Optimization-based frequentist confidence intervals for functionals in constrained inverse problems: Resolving the Burrus conjecture

Pau Batlle[†]
pbatllef@caltech.edu

Pratik Patil[‡]
pratikpatil@berkeley.edu

Michael Stanley[§]
mcstanle@andrew.cmu.edu

Houman Owhadi[†]
owhadi@caltech.edu

Mikael Kuusela[§]
mkuusela@andrew.cmu.edu

September 11, 2024

Abstract

We present an optimization-based framework to construct confidence intervals for functionals in constrained inverse problems, ensuring valid one-at-a-time frequentist coverage guarantees. Our approach builds upon the now-called strict bounds intervals, originally pioneered by [4, 46], which offer ways to directly incorporate any side information about the parameters during inference without introducing external biases. This family of methods allows for uncertainty quantification in ill-posed inverse problems without needing to select a regularizing prior. By tying optimization-based intervals to an inversion of a constrained likelihood ratio test, we translate interval coverage guarantees into type I error control and characterize the resulting interval via solutions to optimization problems. Along the way, we refute the Burrus conjecture, which posited that, for possibly rank-deficient linear Gaussian models with positivity constraints, a correction based on the quantile of the chi-squared distribution with one degree of freedom suffices to shorten intervals while maintaining frequentist coverage guarantees. Our framework provides a novel approach to analyzing the conjecture, and we construct a counterexample employing a stochastic dominance argument, which we also use to disprove a general form of the conjecture. We illustrate our framework with several numerical examples and provide directions for extensions beyond the Rust–Burrus method for nonlinear, non-Gaussian settings with general constraints.

1 Introduction

Advances in data collection and computational power in recent years have led to an increase in the prevalence of high-dimensional, ill-posed inverse problems, especially within the physical sciences. These challenges are particularly evident in domains such as remote sensing and data assimilation, where uncertainty quantification (UQ) in inverse problems is of paramount importance. Many of these inverse problems also come with inherent physical constraints on their parameters. This paper focuses on constrained inverse problems for which the noise model is known, and the forward model, defined on a finite-dimensional parameter space, can be computationally evaluated. Our primary objective is to construct a confidence interval for a functional of the forward model parameters.

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

[‡]Department of Statistics, University of California, Berkeley, CA 94720, USA.

[§]Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA.
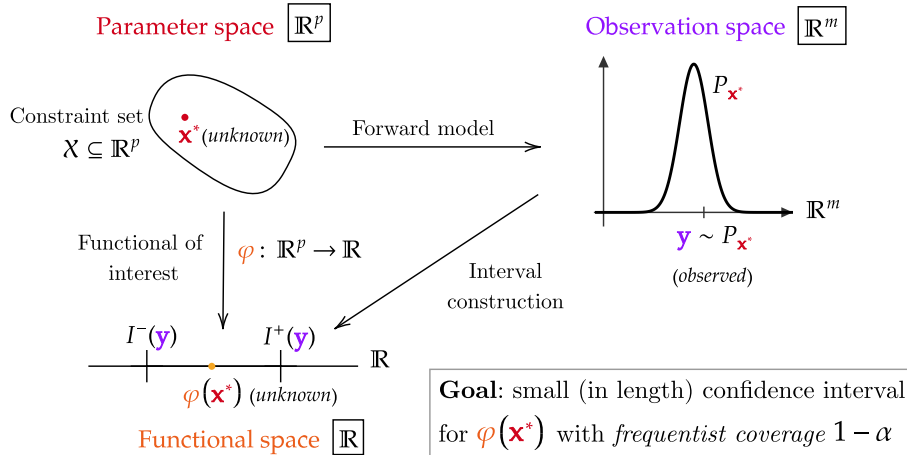
Figure 1.1: Illustration of the problem setup. We seek to construct confidence intervals $[I^-(\boldsymbol{y}), I^+(\boldsymbol{y})] \subseteq \mathbb{R}$ for $\varphi(\boldsymbol{x}^*) \in \mathbb{R}$ from an observation $\boldsymbol{y} \in \mathbb{R}^m$ sampled from $P_{\boldsymbol{x}^*}$ that satisfies a frequentist coverage guarantee in finite sample while being as small (in length) as possible.

Formally, we consider statistical models of the form $\boldsymbol{y} \sim P_{\boldsymbol{x}^*}$, where $\boldsymbol{y} \in \mathbb{R}^m$ is sampled according to a parametric probability distribution. Here $\boldsymbol{x}^* \in \mathbb{R}^p$ is a fixed unknown parameter, which we know a priori lies within the set $\mathcal{X}$; see Figure 1.1 for an illustration. Our goal is to construct confidence intervals for a known one-dimensional functional $\varphi(\boldsymbol{x}^*) \in \mathbb{R}$. Ideally, we want the length of these intervals to be as small as possible, while still maintaining a nonasymptotic frequentist coverage guarantee. In other words, given a prescribed coverage level $1 - \alpha$ for some $\alpha \in (0, 1)$, we want to construct functions of the data $I^-(\boldsymbol{y})$ and $I^+(\boldsymbol{y})$ such that the following coverage guarantee holds in finite sample[1]:

$$\inf_{\boldsymbol{x} \in \mathcal{X}} \mathbb{P}_{\boldsymbol{y} \sim P_{\boldsymbol{x}}}\left(\varphi(\boldsymbol{x}) \in [I^-(\boldsymbol{y}), I^+(\boldsymbol{y})]\right) \geq 1 - \alpha. \tag{1.1}$$

While the requirement (1.1) requires that we maintain at least $1 - \alpha$ coverage, we also want it to be approximately accurate by minimizing the slack in the inequality. Ensuring such *proper calibration*, namely, confidence intervals that do not *undercover* (fail to meet the $1 - \alpha$ guarantee for some $\boldsymbol{x} \in \mathcal{X}$) or *overcover* (are too large and therefore exceed the required coverage) is paramount in practical applications. This is especially true in contexts that require stringent safety and certification standards. Intervals that undercover yield unreliable inferences that may expose the system to unforeseen risks. Conversely, intervals that overcover might lead to excessive economic costs by needing to guard against scenarios that are unlikely to occur.

In many applied contexts, Bayesian methods constitute a primary set of techniques for uncertainty quantification. These methods leverage a prior for regularization, derived either from the intrinsic details of the problem or introduced externally. A key advantage of this regularization approach is the natural UQ that emerges from the Bayesian statistical framework. Specifically, the combination of a predefined prior and data likelihood results in a posterior distribution via Bayes' theorem. This distribution can subsequently be used to derive the intended posterior UQ. However, there is a caveat: Bayesian methods can offer *marginal coverage* (probability over $\boldsymbol{x}$ and $\boldsymbol{y}$) if the prior is correctly specified. They do not necessarily provide *conditional coverage* (probability over $\boldsymbol{y}$ given $\boldsymbol{x}$). The former notion of coverage is weaker (and, in particular, the latter implies the former, but

---

[1]This form of "simple" interval is only for expositional simplicity. One can consider more general forms of confidence sets $\mathcal{I}(\boldsymbol{y})$ beyond simple intervals, which we will do when describing the general framework in Section 3.

the converse may not be true), as it replaces the infimum in the coverage requirement (1.1) with a probability distribution over $\boldsymbol{x}$. Generally, Bayesian methods may not align with the analyst's expectations due to inherent bias [26, 43]. While, in theory, priors present an effective mechanism to incorporate scientific knowledge into UQ, they can inadvertently introduce extraneous information [58] and a lack of robustness in the resulting estimates [39, 40, 38].

On the other hand, we could consider a basic worst-case approach that is rooted in the simple observation that:

$$\varphi(\boldsymbol{x}^*) \in \left[\inf_{\boldsymbol{x} \in \mathcal{X}} \varphi(\boldsymbol{x}), \ \sup_{\boldsymbol{x} \in \mathcal{X}} \varphi(\boldsymbol{x}) \right]. \tag{1.2}$$

Of course, this method is inherently conservative given the absence of assumptions and any specific knowledge regarding data generation. More importantly, the method does not use observations y in any way to calibrate the confidence set. This means that the sets cannot be fine-tuned to approximately achieve the desired $1 - \alpha$ coverage level. Nevertheless, they illustrate the essential idea of constructing a confidence interval based on the outcomes of two boundary optimization problems, an approach that the more sophisticated methods that we will study in this paper build on. We shall henceforth refer to such intervals with the notation:

$$\begin{array}{cc} \inf_{\boldsymbol{x}} / \sup_{\boldsymbol{x}} & \varphi(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X} \end{array} := \left[\inf_{\boldsymbol{x} \in \mathcal{X}} \varphi(\boldsymbol{x}), \ \sup_{\boldsymbol{x} \in \mathcal{X}} \varphi(\boldsymbol{x}) \right].$$

An example of such a more sophisticated method is the so-called "simultaneous" approach [55, 57], which provides intervals with at least $1 - \alpha$ frequentist coverage for the functional of interest $\varphi(\boldsymbol{x}^*)$ from confidence sets for the parameter $\boldsymbol{x}^*$. The approach can be summarized in three steps (see Figure 1.2 for an illustration):

Step 1. Construct a set $\mathcal{C}(\boldsymbol{y}) \in \mathbb{R}^p$ that is a $1 - \alpha$ confidence set for $\boldsymbol{x}^*$.

Step 2. Intersect this set $\mathcal{C}(\boldsymbol{y})$ with the constraint set $\mathcal{X}$.

Step 3. Project this intersection through the functional of interest φ.

The term "simultaneous" refers to Steps 1 and 2 being independent of the quantity of interest φ, so the resulting set from Step 2 can be simultaneously projected to different quantities of interest. Under mild assumptions, the resulting intervals can be equivalently written as:

$$\mathcal{I}_{\mathsf{SSB}}(\boldsymbol{y}) := \left[\inf_{\boldsymbol{x} \in \mathcal{X} \cap \mathcal{C}(\boldsymbol{y})} \varphi(\boldsymbol{x}), \ \sup_{\boldsymbol{x} \in \mathcal{X} \cap \mathcal{C}(\boldsymbol{y})} \varphi(\boldsymbol{x}) \right] = \begin{array}{cc} \inf_{\boldsymbol{x}} / \sup_{\boldsymbol{x}} & \varphi(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X} \cap \mathcal{C}(\boldsymbol{y}). \end{array} \tag{1.3}$$

This illustrates how the simultaneous approach is a refinement of the basic worst-case method (1.2): the observation of the data $\boldsymbol{y}$ shrinks the "pre-data set" $\mathcal{X}$ into a smaller "post-data set" $\mathcal{X} \cap \mathcal{C}(\boldsymbol{y})$, which is then projected through φ in a worst-case manner. Given that this simultaneous framework is broadly encapsulated in [55] as "strict bounds," we label these intervals as "simultaneous strict bounds" or SSB intervals, for short.

Unlike methods that rely on explicit regularization through a prior, the techniques outlined above leverage only the physical constraints and the functional of interest to address the underlying ill-posedness of the inverse problem. This approach allows for uncertainty quantification without the need to assume a prior distribution, circumventing potential biases and miscalibrated coverage issues previously mentioned.
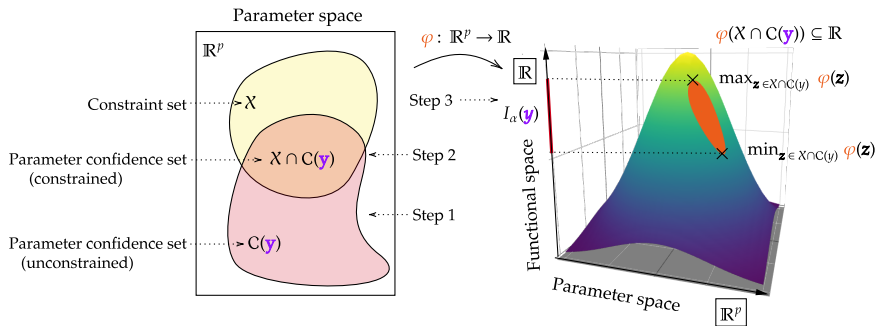
Figure 1.2: Illustration of the simultaneous approach for confidence interval building, which works generically for any $\mathcal{X}$, φ and P. The intersection of $\mathcal{X}$ and $\mathcal{C}(\boldsymbol{y})$ occurs in the original parameter space $\mathbb{R}^p$, and is then projected via the functional of interest function into the real line. The confidence interval is then constructed using the minimum and maximum of the quantity of interest φ over the intersection $\mathcal{X} \cap \mathcal{C}(\boldsymbol{y})$.

Although the interval (1.3) has guaranteed coverage for $\varphi(\boldsymbol{x}^*)$ inherited from the coverage of $\mathcal{C}(\boldsymbol{y})$, this method generally suffers from overcoverage, especially when the dimension of $\mathcal{X}$ is large [43, 54, 28]. This happens due to two main factors: (i) its generality cannot account for the specific structure of P, φ, and $\mathcal{X}$; and (ii) while the set $\mathcal{C}(\boldsymbol{y})$ being a $1 - \alpha$ confidence set is a sufficient condition, it is not necessary for (1.3) to ensure accurate coverage, which implies that smaller sets might also produce valid confidence intervals. Consequently, an important research direction has been constructing confidence intervals that are shorter than the simultaneous approach, but still maintain nominal coverage for a given φ. Sometimes, this is achieved by assuming that P, φ, and $\mathcal{X}$ come from a particular class [57, 46, 61, 43, 54]. In the sequel, we discuss one such special class.

1.1 The Burrus conjecture

The Gaussian linear forward model with nonnegativity constraints and a linear functional of interest is a setting that has attracted significant attention, going back to the works of [4, 46]. These foundational studies consider the applied problem of unfolding gamma-ray and neutron spectra from pulse-height distributions under rank-deficient linear systems. They demonstrated that incorporating the nonnegativity physical constraint allowed for the computation of nontrivial (i.e., finite length) intervals for linear functionals of the parameters. In order to describe the construction of these intervals, consider the canonical form of the Gaussian linear model with nonnegativity constraints, along with a linear functional of interest:

$$\underbrace{\boldsymbol{y} = \boldsymbol{K}\boldsymbol{x}^* + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_m)}_{\text{model}}, \quad \text{with} \quad \underbrace{\boldsymbol{x}^* \geq \boldsymbol{0}}_{\text{constraints}} \quad \text{and} \quad \underbrace{\varphi(\boldsymbol{x}^*) = \boldsymbol{h}^\top \boldsymbol{x}^*}_{\text{functional}}. \tag{1.4}$$

Here, $\boldsymbol{K} \in \mathbb{R}^{m \times p}$ is the forward operator[2], $\boldsymbol{x}^* \in \mathbb{R}^p$ is the true parameter vector, and $\boldsymbol{h} \in \mathbb{R}^p$ contains weights for the functional of interest. In this setting, [4, 46] posed that the following interval construction yields valid $1 - \alpha$ confidence intervals, a result now known as the *Burrus*

---

[2]Note that the forward operator $\boldsymbol{K}$ is allowed to be column rank deficient and the overparameterized setting when $p > m$ is allowed.

conjecture [47]:

$$\mathcal{I}_{\mathsf{OSB}}(\boldsymbol{y}) := \begin{array}{ll} \min_{\boldsymbol{x}} / \max_{\boldsymbol{x}} & \boldsymbol{h}^\top \boldsymbol{x} \\ \text{subject to} & \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2 \le \psi_\alpha^2(\boldsymbol{y}), \\ & \boldsymbol{x} \ge \boldsymbol{0}, \end{array} \tag{1.5}$$

where $\psi_\alpha^2 = z_{\alpha/2}^2 + s^2(\boldsymbol{y})$. Here z_α is the upper quantile of standard normal such that $\mathbb{P}(Z > z_\alpha) = \alpha$ for $Z \sim \mathcal{N}(0,1)$, and $s^2(\boldsymbol{y})$ is defined through an optimization problem as follows:

$$s^2(\boldsymbol{y}) := \begin{cases} \min_{\boldsymbol{x}} & \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2 \\ \text{subject to} & \boldsymbol{x} \ge \boldsymbol{0}. \end{cases} \tag{1.6}$$

Comparison of (1.5) with (1.3) shows that Rust and Burrus proposed a "simultaneous-like" construction. In this construction, the set $\{\boldsymbol{x} : \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2 \le \psi_\alpha^2(\boldsymbol{y})\}$ plays the role of $\mathcal{C}(\boldsymbol{y})$. It typically does not represent a $1 - \alpha$ confidence set for $\boldsymbol{x}^*$, thus relaxing the stringent assumption of the SSB interval construction. Furthermore, a possible simultaneous interval for this setting can be built by observing that $\|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}^*\|_2^2 \sim \chi_m^2$. This yields the following valid $1 - \alpha$ interval:

$$\begin{array}{ll} \min_{\boldsymbol{x}} / \max_{\boldsymbol{x}} & \boldsymbol{h}^\top \boldsymbol{x} \\ \text{subject to} & \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2 \le Q_{\chi_m^2}(1 - \alpha) \\ & \boldsymbol{x} \ge \boldsymbol{0}. \end{array} \tag{1.7}$$

Here, $Q_{\chi_m^2}$ is the quantile function of a χ_m^2 distribution. It should be noted that the data-dependent term $\psi_\alpha^2(\boldsymbol{y})$ in (1.5) could be considerably smaller than $Q_{\chi_m^2}(1 - \alpha)$, especially when m is large and α is small. So if the Burrus conjecture were true, it would provide a significant reduction in the length of the interval for problems in the class (1.4). For instance, assuming $\alpha = 0.05$ (so that we are after a 95% coverage level), [54] observe an expected length reduction of about a factor of two across a variety of functionals in a particle unfolding application. The gain in the interval length originates from the fact that these intervals take into account that we are only required to guarantee coverage for *one specific* functional. Given that intervals of the form (1.5) are designed to provide coverage for one functional at a time, following the nomenclature of [54], we refer to these intervals as "one-at-a-time strict bounds" or OSB intervals, for short[3].

