon) = Hx^* + B^\top \varepsilon \sim \mathcal{N}(\mu, B^\top B)$, so

$$\lambda^2(\mu, y) = \|B^\top y - \mu\|_{B^\top B}^2 \sim \chi_r^2.$$

On the other hand,

$$c(y) = \|(I - P_K)y\|_2^2 = \|(I - P_K)\varepsilon\|_2^2 \sim \chi_{n-R}^2.$$

Because $(I - P_K)\varepsilon$ is orthogonal to $P_K\varepsilon$ and $B^\top \varepsilon$ depends only on $P_K\varepsilon$ (as B^\top maps into $\text{row}(K)$), the random variables $c(y)$ and $\lambda^2(\mu, y)$ are independent. Hence

$$\lambda^1(\mu, y) = c(y) + \lambda^2(\mu, y) \sim \chi_{n-R}^2 + \chi_r^2 \stackrel{d}{=} \chi_{n-R+r}^2.$$

The second claim follows by test inversion using the exact null laws proved in the first claim. \square

Remark 1. Proposition 2.2 implies coverage of the SSB region calibrated by χ_n^2 (2.8), though the region can be shrunk when $r < R$ while maintaining coverage by using the smaller quantile of χ_{n-R+r}^2 .

When $k = 1$, the projection through H corresponds to taking the minimum and maximum of the quantity of interest over a particular data-dependent set. Writing $h^\top x$ for Hx when $k = 1$, (2.3) becomes the classic least squares interval $\underbrace{h^\top (K^\top K)^{-1} Ky}_{\hat{x}} \pm z_{1-\alpha/2}$. Noting that $z_{1-\alpha/2}^2 = Q_{\chi_1^2, 1-\alpha}$ and applying Proposition 2.1 shows that this is equivalent to the interval:

$$\left[\min_{x: \|Kx - y\|_2^2 \leq \min_{x'} \|Kx' - y\|_2^2 + Q_{\chi_1^2, 1-\alpha}} h^\top x, \max_{x: \|Kx - y\|_2^2 \leq \min_{x'} \|Kx' - y\|_2^2 + Q_{\chi_1^2, 1-\alpha}} h^\top x \right]. \quad (2.17)$$

We will henceforth write intervals coming from minimizing and maximizing a quantity over the same constraint set compactly as:

$$\begin{aligned} & \min_x / \max_x \quad h^\top x \\ & \text{subject to} \quad \|Kx - y\|_2^2 \leq \min_{x'} \|Kx' - y\|_2^2 + Q_{\chi_1^2, 1-\alpha}. \end{aligned} \quad (2.18)$$

2.2 Virtues of constraints

In the unconstrained setting, functionals outside the row space of K are unobservable, leading to infinitely wide confidence intervals. By contrast, incorporating physical constraints can make these same functionals identifiable by restricting the parameter space. This is one of the key benefits of constraints: for specific quantities of interest, they can transform an ill-posed problem into a well-posed one. In what follows, we begin with linear constraints $Ax^* \leq b$ and then note the minor changes needed for cone constraints $x^* \in \mathcal{C}$.

Analogously to the unconstrained case, define the oracle compatibility region as:

$$\mathcal{R}^O = \{\mu \in \mathbb{R}^k : \exists x \in \mathbb{R}^p \text{ s.t. } Ax \leq b, Kx = y, Hx = \mu\} = \{Hx : x \in \mathbb{R}^p, Ax \leq b, Kx = y\}. \quad (2.19)$$

We assume that (2.19) is non-empty in this section, which will happen if the true parameter x^* satisfies the imposed constraints. We note that a set similar to (2.19) can also be constructed before observing the data by eliminating the constraint $Kx = y$, which then leads to worst-case methods used in the uncertainty quantification literature.

We next show that the presence of noise does not affect the question of boundedness. Specifically, boundedness of \mathcal{R}^O is equivalent to the existence of a procedure that gives a.s.-bounded $1 - \alpha$ confidence regions, as formalized below.

Lemma 2.3 (Noiseless versus noisy boundedness equivalence). *For any $0 \leq \alpha \leq 1$, there exists a procedure to obtain an almost surely bounded confidence region for Hx^* given the observation of $y = Kx + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, I)$, and the constraint $Ax^* \leq b$ if and only if the oracle compatibility region (2.19) is bounded.*

Proof. We use throughout the proof that a non-empty closed convex set is bounded if and only if its recession cone is the origin. We first note that the boundedness of \mathcal{R}^O is independent of the observation y , since its recession cone is $\mathcal{R} = \{Hd : Ad \leq 0, Kd = 0\}$. To prove that if \mathcal{R}^O is not bounded, a $1 - \alpha$ almost surely bounded confidence region cannot exist, consider a feasible x_0 and points of the form $x_t = x_0 + td$, where $t \in \mathbb{R}$ and $0 \neq Hd \in \mathcal{R}$. Since $Kx_t = Kx_0$ and $Hx_t - Hx_0 = tHd$, there exist points that produce the same data distribution that are arbitrarily far away in functional space. This disallows finite confidence sets almost surely. Formally, fix $\varepsilon \in (0, 1)$; by a.s. finiteness there exists $M < \infty$ with $\mathbb{P}(D(\mathcal{C}(y)) \leq M) \geq 1 - \varepsilon$, where D is the diameter of a set. Choose t so large that $\|Hx_t - Hx_0\| > M$. On the event $\{D(\mathcal{C}(y)) \leq M\}$, the set $\mathcal{C}(Y)$ cannot contain both Hx_0 and Hx_t . Yet the data law under x_t equals that under x_0 , and uniform $1 - \alpha$ -coverage gives $\mathbb{P}(Hx_0 \in \mathcal{C}(Y)) \geq 1 - \alpha$ and $\mathbb{P}(Hx_t \in \mathcal{C}(Y)) \geq 1 - \alpha$. Therefore

$$\mathbb{P}(\{Hx_0 \in \mathcal{C}(Y)\} \cap \{Hx_t \in \mathcal{C}(Y)\} \cap \{D(\mathcal{C}(Y)) \leq M\}) \geq 1 - 2\alpha - \varepsilon. \quad (2.20)$$

For $\alpha < 1/2$ (or for any $\alpha < 1/l$ if we pick l points), this probability is positive, contradicting $\{D(Y) \leq M\}$. Hence no a.s.-bounded $1 - \alpha$ confidence set exists when \mathcal{C}^O is unbounded.

In the other direction, for any $c \geq 0$, $\{\mu \in \mathbb{R}^k : \exists x \in \mathbb{R}^p \text{ s.t. } Ax \leq b, Kx = y, Hx = \mu\}$ is bounded if and only if $\{\mu \in \mathbb{R}^k : \exists x \in \mathbb{R}^p \text{ s.t. } Ax \leq b, \|Kx - y\|_2^2 \leq c, Hx = \mu\}$ is bounded. This is because both sets share the same recession cone. The quantity c can then always be picked to obtain a $1 - \alpha$ confidence set for any required α based on the distribution of the norm of the noise, so a bounded almost surely method exists when \mathcal{R}^O is finite. \square

Theorem 2.4 (Boundedness characterization). *The compatibility region (2.19) is bounded, and therefore a method to obtain an almost surely bounded confidence region exists, if and only if*

$$\forall d \in \mathbb{R}^p : \quad Kd = 0, Ad \leq 0 \Rightarrow Hd = 0. \quad (2.21)$$

Proof. As we have used in the proof of Lemma 2.3, \mathcal{R}^O is bounded if and only if its recession cone $\mathcal{R} = \{Hd : d \in \mathbb{R}^p, Kd = 0, Ad \leq 0\}$ equals the origin. This happens if and only if (2.21) holds. \square

We can furthermore derive a more refined test that can be performed to check whether upper/lower bounded regions can be constructed for each individual functional, by looking at the projection of the recession cone in each coordinate. This can be understood as a compatibility test of the observation model K , constraint set \mathcal{X} , and functionals of interest H , answering the question “Is there enough information in the constraint set \mathcal{X} and the observation K to be able to obtain bounded confidence regions for H ?”

Theorem 2.5 (Coordinate-wise boundedness via the recession cone). *Let $\mathcal{R} = \{Hd : d \in \mathbb{R}^p, Kd = 0, Ad \leq 0\}$ be the recession cone of \mathcal{R}^O and let $\mathcal{D} := \{d \in \mathbb{R}^p : Ad \leq 0, Kd = 0\}$ be its preimage of \mathcal{R} under H . The projection of \mathcal{R}^O into the i -th coordinate, P_i , is a (possibly unbounded) interval that falls into exactly one of:*

$$\left\{ \begin{array}{ll} \forall d \in \mathcal{D} \quad h_i^\top d = 0 & \implies P_i \text{ is a finite interval,} \\ \exists d \in \mathcal{D} : h_i^\top d < 0, \nexists d \in \mathcal{D} : h_i^\top d > 0 & \implies P_i = (-\infty, U_i], \\ \exists d \in \mathcal{D} : h_i^\top d > 0, \nexists d \in \mathcal{D} : h_i^\top d < 0 & \implies P_i = [L_i, \infty), \\ \exists d_+, d_- \in \mathcal{D} : h_i^\top d_+ > 0, h_i^\top d_- < 0 & \implies P_i = (-\infty, \infty). \end{array} \right.$$

The boundedness (though not the numerical values) of the endpoints L_i , U_i depends only on h_i and \mathcal{D} , not on y or on the other rows $\{h_j\}_{j \neq i}$. Moreover, an upper/lower a.s.-bounded confidence set for $h_i^\top x$ exists if and only if P_i is upper/lower bounded.

Note that each functional can be checked independently, and the existence or lack thereof of bounded confidence sets for a given functional is independent of other observed functionals.

Proof. Since the compatibility set \mathcal{R}^O is non-empty, closed, and convex, its linear one-dimensional projection $P_i := \{\mu_i : \mu \in \mathcal{R}^O\}$ is a possibly unbounded interval of the form:

$$\begin{aligned} \left[\inf_{\mu \in \mathcal{R}^O} \mu_i, \sup_{\mu \in \mathcal{R}^O} \mu_i \right] &= \left[\inf_{x \in \mathbb{R}^1: Ax \leq b, Kx=y} (Hx)_i, \sup_{x \in \mathbb{R}^1: Ax \leq b, Kx=y} (Hx)_i \right] \\ &= \left[\inf_{x \in \mathbb{R}^1: Ax \leq b, Kx=y} h_i^\top x, \sup_{x \in \mathbb{R}^1: Ax \leq b, Kx=y} h_i^\top x \right]. \end{aligned}$$

It generally holds for any polyhedron \mathcal{P} and vector h that $\sup_{x \in \mathcal{P}} h^\top x < \infty \iff h^\top d \leq 0$ for all $d \in \text{rec}(\mathcal{P})$ and $\inf_{x \in \mathcal{P}} h^\top x > -\infty \iff h^\top d \geq 0$ for all $d \in \text{rec}(\mathcal{P})$. Since $\text{rec}(\{x \in \mathbb{R}^1 : Ax \leq b, Kx = y\}) = \mathcal{D}$, this concludes the proof of the criterion. One can then use the same argument in Lemma 2.3 to prove the equivalence of this criterion to the ability of finding a.s.-bounded confidence sets for each individual functional h_i . \square

Remark 2. In the unconstrained case (in which A is absent), $\mathcal{D} = \ker K$ is a subspace, so P_i is either a finite interval or $(-\infty, \infty)$. Moreover, P_i is finite if and only if $h_i \in \text{row}(K)$.

Cone constraints. All of the results above extend almost verbatim to cone constraint $x^* \in \mathcal{C}$, where \mathcal{C} is a closed, convex cone. Since $\text{rec}(\mathcal{C}) = \mathcal{C}$, the recession cone of the updated compatibility region

$$\mathcal{R}^O = \{\mu \in \mathbb{R}^k : \exists x \in \mathbb{R}^p \text{ s.t. } x \in \mathcal{C}, Kx = y, Hx = \mu\} = \{Hx : x \in \mathbb{R}^p, x \in \mathcal{C}, Kx = y\} \quad (2.22)$$

is $\text{rec}(\mathcal{R}^O) = \{Hd : d \in \mathcal{C}, Kd = 0\}$. Thus, Theorems 2.4 and 2.5 follow directly by replacing $Ad \leq 0$ with $d \in \mathcal{C}$ throughout.

3 Classical and modern perspectives on the non-negative single-functional case

In this section, we summarize key developments for the case of a single functional under non-negativity constraints. This setting has received the most attention since at least 1964. Following the notation of the previous section, when $k = 1$ we identify Hx with the scalar $h^\top x$.

3.1 Classical strict bounds perspectives and the Burrus conjecture

In his Ph.D. thesis [3, p. 64], Walter Burrus observed the $k = 1$ analogue of Proposition 2.1 in the unconstrained, full column rank case: the usual Gaussian (Wald) interval around the least squares estimator:

$$[h^\top \hat{x} - z_{\alpha/2} \|h\|_{K^\top K}, h^\top \hat{x} + z_{\alpha/2} \|h\|_{K^\top K}] \text{ with } \hat{x} = (K^\top K)^{-1} K^\top y, \quad (3.1)$$

can be obtained as the solution of the pair of optimization problems:

$$\begin{aligned} & \min_x / \max_x \quad h^\top x \\ & \text{subject to} \quad \|Kx - y\|_2^2 \leq \min_{x'} \|Kx' - y\|_2^2 + z_{\alpha/2}^2. \end{aligned} \quad (3.2)$$

On p. 67, he further claimed “if it is known a priori that $x^* \geq 0$, then the confidence level resulting from the solution will not be reduced if only $x \geq 0$ is considered”, and claims

$$\begin{aligned} & \min_x / \max_x \quad h^\top x \\ & \text{subject to} \quad \|Kx - y\|_2^2 \leq \min_{x'} \|Kx' - y\|_2^2 + z_{\alpha/2}^2 \\ & \quad x \geq 0, \end{aligned} \quad (3.3)$$

has exact $1 - \alpha$ coverage even when K is not full column rank. He considered that the fact that the constraint set might be empty for some observations is “a serious conceptual problem” and proposes the more relaxed version:

$$\begin{aligned} & \min_x / \max_x \quad h^\top x \\ & \text{subject to} \quad \|Kx - y\|_2^2 \leq \min_{x' \geq 0} \|Kx' - y\|_2^2 + z_{\alpha/2}^2 \\ & \quad x \geq 0, \end{aligned} \quad (3.4)$$

claiming that this interval will always be non-empty, but will have coverage at least $1 - \alpha$. He believes that “in many cases, a weak interval that always exists is preferable to a stronger interval that may occasionally fail to exist”. It is now known that his claim that the confidence level will not be reduced whenever we add the extra constraint to the optimization problem is incorrect, and furthermore, his relaxed version may not even achieve $1 - \alpha$: the claim that (3.4) has $1 - \alpha$ coverage was later known as the Burrus conjecture, and was open until the disproof first appeared in [2] in 2023. Note that since (3.4) is necessarily a larger interval than (3.3), this also disproves coverage of (3.3).

In the later book by Burrus with Rust [23, p. 200], they provide geometric evidence that (3.3) may not yield a $1 - \alpha$ interval, while also claiming (3.4) has coverage. They correctly noted, however, that coverage can be guaranteed by first constructing any $1 - \alpha$ confidence set $C(y)$ for x^* and then optimizing:

$$\begin{aligned} & \min_x / \max_x \quad h^\top x \\ & \text{subject to} \quad x \in C(y) \\ & \quad x \geq 0. \end{aligned} \quad (3.5)$$

This is the template later formalized and generalized as the strict bounds method [31, 9]. In their instance, $C(y)$ is constructed using the distribution of $(x - \hat{x})^\top K^\top K(x - \hat{x}) = \|K(x - \hat{x})\|_2^2$, where $\hat{x} \in \operatorname{argmin}_x \|Kx - y\|_2^2$ is any least squares estimator. This can be seen as building a confidence set for x^* using the log-likelihood ratio test. We point out here the important equality:

$$\|Kx - y\|_2^2 - \min_x \|Kx - y\|_2^2 = \|Kx - y\|_2^2 - \|K\hat{x} - y\|_2^2 = \|K(x - \hat{x})\|_2^2, \quad (3.6)$$

a consequence of the normal equations $K^\top K\hat{x} = K^\top y$. From this equality, a condition of the form $\|Kx - y\|_2^2 - \min_x \|Kx - y\|_2^2 \leq a$ can be turned into an equivalent condition of the form

$\|K(x - \hat{x})\|_2^2 \leq a$. The book identifies the distribution of $\|K(x - \hat{x})\|_2^2$ as χ_r^2 where $r = \text{rank}(K)$, proposing therefore the two equivalent intervals:

$$\begin{aligned} & \min_x / \max_x \quad h^\top x \\ & \text{subject to} \quad \|K(x - \hat{x})\|_2^2 \leq Q_{\chi_r^2, 1-\alpha} \\ & \quad x \geq 0, \end{aligned} \tag{3.7}$$

and

$$\begin{aligned} & \min_x / \max_x \quad h^\top x \\ & \text{subject to} \quad \|Kx - y\|_2^2 \leq \min_{x'} \|Kx' - y\|_2^2 + Q_{\chi_r^2, 1-\alpha} \\ & \quad x \geq 0. \end{aligned} \tag{3.8}$$

Rust and O’Leary [16] built upon this result to propose a practical algorithm based on (3.8) in 1986. They identify the fact that the solutions of the optimization problems in (3.8) can be found by defining the function $L(\phi) := \min_{h^\top x = \phi, x \geq 0} \|Kx - y\|_2^2$ and solving for the two roots of the equation $L(\phi) = \min_{x'} \|Kx' - y\|_2^2 + Q_{\chi_r^2, 1-\alpha}$, a result that can be seen as a corollary of *Proposition 2.1*. They proved that L is a convex, differentiable, and piecewise quadratic.

In 1994, Rust and O’Leary published an attempted proof of the Burrus conjecture [24], based on the dualization of (3.4). The dual of any interval of the family:

$$\begin{aligned} & \min_x / \max_x \quad h^\top x \\ & \text{subject to} \quad \|Kx - y\|_2^2 \leq g(y) \\ & \quad x \geq 0, \end{aligned} \tag{3.9}$$

where $g(y)$ is a non-negative function, given by:

$$\max_{u^\top K \leq h^\top} u^\top y - \sqrt{g(y)} \|u\|_2, \tag{3.10}$$

$$\min_{u^\top K \geq h^\top} u^\top y + \sqrt{g(y)} \|u\|_2. \tag{3.11}$$

When the constraint $x \geq 0$ is absent, the inequalities of the constraints in (3.10) and (3.11) become equality constraints. Note that this constraint corresponds to the bias of the estimator $u^\top y$, making it, in expectation, either equal or an upper/lower bound of h . We observe here that the constraint allows us to solve the optimization problem in the case in which h is not in the row space of K , for which the set of u satisfying $K^\top u = h$ would be empty.

The proof attempt of the Burrus conjecture in that paper relies on studying this dual problem, under the claim that, since for every fixed vector u , $\frac{u^\top(Kx-y)}{\|u\|_2} \sim \mathcal{N}(0, 1)$, this is also true when u is the solution of the dual problems. However, [32] correctly pointed out in 2007 that the dependence of u on the data y invalidates this “pivotal” argument (no pun intended) and thereby the proof of the conjecture. In modern terms, the proof attempt can be seen as an early attempt to characterize the distribution of a given random function of the data y , which would later be formalized with test inversion theory, and prove that pivotal or almost-pivotal quantities such as the one they intended to define are not easy to find in this problem, making calibration quite challenging.