[46] and subsequently [47] investigated the conjecture posed in [4]. The latter work purported to have found definitive proof for the conjecture's validity. However, this claim was later refuted by [61] through a two-dimensional counterexample. In this work, we demonstrate that, in fact, this two-dimensional counterexample proposed in [61] is not a valid counterexample. However, we present and prove another counterexample that refutes the conjecture and we propose ways to fix the previous faulty results by reinterpreting the conjecture. We achieve this through a novel hypothesis test-based framework that not only revisits but also broadens the scope beyond the linear Gaussian setting paired with positivity constraints in which the conjecture was originally proposed.

---

[3][43, 54] also extend the setting and the conjecture to encompass linear constraints of the form $\boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}$. Such constraints are of interest in practical applications such as X_{CO2} retrieval and particle unfolding. For simplicity, we present only the positivity constraint case here. However, our counterexample based on positivity constraints in Section 4 will also be sufficient to disprove the conjecture in this general case.

1.2 Summary and outline

In this paper, we frame the problem of confidence interval construction for functionals in constrained, ill-posed problems through the inversion of a particular likelihood ratio test. This perspective allows us to reinterpret the interval coverage guarantee in terms of type-I error control associated with the test and, subsequently, the distribution of the log-likelihood ratio under the null hypothesis. We also establish connections between different fields of hypothesis testing with likelihood ratio tests, optimization-based confidence intervals, and chance-constrained optimization. A detailed summary of contributions in this paper along with an outline for the paper is given below.

(1) **Strict bounds intervals from test inversion.** In Section 2, we present a general framework to construct strict bounds intervals through test inversion, resulting in two optimization problems for the interval endpoints. This approach generalizes the Rust–Burrus-type interval technique to potentially nonlinear and non-Gaussian settings. Our main result in Theorem 2.4 proves coverage of the test inversion construction and Proposition 2.5 provides sufficient conditions under which the coverage is tight. Examples in Section 2.4 provide straightforward but concrete analytical illustrations of our framework.

(2) **General interval construction methodology.** In Section 3, we present a general methodology for computing confidence intervals that builds on the framework in Section 2. We outline the methodology in Algorithm 1 and discuss two key components: the chance-constrained optimization problem and the stochastic dominance argument in Section 3.1 and Section 3.2, respectively. The chance-constrained optimization problem allows us to obtain optimal decision values for the proposed framework, while the stochastic dominance argument provides a theoretical tool to find provable upper bounds.

(3) **Refuting the Burrus conjecture.** In Section 4, we demonstrate that our method successfully recovers previously proposed OSB intervals for the linear Gaussian setting. In Theorem 4.1, we leverage this novel interpretation to disprove the Burrus conjecture [46, 47] in the general case, by refuting a previously proposed counterexample and providing a new, provably correct counterexample in Lemma 4.5. Furthermore, we provide a negative result that disproves a natural generalization of the original conjecture in Proposition 4.6. Our proof technique provides a method to detect when the Rust–Burrus approach is effective and when it falls short and introduces a means to rectify the earlier erroneous examples.

(4) **Illustrative numerical examples.** In Section 5, we elucidate our findings through a suite of numerical illustrations. These span various scenarios, including the counterexample to the Burrus conjecture. We show that test inversion-based intervals, which have provable guarantees, achieve better coverage calibration than previous approaches.

1.3 Other related work

Given the effectiveness of the strict bounds methodology in high-dimensional ill-posed inverse problems, this paper seeks to deepen our understanding of these intervals and provide related perspectives by connecting them with the broader statistical literature. Specifically, we relate these intervals to the well-developed areas of likelihood ratio tests, test inversions, and constrained inference, enabling us to make rigorous statements about their properties and generalize the methodology beyond its earlier confines. We provide a brief overview of earlier work in this area below.

Confidence intervals in penalized inverse problems. Various optimization-based strategies exist for constructing confidence intervals for functionals in linear inverse problems with constraints. A first connection between optimization and inference in inverse problems is given by classical approaches seeking to optimize an objective function to balance data misfit with regularization, while adhering to prior constraints [21, 22]. It is then common to use the variability of the minimizer to quantify uncertainty. Another closely related strategy employs Bayesian methods to estimate the posterior distribution of model parameters given a regularizing prior and subsequently constructs credible intervals from marginal distributions [60, 59]. Since these methods effectively quantify uncertainty around the expectation of the regularized estimator, their coverage is highly dependent on the precision of prior information and strength of regularization [43, 26, 27]. A recent line of work starting with [24] attempts to improve the coverage of confidence intervals derived from penalized estimators by "de-biasing" the regularized estimators; however, in practice, guarantees can only be obtained asymptotically and finite-sample performance depends on the choice of tuning parameters.

An alternative line of work in optimization-based confidence intervals focuses on ensuring correct frequentist finite-sample coverage. This approach is more resistant to the aforementioned challenges associated with relying heavily on prior assumptions or de-biasing and offers a robust framework for uncertainty quantification. In the following section, we will describe these optimization-based methods in more detail.

Optimization-based confidence intervals and the Burrus conjecture. This paper is largely motivated by the literature ([46, 37, 47, 61, 43, 54]) surrounding the Burrus conjecture (see Section 4 for further discussion), which makes a claim about how to set a calibration parameter in an optimization-based confidence interval construction so that the resulting interval has a desired level of coverage for a single functional. For intervals with a simultaneous coverage guarantee for an arbitrary collection of functionals, [56] provides the most general optimization-based confidence interval construction. While these intervals provide the desired coverage, they are overly conservative in terms of length when compared to intervals calibrated for a specific functional [54]. These prior works consider only the Gaussian linear inverse problem and can thus be seen as a particular instance of the more general optimization-based confidence intervals treated in this paper.

Inverting likelihood ratio tests and constrained inference. Traditionally, optimization-based confidence interval constructions in inverse problems have developed somewhat independently of the broader statistical literature, often overlooking the duality between confidence intervals and hypothesis testing [5, 65, 29]. Our work reinterprets these optimization-based confidence intervals from the inverse problem literature as inverted hypothesis tests and situates them within the realm of constrained testing and inference; see, e.g., [18, 67, 44, 52, 68, 34], among others.

The constrained inference literature often employs the $\bar{\chi}^2$ distribution, a convex combination of χ^2 distributions with different degrees of freedom, dictated by the problem constraints. Recent work in [69] has extended these constrained testing frameworks to high-dimensional settings with linear inequality constraints, examining both sparse and non-sparse scenarios. Although such tests can be more powerful than their unconstrained counterparts, their definitions typically limit the null hypothesis to linear subspaces, complicating their use in test inversion scenarios [53].

Although there have been applications of constrained test inversion ([13]), these are limited in scope due to grid-based inversion approaches. The statistics literature contains other approaches to inverting likelihood ratio tests (LRTs), which center around sampling procedures [6, 63, 16, 35, 36, 50]. Alternatively, one can sample from the parameter space and the forward model to

generate training data for a quantile regression, which can then be used to invert an LRT ([9, 32, 10, 14]). Since these latter approaches require sampling points in the parameter space, they are practically limited to compact parameter spaces and may encounter difficulties with high-dimensional parameters. In scenarios where the data can be split, approaches such as Universal Inference [66] offer a way to obtain confidence sets for irregular likelihoods with nonasymptotic coverage.

Worst-case and likelihood-free methods. Most of the approaches and methods referenced and described above make strong assumptions about the underlying data-generating distribution (e.g., linear forward model and Gaussian noise). To generalize these assumptions, one can either take a worst-case approach within the model class (e.g., [12] which looks at worst-case confidence intervals for linear inverse problems) or remove distribution assumptions altogether. For example, Optimal Uncertainty Quantification (OUQ) (see, e.g., [41]) does not assume a particular likelihood function to perform statistical inference by relying instead on worst-case bounds. If one is willing to make boundedness assumptions on the parameter space, simulation-based inference approaches such as [19, 62, 9, 32, 10, 7, 8] explore the use of sampling-only access to the likelihood, typically through a simulator, which has found particular relevance in the physical sciences. While these likelihood-free methods are advantageous in contexts where the likelihood is uncertain, unknown or accessible only through a simulator, they tend to yield conservative estimates when a well-defined likelihood is available.

2 Strict bounds intervals from test inversion

Suppose that we observe data $\boldsymbol{y} \in \mathbb{R}^m$ according to a data-generating process $\boldsymbol{y} \sim P_{\boldsymbol{x}^*}$. Here, $P_{\boldsymbol{x}^*}$ is a distribution that depends on a fixed but unknown parameter $\boldsymbol{x}^* \in \mathbb{R}^p$. Furthermore, suppose that we have prior knowledge that this parameter $\boldsymbol{x}^*$ lies in a constraint set $\mathcal{X} \subseteq \mathbb{R}^p$, namely $\boldsymbol{x}^* \in \mathcal{X}$. Given a nominal coverage level $1 - \alpha$, where $\alpha \in (0, 1)$, this paper investigates methods for constructing a $1 - \alpha$ confidence interval for $\varphi(\boldsymbol{x}^*)$, where $\varphi : \mathbb{R}^p \to \mathbb{R}$ is a known one-dimensional quantity of interest[4] (we will also refer to φ as a functional of interest). More precisely, we are interested in constructing an interval $\mathcal{I}_\alpha(\boldsymbol{y}) \subseteq \mathbb{R}$ for $\varphi(\boldsymbol{x}^*)$ that satisfies the following coverage requirement:

$$\mathbb{P}_{\boldsymbol{y} \sim P_{\boldsymbol{x}^*}}\big(\varphi(\boldsymbol{x}^*) \in \mathcal{I}_\alpha(\boldsymbol{y})\big) \geq 1 - \alpha, \quad \text{for all} \quad \boldsymbol{x}^* \in \mathcal{X}. \tag{2.1}$$

Our primary focus lies in intervals that: (i) effectively utilize the information that $\boldsymbol{x}^* \in \mathcal{X}$, (ii) are valid (i.e., satisfying the coverage requirement in (2.1)) in the finite data and noisy regimes (rather than, e.g., in the large system or noiseless limits), (iii) do not make overly restrictive assumptions (e.g., identifiability) about the structure of the parametric model $P_{\boldsymbol{x}^*}$, and (iv) are short in length[5]. We view the observation vector $\boldsymbol{y}$ as a single observation in $\mathbb{R}^m$ drawn from a multivariate distribution $P_{\boldsymbol{x}^*}$. This may include the case of repeated sampling (i.i.d. or not) from an experiment and aggregating the samples in a vector. In this case, $P_{\boldsymbol{x}^*}$ is then defined as the measure that accounts for all the observations.[6]

---

[4]Confidence sets of several functionals of interest with guarantees can be constructed by using the proposed method with, e.g., Bonferroni correction, but studying the performance of that approach is beyond the scope of this work

[5]Note that length will, in general, depend on the unknown parameter $\boldsymbol{x}^*$. There are several notions for the "optimality" of the method with respect to length, such as minimax length [12, 48] or expected length [54], among others.

[6]For example, in the typical case where a d dimensional vector is observed a total of n times, we aggregate the results in an $m = n \times d$ dimensional vector. Throughout, we use m to denote the total dimensionality of the observation vector.

2.1 Review: classical test inversion for simple null hypotheses

We briefly review the concept of test inversion and the duality between hypothesis testing and confidence sets upon which the subsequent subsections will be built. After observing $\boldsymbol{y} \sim P_{\boldsymbol{x}^*}$, two classical statistical tasks emerge: (i) determining whether $\boldsymbol{x}^* = \boldsymbol{x}$ for a particular $\boldsymbol{x} \in \mathcal{X}$ at a significance level α (hypothesis testing), and (ii) constructing a subset of $\mathcal{X}$ that contains $\boldsymbol{x}^*$ with a coverage level $1 - \alpha$ (confidence set building). In hypothesis testing, for a given parameter $\boldsymbol{x}$, one can consider the hypothesis test:

$$H_0 \colon \boldsymbol{x}^* = \boldsymbol{x} \quad \text{versus} \quad H_1 \colon \boldsymbol{x}^* \neq \boldsymbol{x}. \tag{2.2}$$

We then build an acceptance region $A(\boldsymbol{x})$ in the data space (the space in which the observations $\boldsymbol{y}$ live) corresponding to the observations that would not reject H_0, with the condition that H_0 is rejected with probability at most α when it is true. In confidence set building, one builds a subset in parameter space (the space in which the parameters $\boldsymbol{x}$ live) as a function of the data, $\mathcal{C}(\boldsymbol{y})$, such that it contains $\boldsymbol{x}^*$ with probability at least $1 - \alpha$ (over repeated samples of $\boldsymbol{y} \sim P_{\boldsymbol{x}^*}$).

Lifting to the product space of the data and parameter spaces (see Figure 2.1 for an illustration), both tasks amount to the construction of a compatibility region $\mathcal{S}$. For a fixed observation $\boldsymbol{y}$, a confidence set is given by $\mathcal{C}(\boldsymbol{y}) = \{\boldsymbol{x} \colon (\boldsymbol{y}, \boldsymbol{x}) \in \mathcal{S}\}$, and for a fixed parameter $\boldsymbol{x}$, the acceptance region is given by $\mathcal{A}(\boldsymbol{x}) = \{\boldsymbol{y} \colon (\boldsymbol{y}, \boldsymbol{x}) \in \mathcal{S}\}$. Observe that $\mathbb{P}(\boldsymbol{y} \in \mathcal{A}(\boldsymbol{x})) = \mathbb{P}(\boldsymbol{x} \in \mathcal{C}(\boldsymbol{y}))$. Therefore, a procedure that forms confidence sets with coverage $1 - \alpha$ for any possible data $\boldsymbol{y}$ also creates a procedure that yields valid hypothesis tests at the level α for any possible parameter value $\boldsymbol{x}$, and vice versa. This observation can be used to create confidence sets as the set of parameter values that would not be rejected by a hypothesis test, a construction known as test inversion (see, e.g., Chapter 7 of [5] or Chapter 5 of [42]).
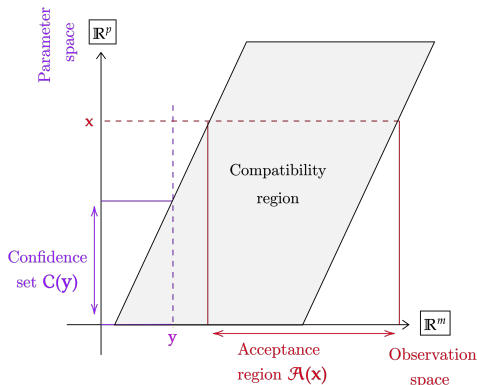

Figure 2.1: Illustration of the classical duality between hypothesis testing and confidence set building as seen in the product space of the data and parameter spaces. Pairs of the dual hypothesis test and the confidence set can be viewed as a set $\mathcal{S}$ in the product space (the compatibility region). For fixed data $\boldsymbol{y}$, a confidence set is given by $\mathcal{C}(\boldsymbol{y}) = \{\boldsymbol{x} : (\boldsymbol{y}, \boldsymbol{x}) \in \mathcal{S}\}$, and for a fixed parameter $\boldsymbol{x}$, the acceptance region is given by $\mathcal{A}(\boldsymbol{x}) = \{\boldsymbol{y} : (\boldsymbol{y}, \boldsymbol{x}) \in \mathcal{S}\}$.

2.2 Formulation and inversion of constrained likelihood ratio tests

The starting point of this work is the inversion of specific hypothesis tests that can incorporate the constraint information $\mathcal{X}$ and the functional of interest φ. We will establish that the test inversion can be achieved by solving two endpoint optimization problems. We note that unlike the simple null versus composite alternative tests (2.2) described in Section 2.1, the tests we will consider have composite nulls. We focus on the continuous case and assume that the Lebesgue measure dominates the set of distributions $\mathcal{P} := \{P_{\boldsymbol{x}} \mid \boldsymbol{x} \in \mathcal{X}\}$. However, a discrete analog can also be constructed using a similar approach as in [13]. Let $L_{\boldsymbol{x}}$ be the density of $P_{\boldsymbol{x}}$, and let $\ell_{\boldsymbol{x}} := \log L_{\boldsymbol{x}}$. For any

$\mu \in \mathbb{R}$, denote the level sets of the quantity of interest φ by Φ_μ. These are defined as follows:

$$\Phi_\mu := \{\boldsymbol{x} : \varphi(\boldsymbol{x}) = \mu\} \subseteq \mathbb{R}^p. \tag{2.3}$$

Subsequently, define a hypothesis test T_μ as follows:

$$H_0 : \boldsymbol{x}^* \in \Phi_\mu \cap \mathcal{X} \quad \text{versus} \quad H_1 : \boldsymbol{x}^* \in \mathcal{X} \setminus \Phi_\mu. \tag{2.4}$$

We can test hypothesis (2.4) (for a fixed μ) with a Likelihood Ratio (LR) test statistic defined as the following function of the observed data $\boldsymbol{y}$:

$$\Lambda(\mu, \boldsymbol{y}) := \frac{\displaystyle\sup_{\boldsymbol{x} \in \Phi_\mu \cap \mathcal{X}} L_{\boldsymbol{x}}(\boldsymbol{y})}{\displaystyle\sup_{\boldsymbol{x} \in \mathcal{X}} L_{\boldsymbol{x}}(\boldsymbol{y})}. \tag{2.5}$$

The corresponding log-likelihood ratio (LLR) statistic $\lambda(\mu, \boldsymbol{y})$ is given by:

$$
\begin{aligned}
\lambda(\mu, \boldsymbol{y}) = -2 \log \Lambda(\mu, \boldsymbol{y}) &= -2 \left\{ \sup_{\boldsymbol{x} \in \Phi_\mu \cap \mathcal{X}} \ell_{\boldsymbol{x}}(\boldsymbol{y}) - \sup_{\boldsymbol{x} \in \mathcal{X}} \ell_{\boldsymbol{x}}(\boldsymbol{y}) \right\} \\
&= \inf_{\boldsymbol{x} \in \Phi_\mu \cap \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}) - \inf_{\boldsymbol{x} \in \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}).
\end{aligned}
\tag{2.6}
$$

As is standard (see, e.g., [5, 65]), we use the supremum over all $\mathcal{X}$ in the denominator of (2.5), instead of over $\mathcal{X} \setminus \Phi_\mu$[7]. The factor of -2 helps connect with the standard likelihood ratio test in the context of Wilks' theorem and is needed, together with the optimization being over the whole space, to reinterpret the previous constrained inference intervals as coming from the inversion of this test (see Section 4).