Tenorio et al. in [32] also provided an (alleged and later disproved in [2]) counterexample to the conjecture and provided one-dimensional and two-dimensional reductions for any non-negativity

constrained problem in which K has full column rank. We defer the exposition of such reductions to Section 6.1.

In parallel to the analysis and developments of the particular methods tailored for the Gaussian linear model, the strict bounds method was developed in the seminal work of [30] as a way to build confidence regions in arbitrary and even infinite-dimensional constrained settings. The basic idea is to construct a $1 - \alpha$ confidence region for x^* in the original space, and then to minimize and maximize an arbitrary number of quantities of interest over this set. When instantiated in our linear setting, this recovers the same structure as (3.5), and the confidence set is usually obtained by bounding the norm of the noise, so that the final interval reads:

$$\begin{aligned} & \min_x / \max_x \quad h^\top x \\ & \text{subject to} \quad \|Kx - y\|_2^2 \leq Q_{\chi_n^2, 1-\alpha} \\ & \quad \quad \quad x \geq 0. \end{aligned} \tag{3.12}$$

This interval was used in [29] in a particle unfolding application.

As pointed out in [30], the coverage in (3.12), and more generally in any interval of the form (3.5) whenever $C(y)$ is a $1 - \alpha$ confidence set for x^* , holds for any number of functionals, in particular for Hx^* for $k > 1$ by building the region defined as the product of the following k intervals:

$$\begin{aligned} & \prod_{i=1}^k \min_x / \max_x \quad h_i^\top x \\ & \quad \quad \quad \text{subject to} \quad \|Kx - y\|_2^2 \leq Q_{\chi_n^2, 1-\alpha} \\ & \quad \quad \quad x \geq 0. \end{aligned} \tag{3.13}$$

This provides a first benchmark method for the constrained $k > 1$ setting that we will further explore in Section 4. Another alternative extension to the $k > 1$ is given by applying Bonferroni-adjusted one-at-a-time intervals: if $\mathbb{P}(h_i^\top x \in I_i) \geq 1 - \alpha_i$ for all $i = 1, \dots, k$, and $\sum_{i=1}^k \alpha_i \leq \alpha$, then $\mathbb{P}(Hx \in \prod_{i=1}^k I_i) \geq 1 - \alpha$. However, note that this method is strictly worse than (3.13) whenever the individual intervals are built using the strict bounds procedure (3.5), as optimizing a $1 - \alpha$ confidence set over x already guarantees joint coverage without the need for the Bonferroni correction.

Finally, [30] explicitly suggested seeking confidence sets $C(y)$ tailored to the specific functional in the $k = 1$ case, rather than using the generic choice $\{x : \|Kx - y\|_2^2 \leq Q_{\chi_r^2, 1-\alpha}\}$. This idea anticipates the modern test-inversion viewpoint and motivates our developments in the sections that follow.

3.2 Modern test-inversion perspectives and resolution of the Burrus conjecture

Recently, [2] presented a theoretical framework based on test inversion that explains the coverage properties of many different optimization-based intervals, including all of those reviewed in this section. Prior to this work, rigorous coverage guarantees were typically available only when the optimization was performed over a preconstructed $1 - \alpha$ confidence set for x^* (i.e., intervals of the form (3.5), including (3.7), (3.8), and (3.12)).

Informally, for intervals obtained by solving

$$\begin{aligned} & \min_x / \max_x \quad h^\top x \\ & \text{subject to} \quad \|Kx - y\|_2^2 \leq f(y) + M \\ & \quad \quad \quad x \in \mathcal{X}, \end{aligned} \tag{3.14}$$

the test inversion framework provides, for each choice of f , a quantity $\mathcal{Q}_{f,1-\alpha}(x)$ (dependent on K , h , and \mathcal{X}) such that the interval has at least $1 - \alpha$ coverage for a given x (for the functional of interest $h^\top x$ under $y \sim P_x$) if and only if $M \geq \mathcal{Q}_{f,1-\alpha}(x)$. Thus, uniform calibration for all $x \in \mathcal{X}$ is achieved by taking $M \geq \sup_{x \in \mathcal{X}} \mathcal{Q}_{f,1-\alpha}(x)$.

Though computing this quantity can itself be computationally challenging. We defer to Section 4 the formal definition of $\mathcal{Q}_{f,1-\alpha}(x)$ and the rest of the test inversion setup. We discuss below important consequences obtained from defining and computing this function $\mathcal{Q}_{f,1-\alpha}(x)$:

- *Tightening known valid methods.* Procedures that already have provable coverage (in particular, (3.7), (3.8), (3.12)) correspond to specific choices of f in (3.14), but with $M > \sup_{x \in \mathcal{X}} \mathcal{Q}_{f,1-\alpha}(x)$. The test inversion machinery can then provide calibrated constants M that retain the nominal coverage while being smaller than the original constants. This is because we obtain coverage in functional space, avoiding the conservativeness of first building a $1 - \alpha$ set in parameter space, which in turn can provide smaller confidence regions in practice.
- *Diagnosing and fixing invalid methods.* Procedures that do not yet have provable coverage (in particular, (3.3), (3.4)) also correspond to specific choices of f in (3.14) but with M not satisfying the condition $M \geq \sup_{x \in \mathcal{X}} \mathcal{Q}_{f,1-\alpha}(x)$. Identifying points x such that $\mathcal{Q}_{f,1-\alpha}(x) > M$ shows that the procedure can undercover (as an instance, this was used to disprove the Burrus conjecture). However, the test inversion machinery can provide an alternative constant (in particular, any $M \geq \sup_{x \in \mathcal{X}} \mathcal{Q}_{f,1-\alpha}(x)$) for intervals of this form such that the uniform coverage is attained.

On the computational side, [28] presented a way to calibrate the interval (3.4) using the theoretical setup in [2] by solving a quantile optimization problem restricted over a suitable confidence set.

In the next section, we present the test inversion setup formally, together with how it can be used to recover all the intervals presented in this section and how it generalizes to $k > 1$ functionals and thresholds M that might depend on x in (3.14). In this work, we focus on the linear Gaussian case, but we note that the exposition in [2] treats a broader class of settings.

4 Test-inversion framework for the constrained multiple-functional case

In this section, we present the test inversion framework which will be used to calibrate confidence regions. We will only focus on the linear Gaussian setting. A more general description of test inversion for constrained single-functional inference can be found in [2], which can be extended to the multiple-functional case as we do below.

4.1 Summary of the single-functional case

This subsection follows the exposition from [2, 28]. Following the duality between hypothesis tests and confidence sets, we define, for every $\mu \in \mathbb{R}$, $\Phi_\mu := \{x \in \mathbb{R}^p : h^\top x = \mu\}$ and consider the family of hypothesis tests that test whether $h^\top x^* = \mu$ while preserving the known constraints of $x^* \in \mathcal{X}$:

$$H_0 : x^* \in \Phi_\mu \cap \mathcal{X} \quad \text{versus} \quad H_1 : x^* \in \mathcal{X} \setminus \Phi_\mu. \quad (4.1)$$

For every parameter μ and data y , we define a test statistic $\lambda(\mu, y)$ that rejects the test when the value is large. We use $\mathcal{X} = \{x \in \mathbb{R}^p : Ax \leq b\}$ throughout the section to connect with

previous methods, but the same results hold for cone constraints by replacing $Ax \leq b$ for $x \in \mathcal{C}$ unless explicitly noted. All of the intervals previously discussed in the literature, including those in Section 3, arise from considering one of three different test statistics to test (4.1), and the analysis of each of them will become the main focus of this paper:

$$\lambda_c^2(\mu, y) = \min_{h^\top x = \mu, Ax \leq b} \|Kx - y\|_2^2 - \min_{Ax \leq b} \|Kx - y\|_2^2, \quad (4.2)$$

$$\lambda_u^2(\mu, y) = \min_{h^\top x = \mu, Ax \leq b} \|Kx - y\|_2^2 - \min_{x \in \mathbb{R}^p} \|Kx - y\|_2^2, \quad (4.3)$$

$$\lambda^1(\mu, y) = \min_{h^\top x = \mu, Ax \leq b} \|Kx - y\|_2^2. \quad (4.4)$$

We follow the naming convention $\lambda^1, \lambda_c^2, \lambda_u^2$ to indicate whether one or two terms are used, and when two terms, whether the second minimization is constrained (c) or unconstrained (u). We note that λ_c^2 corresponds to a re-scaled log likelihood ratio test statistic. It is understood in the above definitions that if there is no x satisfying $h^\top x = \mu$ and the constraint $Ax \leq b$, we take $+\infty$ as the test statistic value so that the hypothesis is rejected even before seeing the data (alternatively, we can consider the test (4.1) only for μ in the image of the constraint set under the quantity of interest to perform test inversion).

For any general test statistic $\lambda(\mu, y)$, we build one of two types of confidence regions, which we define as sliced and global, by obtaining a suitable decision threshold function $d(\mu)$ or constant threshold D and then defining:

$$\mathcal{R}_s = \{\mu \in \mathbb{R} : \lambda(\mu, y) \leq d(\mu)\}, \quad (4.5)$$

$$\mathcal{R}_g = \{\mu \in \mathbb{R} : \lambda(\mu, y) \leq D\}. \quad (4.6)$$

Under this setup, [2, Lemma 2.2] characterizes the optimal (smallest) function $d(\mu)$ and constant D such that $1 - \alpha$ coverage is preserved for all x satisfying $Ax \leq b$: they are obtained as optimum values of quantile optimization problems.

To define such problems, we first define, for each $x \in \mathcal{X}$, the random variable Z_x corresponding to $\lambda(\mu, y)$ when μ is fixed to be $h^\top x$ and $y \sim P_x$, i.e., $y = Kx + \varepsilon$ with $\varepsilon \sim \mathcal{N}(0, I)$. More formally, letting \mathcal{T}_x be the measurable map $\mathcal{T}_x : \varepsilon \rightarrow \lambda(h^\top x, Kx + \varepsilon)$, the random variable Z_x has pushforward law $(\mathcal{T}_x)_*(\mathcal{N}(0, I))$.

We note that the null hypothesis is composite, and while $x_1 \neq x_2$ with $h^\top x_1 = h^\top x_2$ will share the same null, we generally have the laws of Z_{x_1} and Z_{x_2} be different. We let $Q_{x, 1-\alpha}$ be the quantile of Z_x at level $1 - \alpha$, so that $\mathbb{P}_{\varepsilon \sim \mathcal{N}(0, I)}(\lambda(h^\top x, Kx + \varepsilon) \leq Q_{x, 1-\alpha}) = 1 - \alpha$. The next result characterizes the optimal thresholds as functions of these quantiles:

Lemma 4.1 (Lemma 2.2 of [2]). *The optimal (smallest) thresholds ensuring uniform $1 - \alpha$ coverage over \mathcal{X} are:*

$$d^*(\mu) = \mu \mapsto \sup_{h^\top x = \mu, Ax \leq b} Q_{x, 1-\alpha} \quad (4.7)$$

$$D^* = \sup_{Ax \leq b} Q_{x, 1-\alpha}. \quad (4.8)$$

Thus, the desired uniform coverage of (4.5) and (4.6) can be guaranteed at level $1 - \alpha$ if and only if $d(\mu) \geq d^*(\mu)$ and $D \geq D^*$, respectively. Furthermore, coverage at a particular x is achieved at level $1 - \alpha$ if and only if $d(h^\top x)$ and D is at least $Q_{x, 1-\alpha}$.

We therefore identify the quantity $\mathcal{Q}_{f,1-\alpha}(x)$ from the previous section as $Q_{x,1-\alpha}$, the $1-\alpha$ quantile of Z_x . We postpone to Section 5 the discussion of optimization problems and null statistic distributions appearing in (4.7) and (4.8) for the three test statistics discussed, and assume for now that suitable $d(\mu)$ and D ensuring the desired coverage can be computed.

Finally, as a consequence of [2, Theorem 2.4] (which uses the same argument used in Proposition 2.1, but with the added constraint $x \in \mathcal{X}$), when $\lambda(\mu, y)$ is of the form $\min_{h^\top x = \mu, Ax \leq b} \|Kx - y\|_2^2 - f(y)$, we have the following equivalences:

$$\mathcal{R}_s = \{\mu \in \mathbb{R} : \min_{h^\top x = \mu, Ax \leq b} \|Kx - y\|_2^2 - f(y) \leq d(\mu)\} \quad (4.9)$$

$$= \{h^\top x : x \in \mathbb{R}^p, Ax \leq b, \|Kx - y\|_2^2 \leq f(y) + d(h^\top x)\} \quad (4.10)$$

and

$$\mathcal{R}_g = \{\mu \in \mathbb{R} : \min_{h^\top x = \mu, Ax \leq b} \|Kx - y\|_2^2 + f(y) \leq D\} \quad (4.11)$$

$$= \{h^\top x : x \in \mathbb{R}^p, Ax \leq b, \|Kx - y\|_2^2 \leq f(y) + D\}, \quad (4.12)$$

which correspond to the μ -descriptions and x -descriptions of the \mathcal{R}_s and \mathcal{R}_g regions. Furthermore, since (4.11) is a convex set in the real line, it holds that:

$$\begin{aligned} \mathcal{R}_g = \quad & \min_x / \max_x \quad h^\top x \\ & \text{subject to} \quad \|Kx - y\|_2^2 \leq f(y) + D \\ & Ax \leq b. \end{aligned} \quad (4.13)$$

Since (4.9) is not necessarily convex unless d is concave, \mathcal{R}_s is not necessarily an interval, but it is nevertheless included in the following interval:

$$\begin{aligned} \mathcal{R}_s \subseteq \quad & \min_x / \max_x \quad h^\top x \\ & \text{subject to} \quad \|Kx - y\|_2^2 \leq f(y) + d(h^\top x) \\ & Ax \leq b. \end{aligned} \quad (4.14)$$

Combining (4.13) with Lemma 4.1 is the key insight that lets us analyze coverage of a broad class of optimization-based methods and to tighten classical strict bounds methods that first construct a $1 - \alpha$ confidence set for x in parameter space. In Section 5, we connect the test statistics in (4.2)–(4.4) to previously proposed optimization-based methods, together with the consequences of such analysis.

4.2 Generalization to the multiple-functional case

When the quantity of interest is the vector $Hx^* \in \mathbb{R}^k$, the setup of Section 4.1 extends directly. We test whether $Hx^* = \mu$ for each $\mu \in \mathbb{R}^k$ while preserving the known constraints. Defining $\Phi_\mu := \{x \in \mathbb{R}^p : Hx = \mu\}$ analogously as above, the test remains:

$$H_0 : x^* \in \Phi_\mu \cap \mathcal{X} \quad \text{versus} \quad H_1 : x^* \in \mathcal{X} \setminus \Phi_\mu. \quad (4.15)$$

Given a test statistic $\lambda : \mathbb{R}^k \times \mathbb{R}^n \rightarrow \mathbb{R}$ mapping a pair (μ, y) to the real line, the acceptance region is now the region in \mathbb{R}^k defined by either of:

$$\mathcal{R}_s = \{\mu \in \mathbb{R}^k : \lambda(\mu, y) \leq d(\mu)\}, \quad (4.16)$$

$$\mathcal{R}_g = \{\mu \in \mathbb{R}^k : \lambda(\mu, y) \leq D\}, \quad (4.17)$$

for a slice-dependent threshold $d(\cdot)$ or a global constant D . Let $\varepsilon \sim \mathcal{N}(0, I)$ and define $Z_x := \lambda(Hx, Kx + \varepsilon)$ as before. Write $Q_{x, 1-\alpha}$ for the $(1 - \alpha)$ -quantile of Z_x . Then, analogously to Lemma 4.1, the optimal (smallest) thresholds ensuring uniform $1 - \alpha$ coverage over \mathcal{X} are given by:

$$d^*(\mu) = \mu \mapsto \sup_{Hx=\mu, Ax \leq b} Q_{x, 1-\alpha}, \quad (4.18)$$

$$D^* = \sup_{Ax \leq b} Q_{x, 1-\alpha}. \quad (4.19)$$

The three single-functional test statistics (4.2), (4.3), and (4.4) easily generalize to this case:

$$\lambda_c^2(\mu, y) = \min_{Hx=\mu, Ax \leq b} \|Kx - y\|_2^2 - \min_{Ax \leq b} \|Kx - y\|_2^2, \quad (4.20)$$

$$\lambda_u^2(\mu, y) = \min_{Hx=\mu, Ax \leq b} \|Kx - y\|_2^2 - \min_{x \in \mathbb{R}^p} \|Kx - y\|_2^2, \quad (4.21)$$

$$\lambda^1(\mu, y) = \min_{Hx=\mu, Ax \leq b} \|Kx - y\|_2^2. \quad (4.22)$$

When $k > 1$ and for test statistics of the family $\lambda(\mu, y) = \min_{Hx=\mu, Ax \leq b} \|Kx - y\|_2^2 - f(y)$, the global acceptance region (4.17) is a convex set in \mathbb{R}^k that is not necessarily a product of intervals, and the sliced acceptance region (4.16) is a general region of \mathbb{R}^k . Similarly to the $k = 1$ case, we can express these regions as linear images under the quantity of interest matrix H of certain regions in the original parameter space (x -description), which will be helpful to connect with optimization-based methods.

Lemma 4.2 (Equivalence between x -description and μ -description in multiple dimensions). *For any $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $d : \mathbb{R}^K \rightarrow \mathbb{R}$, we have*

$$\mathcal{R}_s = \{\mu \in \mathbb{R}^k : \min_{Hx=\mu, Ax \leq b} \|Kx - y\|_2^2 - f(y) \leq d(\mu)\} \quad (4.23)$$

$$= \{Hx : x \in \mathbb{R}^p, Ax \leq b, \|Kx - y\|_2^2 \leq f(y) + d(Hx)\}, \quad (4.24)$$

and

$$\mathcal{R}_g = \{\mu \in \mathbb{R}^k : \min_{Hx=\mu, Ax \leq b} \|Kx - y\|_2^2 - f(y) \leq D\} \quad (4.25)$$

$$= \{Hx : x \in \mathbb{R}^p, Ax \leq b, \|Kx - y\|_2^2 \leq f(y) + D\}. \quad (4.26)$$

Proof. The result follows directly from the proof of Proposition 2.1, by replacing the single-row h^\top with the multi-row matrix H . The same least squares decomposition and orthogonality argument apply without modification, although, because of the extra constraint $Ax \leq b$, the optimization problem $\min_{Hx=\mu, Ax \leq b} \|Kx - y\|_2^2$ cannot be explicitly solved. \square

Observe from the statement of the previous lemma that both \mathcal{R}_g and \mathcal{R}_s admit a μ -description and an x -description. Since \mathcal{R}_g is not necessarily a product of intervals, there is no direct equality such as (4.13), which we had in the $k = 1$ case. Nevertheless, we can consider, for both the sliced and the global regions, their following x -description bounding boxes in \mathbb{R}^k :

$$\mathcal{R}_s \subseteq \prod_{i=1}^k \begin{array}{l} \min_x / \max_x \quad h_i^\top x \\ \text{subject to} \quad \|Kx - y\|_2^2 \leq f(y) + d(h^\top x) \\ Ax \leq b, \end{array} \quad (4.27)$$

and

$$\mathcal{R}_g \subseteq \prod_{i=1}^k \begin{array}{l} \min_x / \max_x \quad h_i^\top x \\ \text{subject to} \quad \|Kx - y\|_2^2 \leq f(y) + D \\ Ax \leq b. \end{array} \quad (4.28)$$

We call this the x -description bounding box. In particular, this shows that the previously discussed simultaneous method for multiple functionals (3.13) is a bounding box encapsulating a convex region that, while not being a product intervals, can already provide $1 - \alpha$ coverage while being a set of smaller volume.

Remark 3. If $\text{rank}(H) = p$, then testing $Hx^* = \mu$ is equivalent to testing $x^* = x$, because there is enough information in the functionals to recover the full true parameter. Therefore, inverting the test yields a $1 - \alpha$ confidence test for x^* , recovering strict bounds-type methods. Hence, the gains of the methods presented here (with respect to strict bounds-type methods) are largest when $\text{rank}(H) \ll p$.

5 Unifying different test statistics and calibration strategies

As mentioned in the previous section, most optimization-based approaches to constrained confidence regions can be expressed as test inversions of one of three test statistics:

$$\lambda_c^2(\mu, y) = \min_{Hx=\mu, Ax \leq b} \|Kx - y\|_2^2 - \min_{Ax \leq b} \|Kx - y\|_2^2, \quad (5.1)$$

$$\lambda_u^2(\mu, y) = \min_{Hx=\mu, Ax \leq b} \|Kx - y\|_2^2 - \min_{x \in \mathbb{R}^p} \|Kx - y\|_2^2, \quad (5.2)$$

$$\lambda^1(\mu, y) = \min_{Hx=\mu, Ax \leq b} \|Kx - y\|_2^2. \quad (5.3)$$

In this section, we make the mapping to prior methods explicit and then analyze each statistic from a calibration point of view: how to choose thresholds to guarantee $1 - \alpha$ coverage, when optimal constants exist, and where the quantile maximizers lie, highlighting convexity and extremal structure that make some cases tractable.

5.1 Mapping methods to test statistics

For each statistic in (5.1)–(5.3), there are four natural acceptance sets: the sliced set (4.23) or the global set (4.25), each optionally embedded into a product of intervals via (4.27) or (4.28). Note they all live in the same functional space \mathbb{R}^k , and by construction, we always have that the μ -description region is a subset of the x -description bounding box region. Table 1 shows the μ -descriptions and the x -description bounding boxes of the acceptance regions of the three test statistics. We emphasize through the notation $\delta_u^2, \delta_c^2, \delta^1$, that valid thresholds achieving the nominal coverage are not the same for different test statistics. We formulate the corresponding optimization problems that yield the optimum δ in each case in the next subsection. Since there is a pointwise inequality between the three test statistics of the form $\lambda_c^2(\mu, y) \leq \lambda_u^2(\mu, y) \leq \lambda^1(\mu, y)$, it immediately follows that any threshold that ensures valid coverage for λ^1 also ensures valid coverage for λ_u^2 and λ_c^2 ; similarly, a valid threshold for λ_u^2 is valid for λ_c^2 .

However, if we have a valid threshold for one test statistic, using it as a threshold for a smaller test statistic will provide a larger region than using the threshold for the original test statistic, so it is

Table 1: Acceptance regions for the three test statistics (4.20)–(4.22). Here, δ denotes either the global constant D or the sliced threshold $d(\mu)$ (in the μ -description) or $d(Hx)$ (in the x -description).

Statistic	μ -description	x -description (bounding box)
λ_c^2	$\{\mu : \min_{Ax \leq b, Hx = \mu} \ Kx - y\ _2^2 - \min_{Ax \leq b} \ Kx - y\ _2^2 \leq \delta_c^2\}$	$\prod_{i=1}^k \min_x / \max_x h_i^\top x$ s.t. $\ Kx - y\ _2^2 \leq \min_{Ax \leq b} \ Kx' - y\ _2^2 + \delta_c^2$ $Ax \leq b.$
λ_u^2	$\{\mu : \min_{Ax \leq b, Hx = \mu} \ Kx - y\ _2^2 - \min_{x \in \mathbb{R}^p} \ Kx - y\ _2^2 \leq \delta_u^2\}$	$\prod_{i=1}^k \min_x / \max_x h_i^\top x$ s.t. $\ Kx - y\ _2^2 \leq \min_{x' \in \mathbb{R}^p} \ Kx' - y\ _2^2 + \delta_u^2$ $Ax \leq b.$
λ^1	$\{\mu : \min_{Ax \leq b, Hx = \mu} \ Kx - y\ _2^2 \leq \delta^1\}$	$\prod_{i=1}^k \min_x / \max_x h_i^\top x$ s.t. $\ Kx - y\ _2^2 \leq \delta^1$ $Ax \leq b.$

Table 2: Reinterpretation of historical methods for the non-negatively constrained case via the unified test inversion framework. All methods use the global version of test inversion, so $\delta = D$ is a constant.

Method	Equation	Statistic	δ	$1 - \alpha?$	δ optimal?
Burrus unconstrained (1964)	(3.3)	λ_u^2	$Q(\chi_1^2)$	No	N/A
Burrus constrained (1964)	(3.4)	λ_c^2	$Q(\chi_1^2)$	No	N/A
Rust and Burrus (1972)	(3.7), (3.8)	λ_u^2	$Q(\chi_r^2)$	Yes	No
Strict bounds ($k = 1$)	(3.12)	λ^1	$Q(\chi_n^2)$	Yes	No
Strict bounds ($k > 1$)	(3.13)	λ^1	$Q(\chi_n^2)$	Yes	No

generally suboptimal in practice. Nonetheless, the inequality of valid thresholds also implies that the optimum thresholds obey $(\delta_c^2)^* \leq (\delta_u^2)^* \leq (\delta^1)^*$, for both global constants D and, pointwise, for sliced thresholds $d(\mu)$.

Furthermore, using Table 1, the intervals in Section 3 can be interpreted as test inversions with specific choices of statistic and threshold. Table 2 summarizes the mapping, whether the stated δ attains at least $1 - \alpha$ coverage, and whether it is optimal. Two takeaways are: (i) for methods with coverage, the constants can be improved; (ii) the Burrus intervals require calibration. In the next subsection, we address calibration for each statistic. For the non-negative case, we improve on the methods in Table 2: we provide theory and an algorithm to obtain the optimal constant δ^* for λ_u^2 -based regions such as (3.7) or (3.8) (Corollary 5.9) and for λ^1 -based regions such as (3.12) or (3.13) (Corollary 5.13), including the extensions for $k \geq 1$. By contrast, an optimal calibration algorithm of λ_c^2 -based regions such as (3.4) remains open in general.

5.2 The calibration problem

This section compiles the known results on choosing thresholds δ so that the regions in Table 1 achieve $1 - \alpha$ coverage. Test inversion yields optimal constants for the μ -descriptions, which automatically provide coverage for their relaxed x -description bounding boxes. While this already significantly improves prior methods, quantifying how much the bounding box constant can be reduced below the μ -description optimum while preserving coverage remains an open question. In

what follows, we focus on optimal constants for the μ -descriptions.

Recall that for any test statistic λ ,

$$d^*(\mu) = \mu \mapsto \sup_{Hx=\mu, Ax \leq b} Q_{x,1-\alpha} \quad (5.4)$$

$$D^* = \sup_{Ax \leq b} Q_{x,1-\alpha} \quad (5.5)$$

$$\mathbb{P}_{\varepsilon \sim \mathcal{N}(0, I)} (\lambda(Hx, Kx + \varepsilon) \leq Q_{x,1-\alpha}) = 1 - \alpha, \quad (5.6)$$

and we define Z_x to be the random variable $\mathcal{T}_x(\varepsilon) = \lambda(Hx, Kx + \varepsilon)$ for $\varepsilon \sim \mathcal{N}(0, I)$. We therefore will start the analysis of each test statistic by describing $\mathcal{T}_x(\varepsilon)$ and analyzing properties of the quantile function Q_x as a function of $x \in \mathcal{X}$, to understand the maximization problems in (5.4) and (5.5). All of the previously proposed methods we analyze and improve upon here correspond to the global case (5.5); in general, computing (5.4) is typically challenging, since it requires the whole function $d^* : \mathbb{R}^k \rightarrow \mathbb{R}$, so our results primarily target the global case.

Beside properties of the quantile function Q_x , stochastic upper bounds for the test statistics $\lambda(Hx, Kx + \varepsilon)$ are also of interest. This is because if a random variable X stochastically dominates $\lambda(Hx, Kx + \varepsilon)$ for all $x \in \mathcal{X}$, then by stochastic dominance $Q_{x,1-\alpha} \leq Q_{X,1-\alpha}$ for all α and for all x . Therefore, $Q_{X,1-\alpha} \geq D^*$ and it can be used to obtain $1 - \alpha$ coverage (possibly conservative). See [2, Section 3.2] for an in-depth treatment of stochastic dominance in this context. From the pointwise inequalities, there is a chain of stochastic dominance relations between our three test statistics and χ^2 distribution of varying degrees.

Proposition 5.1 (Stochastic dominance chain between test statistics). *Let $\lambda_c^2, \lambda_u^2, \lambda^1$ be as in (5.3)–(5.1). For each $x \in \mathcal{X}$, let $Z_{c,x}^2, Z_{u,x}^2, Z_x^1$ denote $\lambda(Hx, Kx + \varepsilon)$ for $\lambda \in \{\lambda_c^2, \lambda_u^2, \lambda^1\}$ with $\varepsilon \sim \mathcal{N}(0, I)$. Then, we have*

$$Z_{c,x}^2 \preceq Z_{u,x}^2 \preceq Z_x^1 \preceq \chi_n^2, \quad (5.7)$$

and

$$Z_{u,x}^2 \preceq \chi_{\text{rank}(K)}^2 \preceq \chi_n^2. \quad (5.8)$$

Proof. It follows immediately from the pointwise inequalities:

$$\begin{aligned} \lambda_c^2(Hx, Kx + \varepsilon) &= \min_{H\xi=Hx, A\xi \leq b} \|K\xi - Kx - \varepsilon\|_2^2 - \min_{A\xi \leq b} \|K\xi - Kx - \varepsilon\|_2^2 \\ &\leq \min_{H\xi=Hx, A\xi \leq b} \|K\xi - Kx - \varepsilon\|_2^2 - \min_{\xi} \|K\xi - Kx - \varepsilon\|_2^2 = \lambda_u^2(Hx, Kx + \varepsilon) \\ &\leq \min_{H\xi=Hx, A\xi \leq b} \|K\xi - Kx - \varepsilon\|_2^2 = \lambda^1(Hx, Kx + \varepsilon) \\ &\leq \|\varepsilon\|_2^2, \end{aligned}$$

and

$$\begin{aligned} \lambda_u^2(Hx, Kx + \varepsilon) &= \min_{H\xi=Hx, A\xi \leq b} \|K\xi - Kx - \varepsilon\|_2^2 - \min_{\xi} \|K\xi - Kx - \varepsilon\|_2^2 \\ &\leq \|\varepsilon\|_2^2 - \min_{\xi} \|K\xi - Kx - \varepsilon\|_2^2 \\ &= \|\varepsilon\|_2^2 - \min_{\xi'} \|K\xi' - \varepsilon\|_2^2 \\ &\leq \|\varepsilon\|_2^2. \end{aligned}$$

Because the inequalities hold almost surely, this implies stochastic dominance when $\varepsilon \sim \mathcal{N}(0, I)$, in which case $\|\varepsilon\|_2^2 \sim \chi_n^2$ and $\|\varepsilon\|_2^2 - \min_{\xi'} \|K\xi' - \varepsilon\|_2^2 \sim \chi_{\text{rank}(K)}^2$. This concludes the proof. \square

5.2.1 Test statistic with constrained second term λ_c^2

We first analyze $\lambda_c^2(\mu, y) = \min_{Hx=\mu, Ax \leq b} \|Kx - y\|_2^2 - \min_{Ax \leq b} \|Kx - y\|_2^2$, which is a natural choice as it corresponds to the log-likelihood ratio statistic. However, among the three statistics, this has the least structure in terms of the quantile function and developing a general theory about where the maximizer of the quantile optimization resides remains open.

We begin describing the function $\mathcal{T}_{c,x}^2(\varepsilon) := \lambda_c^2(Hx, Kx + \varepsilon)$. Relabeling the optimization variable to ξ and changing coordinates, we have

$$\mathcal{T}_{c,x}^2(\varepsilon) = \min_{H\xi=Hx, A\xi \leq b} \|K\xi - Kx - \varepsilon\|_2^2 - \min_{A\xi \leq b} \|K\xi - Kx - \varepsilon\|_2^2.$$

Now, we define $z := \xi - x$, so that

$$\mathcal{T}_{c,x}^2(\varepsilon) = \min_{Hz=0, Az \leq b-Ax} \|Kz - \varepsilon\|_2^2 - \min_{Az \leq b-Ax} \|Kz - \varepsilon\|_2^2.$$

Let $Q_c^2(x)$ be the $1 - \alpha$ quantile of $\mathcal{T}_{c,x}^2(\varepsilon)$ for $\varepsilon \sim \mathcal{N}(0, I)$. We fix α throughout so we omit its dependence to simplify the notation. We summarize below the known properties and open questions regarding this function:

- *Convexity.* $Q_c^2(x)$ is generally neither a convex or a concave function. As a simple counterexample, consider $H = (1, 1)$, $y = x^* + \varepsilon \in \mathbb{R}^2$, $x^* \geq 0$, $\varepsilon \sim \mathcal{N}(0, I)$. The quantile function can be shown to be neither convex nor concave; see, e.g., [28, Figure 1.1] and the slice $x_2 = 0$.
- *Maximization.* No general result is known for the maximizer of $Q_c^2(x)$ over $Ax \leq b$ (or over $Ax \leq b$, $Hx = \mu$ for sliced versions) for a given choice of problem parameters (K, H, A, b) . Furthermore, the quantile maximizer can (i) escape to ∞ for unbounded regions, and (ii) depend on the quantile level α . (These comments also apply to cone constraints.) The same two-dimensional setup as above exemplifies this behavior, as it can be shown that the maximizer over $x \geq 0$ of the quantile is at $x = 0$ for $1 - \alpha$ sufficiently close to 1 and the quantile function grows towards $x = \infty$ when $1 - \alpha$ is sufficiently close to 0 (note that the optimal value is still finite in both cases). This is a counterexample for both the linear constraints and the cone constraints case.
- *Bounds.* If K has full column rank and there are no constraints, Proposition 2.2 shows $Z_{c,x}^2 \stackrel{d}{=} \chi_{\text{rank}(H)}^2$. Under constraints, it is however not true in general that $Q_c^2(x) \leq Q_{\chi_{\text{rank}(H)}^2}$ for all α (or, in terms of stochastic dominance, that $Z_{c,x}^2 \preceq \chi_{\text{rank}(H)}^2$). The $k = 1$ version of this statement is the Burrus conjecture, which is recently shown to be false [2]. From (5.7) and (5.8), we have $Q_c^2(x) \leq Q_{\chi_{\text{rank}(K)}^2}$. One might ask if this bound is tight. While the bound is not necessarily tight for a given α and problem instance (specific choice of K, H, \mathcal{X}), we show below that if we aim for a bound that holds for all α and problem instances, the stochastic bound $Z_{c,x}^2 \preceq \chi_{\text{rank}(K)}^2$ is tight in the family of gamma distributions $\{\Gamma(t, 2)\}_{t \in \mathbb{R}^+}$ (which includes χ_k^2 , as $\Gamma(k/2, 2)$), a family over which stochastic order is a total order.

Lemma 5.2. *For every $\epsilon > 0$, there exists K, H , and $x^* \geq 0$ such that the non-negatively constrained problem induces a random variable $Z_{c,x^*}^2 \not\preceq \Gamma(\text{rank}(K)/2 - \epsilon, 2)$. Therefore, $\text{rank}(K)$ is the smallest $t > 0$ such that $Z_{c,x}^2 \preceq \Gamma(t/2, 2)$ holds uniformly over x , constraint, and K, H .*

Proof. We will prove that $\Gamma(\text{rank}(K)/2 - \epsilon, 2)$ does not stochastically dominate Z_{c,x^*}^2 by comparing their moment generating functions $\mathbb{E}[e^{tX}]$ at a particular $t > 0$ in a given family of non-negatively

constrained examples of increasing dimension. Let $K = I_n$ ($\text{rank}(K) = n$), $H = (1, \dots, 1, -1)$, and $x^* = (0, \dots, 0, M)$. Consider the following region of \mathbb{R}^n :

$$\mathcal{A} = \{\varepsilon \in \mathbb{R}^n : \varepsilon_i \geq 0, \varepsilon_i + \varepsilon_n \leq 0 \ (i = 1, \dots, n-1)\}.$$

On \mathcal{A} , feasibility is controlled entirely by the last coordinate: if M is sufficiently large, the constraint induced by H is inactive except on \mathcal{A} . Thus, for every fixed $\varepsilon \in \mathcal{A}$, we eventually have $\mathcal{T}_{c,x^*}^2(\varepsilon) = \|\varepsilon\|_2^2$ as $M \rightarrow \infty$. More precisely, for each $\delta > 0$ there exists M large enough such that there is a subset $\mathcal{A}_M \subseteq \mathcal{A}$ with $\mathbb{P}(\mathcal{A}_M) \geq (1 - \delta)\mathbb{P}(\mathcal{A})$ on which $\mathcal{T}_{c,x^*}^2 = \|\varepsilon\|_2^2$. Together with the integrand being non-negative, it follows that:

$$\mathbb{E}_{\varepsilon \sim \mathcal{N}(0, I_n)}[e^{tZ_{c,x^*}^2}] = \int_{\mathbb{R}^n} e^{t\mathcal{T}_{c,x^*}^2(\varepsilon)} f(\varepsilon) \geq \int_{\mathcal{A}_M} e^{t\|\varepsilon\|_2^2} f(\varepsilon) \geq (1 - \delta) \int_{\mathcal{A}} e^{t\|\varepsilon\|_2^2} f(\varepsilon),$$

where $t \in (0, \frac{1}{2})$, f is the density of $\mathcal{N}(0, I_n)$, $\mathcal{A}_M \subseteq \mathcal{A}$ with $1_{\mathcal{A}_M} \uparrow 1_{\mathcal{A}}$ as $M \rightarrow \infty$, and for any fixed $\delta > 0$, there exists M large enough so that the last inequality holds. Then, we have

$$\begin{aligned} \int_{\mathcal{A}} e^{t\|\varepsilon\|_2^2} f(\varepsilon) &= \int_{\mathcal{A}} e^{t\|\varepsilon\|_2^2} \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\|\varepsilon\|_2^2} \\ &= \int_{\mathcal{A}} \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}(1-2t)\|\varepsilon\|_2^2} \\ &= \frac{1}{(1-2t)^{n/2}} \int_{\mathcal{A}} \frac{(1-2t)^{n/2}}{(2\pi)^{n/2}} e^{-\frac{1}{2}(1-2t)\|\varepsilon\|_2^2} \\ &= \frac{1}{(1-2t)^{n/2}} \mathbb{P}(\mathcal{Z} \in \mathcal{A}), \end{aligned}$$

where $\mathcal{Z} \sim N(0, (1-2t)I_n)$. To calculate this value, set $X_i = \varepsilon_i$ for $i < n$, and $Y = -\varepsilon_n$. Since the distribution is spherical, X_1, \dots, X_{n-1}, Y are i.i.d. $\mathcal{N}(0, \sigma^2)$ with $\sigma^2 = 1-2t$. The region can be written as

$$\mathcal{A} = \{Y \geq 0, 0 \leq X_i \leq Y \ (i = 1, \dots, n-1)\}.$$

Conditional on $Y = y \geq 0$, we have:

$$\mathbb{P}(\mathcal{Z} \in \mathcal{A} \mid Y = y) = (\mathbb{P}(0 \leq X \leq y))^{n-1} = (\Phi(y/\sigma) - \frac{1}{2})^{n-1},$$

where Φ is the standard normal CDF. Hence

$$\begin{aligned} \mathbb{P}(\mathcal{Z} \in \mathcal{A}) &= \int_0^\infty (\Phi(y/\sigma) - \frac{1}{2})^{n-1} f_Y(y) dy \\ &= \int_0^\infty (\Phi(u) - \frac{1}{2})^{n-1} \phi(u) du \\ &= \int_0^{1/2} v^{n-1} dv = \frac{1}{n} \left(\frac{1}{2}\right)^n, \end{aligned}$$

where $f_Y(y) = \frac{1}{\sigma} \phi(y/\sigma)$ and ϕ the standard normal PDF, and we used the variable changes $u = y/\sigma$ and $v = \Phi(u) - \frac{1}{2}$. Writing all the chain of inequalities, we obtain:

$$\begin{aligned} \mathbb{E}_{\varepsilon \sim \mathcal{N}(0, I_n)}[e^{tZ_{c,x^*}^2}] &\geq (1 - \delta) \int_{\mathcal{A}} e^{t\|\varepsilon\|_2^2} f(\varepsilon) \\ &= (1 - \delta) \left(\frac{1}{2}\right)^n \frac{1}{n} \frac{1}{(1-2t)^{n/2}} \end{aligned}$$

$$> M_{\Gamma\left(\frac{n-\epsilon}{2}, 2\right)}(t) = \frac{1}{(1-2t)^{\frac{n-\epsilon}{2}}}.$$