Motivation behind the choice of test and test statistic. In addition to the reinterpretation of the previous constrained inference intervals as a result of inverting this test, there are other theoretical and practical reasons that make it a reasonable choice for this work. Theoretically, the LR emerges as the optimal test statistic (resulting in the most powerful level-α test) in the simple versus simple hypothesis testing setting via the Neyman–Pearson Lemma [5, 29]. Although uniformly most powerful tests do not exist in general, LR tests have been effective in several contexts. For example, [64] provides some optimality properties for the likelihood ratio test in terms of its asymptotic average power. Although our test of interest does not fall under the simple versus simple paradigm and we are interested in nonasymptotic properties, these two properties support the sensibility of adopting the LR-based test. Furthermore, the literature on constrained inference [44, 53] extensively uses the LR, deriving both asymptotic and nonasymptotic log-likelihood ratio (LLR) distributions in various scenarios, often leading to the $\bar{\chi}^2$ distribution. These characterizations indicate that, in very specific situations, it is possible to obtain the distribution of the test statistic under the null hypothesis, either exactly or in an asymptotic sense. Our setting extends well beyond those situations because our setup is general, and we do not make particular assumptions on the likelihood model or the constraint set.

The distribution of the LLR and test inversion. In hypothesis testing, we reject the null hypothesis when the values of $\lambda(\mu, \boldsymbol{y})$ exceed a threshold. This indicates that there is substantial evidence against the data being generated by a distribution in the composite null defined by μ. To

---

[7][49] provides conditions for the equality of both test statistics.

choose a rejection region, we next study the distribution of the LLR, denoted as $\lambda(\mu, \boldsymbol{y})$, in the context where $\mu = \varphi(\boldsymbol{x})$ (pertaining to the null hypothesis) and $\boldsymbol{y} \sim P_{\boldsymbol{x}}$, a data sampling model, across various values of $\boldsymbol{x} \in \mathcal{X}$. Let $F_{\boldsymbol{x}}$ denote the distribution of $\lambda(\varphi(\boldsymbol{x}), \boldsymbol{y})$ for any $\boldsymbol{x} \in \mathcal{X}$, where $\boldsymbol{y} \sim P_{\boldsymbol{x}}$. To simplify the notation, we will write $\lambda \sim F_{\boldsymbol{x}}$ to indicate that an LLR is sampled following the procedure described above.

To ensure an α-level test for test inversion, we need to control the distribution of the test statistic under the null hypothesis. Since the null is composite, the false positive rate must hold for any parameter under the null hypothesis H_0.

Suppose that we are conducting a test T_μ to determine whether $\mu^* = \varphi(\boldsymbol{x}^*)$ equals a particular $\mu \in \varphi(\mathcal{X}) \subseteq \mathbb{R}$, that is, $\boldsymbol{x}^* \in \Phi_\mu \cap \mathcal{X}$. We use $\lambda > q_\alpha$ as the rejection region, where q_α is a predetermined decision threshold. Under the null hypothesis, if the decision threshold satisfies:

$$\sup_{\boldsymbol{x} \in \Phi_\mu \cap \mathcal{X}} \mathbb{P}_{\lambda \sim F_{\boldsymbol{x}}} (\lambda > q_\alpha) \le \alpha \tag{2.7}$$

for all $\alpha \in (0, 1)$, then we say T_μ is a *level-α test*.[8]

Inverting the test with respect to μ will require choosing an appropriate q_α for all μ; henceforth we will denote it as $q_\alpha(\mu)$.

We seek to invert this test using a methodology similar to that outlined in Section 2.1, but adapted to accommodate the composite null hypothesis. The acceptance region is formally defined as:

$$\mathcal{A}_\alpha(\mu) := \{\boldsymbol{y} : \lambda(\mu, \boldsymbol{y}) \le q_\alpha(\mu)\}. \tag{2.8}$$

Subsequently, we define the proposed confidence set for $\mu^* = \varphi(\boldsymbol{x}^*) \in \mathbb{R}$ through test inversion as follows:

$$\mathcal{C}_\alpha(\boldsymbol{y}) := \{\mu : \lambda(\mu, \boldsymbol{y}) \le q_\alpha(\mu)\}. \tag{2.9}$$

We prove in Lemma 2.1 that if (2.7) is satisfied for $\mu^* := \varphi(\boldsymbol{x}^*)$ (that is, T_{μ^*} is a level-α test), the resulting confidence set will have the desired $1 - \alpha$ coverage, thus extending the classical test inversion framework to our specific case.

Lemma 2.1 (Coverage of the inverted test). *Let $\alpha \in (0, 1)$. Let $\boldsymbol{x}^*$ be the true parameter value and μ^* its image under φ. If T_{μ^*} is a level-α test, then*

$$\mathbb{P}_{\boldsymbol{y} \sim P_{\boldsymbol{x}^*}} (\mu^* \in \mathcal{C}_\alpha(\boldsymbol{y})) \ge 1 - \alpha.$$

Proof sketch. The proof is based on a straightforward test inversion argument. For a detailed proof, see Appendix A.1. $\qquad\square$

To ensure that condition (2.7) holds in practice, when $\boldsymbol{x}^*$ and therefore μ^* are both unknown, we need to satisfy this condition for all possible null hypotheses. Specifically, we choose an appropriate $q_\alpha(\mu)$ for each μ to ensure that all hypothesis tests T_μ are level-α. Formally, this is expressed as:

$$\sup_{\mu \in \varphi(\mathcal{X})} \sup_{\boldsymbol{x} \in \Phi_\mu \cap \mathcal{X}} \mathbb{P}_{\lambda \sim F_{\boldsymbol{x}}} (\lambda > q_\alpha(\mu)) \le \alpha. \tag{2.10}$$

---

[8]Here q_α is the decision value corresponding to intervals with a coverage probability of 1-α, aligning with classical textbook notation (see, e.g., [5], [65]). For any random variable Z, we will denote with the subscript α the cutoff points that satisfy $\mathbb{P}(Z > z_\alpha) = \alpha$.

This condition is equivalent[9] to:

$$\sup_{\boldsymbol{x}\in\mathcal{X}} \mathbb{P}_{\lambda\sim F_{\boldsymbol{x}}} \left(\lambda > q_\alpha(\varphi(\boldsymbol{x}))\right) \leq \alpha. \tag{2.11}$$

Although (2.11) lacks the interpretation of (2.10) of having hypothesis tests for each different $\mu \in \varphi(\mathcal{X})$, it simplifies the calculations. We refer to a set of values $q_\alpha(\mu)$ that satisfy (2.11) (or equivalently (2.10)) as *valid values*. Since μ in (2.10) is equal to $\varphi(\boldsymbol{x})$ as $\boldsymbol{x} \in \Phi_\mu$, we can use $q_\alpha(\mu)$ and $q_\alpha(\varphi(\boldsymbol{x}))$ interchangeably.

From Lemma 2.1, we know that valid values can be used in (2.9) to construct a confidence set for μ^* with the correct $1 - \alpha$ coverage. Moreover, as argued in the proof of Lemma 2.1, the probability that the set (2.9) covers the unknown μ^* is given by:

$$\mathbb{P}_{\boldsymbol{y}\sim P_{\boldsymbol{x}^*}} \left(\mu^* \in \mathcal{C}_\alpha(\boldsymbol{y})\right) = 1 - \mathbb{P}_{\lambda\sim F_{\boldsymbol{x}^*}} \left(\lambda > q_\alpha(\mu^*)\right), \tag{2.12}$$

which is guaranteed to be at least $1-\alpha$ by the condition (2.10). To obtain intervals with the smallest possible size while maintaining coverage, we aim to find the optimal decision values $q_\alpha(\mu)$, which are solutions to optimization problems involving the quantiles $Q_{F_{\boldsymbol{x}}} : [0,1] \to \mathbb{R}$ of the distributions of the family $\{F_{\boldsymbol{x}}, \boldsymbol{x} \in \mathcal{X}\}$.

Lemma 2.2 (Optimal decision values). *The optimal (smallest valid) value of $q_\alpha(\mu)$ is given by the maximum quantile (MQ) optimization problem:*

$$Q_{\mu,1-\alpha}^{\max} := \sup_{\boldsymbol{x}\in\Phi_\mu\cap\mathcal{X}} Q_{F_{\boldsymbol{x}}}(1-\alpha). \tag{2.13}$$

Furthermore, if one wants to choose a single q_α for all μ, then the optimal value is given by:

$$Q_{1-\alpha}^{\max} := \sup_{\mu\in\varphi(\mathcal{X})} Q_{\mu,1-\alpha}^{\max} = \sup_{\boldsymbol{x}\in\mathcal{X}} Q_{F_{\boldsymbol{x}}}(1-\alpha). \tag{2.14}$$

Proof sketch. It can be directly checked that the proposed quantities are valid and that using any smaller decision value leads to intervals with undercoverage for at least one point. See Appendix A.2 for more details. $\square$

In the event that only an upper bound on the quantities defined in Lemma 2.2 can be obtained, those can also be used as valid decision values, as stated precisely in the following corollary.

Corollary 2.3. *For every $\mu \in \mathbb{R}$, $\nu(\mu) \geq Q_{\mu,1-\alpha}^{\max}$ (as defined in Lemma 2.2) if and only if $\nu(\mu)$ is a valid decision value for T_μ for a particular α. In addition, $\nu \geq Q_{1-\alpha}^{\max}$ if and only if ν is a valid decision value for T_μ for all $\mu \in \mathbb{R}$ for a particular α.*

The result follows immediately from substituting the proposed ν values into the probability statements in Lemma 2.2. Theoretical and computational methods for obtaining valid $q_\alpha(\mu)$ are discussed in Section 3, and we investigate the computation of $\mathcal{C}_\alpha(\boldsymbol{y})$ via optimization techniques in the next subsection, assuming valid $q_\alpha(\mu)$ are known.

---

[9]Note that every $\boldsymbol{x} \in \mathcal{X}$ is accounted for in (2.10) since $\mu = \varphi(\boldsymbol{x})$.

2.3 Characterizing the inverted confidence set via optimization problems

The set defined in (2.9) produces a random collection of real numbers that contains the true functional value with a probability of at least $1 - \alpha$. Although this set is not necessarily an interval, it is contained within an interval whose (possibly infinite) extremes are computable through optimization techniques.

Given a valid $q_\alpha(\mu)$, which satisfies either (2.10) or (2.11), let us define the following sets:

$$\mathcal{D}(\boldsymbol{y}) := \{\boldsymbol{x}: -2\ell_{\boldsymbol{x}}(\boldsymbol{y}) \leq q_\alpha(\varphi(\boldsymbol{x})) + \inf_{\boldsymbol{x}' \in \mathcal{X}} -2\ell_{\boldsymbol{x}'}(\boldsymbol{y})\} \subseteq \mathbb{R}^p, \tag{2.15}$$

$$\bar{\mathcal{X}}_\alpha(\boldsymbol{y}) := \mathcal{X} \cap \mathcal{D}(\boldsymbol{y}). \tag{2.16}$$

If $\bar{\mathcal{X}}_\alpha(\boldsymbol{y}) \neq \emptyset$, we further define:

$$\mathcal{I}_\alpha(\boldsymbol{y}) := \left[\inf_{\boldsymbol{x} \in \bar{\mathcal{X}}_\alpha(\boldsymbol{y})} \varphi(\boldsymbol{x}), \sup_{\boldsymbol{x} \in \bar{\mathcal{X}}_\alpha(\boldsymbol{y})} \varphi(\boldsymbol{x}) \right]. \tag{2.17}$$

If $\bar{\mathcal{X}}_\alpha(\boldsymbol{y}) = \emptyset$, let $\mathcal{I}_\alpha(\boldsymbol{y})$ be the empty interval.

Theorem 2.4 (From test inversion to optimization-based intervals)**.** *For any $\alpha \in (0, 1)$, and for any $\boldsymbol{x} \in \mathcal{X}$, let $\mathcal{I}_\alpha(\boldsymbol{y})$ be the interval constructed according to (2.17). It holds that*

$$\mathbb{P}_{\boldsymbol{y} \sim P_{\boldsymbol{x}}} \left(\varphi(\boldsymbol{x}) \in \mathcal{I}_\alpha(\boldsymbol{y}) \right) \geq 1 - \alpha.$$

In other words, $\mathcal{I}_\alpha(\boldsymbol{y})$ is a valid $1 - \alpha$ confidence interval for $\varphi(\boldsymbol{x}^*)$.

Proof sketch. We prove that

$$\left[\inf_{\lambda(\mu, \boldsymbol{y}) \leq q_\alpha(\mu)} \mu, \sup_{\lambda(\mu, \boldsymbol{y}) \leq q_\alpha(\mu)} \mu \right] = \mathcal{I}_\alpha(\boldsymbol{y}). \tag{2.18}$$

The object on the left-hand side is defined by enclosing $\mathcal{C}_\alpha(\boldsymbol{y})$ within the smallest possible interval that contains it, and therefore, it has guaranteed coverage. The equality arises from the equivalence between the optimization problems under consideration. For a complete proof, see Appendix A.3. ☐

Remark 1 (Comparison with the simultaneous strict bound intervals)**.** Observe that the construction of $\mathcal{I}_\alpha(\boldsymbol{y})$ follows the form outlined in (1.3) for the simultaneous strict bound intervals. However, a key distinction lies in not requiring that $\mathcal{D}(\boldsymbol{y}) \subseteq \mathcal{X}$ serves as a $1 - \alpha$ confidence set for $\boldsymbol{x}$. This relaxation will translate into shorter intervals when q_α is chosen appropriately.

Remark 2 (Handling empty constrained sets)**.** If α is chosen such that $1 - \alpha$ becomes too small, the set $\bar{\mathcal{X}}_\alpha(y)$ can be empty. In that case, we default to the empty interval, under the interpretation that there are no parameter values that simultaneously agree with the constraint and the observed data (at a particular level α). However, the actual interval produced under this circumstance does not compromise the $1 - \alpha$ coverage level provided by the theorem. If a point estimate inside the constraint region is desired, an option is to choose the closest point from $\mathcal{X}$ to $\mathcal{D}$. This point specifically ensures the continuity of the interval with respect to α in many standard scenarios. Generally, an empty set $\bar{\mathcal{X}}_\alpha(\boldsymbol{y})$ should inform one of three possibilities: either (i) an outlier event has been observed, or (ii) the initial assumption that $\boldsymbol{x} \in \mathcal{X}$ is flawed, or (iii) the forward model $P_{\boldsymbol{x}}$ is misspecified. Here, the definition of an "outlier" is intrinsically linked to the choice of α. A larger α will make such events more frequent, as it broadens the range of data considered as outliers.

We also present a partial converse result, stating that the interval coverage implies the validity of q_α, subject to appropriate assumptions on φ, P, and $\mathcal{X}$. This result will be instrumental in refuting the coverage claims of the Rust–Burrus intervals, and consequently, the Burrus conjecture, as discussed in Section 4.

Proposition 2.5 (Coverage implies validity of quantile levels). *Assume that $\mathcal{X}$ forms a convex cone, $\ell_x(y)$ is a concave function, and $\varphi(x)$ is linear. Define $\mathcal{I}_\alpha(y)$ as in Theorem 2.4, for a particular choice of $q_\alpha(\mu)$. If $\mathcal{I}_\alpha(y)$ is a valid $1 - \alpha$ confidence interval for all x, then the values of $q_\alpha(\mu)$ are valid.*

Proof sketch. Generally, the values of $q_\alpha(\mu)$ are valid if and only if $\mathcal{C}_\alpha(y)$ constitutes a $1 - \alpha$ set. Since $\mathcal{I}_\alpha(y)$ is the smallest interval that contains $\mathcal{C}_\alpha(y)$, if $\mathcal{C}_\alpha(y)$ is already an interval, then the result holds. The assumptions on $\mathcal{X}, \ell_x(y)$ and φ ensure that this is the case by the convexity of the function

$$\mu \mapsto \inf_{\substack{\varphi(x)=\mu \\ x \in \mathcal{X}}} -2\ell_x(y)$$

for any y. For a detailed proof, see Appendix A.4. $\qquad\square$

Finally, we remark that the construction presented in this paper provides an approach to uncertainty quantification that does not rely on a specific point estimator, distinguishing it from many other UQ procedures. However, if one wishes to obtain a point estimator, it is worth noting that the midpoint of the interval can be justified from a decision-theoretic perspective. This idea has been discussed in previous works by [33, 2], among others.

2.4 Illustrative examples

To elucidate the general methodology outlined in Theorem 2.4, we offer two simple illustrative examples where the LLR and its distribution are explicitly computable: a one-dimensional constrained Gaussian scenario and an unconstrained linear Gaussian case.

Constrained Gaussian in one dimension. As a tangible example, consider the following one-dimensional model:

$$\underbrace{y = x^* + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0,1)}_{\text{model}} \quad \text{with} \quad \underbrace{x^* \geq 0}_{\text{constraints}} \quad \text{and} \quad \underbrace{\varphi(x^*) = x^*}_{\text{functional}}. \tag{2.19}$$

In this case, the distribution of the LLR is precisely known. Hence, a confidence interval can be constructed without resorting to the techniques introduced in Section 3, which are otherwise necessary when such information is not available.

The form of the hypothesis test T_μ, as given in (2.4), is as follows:

$$H_0 : x^* = \mu \quad \text{versus} \quad H_1 : x^* \neq \mu \text{ and } x^* \geq 0. \tag{2.20}$$

The LLR as defined in (2.6) for the test (2.20) is given by:

$$\begin{aligned} \lambda(\mu, y) &= \inf_{x=\mu, x \geq 0} (y-x)^2 - \inf_{x \geq 0} (y-x)^2 \\ &= \begin{cases} (y-\mu)^2, & y \geq 0, \\ (y-\mu)^2 - y^2, & y < 0. \end{cases} \end{aligned} \tag{2.21}$$

We can also derive its distribution under the null hypothesis (i.e., when $x^* = \mu$, leading to $y = \mu + \varepsilon$) for any $\mu \in [0, \infty)$, as formalized below.

Example 2.6 (Distribution of the LLR statistic for a constrained Gaussian in one dimension)**.** *For $\lambda(\mu, y)$ as defined in* (2.21) *with $\mu \geq 0$, when $y \sim \mathcal{N}(\mu, 1)$ (null hypothesis), for all $c > 0$, we have*

$$\mathbb{P}(\lambda(\mu, y) \leq c) = \begin{cases} \chi_1^2(c) + \frac{1}{2}, & \mu = 0 \\ \chi_1^2(c) \cdot \mathbf{1}\{c < \mu^2\} + \{\Phi(\sqrt{c}) - \Phi((-\mu^2 - c)/(2\mu))\} \cdot \mathbf{1}\{c \geq \mu^2\}, & \mu > 0, \end{cases}$$

where χ_1^2 and Φ are the CDFs of a χ_1^2 and a standard Gaussian, respectively.

Proof. See Appendix A.5. $\qquad\square$

The expression for $\lambda(\mu, y)$, with the appropriately scaled log transformation, is equivalent to Equation (4.3) in [13] where the Neyman confidence interval construction for the same problem is considered. [13] characterizes this quantity as a likelihood ordering for determining an acceptance region.

By virtue of the previous result and Lemma 2.2, we can take $Q_\mu(1 - \alpha)$, where Q_μ is the quantile of the distribution of $\lambda(\mu, y)$ when μ is fixed, as $q_\alpha(\mu)$ satisfying (2.7). A direct computation shows

$$q_\alpha(\mu) = Q_\mu(1 - \alpha) = \begin{cases} Q_{\chi_1^2}(1 - \alpha), & 1 - \alpha < \chi_1^2(\mu^2), \\ r_{\mu, \alpha}, & 1 - \alpha \geq \chi_1^2(\mu^2), \end{cases} \tag{2.22}$$

where $r_{\mu, \alpha}$ is the unique nonnegative root of the function $x \mapsto \Phi(\sqrt{x}) - \Phi((-\mu^2 - x)/(2\mu)) - (1 - \alpha)$, which can be found using numerical methods. Therefore, $\mathcal{D} = \{x : (y - x)^2 \leq q_\alpha(x) + \min_{x' \geq 0}(y - x')^2\}$ and the final form of the confidence interval becomes:

$$\mathcal{I}_\alpha(y) = \left[\min_{\substack{x \in \mathcal{D} \\ x \geq 0}} x, \ \max_{\substack{x \in \mathcal{D} \\ x \geq 0}} x \right].$$

For a numerical comparison of this interval with alternative methods, we refer the reader to Section 5.1.

Unconstrained Gaussian linear model. Consider the following problem setup:

$$\underbrace{y = Kx^* + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, I_m)}_{\text{model}} \quad \text{and} \quad \underbrace{\varphi(x^*) = h^\top x^*}_{\text{functional}}. \tag{2.23}$$

Assume $K \in \mathbb{R}^{m \times p}$ has full column rank. The assumption $\text{Cov}(y) = I_m$ is without loss of generality as it is equivalent to assuming a known positive definite covariance for y and performing a basis change with the Cholesky factor. Note that this setup is the same as (1.4) but the parameter space is not constrained, that is, $\mathcal{X} = \mathbb{R}^p$, and the forward model K is assumed to be full rank.

Using the framework established in Section 2.2, our aim is to invert the following family of hypothesis tests:

$$H_0 : h^\top x^* = \mu \quad \text{versus} \quad H_1 : h^\top x^* \neq \mu. \tag{2.24}$$

The LLR as defined in (2.6) for the test (2.24) takes the form:

$$\lambda(\mu, y) := \min_{x : h^\top x = \mu} \|y - Kx\|_2^2 - \min_x \|y - Kx\|_2^2. \tag{2.25}$$

In this particular scenario, the LLR admits a closed-form expression and has a straightforward distribution, as formalized below:

Example 2.7 (Distribution of the LLR statistic for the unconstrained Gaussian linear model). $\lambda(\mu, \boldsymbol{y})$ *for the unconstrained full column rank Gaussian linear model* (2.25) *can be expressed in closed form as*

$$\lambda(\mu, \boldsymbol{y}) = \frac{(\boldsymbol{h}^\top (\boldsymbol{K}^\top \boldsymbol{K})^{-1} \boldsymbol{K}^\top \boldsymbol{y} - \mu)^2}{\boldsymbol{h}^\top (\boldsymbol{K}^\top \boldsymbol{K})^{-1} \boldsymbol{h}}. \tag{2.26}$$

Furthermore, for any $\boldsymbol{x}^*$, whenever $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{K}\boldsymbol{x}^*, \boldsymbol{I}_m)$, $\lambda(\boldsymbol{h}^\top \boldsymbol{x}^*, \boldsymbol{y})$ is distributed as a chi-squared distribution with 1 degree of freedom.

Proof. See Appendix A.6. □

Using the above results, we can set $q_\alpha(\mu) = Q_{\chi_1^2}(1 - \alpha)$ for all values of μ. Here, $Q_{\chi_1^2}$ represents the quantile function of a chi-squared distribution with 1 degree of freedom. Consequently, we can express the interval in (2.17) as:

$$\mathcal{I}_\alpha(\boldsymbol{y}) = \left[\min_{\boldsymbol{x} \in \mathcal{D}(\boldsymbol{y})} \boldsymbol{h}^\top \boldsymbol{x}, \ \max_{\boldsymbol{x} \in \mathcal{D}(\boldsymbol{y})} \boldsymbol{h}^\top \boldsymbol{x} \right], \tag{2.27}$$

where we define $\mathcal{D}(\boldsymbol{y}) := \{\boldsymbol{x} \colon \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2 \leq Q_{\chi_1^2}(1 - \alpha) + \min_{\boldsymbol{x}'} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}'\|_2^2\}$. Similarly, let us define $z_\alpha = \Phi^{-1}(1 - \alpha)$, where Φ is the cumulative distribution function of the standard normal distribution. Using the equivalence $z_{\alpha/2}^2 = Q_{\chi_1^2}(1 - \alpha)$, we can rewrite the expression in terms of the standard normal. Moreover, as shown in Appendix A of [43], the endpoints of the above interval can be calculated in closed form and are given by:

$$\mathcal{I}_\alpha(\boldsymbol{y}) = \left[\boldsymbol{h}^\top \widehat{\boldsymbol{x}} - z_{\alpha/2} \sqrt{\boldsymbol{h}^\top (\boldsymbol{K}^\top \boldsymbol{K})^{-1} \boldsymbol{h}}, \ \boldsymbol{h}^\top \widehat{\boldsymbol{x}} + z_{\alpha/2} \sqrt{\boldsymbol{h}^\top (\boldsymbol{K}^\top \boldsymbol{K})^{-1} \boldsymbol{h}} \right], \tag{2.28}$$

where we define the least-squares estimator $\widehat{\boldsymbol{x}} = (\boldsymbol{K}^\top \boldsymbol{K})^{-1} \boldsymbol{K}^\top \boldsymbol{y}$. This interval is equivalent to the one derived from observing that $\widehat{\boldsymbol{x}} \sim \mathcal{N}(\boldsymbol{x}^*, (\boldsymbol{K}^\top \boldsymbol{K})^{-1})$. Therefore, we have $\boldsymbol{h}^\top \widehat{\boldsymbol{x}} \sim \mathcal{N}(\boldsymbol{h}^\top \boldsymbol{x}^*, \boldsymbol{h}^\top (\boldsymbol{K}^\top \boldsymbol{K})^{-1} \boldsymbol{h})$. The interval in (2.28) is thus a standard construction of a Gaussian $1 - \alpha$ confidence interval. Our construction therefore coincides with the classical interval in this case where a guaranteed-coverage interval can be obtained with standard manipulations; however, our framework remains valid in constrained, rank-deficient, non-Gaussian and/or nonlinear problems where few alternative approaches are available.

3 General interval construction methodology

In this section, we outline the core practical methodology for constructing intervals derived from Theorem 2.4. To summarize the preceding, Lemma 2.1 asserts that if we know $q_\alpha(\mu)$ satisfying (2.11), we can invert the hypothesis test with a composite null hypothesis defined in (2.4) to yield a valid $1 - \alpha$ confidence interval. Lemma 2.2 poses two optimization problems that, if solved, yield valid decision values. The optimization problems in Lemma 2.2 give rise to two approaches of increasing complexity: (i) finding a single q_α that is valid for any μ and (ii) finding valid $q_\alpha(\mu)$ dependent on $\mu \in \varphi(\mathcal{X})$. While approach (ii) can lead to tighter intervals, it is usually at the cost of more complex theoretical analysis and computations, including the computational complexity of solving the optimization problems in (2.17). In particular, when we reinterpret previously proposed optimization-based methods in Section 4, we will observe that these previously proposed methods are of type (i), which this work expands to accommodate type (ii) generalizations. In this section, we briefly analyze the hardness of the optimization problems (2.13) and (2.14) by connecting them

to the chance-constrained optimization literature. In case solving these problems is impractical, in Section 3.2, we describe using stochastic dominance as a theoretical tool that can be used to create and analyze provable upper bounds to the optimization problems. Stochastic dominance will also be used in Section 4 as the main technique to disprove coverage of the previously proposed Rust–Burrus intervals. This section is summarized in a meta-algorithm, detailed in Section 3.3.

3.1 Maximum quantile problems as chance-constrained optimization

Lemma 2.2 presents optimization problems for finding the maximum quantiles, which are crucial for the proposed hypothesis test inversion procedure. We show that these problems, of the form $\sup_{\boldsymbol{x} \in \Phi_\mu \cap \mathcal{X}} Q_{F_{\boldsymbol{x}}}(1 - \alpha)$ and $\sup_{\boldsymbol{x} \in \mathcal{X}} Q_{F_{\boldsymbol{x}}}(1 - \alpha)$, can be formulated as chance-constrained optimization (CCO) problems (see [17] for a review of theory and applications of CCO).

Lemma 3.1 (Chance constrained characterization of the max quantile problems). *Let $\mathcal{S} \subseteq \mathcal{X}$. Then the max quantile optimization problem $\sup_{\boldsymbol{x} \in \mathcal{S}} Q_{F_{\boldsymbol{x}}}(1 - \alpha)$ can be equivalently written as the chance constrained optimization problem:*

$$
\begin{aligned}
\sup_{q, \boldsymbol{x}} \quad & q \\
\text{subject to} \quad & \boldsymbol{x} \in \mathcal{S} \\
& q \in \mathbb{R} \\
& \mathbb{P}_{u \sim \mathcal{U}([0,1])}(\mathcal{F}(\boldsymbol{x}, u) \leq q) \leq 1 - \alpha
\end{aligned}
\tag{3.1}
$$

where $\mathcal{F}(\boldsymbol{x}, u) = F_{\boldsymbol{x}}^{-1}(u)$, with $F_{\boldsymbol{x}}^{-1}$ being the (possibly generalized) inverse CDF of $F_{\boldsymbol{x}}$

Proof. By using the definition of $(1-\alpha)$-quantile of X as the maximum q such that $\mathbb{P}(X \leq q) \leq 1-\alpha$, and $\lambda \sim F_{\boldsymbol{x}} \stackrel{d}{=} \mathcal{F}(\boldsymbol{x}, u \sim \mathcal{U}([0,1])$ □

Note that Lemma 3.1 applies to both (2.13) and (2.14) by choosing appropriate $\mathcal{S}$. In general, CCO problems are known to be strongly NP-hard [17], even with convexity assumptions for $\mathcal{F}$. Although various algorithms exist for general chance-constrained optimization, we leave the development of an algorithm specific to this problem and comparison with the aforementioned algorithms for future work. In our numerical examples (see Section 5), we solve these problems using gradient-free optimizers that do not exploit the chance-constrained structure but instead see the quantile function as a noisy black-box function to be optimized, with evaluations performed by estimating quantiles from large amount of samples. In higher dimensional scenarios, more advanced techniques tailored to the chance-constrained structure might be required. One can also write optimization problem (3.1) and the interval optimization problem (2.17) jointly as one chance-constrained optimization problem; see Appendix B.2.

3.2 Analytical ways to obtain quantile levels via stochastic dominance

In this subsection, we develop an analytical tool to find valid q_α that allows for a straightforward evaluation for any confidence level $1 - \alpha \in (0, 1)$. We first consider the case where we aim to choose a valid q_α for all μ. We propose taking $q_\alpha = Q_X(1 - \alpha)$, where Q_X is the quantile function of a random variable X with a known, easy-to-compute distribution. We establish that for the resulting confidence interval to maintain a $1-\alpha$ coverage guarantee for any α, X must stochastically dominate the random variable with distribution $F_{\boldsymbol{x}^*}$, i.e., $\lambda(\mu^*, \boldsymbol{y})$ where $\boldsymbol{y}$ is a random variable with distribution $P_{\boldsymbol{x}^*}$. This is denoted as $X \succeq \lambda(\mu^*, \boldsymbol{y})$ or, with slight abuse of notation, as $X \succeq F_{\boldsymbol{x}^*}$.

Following the classical definition of stochastic dominance for real-valued random variables (see, e.g., [51]), we say that $X \succeq Y$ if and only if $\mathbb{P}(X \geq z) \geq \mathbb{P}(Y \geq z)$ [10] for all $z \in \mathbb{R}$

Lemma 3.2 (Valid quantile level via stochastic dominance). $Q_X(1 - \alpha)$ *serves as a valid (in the sense of* (2.11)*) choice for* q_α *for all* α *if and only if* $X \succeq \lambda(\mu^*, \boldsymbol{y})$, *where* $\boldsymbol{y} \sim P_{\boldsymbol{x}^*}$.

Proof. See Appendix B.1. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

Remark 3 (Partial validity of quantile levels). If X does not stochastically dominate $\lambda(\mu^*, \boldsymbol{y})$, a valid q_α can still be identified for specific α levels, provided that certain conditions are met. Specifically, z can serve as a valid q_α where $\alpha = 1 - F_X(z)$ and F_X being the cumulative distribution function of X, if and only if $\mathbb{P}(X \leq z) \leq \mathbb{P}(Y \leq z)$ for some value of z.

Remark 4 (Support restriction). Candidates for X can be restricted to the range $[0, \infty)$ without loss of generality, as $\lambda(\mu^*, \boldsymbol{y})$ is supported on this range by moving the mass a candidate X might have in $(-\infty, 0)$ to 0.

An economic interpretation of our result is that agents with nondecreasing utility functions would prefer a reward drawn from X over one from $\lambda(\mu^*, \boldsymbol{y})$. In practical scenarios where the true parameter $\boldsymbol{x}^*$ is unknown, it is required to establish stochastic dominance for the entire family of distributions $F_{\boldsymbol{x}}$, where $\boldsymbol{x} \in \mathcal{X}$.

Although all stochastically dominant distributions provide correct coverage when used to obtain q_α, a larger stochastic dominance gap provides more conservative bounds. Furthermore, if X_1, X_2 both stochastically dominate the family $F_{\boldsymbol{x}}$ for all $\boldsymbol{x} \in \mathcal{X}$, we can take the pointwise minimum $q_\alpha = \min\{Q_{X_1}(1 - \alpha), Q_{X_2}(1 - \alpha)\}$ which will be no worse than using either X_1 or X_2.