The last inequality holds when

$$t \in \left(\frac{1}{2} - \frac{1}{2} \left(\frac{n2^n}{1-\delta} \right)^{-2/\epsilon}, \frac{1}{2} \right).$$

Since δ can be taken arbitrarily small by increasing M , this proves that for every ϵ there exists a t such that the Gamma distribution of $\text{rank}(K) - \epsilon$ fails to dominate $Z_{c,x}^2$, and therefore $\text{rank}(K)$ is the minimal parameter for which stochastic domination can hold in general. \square

Corollary 5.3. *The non-negatively constrained problem $K = I_2$, $h = (1, -1)$, $x^* = (0, M)$ is a two-dimensional counterexample to the Burrus conjecture $Z_{c,x}^2 \preceq \chi_1^2$ for $M > 0$ large enough.*

Proof. The proof follows by taking $n = p = 2$ in the collection of counterexamples in Lemma 5.2. \square

To the best of our knowledge, this is the first known counterexample to the Burrus conjecture in dimension smaller than 3.

5.2.2 Test statistic with unconstrained second term λ_u^2

Defining $\mathcal{T}_{u,x}^2(\epsilon) := \lambda_u^2(Hx, Kx + \epsilon)$ with $\lambda_u^2(\mu, y) = \min_{Hx=\mu, Ax \leq b} \|Kx - y\|_2^2 - \min_x \|Kx - y\|_2^2$, we obtain after a similar computation as the previous subsection:

$$\mathcal{T}_{u,x}^2(\epsilon) = \min_{Hz=0, Az \leq b-Ax} \|Kz - \epsilon\|_2^2 - \min_z \|Kz - \epsilon\|_2^2.$$

Let $Q_u^2(x)$ be the $1 - \alpha$ quantile of $\mathcal{T}_{u,x}^2(\epsilon)$ for $\epsilon \sim \mathcal{N}(0, I)$. We summarize some relevant properties of this function below.

- *Convexity.* We have the following important convexity result:

Theorem 5.4 (Convexity of the quantile function $Q_u^2(x)$). *For any fixed $0 < \alpha < 1$, the quantile function $Q_u^2(x)$ at fixed level $1 - \alpha$ is a convex function of x , both for linear and cone constraints.*

Proof. It is enough to show that $(x, \epsilon) \mapsto \mathcal{T}_{u,x}^2(\epsilon)$ is jointly convex in (x, ϵ) . If that is the case, [8, Lemma 2.28] directly shows that the quantile function is convex. An alternative reasoning comes from Prékopa's theorem [18, 19] which implies that log-concavity is preserved by marginalization: joint convexity implies $\mathcal{S}_t := \{(x, \epsilon) : \mathcal{T}_{u,x}^2(\epsilon) \leq t\}$ is a convex set (therefore its indicator function is log-concave), and if $f(\epsilon)$ is any log-concave density, $h(x, t, \epsilon) := \mathbf{1}_{\mathcal{S}_t}(x, \epsilon) f(\epsilon)$ is product of log-concave functions and hence log-concave. Then, by the Prékopa's theorem $\varphi(x, t) := \int h(x, t, \epsilon) d\epsilon = \mathbb{P}(\mathcal{T}_{u,x}^2(\epsilon) \leq t)$ is log-concave in (x, t) . Therefore, the superlevel sets $\{(x, t) : \varphi(x, t) \geq 1 - \alpha\}$ are convex, which correspond exactly to the epigraph of the quantile function, making Q_u^2 convex.

We now show that $(x, \epsilon) \mapsto \mathcal{T}_{u,x}^2(\epsilon)$ is jointly convex in (x, ϵ) , finishing the proof. Let $\hat{z}(\epsilon) = K^\dagger \epsilon \in \arg\min_z \|Kz - \epsilon\|_2^2$. Then, for any z , by the Pythagorean theorem and orthogonality of $\epsilon - K\hat{z}$ to the range of K , we have:

$$\|Kz - \epsilon\|_2^2 = \|Kz - K\hat{z}\|_2^2 + \|K\hat{z} - \epsilon\|_2^2,$$

which we used earlier in (3.6) to relate the confidence regions around the least square estimator with inverted regions of λ_u^2 . Therefore, the difference cancels the constant $\|K\hat{z} - \epsilon\|_2^2$, and we have:

$$\mathcal{T}_{u,x}^2(\epsilon) = \min_{Hz=0, Az \leq b-Ax} \|K(z - \hat{z}(\epsilon))\|_2^2.$$

Write $\mathcal{Z}(x) := \{z \mid Hz = 0, Az \leq b - Ax\}$ and $g(u, x) := \min_{z \in \mathcal{Z}(x)} \|K(z - u)\|_2^2$. The map $(z, u) \mapsto \|K(z - u)\|_2^2$ is jointly convex, and the feasible set $\mathcal{Z}(x)$ depends affinely on (z, x) . By standard results on convex analysis, $g(u, x)$ is jointly convex in (u, x) . Since $\hat{z}(\varepsilon)$ is affine in ε , the composition $\mathcal{T}_{u,x}^2(\varepsilon) = g(\hat{z}(\varepsilon), x)$ is jointly convex in (x, ε) . The same argument applies for $\mathcal{Z}(x) := \{z \mid Hz = 0, z + x \in \mathcal{X}\}$ where \mathcal{X} is any convex set, so in particular, the result applies for convex constraints as well. \square

Lemma 5.5. *Let y be in the recession cone of the constraint set ($Ay \leq 0$ for linear constraints $Ax^* \leq b$ and all of the convex cone \mathcal{C} for $x^* \in \mathcal{X}$), and let x be any point. Then, $Q_u^2(x+y) \leq Q_u^2(x)$.*

Proof. For the linearly constrained case, we have, for every fixed ε :

$$\min_{Hz=0, Az \leq b-Ax-Ay} \|Kz - \varepsilon\|_2^2 \leq \min_{Hz=0, Az \leq b-Ax} \|Kz - \varepsilon\|_2^2, \quad (5.9)$$

since the feasible set enlarges when $-Ay \geq 0$ (i.e., $Ay \leq 0$). It therefore directly holds from the definition of $\mathcal{T}_{u,x}^2(\varepsilon)$ that $\mathcal{T}_{u,x+y}^2(\varepsilon) \leq \mathcal{T}_{u,x}^2(\varepsilon)$ for all ε . The quantile inequality then follows.

Similarly, for the cone constraint case, we have:

$$\min_{Hz=0, (z+x+y) \in \mathcal{C}} \|Kz - \varepsilon\|_2^2 \leq \min_{Hz=0, (z+x) \in \mathcal{C}} \|Kz - \varepsilon\|_2^2. \quad (5.10)$$

Because $\mathcal{C} - x \subseteq \mathcal{C} - x - y$ for any $x, y \in \mathcal{C}$, the same conclusion follows. \square

- *Maximization.* Theorem 5.4, together with Lemma 5.5, have important consequences for locating where the quantile optimization problems (5.4) and (5.5) are maximized. We separate into the cases of linear and cone constraints.

Theorem 5.6 (Maximization of $Q_u^2(x)$ with linear constraints). *In the linearly constrained case $Ax^* \leq b$, consider*

$$\lambda_u^2(\mu, y) = \min_{Hx=\mu, Ax \leq b} \|Kx - y\|_2^2 - \min_x \|Kx - y\|_2^2$$

with its corresponding quantile $Q_u^2(x)$. Let \mathcal{P} be a polyhedron of the form $\{x : Ax \leq b, Gx \leq d\}$, where A and b are the original constraints on x , and G, d are arbitrary. Let $\{p_i\}_{i=1}^m$ be the set of extreme points of \mathcal{P} . Then, $\sup_{x \in \mathcal{C}} Q_u^2(x)$ is achieved in \mathcal{P} , i.e.,

$$\sup_{x \in \mathcal{P}} Q_u^2(x) = \max_{i=1:m} Q_u^2(p_i). \quad (5.11)$$

Proof. Let $x \in \mathcal{P}$ and use the Weyl–Minkowski theorem to express

$$x = \sum_{i=1}^m \lambda_i p_i + \sum_{j=1}^{\ell} \mu_j r_j, \quad \lambda_i \geq 0, \sum_{i=1}^m \lambda_i = 1, \quad \mu_j \geq 0, \quad Ar_j \leq 0, \quad Gr_j \leq 0.$$

Since $A \sum_{j=1}^{\ell} \mu_j r_j \leq 0$, using Lemma 5.5 and then Theorem 5.4, we have that

$$\begin{aligned} Q_u^2(x) &= Q_u^2\left(\sum_{i=1}^m \lambda_i p_i + \sum_{j=1}^{\ell} \mu_j r_j\right) \\ &\leq Q_u^2\left(\sum_{i=1}^m \lambda_i p_i\right) \end{aligned}$$

$$\begin{aligned}
&\leq \sum_{i=1}^m \lambda_i Q_u^2(p_i) \\
&\leq \sum_{i=1}^m \lambda_i \max_{j=1:m} Q_u^2(p_j) = \max_{j=1:m} Q_u^2(p_j).
\end{aligned}$$

Since $x \in \mathcal{P}$ was arbitrary, this concludes the proof. \square

Thanks to this result, for linear constraints, we can obtain the optimal thresholds as:

$$\begin{aligned}
d^*(\mu) &= \sup_{Hx=\mu, Ax \leq b} Q_u^2(x) = \max_{p_i \in E(Ax \leq b, Hx=\mu)} Q_u^2(p_i), \\
D^* &= \sup_{Ax \leq b} Q_u^2(x) = \max_{q_i \in E(Ax \leq b)} Q_u^2(q_i).
\end{aligned}$$

Theorem 5.7 (Global maximization of $Q_u^2(x)$ with cone constraints). *In the cone constrained case $x^* \in \mathcal{C}$, consider*

$$\lambda_u^2(\mu, y) = \min_{Hx=\mu, x \in \mathcal{C}} \|Kx - y\|_2^2 - \min_x \|Kx - y\|_2^2$$

with its corresponding quantile $Q_u^2(x)$. Then, $\sup_{x \in \mathcal{C}} Q_u^2(x)$ is achieved at $x = 0$, i.e.,

$$\sup_{x \in \mathcal{C}} Q_u^2(x) = Q_u^2(0). \quad (5.12)$$

Furthermore, letting $\Phi_\mu = \{x \in \mathcal{C} | Hx = \mu\}$, and $\text{rb}(\Phi_\mu)$ denote its relative boundary, we have:

$$\sup_{x \in \Phi_\mu} Q(x) = \sup_{x \in \text{rb}(\Phi_\mu)} Q(x). \quad (5.13)$$

Proof. The global maximization result (5.12) comes directly from (5.10), since the recession cone of \mathcal{C} is itself and Q_u^2 satisfies $Q_u^2(x+y) \leq Q_u^2(x)$ for all $x, y \in \mathcal{C}$. Taking $x = 0$ and $y \in \mathcal{C}$ arbitrary yields the desired result. For the sliced result, if $\Phi_\mu = \text{rb}(\Phi_\mu)$, there is nothing to prove. Otherwise, take any $x \in \Phi_\mu \setminus \text{rb}(\Phi_\mu)$, and let us show that $Q_u^2(x) \leq \sup_{z \in \text{rb}(\Phi_\mu)} Q_u^2(z)$.

Consider a nonzero direction $u \in \text{Ker}(K)$ and the maximal interval $I := \{t \in \mathbb{R} : x + tu \in \Phi_\mu\}$. There are two cases:

- If $I = (t_-, t_+)$ with finite endpoints, then the endpoints $a := x + t_-u$ and $b := x + t_+u$ belong to $\text{rb}(\Phi_\mu)$, and by convexity of Q along the line $x + tu$, $Q(x) \leq \max\{Q(a), Q(b)\} \leq \sup_{z \in \text{rb}(\Phi_\mu)} Q(z)$.
- If one endpoint is $+\infty$ or $-\infty$ (suppose $t_+ = +\infty$ without loss of generality), then $u \in \text{rec}(\Phi_\mu) = \mathcal{C} \cap \text{Ker}(K)$. By the cone monotonicity $Q(x + tu) \leq Q(x)$ for all $t \geq 0$, hence again $Q(x) \leq \sup_{z \in \text{rb}(\Phi_\mu)} Q(z)$ (the other endpoint of I is finite and lies on $\text{rb}(\Phi_\mu)$, or the inequality is trivial if both directions are recession. Since $x \in \Phi_\mu \setminus \text{rb}(\Phi_\mu)$ was arbitrary, $\sup_{z \in \Phi_\mu} Q_u^2(z) \leq \sup_{z \in \text{rb}(\Phi_\mu)} Q_u^2(z)$.

The reverse inequality is immediate because $\text{rb}(\Phi_\mu) \subseteq \Phi_\mu$. This concludes the proof. \square

Corollary 5.8 (Non-negative constraints). *Under non-negativity constraints $x^* \geq 0$, we have*

$$\sup_{x \geq 0} Q_u^2(x) = Q_u^2(0), \quad (5.14)$$

and if $k = 1$ and the functional of interest is $h^\top x^*$,

$$\operatorname{argmax}_{x \geq 0, h^\top x = \mu} Q_u^2(x) \in \{(\mu/h_i)e_i, i : \mu/h_i > 0\} \quad (5.15)$$

where e_i are the standard basis vectors.

Proof. Directly from either Theorem 5.6 or Theorem 5.7, as the non-negative constraints are both linear constraints and cone constraints. \square

A similar result holds with more functionals, but the number of extreme points to check grows as at most $\binom{p}{k}$ with k functionals, which can become unfeasible to manually search even with $k = 2$ if p is large.

Corollary 5.9 (Optimal constant for non-negative λ_u^2 constructions in Table 1). *In the non-negatively constrained case, consider the μ -description confidence region in Table 1:*

$$\{\mu \in \mathbb{R}^k : \min_{\substack{Hx = \mu, \\ x \geq 0}} \|Kx - y\|_2^2 - \min_{x \in \mathbb{R}^p} \|Kx - y\|_2^2 \leq \delta_u^2\}, \quad (5.16)$$

and its related x -description bounding box:

$$\begin{aligned} & \min_x / \max_x \quad h_i^\top x \\ & \prod_{i=1}^k \quad \text{subject to} \quad \|Kx - y\|_2^2 \leq \min_{x' \in \mathbb{R}^p} \|Kx' - y\|_2^2 + \delta_u^2 \\ & \quad \quad \quad x \geq 0. \end{aligned}$$

Then, the $1 - \alpha$ quantile of

$$\min_{Hz=0, z \geq 0} \|Kz - \varepsilon\|_2^2 - \min_z \|Kz - \varepsilon\|_2^2 = \min_{Hz=0, z \geq 0} \|K(z - \hat{z}(\varepsilon))\|_2^2, \quad \hat{z}(\varepsilon) \in \operatorname{argmin}_z \|Kz - \varepsilon\|_2^2 \quad (5.17)$$

with $\varepsilon \sim \mathcal{N}(0, I)$, is the optimal (smallest) δ_u^2 satisfying frequentist coverage for all $x \geq 0$ in (5.16).

Proof. By using Corollary 5.8 and the description of optimal constants given by (4.18) and (4.19). \square

In particular, for $k = 1$, consider the construction (3.7) or (3.8), which is of the form

$$\begin{aligned} & \min_x / \max_x \quad h^\top x \\ & \text{subject to} \quad \|Kx - y\|_2^2 \leq \min_{x'} \|Kx' - y\|_2^2 + \delta \\ & \quad \quad \quad x \geq 0 \end{aligned} \quad (5.18)$$

with $\delta = Q_{\chi_{r,1-\alpha}^2}$. Then, the optimal constant δ such that (5.18) has $1 - \alpha$ frequentist coverage for every $x \geq 0$ is the $1 - \alpha$ quantile of $Z_u^2(x = 0)$, i.e., the quantile of

$$\min_{h^\top z=0, z \geq 0} \|Kz - \varepsilon\|_2^2 - \min_z \|Kz - \varepsilon\|_2^2 = \min_{h^\top z=0, z \geq 0} \|K(z - \hat{z}(\varepsilon))\|_2^2, \quad \hat{z}(\varepsilon) \in \operatorname{argmin}_z \|Kz - \varepsilon\|_2^2 \quad (5.19)$$

with $\varepsilon \sim \mathcal{N}(0, I)$, which is always equal or smaller than $Q_{\chi_{r,1-\alpha}^2}$. In Section 5.2.4, we identify this random variable to follow a chi-bar-squared distribution. An improvement is possible by allowing δ to depend on x , at the expense of destroying the convexity of the optimization problem, by taking $\delta(x) := \max_{z \in \{(h^\top x/h_i)e_i, i: h^\top x/h_i > 0\}} Q_u^2(z)$.

- *Bounds.* We note that Lemma 5.2, together with (5.7), shows that there is no better uniform (in the sense of holding for all K, H, x) stochastic dominance bound than $Z_{u,x}^2 \preceq \chi_{\operatorname{rank}(K)}^2$.