The perspective of stochastic dominance also enables the use of coupling arguments to identify stochastically dominating distributions. For instance, one approach to find stochastically dominating distributions to a given $F_{\boldsymbol{x}}$ is finding a function $g(\varphi(\boldsymbol{x}), \boldsymbol{y})$ such that for all z,

$$\mathbb{P}(g(\varphi(\boldsymbol{x}), \boldsymbol{y}) \geq z) \geq \mathbb{P}(\lambda(\varphi(\boldsymbol{x}), \boldsymbol{y}) \geq z),$$

where the randomness is from $\boldsymbol{y} \sim P_{\boldsymbol{x}}$. A particular case is that of nonrandom bounds. If $g(\varphi(\boldsymbol{x}), \boldsymbol{y}) \geq \lambda(\varphi(\boldsymbol{x}), \boldsymbol{y})$ almost surely in $\boldsymbol{y}$ (as opposed to when $\boldsymbol{y} \sim P_{\boldsymbol{x}}$), then this implies a coupling of random variables once $\boldsymbol{y}$ is sampled that implies stochastic dominance (see e.g. Theorem 4.2.3 in [45]).

This technique can be generalized to find $q_\alpha(\mu)$. Instead of finding a stochastic dominant variable X such that $X \succeq F_{\boldsymbol{x}}$ for all $\boldsymbol{x} \in \mathcal{X}$, we aim to find a distribution X_μ for each μ, such that $X_\mu \succeq F_{\boldsymbol{x}}$ for all $\boldsymbol{x} \in \Phi_\mu \cap \mathcal{X}$, and then set $q_\alpha(\mu) = Q_{X_\mu}(1 - \alpha)$. This ensures that $Q_{X_{\mu^*}} \succeq F_{\boldsymbol{x}^*}$, providing the desired coverage guarantees.

As an illustration, we revisit the one-dimensional constrained example discussed in Section 2.4. We consider the model $y = x^* + \varepsilon$, where $\varepsilon \sim \mathcal{N}(0, 1)$, $x^* \geq 0$, and $\varphi(x) = x$. We recall that we have $\lambda(\mu, y) = (y - \mu)^2 - \mathbf{1}(y < 0)y^2$ and an analytical solution for the quantile of the distribution of $\lambda(\mu, y)$ for every μ, which we can use as valid q_α. We extend the results in Example 2.6 below to prove that the distribution of $\lambda(\mu, y)$ is stochastically dominated by a χ_1^2. See Figure 3.1 for an illustration.

---

[10] One can equivalently define stochastic dominance with strict inequalities $X > z$ and $Y > z$, see [51]
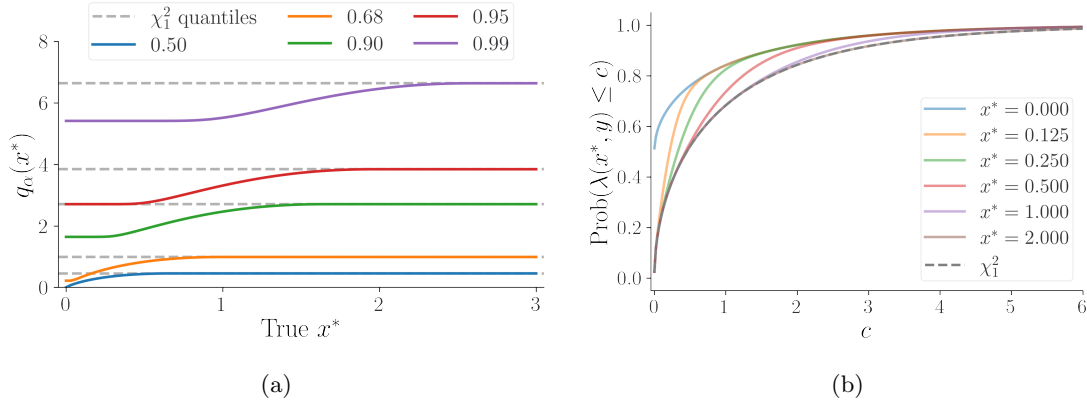
(a)   (b)

Figure 3.1: Comparison of quantile functions and CDFs for LLRs with different true parameter values. The **left** panel provides the true values of the quantile function as a function of x^* across different confidence levels. As proven in Example 3.3, the quantile of χ_1^2 is greater than the true quantile for all x^* and all levels. The **right** panel shows the CDFs for LLRs under different values of the true parameters, x^*. From Example 3.3, as the true parameter increases, the CDF is increasingly dominated by its χ_1^2 component, so it follows that as x^* increases, the CDF approaches the χ_1^2 CDF. This figure also provides a visual explanation of why using the true quantile or the true quantile function to compute the interval in (2.17) produces shorter intervals compared to those computed with the χ_1^2 quantile.

Example 3.3 (Stochastic dominance for LLR for constrained Gaussian in one dimension). *For the LLR $\lambda(\mu, y)$, when $y \sim \mathcal{N}(\mu, 1)$ under the null hypothesis, we have that, for $Z \sim \chi_1^2$, $Z \succeq \lambda(\mu, y)$ for all $\mu \geq 0$.*

Proof. See Appendix B.3. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

For this example, given the stochastic dominance result, we can define $1 - \alpha$ confidence intervals using $\chi_{1,1-\alpha}^2$ instead of using $q_\alpha(\mu)$. This produces larger intervals than using the true quantile, but the true quantile in the closed form will generally be unavailable in more complex examples, while the presented stochastic dominance tools can still be used. The intervals using the χ_1^2 quantile are:

$$
\mathcal{I}_\alpha(y) := \begin{array}{ll} \min / \max\limits_{x} & x \\ \text{subject to} & x \geq 0 \\ & (x - y)^2 \leq \chi_{1,1-\alpha}^2 + \min\limits_{x' \geq 0} \ (x' - y)^2. \end{array} \tag{3.2}
$$

3.3   General confidence interval construction

In this section, we present our meta-algorithm that uses the methodologies described in the preceding sections. The goal of this meta-algorithm is to construct a $1 - \alpha$ confidence interval for a given quantity of interest $\varphi(\boldsymbol{x}^*)$. The algorithmic steps are outlined in Algorithm 1.

It is worth noting that the optimization problems defined in (3.4) and (3.5) may not always be convex or straightforward to solve. However, their dual formulations can be constructed, offering provably valid confidence intervals for any feasible dual solution [56]. We defer the exploration of specialized optimization techniques specifically tailored to solve (3.4) and (3.5) to future work.

19

Algorithm 1 Meta-algorithm for confidence interval construction

---

Input: Observed data $\boldsymbol{y}$, log-likelihood model $\ell_{\boldsymbol{x}}(\boldsymbol{y})$, quantity of interest functional φ, constraint set $\mathcal{X}$, miscoverage level α.

1: **Test statistic**: Write down the LLR test statistic

$$\lambda(\mu, \boldsymbol{y}) \;=\; \inf_{\boldsymbol{x} \in \Phi_\mu \cap \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}) - \inf_{\boldsymbol{x} \in \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}). \tag{3.3}$$

2: **Distribution control**: Control $F_{\boldsymbol{x}}$, the distribution of $\lambda(\varphi(\boldsymbol{x}), \boldsymbol{y})$ where $\boldsymbol{y} \sim P_{\boldsymbol{x}}$, for all $\boldsymbol{x} \in \mathcal{X}$, by either:

    A. *Explicit solution*: Obtain $F_{\boldsymbol{x}}$ explicitly, and let $q_\alpha(\mu) := \sup_{\boldsymbol{x} \in \Phi_\mu \cap \mathcal{X}} Q_{F_{\boldsymbol{x}}}(1 - \alpha)$.

    B. *Computational way to directly find valid q_α* (Section 3.1): Solve $\sup_{\boldsymbol{x} \in \Phi_\mu \cap \mathcal{X}} Q_{F_{\boldsymbol{x}}}(1 - \alpha)$ (to set $q_\alpha(\mu)$) or $\sup_{\boldsymbol{x} \in \mathcal{X}} Q_{F_{\boldsymbol{x}}}(1 - \alpha)$ (to set q_α) numerically.

    C. *Analytical way using stochastic dominance* (Section 3.2): Construct a distribution X that stochastically dominates $F_{\boldsymbol{x}}$ for all $\boldsymbol{x} \in \mathcal{X}$, and let $q_\alpha := Q_X(1 - \alpha)$, or construct distributions X_μ that stochastically dominate $F_{\boldsymbol{x}}$ for all $\boldsymbol{x} \in \Phi_\mu \cap \mathcal{X}$ and let $q_\alpha(\mu) := Q_{X_\mu}(1 - \alpha)$.

3: **Confidence interval calculation**: Obtain the confidence intervals by solving the pair of optimization problems that is easier in the particular case:

    I. Parameter space formulation:

$$\begin{aligned} \min_{\boldsymbol{x}} / \max_{\boldsymbol{x}} \quad & \varphi(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{x} \in \mathcal{X} \\ & -2\ell_{\boldsymbol{x}}(\boldsymbol{y}) \leq q_\alpha(\varphi(\boldsymbol{x})) + \inf_{\boldsymbol{x}' \in \mathcal{X}} -2\ell_{\boldsymbol{x}'}(\boldsymbol{y}). \end{aligned} \tag{3.4}$$

    II. Functional space formulation:

$$\begin{aligned} \min_{\mu} / \max_{\mu} \quad & \mu \\ \text{subject to} \quad & \mu \in \varphi(\mathcal{X}) \subseteq \mathbb{R} \\ & \inf_{\boldsymbol{x} \in \Phi_\mu \cap \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}) - \inf_{\boldsymbol{x} \in \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}) \leq q_\alpha(\mu). \end{aligned} \tag{3.5}$$

Output: Confidence interval with coverage $1 - \alpha$.

---

4 Refuting the Burrus conjecture

As discussed in Section 1, the family of constrained problems that has received the most attention is the positivity-constrained version of the problem as described in Section 2.4. To recap, the model is defined as follows:

$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{K}\boldsymbol{x}^*, \boldsymbol{I}_m) \quad \text{with} \quad \mathcal{X} = \{\boldsymbol{x} : \boldsymbol{x} \geq \boldsymbol{0}\} \quad \text{and} \quad \varphi(\boldsymbol{x}^*) = \boldsymbol{h}^\top \boldsymbol{x}^*. \tag{4.1}$$

Here $\boldsymbol{K} \in \mathbb{R}^{m \times p}$ is the forward linear operator. We again emphasize here that $\boldsymbol{K}$ need not have full column rank, so we can for example have $p > m$. It was initially conjectured in [4, 46] that a

valid $1 - \alpha$ confidence interval could be obtained as

$$
\begin{aligned}
\min_{\boldsymbol{x}} / \max_{\boldsymbol{x}} \quad & \boldsymbol{h}^\top \boldsymbol{x} \\
\text{subject to} \quad & \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2 \leq \psi_\alpha^2 \\
& \boldsymbol{x} \geq \boldsymbol{0}.
\end{aligned} \tag{4.2}
$$

Here, $\psi_\alpha^2 = z_{\alpha/2}^2 + s^2(\boldsymbol{y})$, with $z_{\alpha/2}$ being the previously defined standard Gaussian quantile, and $s^2(\boldsymbol{y})$ is defined as the optimal value of

$$
\begin{aligned}
\min_{\boldsymbol{x}} \quad & \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2 \\
\text{subject to} \quad & \boldsymbol{x} \geq \boldsymbol{0}.
\end{aligned}
$$

Although initially believed to be proved in [47], an error in the proof was later identified in [61], along with a counterexample. However, we demonstrate that this counterexample actually satisfies the conjecture, leaving the conjecture unresolved until now prior to our work, to the best of our knowledge.

The main result of this section is the construction of a new valid counterexample using the test inversion perspective developed in Section 2 and the stochastic dominance approach of Section 3.2, disproving the conjecture.

Theorem 4.1 (Refutation of the Burrus conjecture). *The Burrus conjecture is false in general. The two-dimensional example previously proposed of a particular instance of* (4.1) *in [61],*

$$
\boldsymbol{K} = \boldsymbol{I}_2 \quad and \quad \boldsymbol{h} = (1, -1)^\top \quad with \quad \boldsymbol{x}^* = (a, a)^\top \text{ such that } a \geq 0,
$$

does not constitute a valid counterexample to the Burrus conjecture. However, the following constitutes a valid counterexample for the Burrus conjecture:

$$
\boldsymbol{K} = \boldsymbol{I}_3 \quad and \quad \boldsymbol{h} = (1, 1, -1)^\top \quad with \quad \boldsymbol{x}^* = (0, 0, 1)^\top.
$$

The main idea of the proof is to first connect the conjecture to our framework, identifying the conjectured intervals as a particular case of our construction with a particular choice of q_α. We then apply Proposition 2.5 to show that coverage is equivalent to a valid choice of q_α. Finally, we present a counterexample to prove that the proposed q_α is not universally valid. The proof is divided into several lemmas for clarity.

Our approach is novel in that it diverges from previous geometric perspectives on the Gaussian likelihood [46, 47, 37], instead leveraging the test inversion and stochastic dominance perspectives developed in Section 2 and Section 3.2.

4.1 Proof outline of Theorem 4.1

This subsection provides a structured outline of the proof for Theorem 4.1, which refutes the Burrus conjecture. We break down the proof into several key lemmas.

Lemma 4.2 (Framing the Burrus conjecture as test inversion). *The construction of intervals in* (4.2) *for a particular instance of the problem* $(\boldsymbol{x}^*, \boldsymbol{K}, \boldsymbol{h})$ *is equivalent to the general construction in Theorem 2.4 for the model* $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{K}\boldsymbol{x}^*, \boldsymbol{I}_m)$, *with* $\boldsymbol{x}^* \geq \boldsymbol{0}$ *component wise, and* $\varphi(\boldsymbol{x}) = \boldsymbol{h}^\top \boldsymbol{x}$, *using*

the threshold $q_\alpha(\mu) = z^2_{\alpha/2}$ independent of μ. Therefore, it is equivalent to inverting a hypothesis test $H_0 : \boldsymbol{h}^\top \boldsymbol{x} = \mu$ versus $H_1 : \boldsymbol{h}^\top \boldsymbol{x} \neq \mu$ with LLR

$$\lambda(\mu, \boldsymbol{y}) := \min_{\substack{\boldsymbol{h}^\top \boldsymbol{x} = \mu \\ \boldsymbol{x} \geq \boldsymbol{0}}} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2 - \min_{\boldsymbol{x} \geq \boldsymbol{0}} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2. \tag{4.3}$$

Furthermore, the interval has correct coverage if and only if $q_\alpha = z^2_{\alpha/2}$ is valid in the sense of satisfying the false positive guarantee (2.10).

Proof. See Appendix C.2. $\qquad\qquad\square$

Lemma 4.3 (Reducing the Burrus conjecture to stochastic dominance)**.** *The construction of intervals in* (4.2) *has the right coverage for any α (and hence the conjecture holds) for a particular instance of the problem $(\boldsymbol{x}^*, \boldsymbol{K}, \boldsymbol{h})$ if and only if the log-likelihood ratio test statistic*

$$\lambda(\mu = \boldsymbol{h}^\top \boldsymbol{x}^*, \boldsymbol{y}) := \min_{\substack{\boldsymbol{h}^\top \boldsymbol{x} = \boldsymbol{h}^\top \boldsymbol{x}^* \\ \boldsymbol{x} \geq \boldsymbol{0}}} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2 - \min_{\boldsymbol{x} \geq \boldsymbol{0}} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2$$

is stochastically dominated by a χ_1^2 distribution whenever $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{K}\boldsymbol{x}^*, \boldsymbol{I}_m)$.

Proof. See Appendix C.3. $\qquad\qquad\square$

As an example, the constrained one-dimensional example considered in Section 2.4 satisfies the stochastic dominance result and hence the conjecture. Furthermore, using Example 2.7, an alternative characterization of the conjecture is the stochastic dominance of the unconstrained LLR test statistic $\min_{\boldsymbol{h}^\top \boldsymbol{x} = \boldsymbol{h}^\top \boldsymbol{x}^*} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2 - \min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2$ over the constrained test statistic $\min_{\substack{\boldsymbol{h}^\top \boldsymbol{x} = \boldsymbol{h}^\top \boldsymbol{x}^* \\ \boldsymbol{x} \geq \boldsymbol{0}}} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2 - \min_{\boldsymbol{x} \geq \boldsymbol{0}} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2$.

We use Lemma 4.3 to prove both that the example in [61] obeys the conjecture and that our new counterexample does not.

Invalidity of a previous counterexample in two dimensions. The previously proposed counterexample from [61] is a two-dimensional problem with $\boldsymbol{K} = \boldsymbol{I}_2$, $\boldsymbol{x}^* = (a, a)^\top$ with $a \geq 0$, $\boldsymbol{h} = (1, -1)^\top$ (and therefore $\mu^* = \boldsymbol{h}^\top \boldsymbol{x}^* = 0$). The LLR test statistic is

$$\lambda(\mu^* = 0, \boldsymbol{y}) = \min_{\substack{x_1 = x_2 \\ \boldsymbol{x} \geq \boldsymbol{0}}} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 - \min_{\boldsymbol{x} \geq \boldsymbol{0}} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2$$

which, after solving the optimization problems, is equal to

$$\lambda(\mu^*, \boldsymbol{y}) = \begin{cases} y_1^2 + y_2^2 - (y_1 - \max(y_1, 0))^2 - (y_2 - \max(y_2, 0))^2, & y_1 + y_2 < 0, \\ \frac{1}{2}(y_1 - y_2)^2 - (y_1 - \max(y_1, 0))^2 - (y_2 - \max(y_2, 0))^2, & y_1 + y_2 \geq 0, \end{cases}$$

which we can equivalently write as

$$\lambda(\mu^*, \boldsymbol{y}) = (y_1^2 + y_2^2)\mathbb{1}\{y_1 + y_2 < 0\} + \frac{1}{2}(y_1 - y_2)^2 \mathbb{1}\{y_1 + y_2 \geq 0\} \tag{4.4}$$
$$- y_1^2 \mathbb{1}\{y_1 < 0\} - y_2^2 \mathbb{1}\{y_2 < 0\}.$$

Lemma 4.4 (Invalidity of a previous counterexample). *The LLR statistic $\lambda(\mu^*, \boldsymbol{y})$ in (4.4) is stochastically dominated by a χ_1^2 random variable whenever $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{x}^*, \boldsymbol{I}_2)$, $\boldsymbol{x}^* = (a, a)^\top$ for $a \geq 0$, and $\boldsymbol{h} = (1, -1)^\top$. Therefore, it does not constitute a valid counterexample to the conjecture.*

Proof sketch. The proof follows from a coupling argument between the LLR and a χ_1^2 random variable. See Appendix C.4 for proof details. $\qquad\square$

In summary, we used Lemma 4.3 to demonstrate that the previously proposed counterexample actually satisfies the conjecture.