5.2.3 One term test statistic λ^1

Our final test statistic under consideration is $\lambda^1(\mu, y) = \min_{Hx=\mu, Ax \leq b} \|Kx - y\|_2^2$, which we will show has many similar properties to λ_u^2 . It easily follows from the same computations in the previous subsections that

$$\mathcal{T}_x^1(\varepsilon) = \min_{Hz=0, Az \leq b-Ax} \|Kz - \varepsilon\|_2^2.$$

Let $Q^1(x)$ as the $1-\alpha$ quantile of $\mathcal{T}_x^1(\varepsilon)$ for $\varepsilon \sim \mathcal{N}(0, I)$. We summarize below that similar properties of this function as with Q_u^2 .

- *Convexity.* We have the following convexity result, similar to Theorem 5.4.

Theorem 5.10. *For any fixed $0 < \alpha < 1$, the quantile function $Q^1(x)$ is a convex function of x .*

Proof. Using the same proof argument as in Theorem 5.4, it suffices to show that $(x, \varepsilon) \mapsto \mathcal{T}_x^1(\varepsilon)$ is jointly convex in (x, ε) . This is proven similarly to \mathcal{T}_u^2 . Define $\mathcal{Z}(x) := \{z \mid Hz = 0, Az \leq b - Ax\}$ and $\mathcal{T}_x^1(\varepsilon) := \min_{z \in \mathcal{Z}(x)} \|Kz - \varepsilon\|_2^2$. The map $(z, \varepsilon) \mapsto \|Kz - \varepsilon\|_2^2$ is jointly convex, and the feasible set $\mathcal{Z}(x)$ depends affinely on (z, x) , so the result follows. \square

Lemma 5.11. *Let y be in the recession cone of the constraint set ($Ay \leq 0$ for linear constraints $Ax^* \leq b$ and all of the convex cone \mathcal{C} for $x^* \in \mathcal{X}$), and let x be any point. Then, $Q^1(x+y) \leq Q^1(x)$.*

Proof. Follows directly from the inequalities (5.9) and (5.10) established in the proof of Lemma 5.5. \square

- *Maximization.* Theorem 5.10 along with Lemma 5.11 lead to the following consequence regarding the locations where the underlying quantile optimization problems are maximized.

Theorem 5.12. *Theorems 5.6 and 5.7 apply verbatim for the maximizers of Q^1 in the linearly constrained and cone constrained problems, respectively.*

Proof. The proof of Theorems 5.6 and 5.7 only use convexity of $Q_u^2(x)$ (Theorem 5.4) and the recession property in Lemma 5.5, properties which also apply to Q^1 (Theorem 5.10 and Lemma 5.11, respectively). \square

Thus, we can directly invoke the equivalent of Corollary 5.9 for constructions based on λ^1 .

Corollary 5.13 (Optimal constant for non-negative λ^1 constructions in Table 1). *In the non-negatively constrained case, consider the μ -description confidence region in Table 1:*

$$\{\mu \in \mathbb{R}^k : \min_{\substack{Hx=\mu, \\ x \geq 0}} \|Kx - y\|_2^2 \leq \delta^1\}, \quad (5.20)$$

and its related x -description bounding box:

$$\prod_{i=1}^k \min_x / \max_x h_i^\top x \quad \text{subject to} \quad \|Kx - y\|_2^2 \leq \delta^1, \quad x \geq 0.$$

Then, the $1 - \alpha$ quantile of $\min_{Hz=0, z \geq 0} \|Kz - \varepsilon\|_2^2$ with $\varepsilon \sim \mathcal{N}(0, I)$, is the optimal (smallest) δ^1 satisfying frequentist coverage for all $x \geq 0$ in (5.20).

Proof. By using Theorem 5.12, we can show that Corollary 5.8 applies for $Q^1(x)$, then the description of optimal constants given by (4.18) and (4.19) gives the desired result. \square

In particular, for $k = 1$, consider the construction (3.12) which is of the form

$$\begin{aligned} & \min_x / \max_x \quad h^\top x \\ & \text{subject to} \quad \|Kx - y\|_2^2 \leq \delta \\ & \quad \quad \quad x \geq 0 \end{aligned} \tag{5.21}$$

with $\delta = Q_{\chi_{n,1-\alpha}^2}$. Then, the optimal constant δ such that (5.21) has $1 - \alpha$ frequentist coverage for every $x \geq 0$ is the $1 - \alpha$ quantile of $Z^1(x = 0)$, i.e., the quantile of $\min_{h^\top z=0, z \geq 0} \|Kz - \varepsilon\|_2^2$ with $\varepsilon \sim \mathcal{N}(0, I)$, which is always equal or smaller than $Q_{\chi_{n,1-\alpha}^2}$.

5.2.4 Comparison between λ_u^2 and λ^1 , and their distributions at the origin

As derived in the previous subsections, both λ_u^2 and λ^1 have identical behavior with respect to the locations of maximizers of their induced quantile functions. We now give a preliminary comparison to guide the choice between them in practice.

When K is surjective (full row rank), we have $\min_x \|Kx - y\|_2^2 = 0$ for all y , so $\lambda_u^2 = \lambda^1$. We therefore focus on the non-surjective case, under cone constraints, and for global quantile optimization. A comparison under general linear constraints $Ax^* \leq b$ and/or in the sliced setting remains open.

For a fixed y , the resulting confidence regions are:

$$\begin{aligned} \mathcal{R}^1 &= \{\mu : \min_{Hx=\mu, x \in \mathcal{C}} \|Kx - y\|_2^2 \leq D^1\}, \\ \mathcal{R}_u^2 &= \{\mu : \min_{Hx=\mu, x \in \mathcal{C}} \|Kx - y\|_2^2 \leq \min_x \|Kx - y\|_2^2 + D_u^2\}, \end{aligned}$$

for given constants D^1 and D_u^2 . Coverage holds whenever $D^1 \geq \max_{x \in \mathcal{C}} Q^1(x) = Q^1(0)$ and $D_u^2 \geq \max_{x \in \mathcal{C}} Q_u^2(x) = Q_u^2(0)$. For simplicity, we begin analyzing not the optimal constants given by their quantiles at $x = 0$, but their upper bounds independent of K and H given by the stochastic dominance chain (5.7), i.e., $D^1 = Q_{\chi_{n,1-\alpha}^2}$, and $D_u^2 = Q_{\chi_{r,1-\alpha}^2}$.

Observe that the distribution of $\min_x \|Kx - y\|_2^2$ when $y \sim \mathcal{N}(Kx^*, I)$ is the same for all $x^* \in \mathcal{C}$. Indeed, writing $y = Kx^* + \varepsilon$, we can translate the optimization variables without changing the optimal value. We therefore may take $x^* = 0$, and study the distribution of $\min_x \|Kx - \varepsilon\|_2^2$ with $\varepsilon \sim \mathcal{N}(0, I)$. A standard argument shows that this distribution is χ^2 distribution with $n - r$ degrees of freedom. Thus, with $\xi_{n,r} \sim \chi_{n-r}^2$, we have $\mathcal{R}^1 \subseteq \mathcal{R}_u^2$ if and only if $Q_{\chi_{n,1-\alpha}^2} \leq \xi_{n,r} + Q_{\chi_{r,1-\alpha}^2}$.

Define the quantile gap $\Delta(n, r, \alpha) = Q_{\chi_{n,1-\alpha}^2} - Q_{\chi_{r,1-\alpha}^2}$. With $\xi_{n,r} \sim \chi_{n-r}^2$, we have $\mathcal{R}^1 \subseteq \mathcal{R}_u^2$. This is equivalent to $\xi_{n,r} \geq \Delta(n, r, \alpha)$, which happens with probability $p(n, r, \alpha)$. Using the Cornish-Fisher expansion [5, 6], we can approximate this quantity as $Q_{\chi_{k,1-\alpha}^2} = k + z_{1-\alpha} \sqrt{2k} + O(1)$ (with $z_{1-\alpha}$ the $1 - \alpha$ normal quantile). This yields the approximation:

$$p(n, r, \alpha) \approx 1 - \Phi(z_{1-\alpha} c(\rho)), \tag{5.22}$$

valid up to $O(n^{-1/2})$ corrections, where $\rho := r/n$ and $c(\rho) := \frac{1-\sqrt{\rho}}{\sqrt{1-\rho}}$. From (5.22), we read off three asymptotic regimes for $p(n, r, \alpha)$: (i) if r is fixed and $n \rightarrow \infty$, then $p(n, r, \alpha) \rightarrow \alpha$, (ii) if $r/n \rightarrow \rho \in (0, 1)$, then $p(n, r, \alpha) \rightarrow 1 - \Phi(z_{1-\alpha} c(\rho))$, and (iii) if the difference $s = n - r$ remains fixed while $n, r \rightarrow \infty$, then $p(n, r, \alpha) \rightarrow \frac{1}{2}$. In all cases, for small α , we typically have $p \leq 1/2$, so the random threshold \mathcal{R}_u^2 , $Q_{\chi_{r,1-\alpha}^2} + \xi_{n,r}$ is often smaller than the fixed threshold $Q_{\chi_{n,1-\alpha}^2}$ in \mathcal{R}^1 , at the cost of a worse worst-case over the data. This behavior is illustrated in Figure 5.1: for small

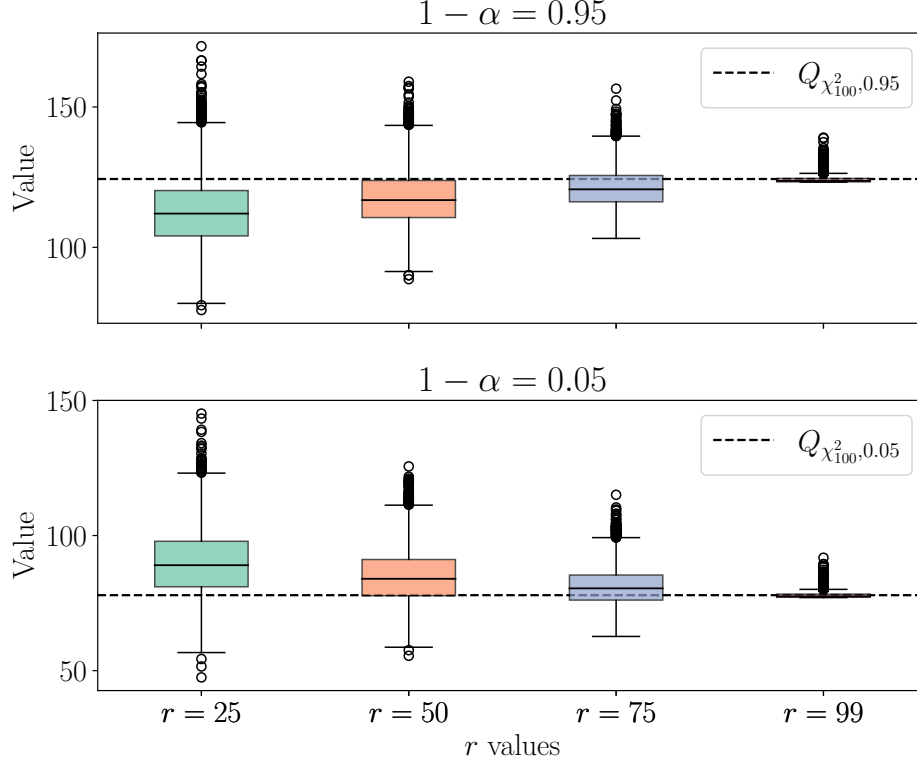


Figure 5.1: Comparison between the fixed chi-square quantile $Q_{\chi^2_n, 1-\alpha}$ (dashed line) and the distribution $Q_{\chi^2_n, 1-\alpha} + \xi_{n,r}$ (boxplots), for $n = 100$ and $r \in \{25, 50, 75, 99\}$. The top panel corresponds to $1 - \alpha = 0.05$, and the bottom panel to $1 - \alpha = 0.25$.

α , most of the mass of the distribution lies below the fixed constant, and for larger α , this relation flips.

We next show that, at the origin $x = 0$ (where the quantiles are maximized over closed convex cones), the distributions of $Z_{u,x}^2$ and Z_x^1 to chi-bar-squared distributions, mixtures of χ^2 distributions with different degrees of freedom. Importantly, their weights are related, so some conclusions of the previous subsection carry over to this case, as the distribution at the origin for λ^1 is a mixture of χ^2 distributions whose degrees of freedom (but not weights) get shifted. We begin with a simple two-dimensional example to illustrate:

Example 5.14 (Two-dimensional example). For $n = 2$, $p = 1$, $K = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, and $F = \mathbb{R}_+$, the cone $S = \{(t, 0) : t \geq 0\}$ gives

$$Z_{u,0}^2 \sim \frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2, \quad Z_0^1 \sim \frac{1}{2}\chi_1^2 + \frac{1}{2}\chi_2^2.$$

Therefore, when using the bounds, the choice would be between the quantile of a χ_2^2 and the quantile of χ_1^2 plus a draw of a χ_1^2 , and with the exact distributions, the fixed threshold of λ^1 is the quantile of $\frac{1}{2}\chi_1^2 + \frac{1}{2}\chi_2^2$, and the random threshold of λ_u^2 is the quantile of $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ plus a draw of a χ_1^2 . We prove below a general statement about the laws of Z_0^1 and $Z_{u,0}^2$. While we expect similar conclusions to the geometry-independent bounds to apply, we defer their theoretical analysis to future work.

Lemma 5.15 (Chi-bar-squared laws for Z_0^1 and $Z_{u,0}^2$). Let $L := \{x \in \mathbb{R}^p : Hx = 0\}$ and $F := \mathcal{C} \cap L$ be a closed convex cone. Let $R := \text{range}(K) \subseteq \mathbb{R}^n$ with $r := \dim(R)$, and define the image cone

$S := K(F) \subseteq R$. Consider

$$\mathcal{T}_0^1(\varepsilon) := \min_{x \in F} \|Kx - \varepsilon\|_2^2, \quad \mathcal{T}_{u,0}^2(\varepsilon) := \min_{x \in F} \|Kx - \varepsilon\|_2^2 - \min_{x \in \mathbb{R}^p} \|Kx - \varepsilon\|_2^2,$$

and let $Z_0^1, Z_{u,0}^2$ be the corresponding random variables for $\varepsilon \sim \mathcal{N}(0, I_n)$. Then, we have:

(i)

$$\mathcal{T}_0^1 = \text{dist}^2(\varepsilon, S), \quad Z_0^1 \stackrel{d}{=} \sum_{j=0}^n v_{n-j}^{(\mathbb{R}^n)}(S) \chi_j^2,$$

where $\{v_k^{(\mathbb{R}^n)}(S)\}_{k=0}^n$ are the conic intrinsic volumes of S as a subset of \mathbb{R}^n , which are constants depending on \mathcal{C}, H , and K .

(ii) Writing $\varepsilon_R := P_R \varepsilon \sim \mathcal{N}(0, I_r)$, we have:

$$\mathcal{T}_{u,0}^2 = \text{dist}^2(\varepsilon_R, S), \quad Z_{u,0}^2 \stackrel{d}{=} \sum_{j=0}^r v_{r-j}^{(R)}(S) \chi_j^2,$$

where $\{v_k^{(R)}(S)\}_{k=0}^r$ are the intrinsic volumes of S computed in the ambient space R . Moreover, the two sets of intrinsic volumes satisfy:

$$v_k^{(\mathbb{R}^n)}(S) = v_k^{(R)}(S) \quad \text{for } k = 0, \dots, r, \quad v_k^{(\mathbb{R}^n)}(S) = 0 \quad \text{for } k > r.$$

Proof. Because $\{Kx : x \in F\} = S$ and $\text{range}(K) = R$, we have:

$$\mathcal{T}_0^1(\varepsilon) = \min_{x \in F} \|Kx - \varepsilon\|_2^2 = \text{dist}^2(\varepsilon, S), \quad \min_{x \in \mathbb{R}^p} \|Kx - \varepsilon\|_2^2 = \text{dist}^2(\varepsilon, R) = \|(I - P_R)\varepsilon\|_2^2. \quad (5.23)$$

(i) Since S is a closed convex cone in \mathbb{R}^n , Moreau's decomposition gives $\varepsilon = \Pi_S \varepsilon + \Pi_{S^\circ} \varepsilon$ with orthogonal parts, hence $\text{dist}^2(\varepsilon, S) = \|\Pi_{S^\circ} \varepsilon\|_2^2$. For Gaussian ε , the distribution of $\|\Pi_{S^\circ} \varepsilon\|_2^2$ is a chi-bar-squared mixture:

$$\text{dist}^2(\varepsilon, S) \sim \sum_{j=0}^n v_{n-j}^{(\mathbb{R}^n)}(S) \chi_j^2, \quad (5.24)$$

where the weights are the conic intrinsic volumes of S in \mathbb{R}^n . (For more details about the chi-bar-squared distribution, see [26].)

(ii) Since $S \subseteq R$, subtracting the identities in (5.23) gives

$$\mathcal{T}_{u,0}^2(\varepsilon) = \text{dist}^2(\varepsilon, S) - \text{dist}^2(\varepsilon, R) = \min_{s \in S} \|\varepsilon - s\|_2^2 - \|\varepsilon_{R^\perp}\|_2^2 = \min_{s \in S} \|\varepsilon_R - s\|_2^2 = \text{dist}^2(\varepsilon_R, S), \quad (5.25)$$

where we have written the orthogonal decomposition $\mathbb{R}^n = R \oplus R^\perp$ and $\varepsilon_R \sim \mathcal{N}(0, I_r)$ and $\varepsilon_{R^\perp} \sim \mathcal{N}(0, I_{n-r})$. Viewing S as a cone in $R \simeq \mathbb{R}^r$, the same reasoning as in (i) yields

$$\text{dist}^2(\varepsilon_R, S) \sim \sum_{j=0}^r v_{r-j}^{(R)}(S) \chi_j^2.$$

It remains to relate $v^{(\mathbb{R}^n)}$ and $v^{(R)}$. From (5.25),

$$\text{dist}^2(\varepsilon, S) = \min_{s \in S} \|\varepsilon_R - s\|_2^2 + \|\varepsilon_{R^\perp}\|_2^2.$$

Table 3: Location of the maximizers of quantiles for different test statistic and constraint sets

Statistic	Global optimization (D^*)	Sliced optimization ($d^*(\mu)$)
λ_c^2 (all cases)	General results not available	General results not available
λ^1, λ_u^2 ($x \geq 0$)	$x = 0$	Extreme point of $x \geq 0, Hx = \mu$ (at most $\binom{p}{k}$)
λ^1, λ_u^2 ($Ax \leq b$)	Extreme point of $Ax \leq b$	Extreme point of $Ax \leq b, Hx = \mu$
λ^1, λ_u^2 ($x \in \mathcal{C}$)	$x = 0$	Relative boundary of $x \in \mathcal{C}, Hx = \mu$

Table 4: Comparative summary of the test statistics.

Property	λ_c^2	λ_u^2	λ^1
Quantile convexity Q_x	No (in general)	Yes (Gaussian noise) [Theorem 5.4]	Yes (log-concave noise) [Theorem 5.10]
Acceptance region emptiness	Always non-empty	May be empty for some y	May be empty for some y
Calibration ease	Typically challenging	Typically easier	Typically easier
Calibration status (non-negative constraints)	Optimal D^* unknown	Optimal D^* known [Corollary 5.9]	Optimal D^* known [Corollary 5.13]
Average vs. worst-case length (small α)		Often shorter on average; worse worst-case than λ^1 (random radius)	Often longer on average; better worst-case than λ_u^2 (fixed radius)

When ε is randomized, $\|\varepsilon_{R^\perp}\|_2^2 \sim \chi_{n-r}^2$, $\text{dist}^2(\varepsilon, S) \sim \sum_{j=0}^r v_{r-j}^{(R)}(S) \chi_j^2$. Using the convolution identity $\chi_j^2 + \chi_{n-r}^2 \stackrel{d}{=} \chi_{j+(n-r)}^2$, we obtain:

$$\text{dist}^2(\varepsilon, S) \sim \sum_{j=0}^r v_{r-j}^{(R)}(S) \chi_{j+(n-r)}^2.$$

Re-indexing with $\ell = j + (n - r)$ (so $\ell = n - r, \dots, n$) and comparing with (5.24) gives precisely the ambient- \mathbb{R}^n mixture:

$$\text{dist}^2(\varepsilon, S) \stackrel{d}{=} \sum_{\ell=0}^n v_{n-\ell}^{(\mathbb{R}^n)}(S) \chi_\ell^2, \quad \text{with} \quad v_{n-\ell}^{(\mathbb{R}^n)}(S) = \begin{cases} v_{r-(\ell-(n-r))}^{(R)}(S) = v_{n-\ell}^{(R)}(S), & \ell \geq n - r, \\ 0, & \ell < n - r. \end{cases}$$

Equivalently,

$$v_k^{(\mathbb{R}^n)}(S) = v_k^{(R)}(S) \text{ for } k = 0, \dots, r, \quad v_k^{(\mathbb{R}^n)}(S) = 0 \text{ for } k > r.$$

This establishes the claimed relationship. \square

We summarize the results of this section regarding the test statistics λ^1 , λ_u^2 , and λ_c^2 in Tables 3 and 4. Table 3 summarizes the results of quantile maximizations, including Theorem 5.6, Theorem 5.7, and Theorem 5.12. Table 4 provides a summary of the advantages and disadvantages of using each test statistic.

6 Practical methods for high-dimensional problems

This section uses the results developed in Section 5 to propose two approaches that scale to high-dimensional problems: (i) reductions for cases in which K has full column rank³, and (ii) a split technique that separates a full column rank part and a null space part for cases in which K is rank-deficient.

6.1 Generalizing and improving TFM reductions

A TFM reduction (which we name after Tenorio, Fleck, and Moses, who first introduced them in [32]) maps a broad class of problems to a single canonical form. In their original formulation, the family

$$y = Kx^* + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, I), \quad x \geq 0, \quad K \in \mathbb{R}^{n \times p}, \quad h \in \mathbb{R}^p,$$

with arbitrary K and a scalar functional $h^\top x^*$, is reduced to a two-dimensional problem:

$$y = x^* + \varepsilon \in \mathbb{R}^2, \quad \varepsilon \sim \mathcal{N}(0, \Sigma), \quad x^* \geq 0, \quad h = (1, -1),$$

which is independent of (K, h) except through the covariance Σ of the transformed noise. In this example, we allow Σ to depend on (K, h) , even though the original proposal never requires computing it explicitly. The problem might be further reduced to one dimension if $h \geq 0$ or $h \leq 0$, as we formalize below. Such reductions lead to computationally inexpensive algorithms that remain valid in high dimensions, at the cost of discarding some constraint information during the reduction. In what follows, we first recall the original reduction of [32], then use our results in Section 5 to improve it and to generalize it to multiple functionals and box constraints.

6.1.1 Review and improvements to TFM reduction for the single-functional case

Let K be a full column rank matrix, and consider $k = 1$. We consider two cases depending on h .

When $h \geq 0$ or $h \leq 0$. Without loss of generality, suppose $h \geq 0$. Given $y = Kx^* + \varepsilon$, the reduction in [32] constructs the one-dimensional vector:

$$\tilde{y} = h^\top (K^\top K)^{-1} K^\top y = h^\top x^* + \tilde{\varepsilon} =: \tilde{x}^* + \tilde{\varepsilon}, \quad (6.1)$$

with $\tilde{x}^* = h^\top x^*$ and $\tilde{\varepsilon} \sim \mathcal{N}(0, \sigma^2)$ with $\sigma^2 = h^\top (K^\top K)^{-1} h$. Observe that $\tilde{x}^* \geq 0$. Thus, this yields a one-dimensional constrained problem. The original paper then builds the interval:

$$\begin{aligned} & \min_{\tilde{x}} / \max_{\tilde{x}} \quad \tilde{x} \\ & \text{subject to} \quad \frac{1}{\sigma^2} (\tilde{x} - y)^2 \leq \frac{1}{\sigma^2} \min_{\tilde{x}' \geq 0} (\tilde{x}' - y)^2 + Q_{\chi_1^2, 1-\alpha} \\ & \quad \tilde{x} \geq 0, \end{aligned} \quad (6.2)$$

which is a valid $1 - \alpha$ confidence interval for the original $h^\top x^*$.

When h is general. Decompose $h = h^+ - h^-$ with $h^+, h^- \geq 0$.

Given $y = Kx^* + \varepsilon$, the reduction in [32] constructs the two-dimensional vector:

$$\tilde{y} = \begin{bmatrix} h^{+, \top} (K^\top K)^{-1} K^\top y \\ h^{-, \top} (K^\top K)^{-1} K^\top y \end{bmatrix} = \begin{bmatrix} h^{+, \top} x^* \\ h^{-, \top} x^* \end{bmatrix} + \tilde{\varepsilon} =: \tilde{x}^* + \tilde{\varepsilon} \quad (6.3)$$

³Depending on the specific alignment of H and K , reductions might also be used for some non-full column rank K , as we explain below

with $(1, -1)^\top \tilde{x}^* = h^\top x^*$ and $\tilde{\varepsilon} \sim \mathcal{N}(0, \Sigma)$ with

$$\Sigma = \begin{pmatrix} h^{+, \top} (K^\top K)^{-1} h^+ & h^{+, \top} (K^\top K)^{-1} h^- \\ h^{-, \top} (K^\top K)^{-1} h^+ & h^{-, \top} (K^\top K)^{-1} h^- \end{pmatrix}.$$

Observe that $\tilde{x}^* \geq 0$. The original paper then builds the interval:

$$\begin{aligned} & \min_{\tilde{x}} / \max_{\tilde{x}} \quad (1, -1)^\top \tilde{x} \\ & \text{subject to} \quad \|\tilde{x} - y\|_\Sigma^2 \leq \min_{\tilde{x}' \geq 0} \|\tilde{x}' - y\|_\Sigma^2 + Q_{\chi_2^2, 1-\alpha} \\ & \quad \tilde{x} \geq 0, \end{aligned} \tag{6.4}$$

where $\|v\|_\Sigma = v^\top \Sigma^{-1} v$. This is a valid $1 - \alpha$ confidence interval for the original $h^\top x^*$.

It is important to note that some information is lost in these reductions. While it is true that $x^* \geq 0$ implies $\tilde{x}^* \geq 0$, the reverse is not necessarily true. Therefore, we have weakened the constraint from $\{x : x \geq 0\}$ to $\{x : \min(h^{+, \top} x, h^{-, \top} x) \geq 0\}$. Consequently, we expect to pay a price in interval length and overcoverage in exchange for the reduction that simplified the computations.

Both (6.2) and (6.4) can be identified as λ_c^2 test inversion regions with thresholds taken from the χ^2 bounds in (5.7), which are not tight in general. Moreover, in (6.4), the term $\min_{x' \geq 0} \|x' - y\|_\Sigma^2$ is unnecessary for coverage, since $\{x : \|x - y\|_\Sigma^2 \leq Q_{\chi_2^2, 1-\alpha}\}$ is already a $1 - \alpha$ confidence set for x . Two types of refinements can be done, depending on whether we wish to keep the term $\min_{\tilde{x}' \geq 0} \|\tilde{x}' - y\|_\Sigma^2$, which we briefly discuss below.

If one keeps the extra term, we work with region coming from λ_c^2 inversion for analysis. This is possible for the one-dimensional $h \geq 0$ or $h \leq 0$ case, but becomes challenging in two dimensions due to the arbitrary covariance matrix Σ . Use of λ_c^2 for the one-dimensional problem $y = x^* + \varepsilon$ with $x^* \geq 0$ was studied in [2, Section 2.4] for unit variance noise, but the results apply easily after rescaling to the non-unit variance case. Define

$$q_\alpha(\mu) = Q_{\mu, 1-\alpha} = \begin{cases} Q_{\chi_1^2, 1-\alpha}, & 1 - \alpha < \chi_1^2(\mu^2), \\ r_{\mu, \alpha}, & 1 - \alpha \geq \chi_1^2(\mu^2), \end{cases}$$

where $r_{\mu, \alpha}$ is the unique non-negative root of the function $x \mapsto \Phi(\sqrt{x}) - \Phi\left(\frac{-\mu^2 - x}{2\mu}\right) - (1 - \alpha)$, which can be found using numerical methods. We can then build the improved region $\{x : \frac{1}{\sigma^2}(y - x)^2 \leq q_\alpha(x) + \frac{1}{\sigma^2} \min_{x' \geq 0} (y - x')^2\}$, which is strictly included in (6.2).

If one drops the extra term, then one obtains regions coming from λ^1 or λ_u^2 (here $\lambda^1 = \lambda_u^2$ as $K = I$) inversion for analysis. The constant $Q_{\chi_2^2, 1-\alpha}$ can then be improved according to Corollary 5.13, with the optimal global constant being the quantile at zero, yielding the following intervals for the one-dimensional and two-dimensional reductions, respectively:

$$\begin{aligned} & \min_x / \max_x \quad x \\ & \text{subject to} \quad \frac{1}{\sigma^2}(x - y)^2 \leq q_1 \\ & \quad x \geq 0, \end{aligned}$$

and

$$\begin{aligned} & \min_x / \max_x \quad (1, -1)^\top x \\ & \text{subject to} \quad \|x - y\|_\Sigma^2 \leq q_2 \\ & \quad x \geq 0, \end{aligned}$$

with

$$q_1 := \frac{1}{\sigma^2} Q_{1-\alpha} \left(\min_{z=0} (z - \varepsilon)^2 \right) = Q_{\chi_1^2, 1-\alpha},$$

and

$$q_2 := Q_{1-\alpha} \left(\min_{z_1=z_2, z_1 \geq 0, z_2 \geq 0} \|z - \varepsilon\|_\Sigma^2 \right),$$

with $\varepsilon \sim \mathcal{N}(0, \Sigma)$. In both of these settings, one can also show that the optimal sliced function equals the global constant $d^*(\mu) = D^*$ for all μ , so no further improvement is possible by slicing.

Extension to some rank-deficient cases. While we have assumed so far that K has full column rank so that the method is valid for *every* quantity of interest h , the only requirement on a given K and h for this construction to work is to be able to write $h = h_1 - h_2$, with $h_i \geq 0, h_i \in \text{row}(K), i = 1, 2$. Note that h_1 and h_2 can, but do not necessarily need to be, the positive and negative parts of the vector h and that $h \in \text{row}(K)$ is not sufficient for this condition to hold when K is not full column rank. The existence of the vectors h_1 and h_2 can be checked by solving the following linear feasibility program:

$$\begin{aligned} & \text{find } v_1, v_2 \in \mathbb{R}^n \\ & \text{s.t. } (v_1 - v_2)^\top K = h \\ & \quad v_1^\top K \geq 0 \\ & \quad v_2^\top K \geq 0. \end{aligned}$$

Whenever this program has a solution, one sets $h_i = v_i^\top K$ and replaces $(K^\top K)^{-1}$ by $(K^\top K)^\dagger$ throughout (so $h_i^\top (K^\top K)^\dagger K^\top y = h_i^\top K^\dagger y$).

6.1.2 Generalizing TFM reduction to the multiple-functional case

Consider now $H \in \mathbb{R}^{k \times p}$ and K having full column rank. Let us assume that out of these k functionals, $k_1 \leq k$ have positive and negative parts, and $k_2 = k - k_1$ are non-negative (or non-positive, but we consider them non-negative without loss of generality), and sort H so that the first k_1 are the ones with positive and negative parts. After observing $y = Kx + \varepsilon$, form the $2k_1 + k_2$ dimensional reduced observation:

$$\tilde{y} = \begin{bmatrix} h_1^{+, \top} (K^\top K)^{-1} K^\top y \\ h_1^{-, \top} (K^\top K)^{-1} K^\top y \\ \dots \\ h_{k_1}^{+, \top} (K^\top K)^{-1} K^\top y \\ h_{k_1}^{-, \top} (K^\top K)^{-1} K^\top y \\ \dots \\ h_{k_1+1}^\top (K^\top K)^{-1} K^\top y \\ \dots \\ h_k^\top (K^\top K)^{-1} K^\top y \end{bmatrix} = \begin{bmatrix} h_1^{+, \top} x^* \\ h_1^{-, \top} x^* \\ \dots \\ h_{k_1}^{+, \top} x^* \\ h_{k_1}^{-, \top} x^* \\ \dots \\ h_{k_1+1}^\top x^* \\ \dots \\ h_k^\top x^* \end{bmatrix} + \tilde{\varepsilon} =: \tilde{x}^* + \tilde{\varepsilon}. \quad (6.5)$$

And now we have $\tilde{y} = \tilde{x} + \tilde{\varepsilon}$, with $\tilde{x} \geq 0$ and $\tilde{\varepsilon} \sim \mathcal{N}(0, \Sigma)$. Letting $h^{(1)}, \dots, h^{(2k_1+k_2)}$ denote the expanded collection of vectors obtained by replacing each functional h_ℓ by (h_ℓ^+, h_ℓ^-) if it has mixed signs, and by h_ℓ itself if it is non-negative, the covariance of the reduced noise Σ is:

$$\Sigma_{ij} = h^{(i)\top} (K^\top K)^{-1} h^{(j)}, \quad i, j = 1, \dots, 2k_1 + k_2.$$

The new functional of interest matrix \tilde{H} equals:

$$\tilde{H} = \begin{pmatrix} I_{k_1} \otimes (1, -1) & \mathbf{0}_{k_1 \times k_2} \\ \mathbf{0}_{k_2 \times 2k_1} & I_{k_2} \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & 1 & -1 & 0 & \cdots \\ \hline 0 & \cdots & 0 & 0 & 0 & 0 & 1 & 0 \\ \vdots & & & & & & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \in \mathbb{R}^{k \times (2k_1+k_2)}. \quad (6.6)$$

In words, \tilde{H} is block diagonal: it has a $[1, -1]$ row for each of the k_1 sign-split functionals, followed by k_2 identity rows for the non-negative functionals. Note that the new dimension $2k_1 + k_2$ being in between k and $2k$ means that the reduction is the most attractive when $k \ll p$. The maximum quantile over $x \geq 0$ is at $x = 0$. The random variable is $\inf_{\tilde{H}z=0, z \geq 0} \|z - \varepsilon\|_\Sigma^2$ with $\varepsilon \sim \mathcal{N}(0, \Sigma)$. Since that sets the last k_2 variables to 0 directly and then every pair of variables corresponding to the same functional collapse to one, this is effectively a k_1 -dimensional problem. Let Q_0 be the $1 - \alpha$ quantile of this random variable. Note that this is a multiple functional problem, so as in the general case, the box given by the k problems:

$$\begin{aligned} \min_x / \max_x \quad & \tilde{H}_i^\top x \\ \text{subject to} \quad & \|x - y\|_\Sigma^2 \leq Q_0 \\ & x \geq 0, \end{aligned}$$

encapsulates the convex region $\{\mu \in \mathbb{R}^{2k_1+k_2} : \lambda(\mu, y) \leq Q_0\}$. Furthermore, there is at most one point in the polyhedron $\tilde{H}x = \mu, x \geq 0$, and it can be shown that the maximum over any polyhedron of the form $\tilde{H}x = \mu, x \geq 0$ equals Q_0 .

Analogous to the $k = 1$ case, a slightly weaker condition than K being full column rank is the collection of expanded functionals $h^{(i)}$ all being in the row space of K , in which case, one can replace the inverses by pseudoinverses. In the collection of expanded functionals, we allow for any split $h = h^1 - h^2$ such that $h^1 \geq 0, h^2 \geq 0$ as long as $h^1, h^2 \in \text{row}(K)$ for the reduction to work.

6.1.3 Special case of box constraints

We next develop an explicit algorithm for the special case of box constraints that arises frequently in practice. These constraints are $x_l \leq x \leq x_u$, where some components of x_l or x_u may be infinite (formally $-\infty$ or $+\infty$), so that variables may be two-sided bounded, one-sided bounded, or unconstrained. As a first step, we make an affine change of coordinates $z = Tx + \delta$ with invertible T such that variables originally bounded from above and below are mapped to $[0, 1]$, variables

bounded only on one side are mapped to be non-negative, and unconstrained variables remain free in \mathbb{R} . After this change of coordinates, we obtain

$$y = KT^{-1}(z - \delta) + \varepsilon \quad \Rightarrow \quad y' = K'z + \varepsilon, \quad (6.7)$$

with $y' = y + KT^{-1}\delta$, $K' = KT^{-1}$, and new constraints given by:

$$0 \leq z_i \leq 1 \quad (i = 1, \dots, s), \quad z_i \geq 0 \quad (i = s+1, \dots, s+t), \quad z_i \in \mathbb{R} \quad (i = s+t+1, \dots, p).$$

Write the index sets $A := \{1, \dots, s\}$ (two-sided bounded), $B := \{s+1, \dots, s+t\}$ (one-sided bounded, $z_i \geq 0$), and $U := \{s+t+1, \dots, p\}$ (unconstrained). Depending on the dimensionality of the problem, it might be feasible to calibrate regions of λ^1 or λ_u^2 by finding the maximum quantile numerically. We therefore start by extending the results for quantile maximization, such as (5.6) and (5.12) to the case in which some variables are unconstrained. We then use these results to develop reductions for high-dimensional cases.