A new provably valid counterexample in three dimensions. We now present a new counterexample in $\mathbb{R}^3$ to refute the Burrus conjecture. Specifically, we consider $\boldsymbol{K} = \boldsymbol{I}_3$, $\boldsymbol{x}^* = (0, 0, 1)^\top$, and $\boldsymbol{h} = (1, 1, -1)^\top$, yielding $\mu^* = -1$. We prove that χ_1^2 does not stochastically dominate $\lambda(\mu^*, \boldsymbol{y})$, which in this case is

$$\lambda(\mu^* = -1, \boldsymbol{y}) = \min_{\substack{x_1 + x_2 - x_3 = -1 \\ \boldsymbol{x} \geq \boldsymbol{0}}} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 - \min_{\boldsymbol{x} \geq \boldsymbol{0}} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2. \tag{4.5}$$

We prove that $\mathbb{E}[\lambda(\mu^*, \boldsymbol{y})] > \mathbb{E}[\chi_1^2] = 1$. Here, the expectation is taken with respect to $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{x}^*, \boldsymbol{I}_3)$, and the inequality is a general sufficient condition to refute stochastic dominance and hence for the conjecture to break.

Lemma 4.5 (Validity of a new counterexample). *$\lambda(\mu^*, \boldsymbol{y})$ in (4.5) is not stochastically dominated by a χ_1^2 random variable whenever $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{x}^*, \boldsymbol{I}_3)$ with $\boldsymbol{x}^* = (0, 0, 1)^\top$. Therefore, it constitutes a valid counterexample to the general conjecture.*

Proof sketch. We compute the expected value and show that it is greater than 1 (the expected value of a χ_1^2), therefore proving stochastic dominance. See Appendix C.5 for the proof details. $\quad\square$

Remark 5 (A more general counterexample). The validity of the counterexample does not hinge on $\boldsymbol{x}^*$ being on the boundary of the constraint set. In fact, the example remains valid for $\boldsymbol{x}^* = (\varepsilon, \varepsilon, 1)^\top$ with $\varepsilon > 0$ sufficiently small. We choose $\varepsilon = 0$ for the simplicity of the proof. See Figure 5.5 for numerical evidence, where the quantiles over the dashed line correspond to valid counterexamples.

Figure 4.1 shows the difference between the two examples. By plotting the difference between the CDF of λ (obtained numerically with $N = 10^6$ samples) and the CDF of a χ_1^2 distribution, we observe stochastic dominance for the two-dimensional example in Figure 4.1 (left panel) and no stochastic dominance (hence breaking of the conjecture) for the three-dimensional example in Figure 4.1 (right panel). Section 5 contains numerical coverage studies for both scenarios agreeing with the observation made here.

4.2 A negative result in high dimensions

After establishing that the χ_1^2 distribution fails to stochastically dominate the constrained log-likelihood ratio, a natural question arises: Is there another distribution, possibly within the χ_k^2 family, that can stochastically dominate the constrained LLR? If such a distribution exists, it would allow us to redefine ψ_α^2 in (4.2) as $s^2 + Q_X(1 - \alpha)$, making the Q_X term in the optimization problem dimension independent, leading to intervals with shorter length in large dimensions. It is worth noting that in the unconstrained scenario, the LLR distribution is precisely χ_1^2, regardless of the
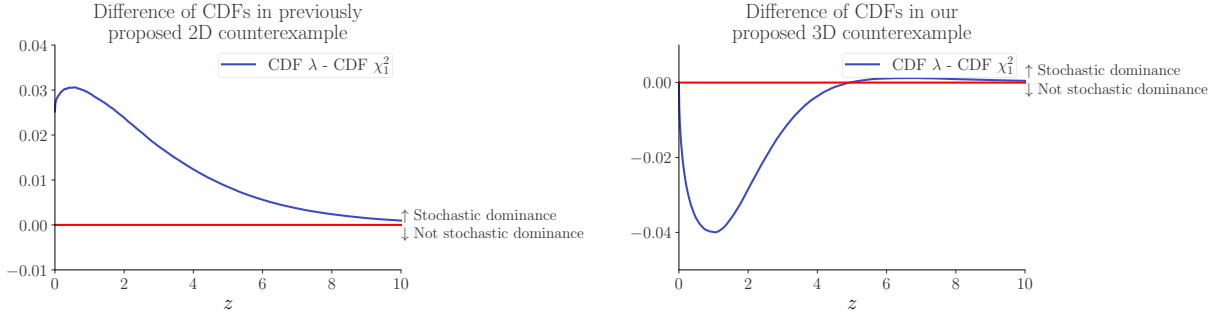
Figure 4.1: Difference of cumulative distribution functions between the LLR test statistic and χ_1^2 distribution for the statistics defined in (4.4) **(left)** and (4.5) **(right)**. Stochastic dominance, which is equivalent to the Burrus conjecture, is broken in the right example only. There is a direct correspondence between the points at which the CDF difference is negative and confidence levels $1 - \alpha$ that fail to hold (see Remark 3).

dimensionality of the problem. However, the following proposition shows that no such dimension-independent distribution exists for the constrained case.

Proposition 4.6 (A negative result in high dimensions)**.** *The family of constrained LLRs for general $\boldsymbol{K}, \boldsymbol{h}$ in arbitrary dimensions, defined as*

$$\lambda(\mu = \boldsymbol{h}^\top \boldsymbol{x}^*, \boldsymbol{y}) \;=\; \min_{\substack{\boldsymbol{h}^\top \boldsymbol{x} = \boldsymbol{h}^\top \boldsymbol{x}^* \\ \boldsymbol{x} \geq \boldsymbol{0}}} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2 - \min_{\boldsymbol{x} \geq \boldsymbol{0}} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2,$$

cannot be stochastically dominated a dimension-independent way by any finite-mean distribution (including all χ_k^2 for $k \geq 1$).

Proof sketch. We construct a sequence of examples with increasing dimensions and demonstrate that the expected value of the constrained LLR grows unbounded as the dimension increases. This result negates the possibility of stochastic dominance by any finite-mean distribution. For a detailed proof, see Appendix C.6. $\qquad\square$

5 Numerical examples

In this section, we provide numerical illustrations for the procedures and theoretical results described above. In particular, we analyze coverage properties of four types of intervals. The first two intervals come from previous works: $\mathcal{I}_{\mathsf{SSB}}$ (1.3), which has provably correct coverage but is known to overcover when inferring a single functional, and $\mathcal{I}_{\mathsf{OSB}}$ (1.5), which comes from the Burrus conjecture and, as we proved in this work, does not correctly cover in general. The last two are the interval constructions described in this work by solving the max quantile problems (2.13) and (2.14) to find $q_\alpha(\mu)$ (Section 3.3): $\mathcal{I}_{\mathsf{MQ}}$ (where q_α does not depend on μ) and the more refined $\mathcal{I}_{\mathsf{MQ}\mu}$ (where q_α depends on μ). Both of these intervals have provable coverage.

We analyze four settings in which the Burrus conjecture applies, including a one-dimensional example in Section 5.1, the previously proposed invalid counterexample in Section 5.2 (for which we prove in Lemma 4.4 that $\mathcal{I}_{\mathsf{OSB}}$ correctly covers), our proposed counterexample in Section 5.3 (for which we prove in Lemma 4.5 that $\mathcal{I}_{\mathsf{OSB}}$ does not cover for at least some level $1 - \alpha$ and a certain $\boldsymbol{x}^*$), and a setting with a bounded constraint set[11] $\mathcal{X}$ in Section 5.4.

---

[11]Strictly speaking, this setting is not included in the original Burrus conjecture but is later extended in [43, 54].

Throughout the examples, the observed coverage (or lack thereof) agrees with the developed theoretical results, and we observe that our interval $\mathcal{I}_{\mathsf{MQ}\mu}$ consistently fixes the miscalibration of the other interval types: when $\mathcal{I}_{\mathsf{OSB}}$ undercovers, $\mathcal{I}_{\mathsf{MQ}\mu}$ is on average longer than $\mathcal{I}_{\mathsf{OSB}}$ to obtain coverage, and when $\mathcal{I}_{\mathsf{OSB}}$ overcovers, $\mathcal{I}_{\mathsf{MQ}\mu}$ is on average shorter than $\mathcal{I}_{\mathsf{OSB}}$ with coverage closer to the prescribed level $1 - \alpha$.

5.1 Constrained Gaussian in one dimension

We revisit the constrained Gaussian model in one dimension (2.19) described in Section 2.4, $y = x^* + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, 1)$, $x^* \geq 0$ and $\varphi(x) = x$ We perform a simulation experiment using six true parameter settings of $x^* \in \{0, 2^{-3}, 2^{-2}, 2^{-1}, 2^0, 2^1\}$. We focus on settings closer to the boundary since that is where the biggest differences between the considered intervals exist. For each of these settings, we simulate 10^5 observations according to the model (2.19) and compute three different 95% confidence intervals for each sample: interval (2.17) using the actual quantile function given in (2.22) ($\mathcal{I}_{\mathsf{MQ}_\mu}$[12]), interval (2.17) using the stochastically dominating $Q_{\chi_1^2}(1 - \alpha)$ quantile ($\mathcal{I}_{\mathsf{OSB}}$, which in this problem is equal to $\mathcal{I}_{\mathsf{MQ}}$), and the standard Truncated Gaussian interval, which equals the SSB interval in this case ($\mathcal{I}_{\mathsf{SSB}}$). The intervals computed with the true quantile function are characterized by:

$$
\begin{aligned}
\mathcal{I}_{\mathsf{MQ}_\mu}(y) := \quad & \min_x / \max_x \quad x \\
& \text{subject to} \quad x \geq 0 \\
& \qquad\qquad (x - y)^2 \leq q_\alpha(x) + \min_{x \geq 0}(x - y)^2,
\end{aligned}
\tag{5.1}
$$

where $q_\alpha(x)$ is given by (2.22). For the stochastically dominating $Q_{\chi_1^2}(1 - \alpha)$, the interval in (2.17) becomes:

$$
\begin{aligned}
\mathcal{I}_{\mathsf{OSB}}(y) = \quad & \min_x / \max_x \quad x \\
& \text{subject to} \quad x \geq 0 \\
& \qquad\qquad (x - y)^2 \leq Q_{\chi_1^2}(1 - \alpha) + \min_{x \geq 0}(x - y)^2.
\end{aligned}
\tag{5.2}
$$

Finally, the truncated Gaussian interval, which is shown below to be equivalent to the SSB interval in this case, is defined as:

$$
\mathcal{I}_{\mathsf{SSB}}(y) := \left[y - z_{\alpha/2}, \, y + z_{\alpha/2} \right] \cap \mathbb{R}_{\geq 0}.
\tag{5.3}
$$

Observe that (5.2) admits an explicit solution:

$$
\mathcal{I}_{\mathsf{OSB}}(y) = \begin{cases}
[y - \sqrt{Q_{\chi_1^2}(1 - \alpha)}, \, y + \sqrt{Q_{\chi_1^2}(1 - \alpha)}] \cap \mathbb{R}_{\geq 0}, & y \geq 0 \\
[y - \sqrt{Q_{\chi_1^2}(1 - \alpha) + y^2}, \, y + \sqrt{Q_{\chi_1^2}(1 - \alpha) + y^2}] \cap \mathbb{R}_{\geq 0}, & y < 0.
\end{cases}
\tag{5.4}
$$

Furthermore, note that $\sqrt{Q_{\chi_1^2}(1 - \alpha)} = z_{\alpha/2}$, so that (5.4) is always larger than or equal to (5.3). Conversely, we can express (5.3) as the solution to optimization problems, illustrating that the truncated Gaussian interval is equivalent to the SSB interval for this case:

$$
\begin{aligned}
\mathcal{I}_{\mathsf{SSB}}(y) = \quad & \min_x / \max_x \quad x \\
& \text{subject to} \quad x \geq 0 \\
& \qquad\qquad (x - y)^2 \leq z_{\alpha/2}^2.
\end{aligned}
\tag{5.5}
$$

---

[12]Since this is a one-dimensional problem with $\varphi(x) = x$, this is equivalent to being able to solve all the $\mu-$dependant quantile optimization problems
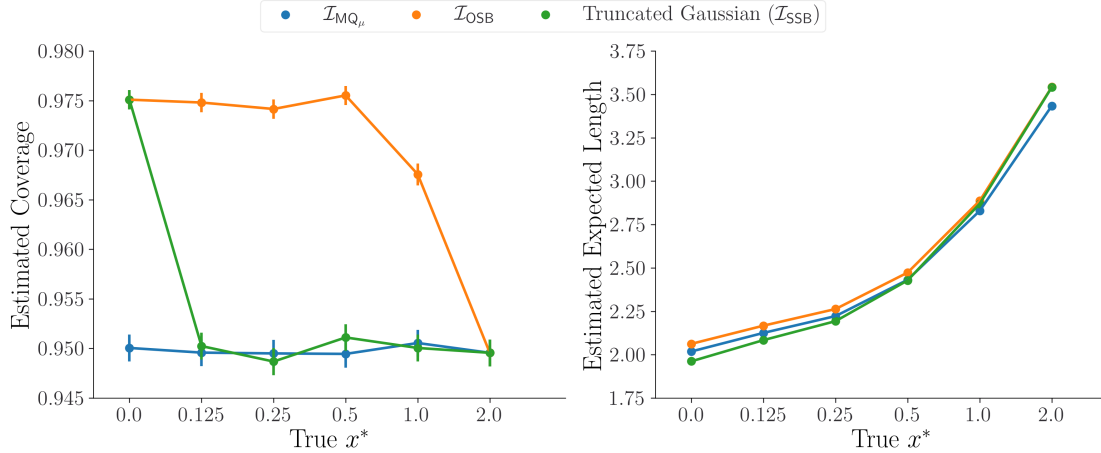
Figure 5.1: The **(left)** figure shows estimated coverage for each 95% confidence interval and each true x^* for the one-dimensional constrained Gaussian model. Both the Truncated Gaussian (SSB) and OSB intervals overcover when $x^* = 0$ while the MQμ interval predictably achieves nominal coverage. All the coverage values converge as x^* gets larger, as the problem moves toward the unconstrained problem where all the intervals are effectively the same. The intervals surrounding the estimated values are 95% Clopper–Pearson intervals, expressing the Monte Carlo uncertainty of each coverage estimate. The **(right)** figure shows an estimate of the expected interval length for each method (with 95% confidence intervals that are nearly length zero since the standard error of each estimate is nearly zero with 10^5 realizations each). Similarly to the coverage results in the left panel, as x^* gets larger, the expected OSB and MQμ interval lengths converge while the SSB intervals remain slightly larger.

To empirically estimate coverage, for each x^* setting and each interval type, we compute 10^5 intervals and keep track of their coverage of the true parameter. The left panel in Figure 5.1 shows how $\mathcal{I}_{\mathsf{OSB}}$ based on $Q_{\chi_1^2}(1-\alpha)$ over-covers when the true parameter is on the boundary, which makes sense as this setting of q_α holds for all x^*, and therefore is a conservative quantile. As expected, the interval computed with $q_\alpha(x)$ maintains the nominal 95% coverage over all considered x^* values. This shows that knowing the quantile function means that we can compute an interval with exact nominal coverage that is adaptive to the unknown true parameter value. Additionally, we note that as x^* grows, the estimated coverage values across these methods converge, illustrating the intuition that when x^* gets sufficiently far from the constraint boundary, the problem is essentially unconstrained, and all considered methods produce nearly identical results. The right panel in Figure 5.1 shows each interval's expected length as a function of x^*. Again, we observe the tightness of the interval (2.17) constructed with the true quantile function $q_\alpha(x)$ compared to the interval constructed with the stochastically dominating quantile, $Q_{\chi_1^2}(1 - \alpha)$. Similarly to coverage, as x^* grows, the expected interval lengths of $\mathcal{I}_{\mathsf{MQ}_\mu}$ and $\mathcal{I}_{\mathsf{OSB}}$ converge, and the methods become indistinguishable. Also, observe that the truncated Gaussian intervals have a smaller expected length compared to the intervals computed with $q_\alpha(x)$. We note that this length observation is particular to this one-dimensional example, as OSB intervals have been shown to be shorter than SSB intervals in higher dimensional problems [37, 46, 54].