Lemma 6.1. *Let $x^* = (x_c^*, x_u^*)$ with $x_c^* \in \mathbb{R}^l$ and $x_u^* \in \mathbb{R}^{p-l}$, and consider constraint set \mathcal{X} of the form $\{x \in \mathbb{R}^p : Ax_c \leq b\}$. Then, the distributions of the random variables $Z_{u,x}^2$ and Z_x^1 do not depend on the value of x_u^* . In particular, by Theorem 5.6, the quantile global maximum is attained at a point of the form $\hat{x} = (\hat{x}_c, \hat{x}_u)$, where \hat{x}_c is an extreme point of $Ax_c \leq b$ viewed as a polyhedron in \mathbb{R}^l , and \hat{x}_u is arbitrary.*

Proof. A direct computation after defining the test statistic λ^1 with the new constraint set yields:

$$\mathcal{T}_x^1(\varepsilon) = \lambda^1(Hx, Kx + \varepsilon) = \inf_{H\xi = Hx^*, A\xi_c \leq b, \xi_u \in \mathbb{R}^{p-l}} \|K(x - x^*) - \varepsilon\|_2^2.$$

Setting $(z_c, z_u) := (x_c, x_u) - (\xi_c, \xi_u)$ as the new optimization variable, we get:

$$\mathcal{T}_x^1(\varepsilon) = \inf_{Hz=0, Az_c \geq Ax_c^* - b, z_u \in \mathbb{R}^{p-l}} \|Kz - \varepsilon\|_2^2,$$

which is independent of x_u . The same computation with the extra term $\inf_z \|Kz - \varepsilon\|_2^2$, which is independent of both x_c^* and x_u^* , proves the result for λ_u^2 . One then shows joint convexity of $\mathcal{T}_x^1(\varepsilon)$ and $\mathcal{T}_{u,x}^2(\varepsilon)$ and monotonicity across recessing directions $Ay_c \leq 0$, and the same results as in Theorem 5.6 can be obtained over x_c . \square

In particular, the quantile maximizer over z , \hat{z} is one of the 2^s points of the form:

$$\hat{z}_i \in \{0, 1\} \quad (i = 1, \dots, s), \quad \hat{z}_i = 0 \quad (i = s+1, \dots, s+t), \quad \hat{z}_i \in \mathbb{R} \quad (i = s+t+1, \dots, p).$$

We propose the following reduction in the case it is unfeasible to compute the quantiles at 2^s points. Assume the model $y' = K'z + \varepsilon$ from (6.7), and rename from now on (K', y') to (K, y) for simplicity. Assume K has full column rank and $k = 1$. Split the variables by constraint type as $x = x_A + x_B + x_U$, where $x_A = (x_1, \dots, x_s, 0, \dots, 0)$ with $0 \leq x_i \leq 1$, $x_B = (0, \dots, 0, x_{s+1}, \dots, x_{s+t}, 0, \dots, 0)$ with $x_i \geq 0$, and $x_U = (0, \dots, 0, x_{s+t+1}, \dots, x_p)$ unconstrained.

Write $h = h^+ - h^-$ with $h^+, h^- \geq 0$. After observing $y = Kx^* + \varepsilon$, define the reduced observation:

$$\tilde{y} = \begin{bmatrix} h^{+, \top} (K^\top K)^{-1} K^\top y \\ h^{-, \top} (K^\top K)^{-1} K^\top y \end{bmatrix} = \begin{bmatrix} h^{+, \top} x_A + h^{+, \top} x_B + h^{+, \top} x_U \\ h^{-, \top} x_A + h^{-, \top} x_B + h^{-, \top} x_U \end{bmatrix} + \tilde{\varepsilon}. \quad (6.8)$$

Introduce the six-dimensional vector $\tilde{x} := (h^{+, \top} x_A, h^{+, \top} x_B, h^{+, \top} x_U, h^{-, \top} x_A, h^{-, \top} x_B, h^{-, \top} x_U) \in \mathbb{R}^6$, so that

$$\tilde{y} = \underbrace{\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}}_{=: \tilde{K} \in \mathbb{R}^{2 \times 6}} \tilde{x} + \tilde{\varepsilon}, \quad (6.9)$$

with $\tilde{\varepsilon} \sim \mathcal{N}(0, \Sigma)$ where covariance matrix Σ given by: $\Sigma = \begin{bmatrix} (h^+)^{\top} (K^{\top} K)^{-1} h^+ & (h^+)^{\top} (K^{\top} K)^{-1} h^- \\ (h^-)^{\top} (K^{\top} K)^{-1} h^+ & (h^-)^{\top} (K^{\top} K)^{-1} h^- \end{bmatrix}$.

The functional of interest is $\tilde{h} = (1, 1, 1, -1, -1, -1)$. The new constraints on \tilde{x} are:

$$0 \leq \tilde{x}_1 \leq M_+^A, \quad \tilde{x}_2 \geq 0, \quad \tilde{x}_3 \in \mathbb{R}, \quad 0 \leq \tilde{x}_4 \leq M_-^A, \quad \tilde{x}_5 \geq 0, \quad \tilde{x}_6 \in \mathbb{R},$$

where $M_+^A := \sum_{i \in A} h_i^+$ and $M_-^A := \sum_{i \in A} h_i^-$. Note that \tilde{h} lies in the row space of the reduced system \tilde{K} , hence the resulting interval is finite.

For calibration, Lemma 6.1 gives us directly that only four points need to be checked for the maximum quantiles of Z_x^1 and $Z_{u,x}^2$ (they are the same in this case since K is surjective):

$$(0, 0, 0, 0, 0, 0), (M_+^A, 0, 0, 0, 0, 0), (0, 0, 0, M_-^A, 0, 0), (M_+^A, 0, 0, M_-^A, 0, 0),$$

where the third and sixth coordinates were set to 0 arbitrarily. Let D be the maximum $1 - \alpha$ quantile of $\inf_{z_1 \geq -x_1, z_4 \geq -x_4, z_i \geq 0, i \in 2, 3, 4, 6} \|\tilde{K}z - \varepsilon\|_2^2$ with $\varepsilon \sim \mathcal{N}(0, \Sigma)$ over $x_1 \in \{0, M_+^A\}$, $x_4 \in \{0, M_-^A\}$. The final interval is then given by:

$$\begin{aligned} & \min_{\tilde{x}} / \max_{\tilde{x}} \quad \tilde{h}^{\top} \tilde{x} \\ & \text{subject to} \quad \|\tilde{K} \tilde{x} - \tilde{y}\|_{\Sigma}^2 \leq D \\ & \quad 0 \leq \tilde{x}_1 \leq M_+^A \\ & \quad \tilde{x}_2 \geq 0 \\ & \quad \tilde{x}_3 \in \mathbb{R} \\ & \quad 0 \leq \tilde{x}_4 \leq M_-^A \\ & \quad \tilde{x}_5 \geq 0 \\ & \quad \tilde{x}_6 \in \mathbb{R}. \end{aligned}$$

The same construction extends directly to the case in which K is not full column rank but h^+ and h^- are in its row space, and also to the case of multiple functionals. After the affine change of variables described above, each functional h_i is split into non-negative parts h_i^+, h_i^- as in the previous section (if needed), and the resulting expanded system $(\tilde{y}, \tilde{x}, \tilde{\varepsilon})$ has the same structure with block matrix \tilde{H} as in (6.6). The same procedure of splitting x into three applies, and the feasible region for \tilde{x} is now determined by the box constraints: coordinates coming from two-sided bounded variables are restricted to $[0, 1]$, those from one-sided bounds are constrained to be non-negative, and those from unconstrained variables remain free in \mathbb{R} .

6.2 Row and null splitting for arbitrary forward models K

Methods such as the TFM reductions in Section 6.1 require that positive/negative splits of each functional lie in $\text{row}(K)$, which holds if K has full column rank. To handle high-dimensional problems where this fails, we develop a “split” technique to decompose the problem into two components:

a full column rank component and a null space part. This construction is completely general and makes no rank assumption on $K \in \mathbb{R}^{n \times p}$.

Let $r \leq \min(n, p)$ be the rank of K and let K^\dagger be its Moore-Penrose pseudoinverse. Set $M := K^\dagger K$, the orthogonal projector onto the row space of K . For any $H \in \mathbb{R}^{k \times p}$, write

$$H_\parallel := HM, \quad H_\perp := H(I - M), \quad \text{thus } Hx = H_\parallel x + H_\perp x.$$

We construct a $1 - \alpha$ confidence region for Hx as a Minkowski sum:

$$\mathcal{R}_\alpha(y) = \mathcal{R}_{\parallel, \alpha_1}(y) \oplus \mathcal{R}_{\perp, \alpha_2}(y), \quad \alpha_1 + \alpha_2 = \alpha,$$

where \oplus denotes the (set) Minkowski sum. The default split is $\alpha_1 = \alpha_2 = \alpha/2$.

Row space part (\parallel). We have

$$HK^\dagger y \sim \mathcal{N}(H_\parallel x, \Sigma_\parallel), \quad \Sigma_\parallel := HK^\dagger (K^\dagger)^\top H^\top,$$

together with the constraint $x \geq 0$. Observe that $\Sigma_\parallel = (HK^\dagger)(HK^\dagger)^\top$ is positive semidefinite of rank at most $r = \text{rank}(K)$. If $\text{rank}(HK^\dagger) = k$, then Σ_\parallel is positive definite and the natural confidence set that ignores the constraints

$$\{\mu \in \mathbb{R}^k : \|\mu - HK^\dagger y\|_{\Sigma_\parallel}^2 \leq Q_{\chi_{k, 1-\alpha_1}^2}\}. \quad (6.10)$$

is a bounded ellipsoid. If $\text{rank}(HK^\dagger) < k$, Σ_\parallel is singular and the set above should be interpreted using the Moore-Penrose pseudoinverse Σ_\parallel^\dagger ; the set is then unbounded along $\ker(\Sigma_\parallel)$, corresponding to redundant functionals that vanish on $\text{row}(K)$.

Different options exist to build the $1 - \alpha_1$ confidence region for the row space part. First and simplest, one can ignore the constraint $Ax \leq b$ and report directly (6.10). This unconstrained part also admits the formulation (2.8), which in this case gives

$$\{\mu \in \mathbb{R}^k : \|\mu - HK^\dagger y\|_{\Sigma_\parallel}^2 \leq Q_{\chi_{n, 1-\alpha_1}^2} - \min_{x'} \|Kx' - y\|_2^2\}. \quad (6.11)$$

To add the constraint, one can view this as a constrained problem of the form $\tilde{y} = \tilde{K}x + \tilde{\varepsilon}$ with $\tilde{K} = H_\parallel$, $\tilde{H} = \tilde{K}$, and $\tilde{\varepsilon} \sim \mathcal{N}(0, \Sigma_\parallel)$. In this particular case, where the functionals of interests exactly match the observation matrix, the test inversion formalism yields, after a calculation, the intersection of (6.10) with $\{H_\parallel x : Ax \leq b\}$. This may lead to a possible improvement in the volume of the confidence region, depending on the geometry of the problem. If the row-space part satisfies the feasibility conditions for a TFM reduction, then this component may likewise be treated using a TFM-based approach.

Null space part (\perp). Whether the null space part will be bounded depends on the strength of the constraints. One can build the strict bounds region:

$$\{H_\perp x : x \in \mathbb{R}^p, Ax \leq b, \|Kx - y\|_2^2 \leq Q_{\chi_{n, 1-\alpha_2}^2}\} = \{\mu \in \mathbb{R}^k : \inf_{H_\perp x = \mu, Ax \leq b} \|Kx - y\|_2^2 \leq Q_{\chi_{n, 1-\alpha_2}^2}\},$$

where the constant $Q_{\chi_{n, 1-\alpha_2}^2}$ can be improved, according to Corollary 5.13, to the $(1 - \alpha_2)$ quantile of $\inf_{H_\perp z=0, z \geq 0} \|Kz - \varepsilon\|_2^2$ with $\varepsilon \sim \mathcal{N}(0, I)$.

Testing set membership. With $\alpha_1 + \alpha_2 t = \alpha$, $\mathcal{R}_\alpha(y) := \mathcal{R}_{\parallel, \alpha_1}(y) \oplus \mathcal{R}_{\perp, \alpha_2}(y)$ is a $1 - \alpha$ confidence region for Hx .

An important question is whether it is computationally easy to check whether a particular point μ belongs to the Minkowski sum set $A \oplus B$. If \mathcal{R}_{\parallel} and \mathcal{R}_{\perp} are given by expressions of the form $\mathcal{R}_{\parallel,\perp} = \{\mu : g_{\parallel,\perp}(\mu, y) \leq 0\}$ for convex, easy-to-evaluate functions g_{\parallel} and g_{\perp} (as is the case for all the TFM reductions and our global methods), then the Minkowski sum is convex and testing membership for a given μ is equivalent to solving the following convex feasibility program:

$$\text{find } a \quad \text{such that} \quad g_{\parallel}(a, y) \leq 0, \quad g_{\perp}(\mu - a, y) \leq 0.$$

If $g_{\parallel}(\mu) = \inf_{H_1 x = \mu, x \geq 0} \|Kx - y\|_2^2 - d_1(y)$ and $g_{\perp}(\mu) = \inf_{H_2 \tilde{x} = \mu, \tilde{x} \geq 0} \|\tilde{K}\tilde{x} - \tilde{y}\|_2^2 - d_2(\tilde{y})$, then to check whether μ is in the Minkowski sum of the two sets, we can solve the following joint feasibility convex quadratic program:

$$\begin{aligned} & \text{find } a, b, x, \tilde{x} \\ & \text{s.t. } a + b = \mu \\ & \quad Hx = a, \quad x \geq 0, \quad \|Kx - y\|_2^2 \leq d_1(y) \\ & \quad \tilde{H}\tilde{x} = b, \quad \tilde{x} \geq 0, \quad \|\tilde{K}\tilde{x} - \tilde{y}\|_2^2 \leq d_2(\tilde{y}). \end{aligned}$$

This region can also be embedded into a hyper-rectangle by solving for the minimum and maximum of $h_i^{\top} x$ for $i = 1, \dots, k$ over this convex, easy-to-describe region.

7 Numerical comparisons

To illustrate the methods in this work, we consider the following non-negatively constrained problem:

$$y = Kx + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, I), \quad x^* \geq 0, \quad K = \begin{pmatrix} 2, 1, 1 \\ 0, 1, 1 \end{pmatrix}, \quad H = \begin{pmatrix} 1, -1, 0 \\ 0, 1, -1 \end{pmatrix}. \quad (7.1)$$

Note that neither functional in H lies in the row space of K . Hence, the non-negativity constraint is essential to obtain finite-area confidence regions in \mathbb{R}^2 . By checking the condition in Theorem 2.4, we confirm that the only $d \geq 0$ with $Kd = 0$ is the origin $d = 0$, which implies $Hd = 0$ and thus guarantees finite regions under the added constraints. We construct 68% and 95% confidence regions from a single observation of y . We focus on regions built from λ_1 or λ_u^2 (which are identical in this example since K is surjective), and we assess two improvements: (i) using the μ -description instead of the x -description bounding box, and (ii) using the optimal constants from Corollaries 5.9 and 5.13 instead of the naive χ^2 thresholds.

7.1 Details of competing methods

Our first benchmark is given by the Simultaneous Strict Bounds (SSB) method (3.13), which constructs the product of intervals:

$$\prod_{i=1}^2 \begin{aligned} & \min_x / \max_x \quad h_i^{\top} x \\ & \text{subject to} \quad \|Kx - y\|_2^2 \leq Q_{\chi_2^2, 1-\alpha} \\ & \quad x \geq 0. \end{aligned} \quad (7.2)$$

We refer to this method as SSB_x , since it is the x -description bounding box of the λ^1 inversion region. Two improvements are immediate. First, consider the μ -description, i.e., the convex set:

$$\{\mu \in \mathbb{R}^2 : \inf_{Hx = \mu, x \geq 0} \|Kx - y\|_2^2 \leq Q_{\chi_2^2, 1-\alpha}\} = \{Hx : x \in \mathbb{R}^3, x \geq 0, \|Kx - y\|_2^2 \leq Q_{\chi_2^2, 1-\alpha}\}, \quad (7.3)$$

which we call SSB_μ and which $\text{SSB}_\mu \subseteq \text{SSB}_x$ by construction. Second, tighten the constant $Q_{\chi_2^2, 1-\alpha}$ by replacing it (in either description) with the $1 - \alpha$ quantile of Z_0^1 per Corollary 5.13, i.e., the quantile of $\inf_{Hx=0, z \geq 0} \|Kz - \varepsilon\|_2^2$ with $\varepsilon \sim \mathcal{N}(0, I)$. As noted in Section 5.2.4, this is a chi-bar-squared distribution. A direct analysis shows that for this example, the distribution of Z_0^1 is $\frac{1}{2}\chi_1^2 + \frac{1}{2}\chi_2^2$, so its quantiles can be computed by bisection on the cumulative density function. In higher dimensions, we estimate these quantiles by Monte Carlo sampling of ε . At 68% and 95%, the corresponding quantiles are 1.644 and 5.139, compared to 2.279 and 5.991 for the χ_2^2 distribution. We denote the methods with improved constants in (7.2) and (7.3) as QuantileZero_x and QuantileZero_μ , respectively.

We also compare against a Bonferroni-corrected product of intervals by constructing marginal intervals for $h_1^\top x$ and $h_2^\top x$ using a Bonferroni-corrected quantiles at zero of $\inf_{h_i^\top x=0, z \geq 0} \|Kz - \varepsilon\|_2^2, i = 1, 2$ to achieve joint coverage at levels 68% and 95%.

Finally, we apply the split technique of Section 6.1 by decomposing $H = H_\parallel + H_\perp$ using the projectors onto the row and null spaces of K , yielding

$$H_\parallel = \begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 0 \end{pmatrix}, \quad H_\perp = \begin{pmatrix} 0 & -\frac{1}{2} & \frac{1}{2} \\ 0 & 1 & -1 \end{pmatrix}.$$

The row space contribution reduces to $HK^\dagger y \sim \mathcal{N}(H_\parallel x, \Sigma_\parallel)$ with $\Sigma_\parallel = \text{diag}(5/4, 0)$, so only the first component is noisy: $(HK^\dagger y)_1 = \frac{1}{2}y_1 - y_2 \sim \mathcal{N}((H_\parallel x)_1, 5/4)$, while the second component vanishes deterministically. This yields a simple $(1 - \alpha_1)$ Gaussian interval for $(H_\parallel x)_1$. The null space contribution enforces $x_2 = x_3$, so feasible x take the form (x_1, t, t) with $x_1, t \geq 0$. A $(1 - \alpha_2)$ feasible set for $H_\perp x$ is obtained as $\{H_\perp x : x \geq 0, \|Kx - y\|_2^2 \leq R_{1-\alpha_2}\}$, with $R_{1-\alpha_2}$ chosen either as a χ_2^2 quantile (which we call the naive version) or the quantile at zero of a λ^1 test statistic (which we call refined). Combining both parts with $\alpha_1 + \alpha_2 = \alpha$, the final $1 - \alpha$ confidence region is the Minkowski sum $\mathcal{R}_\alpha(y) = \mathcal{R}_{\parallel, \alpha_1}(y) \oplus \mathcal{R}_{\perp, \alpha_2}(y)$, where the first factor is a one-dimensional segment and the second a line-constrained residual set. For this example, we take $\alpha_1 = \alpha_2 = \alpha/2$, though this split could be optimized.

7.2 Visualization of confidence regions for fixed y

We fix y and compare the resulting confidence regions $\mathcal{R}_\alpha(y)$ for the different methods in Section 7.1, deferring empirical coverage under resampling of y to Section 7.3. Figure 7.1 shows results for $y = (0, 0)$ (left) and $y = (20, 10)$ (right), the noiseless observations of $x^* = (0, 0, 0)$ and $x^* = (5, 5, 5)$.

We observe how the μ -descriptions produce non-rectangular convex sets whose bounding boxes are precisely the x -description bounding boxes, hence they occupy smaller area. Furthermore, the tightening the constant from χ_2^2 to the quantile at the origin shrinks both the bounding boxes and the μ -descriptions. The Bonferroni method is additionally conservative due to both the correction and its restriction to axis-aligned products of intervals. The split methods are Minkowski sums of two line segments in \mathbb{R}^2 ; they are fairly conservative when $y = (0, 0)$ in area, but they match the μ -descriptions more closely when $y = (20, 10)$.

7.3 Comparison of areas and coverages

We evaluate empirical coverage and area distributions of different methods that construct $\mathcal{R}_\alpha(y)$ for $1 - \alpha = 0.68$, by sampling $y \sim \mathcal{N}(Kx^*, I)$ a total of $N = 10^5$ times and constructing all regions for $x^* = (0, 0, 0)$ and $x^* = (5, 5, 5)$. For each region and y , we compute its area using polar quadrature

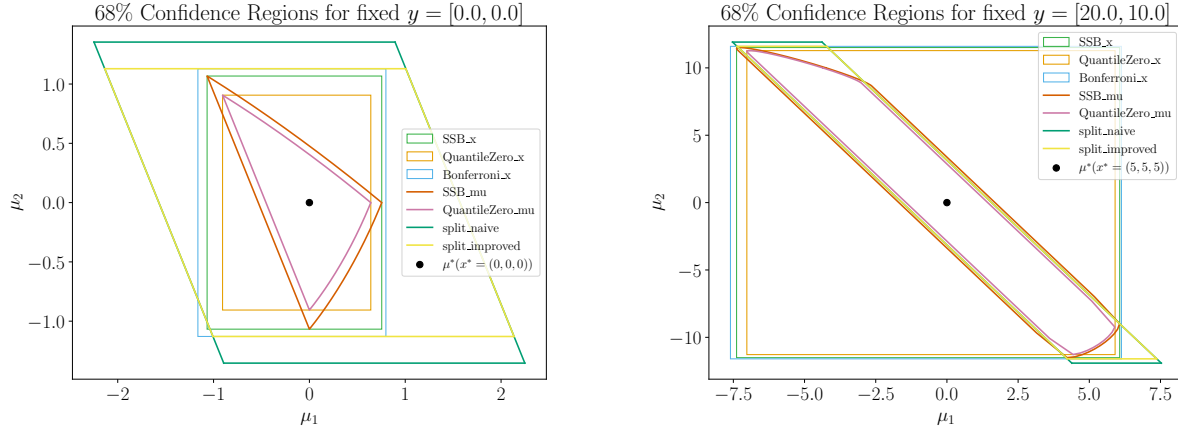


Figure 7.1: Confidence regions of at level $1 - \alpha = 68\%$ for Hx^* in the problem setup (7.1), comparing the different methods in Section 7.1 when $y = (0, 0)$ (left) and $y = (20, 10)$ (right), which are the noiseless observations for $x^* = (0, 0, 0)$ and $x^* = (5, 5, 5)$, respectively.

numerical integration, and check if they contain Hx^* . Figure 7.2 reports empirical coverage (left) and area distributions (right) for $x^* = (0, 0, 0)$ (top) and $x^* = (5, 5, 5)$ (bottom).

We observe that results align well with our theoretical predictions and the fixed- y comparison in Section 7.2. Because the λ^1 quantile is maximized at $x^* = (0, 0, 0)$, the μ -description with the correct constant (quantile at the origin) attains near-exact coverage at this point and also achieves the smallest average area. For $x^* > 0$, including our example of $(5, 5, 5)$, this method overcovers, since the test statistic is not pivotal and the quantile outside of the origin is smaller. When x^* is away from the origin, areas increase substantially for all methods, but the μ -descriptions retain a sizable advantage in average area over their x -description counterparts. We expect this improvement in terms of area to also hold and be typically magnified for higher-dimensional problems as the number of functionals k grows.

8 Discussion

In this paper, we introduced a unified test-inversion framework for confidence regions in linear inverse problems with linear or conic constraints. Our analysis ties three statistics, λ_c^2 , λ_u^2 , and λ^1 , to concrete optimization templates and calibration rules. We established convexity and recession-monotonicity for λ_u^2 and λ^1 , identified where their worst-case quantiles occur, and gave tight stochastic dominance relations. Over closed convex cones, we proved that the maximizing quantiles for λ_u^2 and λ^1 occur at $x = 0$ and admit $\bar{\chi}^2$ laws with linked intrinsic-volume weights. We developed practical high-dimensional methods, including refined TFM reductions, multi-functional generalizations, box-constraint reductions, and a row/null splitting strategy. We supplied optimal global thresholds for nonnegative problems for λ_u^2 and λ^1 , improving classical χ^2 bounds and Burrus-style intervals.

The convex-analytic parts of our framework extend beyond Gaussian noise. In particular, for general log-concave likelihoods, by taking test statistics to be based on the negative log likelihood $-2\log p(y|x)$ as in [2], the statistic λ^1 and its induced quantile remain convex, since its proof of convexity relies only on log-concavity and Prékopa's theorem. By contrast, λ_u^2 is no longer convex in general, since the Gaussian-specific structure of least squares projections and the Pythagorean

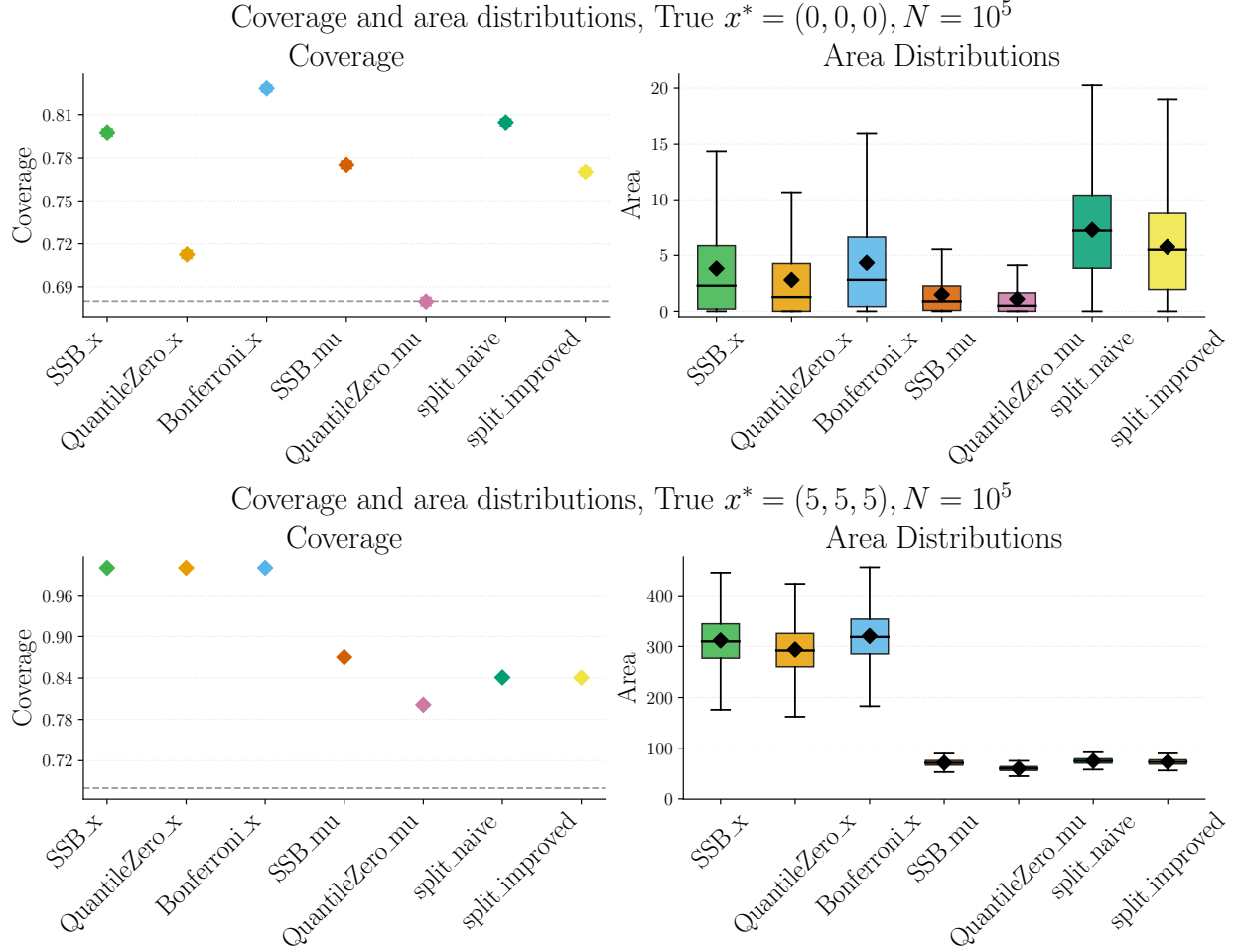


Figure 7.2: Coverage probabilities and area distributions of the constructed confidence sets for two scenarios: (top) $x^* = (0, 0, 0)$, and (bottom) $x^* = (5, 5, 5)$. Each panel shows the empirical coverage (left) and the distribution of the areas (right) for different methods, with $N = 10^5$ samples of $y \sim \mathcal{N}(Kx^*, I)$. In each plot, the diamond shows the empirical average.

decomposition are lost. Thus, in the log-concave setting, one is restricted to λ^1 for tractable convex confidence regions. The structural results concerning cones, linear constraints, monotonicity along recession directions, and boundedness, however, do not depend on Gaussianity and therefore continue to hold for arbitrary log-concave noise distributions. What does not extend are the exact chi-square and chi-bar-squared distributional laws, whose intrinsic-volume decompositions rely on Gaussian orthogonal invariance. Nevertheless, the distributions at $x = 0$ that maximize the quantiles for closed convex cones can still be sampled by solving convex programs.

Although this paper concentrates on the multiple-functional case, there remain many interesting future research directions for the single-functional case. We focused on test statistics amenable to closed-form analysis of their distributional properties, but it is possible to directly attack the calibration of the original λ_c^2 statistic by solving the related chance-constrained problem (CCP) specified in [2]. CCP's are generally difficult to solve, but there are a variety of computational approaches one can take. Scenario approaches such as those explored in [14, 4] provide one computational approach. "Scenarios" in this context refer to realizations of the underlying random process, which is known in the theoretical setup of this paper. The realized scenarios are then used to approximate

the true CCP and theoretical results provide certainty that the result of the scenario optimization is feasible to the original. Since scenarios add constraints to the approximate optimization, there is a trade-off between the certainty in the solution feasibility and the optimality. [4] provides an interesting approach to characterize this relationship. A second possible computational approach is to use a convex approximation of the CCP. Since the CCP for this problem is essentially that of finding a largest quantile, it can be framed as a “Value-at-Risk” (VaR) optimization problem. [22] considers the relaxation of VaR to “Conditional Value-at-Risk” (CVaR) which is convex and thus easier to optimize. [15] generalizes the notion of convex approximations to CCP and provides several computational options.

Since it is often difficult to verify or be sure that the above approaches produce a valid quantile, there is still ample motivation to discover new analytical characterizations. The original approach to generating functional confidence sets was based on the likelihood ratio test statistic [2]. However, as we have explored in this paper, it is possible to use others which come with analytical upside. Unlike this paper which has focused on modifications of the likelihood ratio test statistic, it may be possible to consider test statistics for which the resulting CCP is more easily solvable. For example, [20] shows that when ε follows a log-concave distribution and the test statistic is quasi-concave in the parameter that solving the CCP is equivalent to solving a convex problem. Since our Gaussian assumption on ε satisfies the log-concavity criterion, only mathematical creativity stands in the way of finding a quasi-concave test statistic with good statistical properties. Although λ^1 defined in this paper satisfies this property, we hypothesize that this test statistics is one of many possibilities.

In addition to generating altogether new test statistics with friendly mathematical properties, there is room to explore particular configurations of the test statistics presented in this paper. For example, although the Burrus conjecture was refuted in [2], we conjecture that there are *some* scenarios in which the χ_1^2 quantiles are valid due to the multiplicity of empirical circumstances in which the resulting intervals were valid [17, 29, 28]. It is possible that there exists a class of (K, h) such that χ_1^2 achieves finite-sample validity. Additionally, there may be asymptotic senses in which these quantiles or quantiles from a different distribution are valid. Situations in which one or both the number of observations (n) and the parameter dimension (p) are large could be interesting situations to investigate. Or, since parameter settings violating the Burrus conjecture were typically found along the constraint boundary, it may be possible to prove validity additionally assuming that the true parameter is sufficiently far from the constraint boundary.

As noted in Section 5, even partial results about maximizing the quantile function induced by the natural log-likelihood ratio test λ_c^2 remain elusive. We conjecture that the correctly calibrated λ_c^2 tests could yield regions better than those obtained from λ_u^2 or λ^1 , but this problem remains open. A deeper analysis and comparison between λ_u^2 and λ^1 , in particular about their $\bar{\chi}^2$ distributions at the origin would also be of interest, leading to a recommendation of one over the other depending on the geometry of particular problems.

Finally, although this work has been motivated by inverse problems in the physical sciences, we believe these results may be useful to the more general statistical literature on constrained and shape-restricted methods. For example, [7] explores finite-sample confidence envelopes for densities that are known to be monotonic or have k modes relative to a positive weight function. The confidence envelopes are defined via the strict bounds approach from [30] and thus rely on simultaneous confidence sets of the distribution. It is possible that our approach could reduce the conservatism inherent in confidence bands constructed in this way. Furthermore, the constrained inference literature centered on the chi-bar-squared distribution could also potentially benefit from our theoretical insights, as we have treated scenarios outside of the Type A and B problems to

which the approaches in [26] are confined.

Acknowledgments

PB, JRL, and HO acknowledge support from the Air Force Office of Scientific Research under MURI awards number FA9550-20-1-0358 (Machine Learning and Physics-Based Modeling and Simulation), FOA-AFRL-AFOSR-2023-0004 (Mathematics of Digital Twins), the Department of Energy under award number DE-SC0023163 (SEA-CROGS: Scalable, Efficient, and Accelerated Causal Reasoning Operators, Graphs and Spikes for Earth and Embedded Systems), the National Science Foundations under award number 2425909 (Discovering the Law of Stress Transfer and Earthquake Dynamics in a Fault Network using a Computational Graph Discovery Approach) and the DoD Vannevar Bush Faculty Fellowship Program.

References

- [1] Anastasios N. Angelopoulos and Stephen Bates. A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *Foundations and Trends in Machine Learning*, 16(4):494–591, 2023.
- [2] Pau Batlle, Michael Stanley, Pratik Patil, Mikael Kuusela, and Houman Owadi. Optimization-based frequentist confidence intervals for functionals in constrained inverse problems: Resolving the burrus conjecture. *arXiv preprint arXiv:2310.02461*, 2023.
- [3] Walter Ross Burrus. *Utilization of a Priori Information in the Statistical Interpretation of Measured Distribution*. PhD thesis, The Ohio State University, 1964.
- [4] M. C. Campi and S. Garatti. A sampling-and-discarding approach to chance-constrained optimization: Feasibility and optimality. *Journal of Optimization Theory and Applications*, 148(2):257–280, 2011.
- [5] Edmund A. Cornish and Ronald A. Fisher. Moments and cumulants in the specification of distributions. *Revue de l’Institut international de Statistique*, pages 307–320, 1938.
- [6] Ronald A. Fisher and Edmund A. Cornish. The percentile points of distributions having known cumulants. *Technometrics*, 2(2):209–225, 1960.
- [7] Nicolas W. Hengartner and Philip B. Stark. Finite-sample confidence envelopes for shape-restricted densities. *The Annals of Statistics*, 23(2):525–550, 1995.
- [8] Andrey I. Kibzun and Yuri S. Kan. *Stochastic programming problems with probability and quantile functions*. Wiley Interscience Series in Systems and Optimization. John Wiley & Sons, 1995.
- [9] Mikael Kuusela and Philip B. Stark. Shape-constrained uncertainty quantification in unfolding steeply falling elementary particle spectra. *Annals of Applied Statistics*, 2017.
- [10] Jing Lei, Max G’Sell, Alessandro Rinaldo, Ryan J. Tibshirani, and Larry Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523):1094–1111, 2018.

- [11] Jing Lei and Larry Wasserman. Distribution-free prediction bands for non-parametric regression. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76(1):71–96, 2014.
- [12] Jessie Li. The proximal bootstrap for constrained estimators. *Journal of Statistical Planning and Inference*, 236:106245, 2025.
- [13] Rupert G. Miller. *Simultaneous Statistical Inference*. Springer, 1981.
- [14] Arkadi Nemirovski and Alexander Shapiro. *Scenario Approximations of Chance Constraints*, pages 3–47. Springer London, London, 2006.
- [15] Arkadi Nemirovski and Alexander Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4):969–996, 2007.
- [16] Dianne P. O’Leary and Bert W. Rust. Confidence intervals for inequality-constrained least squares problems, with applications to ill-posed problems. *SIAM Journal on Scientific and Statistical Computing*, 7(2):473–489, 1986.
- [17] Pratik Patil, Mikael Kuusela, and Jonathan Hobbs. Objective frequentist uncertainty quantification for atmospheric CO₂ retrievals. *SIAM/ASA Journal on Uncertainty Quantification*, 10(3):827–859, 2022.
- [18] A. Prékopa. Logarithmic concave measures with application to stochastic programming. *Acta Sci. Math. (Szeged)*, 32:301–316, 1971.
- [19] A. Prékopa. On logarithmic concave measures and functions. *Acta Sci. Math. (Szeged)*, 34:335–343, 1973.
- [20] András Prékopa. Probabilistic programming. *Handbooks in Operations Research and Management Science*, 10:267–351, 2003.
- [21] Tim Robertson, R. T. Wright, and R. L. Dykstra. *Order Restricted Statistical Inference*. John Wiley and Sons, 1988.
- [22] R. Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42, 2000.
- [23] B. W. Rust and W. R. Burrus. *Mathematical Programming and the Numerical Solution of Linear Equations*. Elsevier Science, 1972.
- [24] Bert W. Rust and Dianne P. O’leary. Confidence intervals for discrete approximations to ill-posed problems. *Journal of Computational and Graphical Statistics*, 3(1):67–96, 1994.
- [25] George A. F. Seber and Alan J. Lee. *Linear Regression Analysis*. Wiley, 2003.
- [26] Mervyn J. Silvapulle and Pranab K. Sen. *Constrained Statistical Inference: Inequality, Order, and Shape Restrictions*. Wiley, 2001.
- [27] Michael Stanley. *Uncertainty Quantification for Ill-Posed Inverse Problems in the Physical Sciences: Confidence Intervals via Optimization and Sampling*. PhD thesis, Carnegie Mellon University, 2024.
- [28] Michael Stanley, Pau Batlle, Pratik Patil, Houman Owhadi, and Mikael Kuusela. Confidence intervals for functionals in constrained inverse problems via data-adaptive sampling-based calibration. *arXiv preprint arXiv:2502.02674*, 2025.

- [29] Michael Stanley, Pratik Patil, and Mikael Kuusela. Uncertainty quantification for wide-bin unfolding: one-at-a-time strict bounds and prior-optimized confidence intervals. *Journal of Instrumentation*, 17(10):P10013, 2022.
- [30] Philip B. Stark. Inference in infinite-dimensional inverse problems: Discretization and duality. *Journal of Geophysical Research: Solid Earth*, 97(B10):14055–14082, 1992.
- [31] Philip B. Stark and Robert L. Parker. Rigorous velocity bounds from soft $\tau(p)$ and $x(p)$ data. *Geophysical Journal International*, 89(3):987–996, 1987.
- [32] L. Tenorio, A. Fleck, and K. Moses. Confidence intervals for linear discrete inverse problems with a non-negativity constraint. *Inverse Problems*, 23(2):669, 2007.
- [33] Vladimir Vovk, Alexander Gammernan, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer, 2005.
- [34] Larry Wasserman, Aaditya Ramdas, and Sivaraman Balakrishnan. Universal inference. *Proceedings of the National Academy of Sciences*, 117(29):16880–16890, 2020.