5.2 Constrained Gaussian in two dimensions

We consider the Gaussian linear model in (1.4) with $\boldsymbol{K} = \boldsymbol{I}_2$, $\varphi(\boldsymbol{x}) = \boldsymbol{h}^\top \boldsymbol{x} = x_1 - x_2$ and $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^2 : \boldsymbol{x} \geq 0\}$. The work [61] proposes this scenario as a counterexample to the Burrus conjecture, but as shown in Lemma 4.4, it is in fact a case where the χ_1^2 distribution stochastically dominates
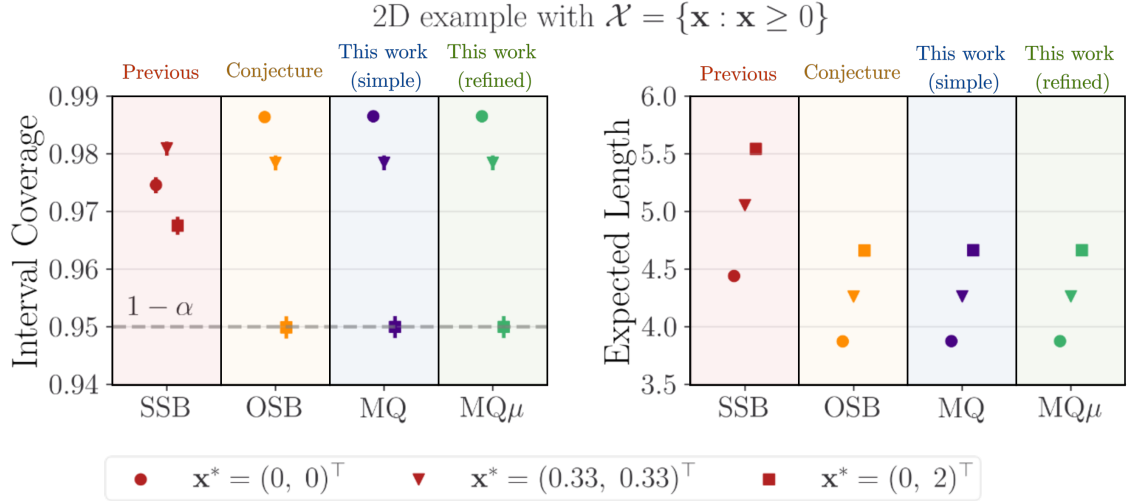
Figure 5.2: Estimated interval coverage **(left)** and expected lengths **(right)** for 95% intervals resulting from the SSB, OSB, MQ and MQμ methods for the Gaussian linear model in (1.4) with $\boldsymbol{K} = \boldsymbol{I}_2$, $\varphi(\boldsymbol{x}) = \boldsymbol{h}^\top \boldsymbol{x} = x_1 - x_2$ and $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^2 : \boldsymbol{x} \geq 0\}$.

the LLR for all true $\boldsymbol{x}^* \in \{\boldsymbol{x} : \boldsymbol{h}^\top \boldsymbol{x} = 0, \, \boldsymbol{x} \geq 0\}$, so the OSB intervals proposed by the conjecture have provably correct coverage for $\boldsymbol{x}^*$ in this set.

We estimate interval coverage with $1 - \alpha = 0.95$ for the four types of intervals ($\mathcal{I}_{\mathsf{SSB}}$, $\mathcal{I}_{\mathsf{OSB}}$, $\mathcal{I}_{\mathsf{MQ}}$ and $\mathcal{I}_{\mathsf{MQ}\mu}$) for three true parameter values, two of them inside the $\{\boldsymbol{x} : \boldsymbol{h}^\top \boldsymbol{x} = 0, \, \boldsymbol{x} \geq 0\}$ region in which Lemma 4.4 applies, and one outside. In this and all the examples that follow, we solve the max quantile optimization problems (2.13), (2.14) using the Bayesian Optimization package BayesOpt [31] (see fig. 5.3 for an illustration of the quantile function being optimized in this particular example). We solve the outer optimization problems with the convex optimization package CVXPY [11, 1] (for the convex problems in $\mathcal{I}_{\mathsf{SSB}}$, $\mathcal{I}_{\mathsf{OSB}}$ and $\mathcal{I}_{\mathsf{MQ}}$) or root-finding numerical algorithms (in the case of $\mathcal{I}_{\mathsf{MQ}\mu}$, where we use the functional space formulation in algorithm 1 and look for those μ satisfying the constraint as equality).

For each parameter value and all intervals, coverage and expected length are estimated by drawing 5×10^4 observations from the data generating process, computing all interval types for each generated observation, and then checking coverage and length. The coverage confidence intervals are 95% Clopper–Pearson intervals for a binomial parameter, whereas the length confidence intervals are standard asymptotic Gaussian intervals using sample means and standard errors.

The results are shown in fig. 5.2. We observe correct coverage for the OSB intervals, in agreement with Lemma 4.4 (which applies for $\boldsymbol{x}^* = (0,0)^\top$ and $\boldsymbol{x}^* = (0.33, 0.33)^\top$). We observe that the SSB intervals are the longest on average and tend to overcover and that the MQ and MQμ intervals have nearly identical properties to the OSB intervals. This is because, for this problem setting, solving the optimization problems (2.13) and (2.14) recovers the χ_1^2 quantile (up to the numerical precision of the optimization solvers) of the Burrus conjecture as the maximum quantile (both for all $\mathcal{X}$ and for $\boldsymbol{h}^\top \boldsymbol{x} = \mu$ for any μ), so both of those intervals actually recover the OSB intervals in this case.
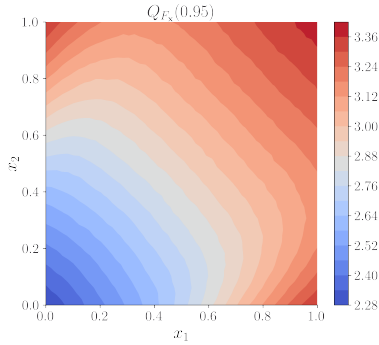
Figure 5.3: Estimated 95th quantiles from the LLR test statistic null distributions in the region $[0,1]^2 \subset \mathbb{R}_+^2$ where color shows the estimated quantiles. In contrast to the unconstrained case, the quantile is dependent on the true parameter, and the quantile surface is non-trivial.

5.3 Constrained Gaussian in three dimensions

We use the three-dimensional counterexample of the Burrus conjecture from Section 4 to numerically show that the $\mathcal{I}_{\mathsf{OSB}}$ intervals undercover in this example and that the max-quantile intervals are able to fix the undercoverage. Concretely, we consider the Gaussian linear model in (1.4) with $\boldsymbol{K} = \boldsymbol{I}_3$, $\varphi(\boldsymbol{x}) = \boldsymbol{h}^\top \boldsymbol{x} = x_1 + x_2 - x_3$ and $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^3 : \boldsymbol{x} \geq 0\}$. We repeat the same experimental setup as in Section 5.2, comparing the four interval types for $\boldsymbol{x}^* = (0,0,0)^\top$ and $\boldsymbol{x}^* = (0,0,1)^\top$; this last parameter value is the one analyzed in Lemma 4.5 and for which we know the OSB interval can undercover for some α. Figure 5.4 shows the results for $1 - \alpha = 0.68$ (one sigma interval coverage), and we include in Appendix D the results for $1 - \alpha = 0.95$, which lead to the same conclusions. Furthermore, to illustrate that the conjecture breaks in an area around the studied point $(0,0,1)^\top$ in Figure 5.5, we plot the numerically estimated 68% LLR test statistic quantiles for parameters of the form $(t,t,1)$, showing that there is a range of points with a larger quantile than the χ_1^2 quantile, implying undercoverage.

The results agree with our theoretical findings in Lemma 4.5, as we observe that the OSB intervals undercover both at 95% and 68% confidence levels, thus invalidating the Burrus conjecture. In contrast, the SSB, MQ and MQμ intervals all have provable coverage in this scenario, which is reflected in the estimated coverage values. Notably, the MQμ intervals are not much longer than the OSB intervals, but they obtain the required coverage, enlarging the conjectured intervals just enough. The simpler MQ intervals overcover a bit more, which illustrates the benefit of solving (2.14) over (2.13) when computationally feasible. Nevertheless, their length is not much larger than MQμ and significantly smaller than for SSB, the other simple method with coverage guarantees.

5.4 Bounded constraint set in two dimensions

As a last case study, we consider a modification of the example in Section 5.2 in which the constraint set is chosen to be the bounded set $\mathcal{X} = [0,1]^2$. While not in the original scope of the Burrus conjecture, which only considers $\mathcal{X} = \{\boldsymbol{x} : \boldsymbol{x} \geq 0\}$, as mentioned in Section 1.1, the $\mathcal{I}_{\mathsf{OSB}}$ interval construction has since then been used with constraints of the form $\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}$ by replacing $\boldsymbol{x} \geq 0$ with $\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}$ in the optimization problems (1.5) and (1.6) [43, 54]. We use the same experimental setup as in Section 5.2, taking into account that the change in $\mathcal{X}$ affects both the interval optimization problem and the optimizations required for $q_\alpha(\mu)$ (since the LLR statistic changes as well).

The results are shown in Figure 5.6. We observe that in this setting, as opposed to the previous three-dimensional problem, the OSB intervals *overcover*, because the maximum quantile of the LLR test statistic over the constraint set is *smaller* than the χ_1^2 quantile. Furthermore, our intervals MQ, and especially MQμ, are able to exploit this fact to obtain shorter intervals than OSB with coverage
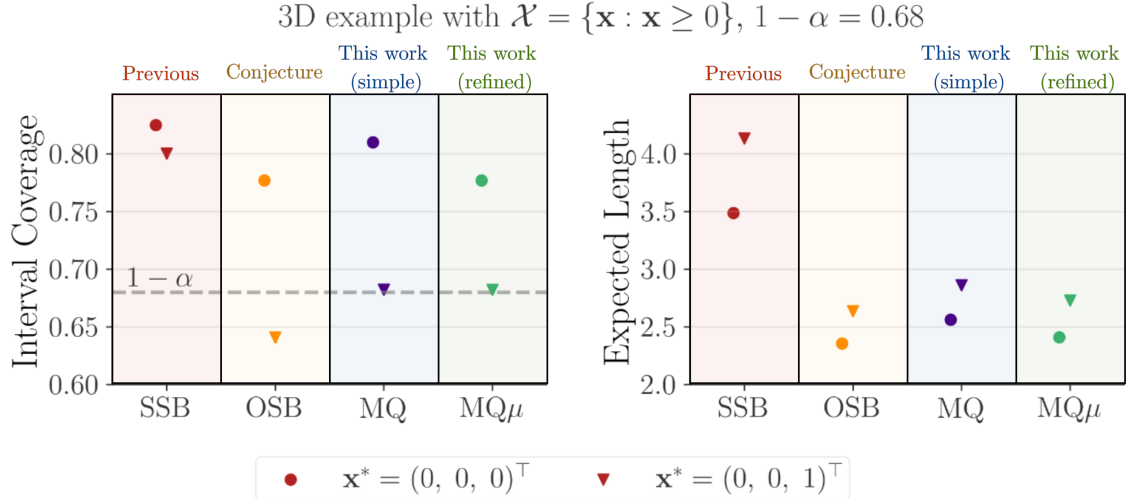
Figure 5.4: Estimated interval coverage **(left)** and expected lengths **(right)** for 68% intervals resulting from the SSB, OSB, MQ and MQμ methods for the Gaussian linear model in (1.4) with $\boldsymbol{K} = \boldsymbol{I}_3$, $\varphi(\boldsymbol{x}) = \boldsymbol{h}^\top \boldsymbol{x} = x_1 + x_2 - x_3$ and $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^3 : \boldsymbol{x} \geq 0\}$.
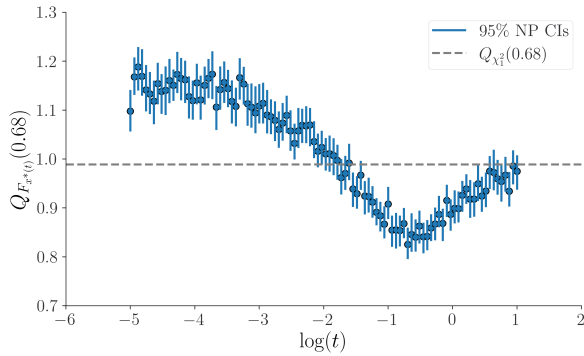

Figure 5.5: For the numerical example in Section 5.3, considering $\boldsymbol{x}^* \in \{\boldsymbol{x}^*(t) = (t, t, 1)^\top : 0 < t \leq e\} \subset \mathbb{R}^3$, we estimate the 68% LLR test statistic quantiles along with 95% nonparametric (NP) confidence intervals for percentiles [20]. The test statistic quantiles exceeding $\chi^2_{1,0.32}$ correspond to the Burrus conjecture failing in this scenario. We note that for this example, the Burrus conjecture fails close to the constraint boundary while the $Q_{\chi^2_1}(0.68)$ quantile becomes valid once sufficiently far from the boundary.

closer to $1 - \alpha$ by using the actual max quantiles over $\mathcal{X}$ instead of the χ^2_1 quantile used by OSB.

6   Discussion

This paper presents a framework for constructing confidence intervals with guaranteed frequentist coverage for a given functional of forward model parameters in the presence of constraints. For the specific case of the Gaussian linear forward model with nonnegativity constraints, we refute the Burrus conjecture [4] by providing a counterexample and propose a more general approach for interval construction. Our approach hinges on the inversion of a specific likelihood ratio test, and we offer theoretical and practical insights into the properties of the constructed intervals via illustrative examples. Our framework is versatile, accommodating potentially nonlinear, non-Gaussian, and rank-deficient settings.

At a high level, the practical effectiveness of UQ methods depends on the (sometimes implicit) assumptions of the method. Different methods come into play depending on what we assume or know, be it the likelihood, the constraints, or the prior in Bayesian settings. In classical statistics, confidence intervals serve as a valuable tool for UQ, especially for one-dimensional quantities of
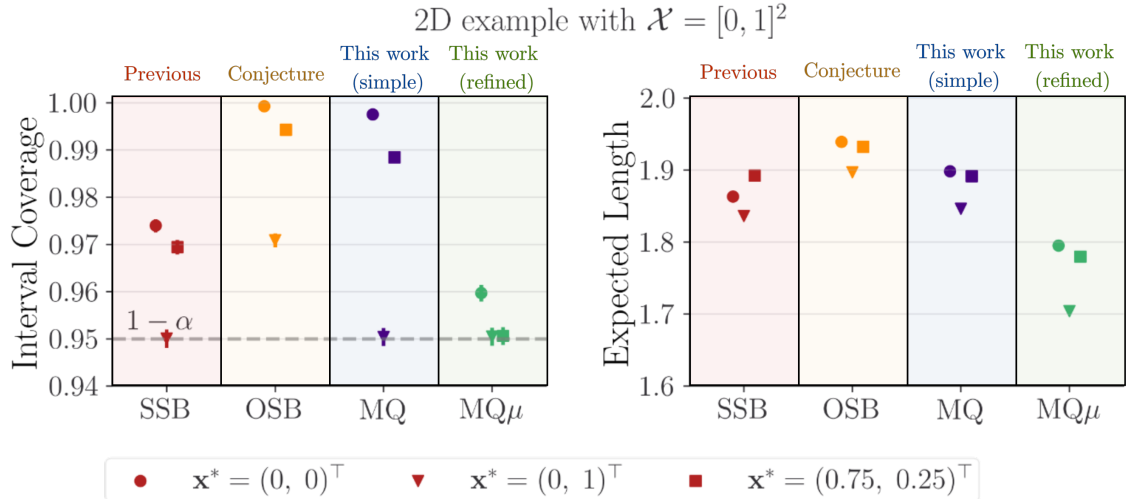
Figure 5.6: Estimated interval coverage **(left)** and expected lengths **(right)** for 95% intervals resulting from the SSB, OSB, MQ and MQμ methods for the Gaussian linear model in (1.4) with $\boldsymbol{K} = \boldsymbol{I}_2$, $\varphi(\boldsymbol{x}) = \boldsymbol{h}^\top \boldsymbol{x} = x_1 - x_2$ and $\mathcal{X} = [0,1]^2 \subset \mathbb{R}^2$.

interest. These intervals are constructed to offer guaranteed coverage under repeated sampling, aligning with frequentist principles. While frequentist coverage guarantees are a useful criterion, especially in contexts where repeatability is essential, we acknowledge that the "best" UQ method is often context-dependent. For example, this frequentist approach is the most natural in applications like remote sensing ([43]), where repeatability is a key requirement. Conversely, when it is natural to think of the parameter as arising from a prior distribution, Bayesian methods are well-motivated and have desirable properties.

A key aim of this paper is to serve as a basis for the future development of these UQ procedures. We conclude this paper by discussing a few possible directions for future work:

- **Data-adaptive calibration procedure.** We saw in Section 5 that MQ is valid where OSB is not and can leverage smaller quantiles to produce tighter intervals. In the event that one does not have the assumed true parameter bounding box, it may be possible to create a data-generated one and adjust the error budget accordingly. Such a procedure would enable us to expand the use of the MQ intervals to scenarios with unbounded parameter constraints. We are currently investigating this approach, which will be the subject of a future follow-up paper.

- **Exploration of high-dimensional problems.** A key benefit of the optimization-based interval construction is that it promises to provide a solution to uncertainty quantification in high-dimensional and ultra-high-dimensional problems, where most alternative approaches (including sampling-based ones) are infeasible. For example, in [54], we applied a precursor of the methods presented here in a problem where $p = 80$, and we are currently exploring the use of these methods in a data-assimilation setting where $p \approx 10^4$. Indeed, optimization is one of the only computational techniques known to work in ultra-high-dimensional problems, such as those in 4DVar data assimilation [25, 15, 30]. While optimization has been successfully used for point estimation in such problems, the approaches developed here may enable modifying the existing programs to obtain confidence intervals in addition to point estimates.

- **Joint confidence sets for multiple functionals.** Since our framework is devised for UQ of

30

a single functional, its application to collections of functionals (a higher-dimensional quantity of interest), would be a natural and desirable extension. Trivially, given a collection of K functionals, one could apply this methodology K times and use the Bonferroni correction to adjust the confidence levels so that they all cover at the desired coverage level. Although this approach might be practically reasonable when K is small, it becomes markedly inefficient as K becomes large. Furthermore, this approach would create a K-dimensional hyper-rectangle for the quantity of interest, which may not be the optimal geometry for bounding the quantity of interest. As such, extending the framework of Section 3 to simultaneously consider the K functionals of interest would be the first step to creating a more nuanced approach. One way this can be achieved is by appropriately adjusting the definition of H_0 in the hypothesis test in (2.4).

- **Choice of test statistics beyond LLR.** The log-likelihood ratio test statistic considered in this work connects with the Rust–Burrus intervals and is observed to perform well in practice, but other choices can be explored in future work. While the LLR is a natural choice for the generic problem, improving the interval length on particular families of problems with different test statistics might be possible. Since the main theoretical machinery comes from the test inversion framework, which is independent of the actual form of the test statistic, alternate versions of Theorem 2.4 can be constructed as long as the test statistic constructs valid level-α hypothesis tests; the resulting intervals of which could be explored theoretically and numerically. For instance, the construction of confidence intervals in Section 2.3, originally written for $\lambda(\mu, \boldsymbol{y}) = \inf_{\boldsymbol{x} \in \Phi_\mu \cap \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}) - \inf_{\boldsymbol{x} \in \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y})$, readily generalizes to test statistics of the form $\lambda_{f,g}(\mu, \boldsymbol{y}) = \inf_{\boldsymbol{x} \in \Phi_\mu \cap \mathcal{X}} f(\boldsymbol{x}, \boldsymbol{y}) - g(\boldsymbol{y})$. In that case, (2.15) becomes $\{\boldsymbol{x} : f(\boldsymbol{x}, \boldsymbol{y}) \leq q_\alpha(\varphi(\boldsymbol{x})) + g(\boldsymbol{y})\}$, where q_α must be valid for the particular choice of f and g, and can be obtained by analyzing the distribution of $\lambda_{f,g}$.

- **Generalization to simulation-based problems.** An extension of our methodology to settings in which the likelihood is not exactly known can be considered, ranging from only partial knowledge of the form of the likelihood to full simulation-based (likelihood-free) settings where the likelihood is not known explicitly but can be sampled from. A possible avenue is to develop robust worst-case approaches with respect to possible likelihoods. In a fully likelihood-free setting, approaches such as [9, 23, 32, 10] provide ways to invert hypothesis tests to obtain confidence sets in these scenarios in multidimensional parameter spaces. Projections of these sets could produce confidence intervals for a functional of the model parameters, as seen for the SSB intervals. However, as we have explored, orienting the hypothesis test to the given functional of interest can have dramatic length benefits for the resulting confidence interval (as seen for the OSB, HSB, and MQ intervals). Since the log-likelihood plays a key role in the definition of our intervals, extensions providing ways to relax that dependence would be a necessary first step.

Acknowledgments

References

[1] Agrawal, A., Verschueren, R., Diamond, S., and Boyd, S. (2018). A rewriting system for convex optimization problems. *Journal of Control and Decision*, 5(1):42–60.

[2] Bajgiran, H. H., Batlle, P., Owhadi, H., Samir, M., Scovel, C., Shirdel, M., Stanley, M., and Tavallali, P. (2022). Uncertainty quantification of the 4th kind: Optimal posterior accuracy-uncertainty tradeoff with the minimum enclosing ball. *Journal of Computational Physics*, 471.

[3] Boucheron, S. and Thomas, M. (2012). Concentration inequalities for order statistics. *Electronic Communications in Probability*, 17:1–12.

[4] Burrus, W. R. (1965). *Utilization of a Priori Information by Means of Mathematical Programming in the Statistical Interpretation of Measured Distributions*. Oak Ridge National Laboratory.

[5] Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Duxbery, second edition.

[6] Cash, W. (1979). Parameter estimation in astronomy through application of the likelihood ratio. *The Astrophysical Journal*, 228:939–947.

[7] Cranmer, K., Brehmer, J., and Louppe, G. (2020). The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062.

[8] Cranmer, K., Pavez, J., and Louppe, G. (2016). Approximating likelihood ratios with calibrated discriminative classifiers. arXiv:1506.02169 [stat.AP].

[9] Dalmasso, N., Izbicki, R., and Lee, A. B. (2020). Confidence sets and hypothesis testing in a likelihood-free inference setting. In *International Conference on Machine Learning*, pages 2323–2334.

[10] Dalmasso, N., Masserano, L., Zhao, D., Izbicki, R., and Lee, A. B. (2023). Likelihood-free frequentist inference: Bridging classical statistics and machine learning for reliable simulator-based inference. *arXiv preprint arXiv:2107.03920*.

[11] Diamond, S. and Boyd, S. (2016). CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5.

[12] Donoho, D. L. (1994). Statistical estimation and optimal recovery. *The Annals of Statistics*, 22(1):238–270.

[13] Feldman, G. J. and Cousins, R. D. (1998). Unified approach to the classical statistical analysis of small signals. *Physical Review D*, 57(7):3873–3889.

[14] Fisher, E., Schweiger, R., and Rosset, S. (2020). Efficient construction of test inversion confidence intervals using quantile regression. *Journal of Computational and Graphical Statistics*, 29(1):140–148.

[15] Forget, G., Campin, J., Heimbach, P., Hill, C. N., Ponte, R. M., and Wunsch, C. (2015). ECCO version 4: an integrated framework for non-linear inverse modeling and global ocean state estimation. *Geoscientific Model Development*.

[16] Garthwaite, P. H. and Buckland, S. T. (1992). Generating Monte Carlo confidence intervals by the Robbins–Monro process. *Journal of the Royal Statistical Society*, 41(1):159–171.

[17] Geng, X. and Xie, L. (2019). Data-driven decision making in power systems with probabilistic guarantees: Theory and applications of chance-constrained optimization. *Annual Reviews in Control*, 47:341–363.

[18] Gouriéroux, C., Holly, A., and Monfort, A. (1982). Likelihood ratio test, Wald test, and Kuhn–Tucker test in linear models with inequality constraints on the regression parameters. *Econometrica*, 50:63–80.

[19] Gutmann, M. U. and Corander, J. (2015). Bayesian optimization for likelihood-free inference of simulator-based statistical models.

[20] Hahn, G. J. and Meeker, W. Q. (1991). *Statistical Intervals*. Wiley.

[21] Hansen, P. C. (1992). Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review*, 34(4):561–580.

[22] Hansen, P. C. and O'Leary, D. P. (1993). The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM Journal on Scientific Computing*, 14(6):1487–1503.

[23] Heinrich, L. (2022). Learning optimal test statistics in the presence of nuisance parameters. *arXiv preprint arXiv:2203.13079*.

[24] Javanmard, A. and Montanari, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *Journal of Machine Learning Research*, 15(82):2869–2909.

[25] Kalmikov, A. G. and Heimbach, P. (2014). A Hessian-based method for uncertainty quantification in global ocean state estimation. *SIAM Journal on Scientific Computing*, 36:S267 – S295.

[26] Kuusela, M. (2016). *Uncertainty quantification in unfolding elementary particle spectra at the Large Hadron Collider*. PhD thesis, École Polytechnique Fédérale de Lausanne.

[27] Kuusela, M. and Panaretos, V. M. (2015). Statistical unfolding of elementary particle spectra: Empirical Bayes estimation and bias-corrected uncertainty quantification. *The Annals of Applied Statistics*, 9(3):1671–1705.

[28] Kuusela, M. and Stark, P. B. (2017). Shape-constrained uncertainty quantification in unfolding steeply falling elementary particle spectra. *Annals of Applied Statistics*.

[29] Lehmann, E. L. and Romano, J. P. (2008). *Testing Statistical Hypotheses*. Springer.

[30] Liu, J., Bowman, K., Meemong, L., et al. (2016). Carbon monitoring system flux estimation and attribution: impact of ACOS-GOSAT X_{CO_2} sampling on the inference of terrestrial biospheric sources and sinks. *Tellus B: Chemical and Physical Meteorology*, 66.

[31] Martinez-Cantin, R. (2014). Bayesopt: A bayesian optimization library for nonlinear optimization, experimental design and bandits.

[32] Masserano, L., Dorigo, T., Izbicki, R., Kuusela, M., and Lee, A. (2023). Simulation-based inference with WALDO: Confidence regions by leveraging prediction algorithms or posterior estimators for inverse problems. In *The International Conference on Artificial Intelligence and Statistics*.

[33] Micchelli, C. A. and Rivlin, T. J. (1977). A survey of optimal recovery. *Optimal estimation in approximation theory*, pages 1–54.

[34] Molenberghs, G. and Verbeke, G. (2007). Likelihood ratio, score, and Wald tests in a constrained parameter space. *The American Statistician*, 61(1):22–27.

[35] Murphy, S. A. (1995). Likelihood ratio-based confidence intervals in survival analysis. *Journal of the American Statistical Association*, 90(432):1399–1405.

[36] Neale, M. C. and Miller, M. B. (1997). The use of likelihood-based confidence intervals in genetic models. *Behavior Genetics*, 27:113–120.

[37] O'Leary, D. P. and Rust, B. W. (1986). Confidence intervals for inequality-constrained least squares problems, with applications to ill-posed problems. *SIAM Journal on Scientific Computing*, 7(2):473–489.

[38] Owhadi, H. and Scovel, C. (2017). Qualitative robustness in Bayesian inference. *ESAIM: Probability and Statistics*, 21:251–274.

[39] Owhadi, H., Scovel, C., and Sullivan, T. (2015a). Brittleness of Bayesian inference under finite information in a continuous world. *Electronic Journal of Statistics*, 9:1–79.

[40] Owhadi, H., Scovel, C., and Sullivan, T. (2015b). On the brittleness of Bayesian inference. *SIAM Review*, 57(4):566–582.

[41] Owhadi, H., Scovel, C., Sullivan, T. J., McKerns, M., and Ortiz, M. (2013). Optimal uncertainty quantification. *SIAM Review*, 55(2):271–345.

[42] Panaretos, V. M. (2016). *Statistics for Mathematicians*. Springer.

[43] Patil, P., Kuusela, M., and Hobbs, J. (2022). Objective frequentist uncertainty quantification for atmospheric CO_2 retrievals. *SIAM/ASA Journal on Uncertainty Quantification*, 10.

[44] Robertson, T., Wright, F. T., and Dykstra, R. (1988). *Order Restricted Statistical Inference*. Wiley.

[45] Roch, S. (2024). *Modern Discrete Probability: An Essential Toolkit*. To be published by Cambridge University Press.

[46] Rust, B. W. and Burrus, W. R. (1972). *Mathematical Programming and the Numerical Solution of Linear Equations*. American Elsevier.

[47] Rust, B. W. and O'Leary, D. P. (1994). Confidence intervals for discrete approximations to ill-posed problems. *Journal of Computational and Graphical Statistics*, 3(1):67–96.

[48] Schafer, C. M. and Stark, P. B. (2009). Constructing confidence regions of optimal expected size. *Journal of the American Statistical Association*, 104(487):1080–1089.

[49] Schervish, M. J. (1995). *Theory of Statistics*. Springer.

[50] Schweiger, R., Fisher, E., Rahmani, E., Shenhav, L., Rosset, S., and Halperin, E. (2018). Using stochastic approximation techniques to efficiently construct confidence intervals for heritability. *Journal of Computational Biology*, 25(7):794–808.

[51] Shaked, M. and Shanthikumar, J. (2007). *Stochastic Orders*. Springer Series in Statistics. Springer New York.

[52] Shapiro, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *International Statistical Review*, pages 49–62.

[53] Silvapulle, M. J. and Sen, P. K. (2011). *Constrained Statistical Inference: Order, Inequality, and Shape Constraints*. Wiley.

[54] Stanley, M., Patil, P., and Kuusela, M. (2022). Uncertainty quantification for wide-bin unfolding: one-at-a-time strict bounds and prior-optimized confidence intervals. *Journal of Instrumentation*, 17.

[55] Stark, P. B. (1992a). Inference in infinite-dimensional inverse problems: Discretization and duality. *Journal of Geophysical Research*, 97(B10):14055–14082.

[56] Stark, P. B. (1992b). Inference in infinite-dimensional inverse problems: discretization and duality. *Journal of Geophysical Research: Solid Earth*, 97(B10):14055–14082.

[57] Stark, P. B. (1994). Simultaneous confidence intervals for linear estimates of linear functionals. Technical report, Citeseer.

[58] Stark, P. B. (2015). Constraints versus priors. *SIAM/ASA Journal on Uncertainty Quantification*, 3.

[59] Stuart, A. M. (2010). Inverse problems: A Bayesian perspective. *Acta Numerica*, 19:451–559.

[60] Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. SIAM.

[61] Tenorio, L., Fleck, A., and Moses, K. (2007). Confidence intervals for linear discrete inverse problems with a non-negativity constraint. *Inverse Problems*, 23(2):669.

[62] Thomas, O., Dutta, R., Corander, J., Kaski, S., and Gutmann, M. U. (2022). Likelihood-Free Inference by Ratio Estimation. *Bayesian Analysis*, 17(1):1–31.

[63] Venzon, D. J. and Moolgavkar, S. H. (1988). A method for computing profile-likelihood-based confidence intervals. *Journal of the Royal Statistical Society: Series C*, 37(1):87–94.

[64] Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54:426–482.

[65] Wasserman, L. (2004). *All of Statistics: A Concise Course in Statistical Inference*. Springer.

[66] Wasserman, L., Ramdas, A., and Balakrishnan, S. (2020). Universal inference. *PNAS*, 117:16880–16890.

[67] Wolak, F. A. (1987). An exact test for multiple inequality and equality constraints in the linear regression model. *Journal of the American Statistical Association*, 82:782–793.

[68] Wolak, F. A. (1989). Testing inequality constraints in linear econometric models. *Journal of Econometrics*, 41:205–235.

[69] Yu, M., Gupta, V., and Kolar, M. (2019). Constrained high dimensional statistical inference. *arXiv preprint arXiv:1911.07319*.

Supplementary Material

This document serves as a supplement to the paper entitled "Optimization-based frequentist confidence intervals for functionals in constrained inverse problems: Resolving the Burrus conjecture". We outline below the structure of the supplement and summarize key notation used both in this supplement and the main paper.

Organization

The organization this supplement is summarized in Table 1.

Appendix	Description
Appendix A	Proofs of Lemma 2.1, Theorem 2.4, Proposition 2.5, and Examples 2.6 and 2.7
Appendix B	Proofs of Lemma 3.2 and example 3.3
Appendix C	Proofs of Theorem 4.1, Lemmas 4.2 to 4.5, and Proposition 4.6
Appendix D	Additional numerical illustrations in Section 5

Table 1: Roadmap of the supplement.

Notation

Some of the general notation used throughout this paper summarized in Table 2. (Any specific notation needed is explained in respective sections as necessary.)

A note about $\min / \max$ versus $\inf / \sup$: We use $\min / \max$ when the optimal value of an optimization problem is attained, otherwise we use $\inf / \sup$.

A note about uniqueness of optimization problems: When we express an equality involving the minimizer of an optimization problem, this is intended to signify a set inclusion. The guarantees presented in our paper are applicable to all solutions, and we make no distinction among multiple solutions.

Notation	Description
Non-bold lower or upper case	Denotes scalars (e.g., α, μ, Q).
Bold lower case	Denotes vectors (e.g., $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{h}$).
Bold upper case	Denotes matrices (e.g., $\boldsymbol{K}$, $\boldsymbol{I}$).
Calligraphic font	Denotes sets (e.g., $\mathcal{X}$, $\mathcal{C}$, $\mathcal{D}$).
$\mathbb{R}$	Set of real numbers.
$\mathbb{R}_{\geq 0}$	Set of non-negative real numbers.
$[n]$	Set $\{1, \ldots, n\}$ for a positive integer n.
(x, y, z)	(Ordered) tuple of elements x, y, z.
$\{x, y, z\}$	Set of elements x, y, z.
$\lceil x \rceil$, $\lfloor x \rfloor$, $(x)_+$	Ceiling, floor, and positive part of a real number x.
$\mathbf{1}\{A\}$, $\mathbb{P}(A)$	Indicator random variable associated with an event A and probability of A.
$\mathbb{E}[X]$, $\mathrm{Var}(X)$	Expectation and variance of a random variable X.
$\boldsymbol{A}^{-1}$	Inverse of a square invertible matrix $\boldsymbol{A} \in \mathbb{R}^{m \times m}$.
$\boldsymbol{b}^\top$, $\boldsymbol{B}^\top$	Transpose of a column vector $\boldsymbol{b} \in \mathbb{R}^m$ and a rectangular matrix $\boldsymbol{B} \in \mathbb{R}^{m \times p}$.
$\mathrm{rank}(\boldsymbol{C})$	Rank of a matrix $\boldsymbol{C} \in \mathbb{R}^{m \times p}$
$\mathrm{null}(\boldsymbol{C})$	Nullity of a matrix $\boldsymbol{C} \in \mathbb{R}^{m \times p}$
$\boldsymbol{D}^{1/2}$	Principal square root of a positive semidefinite matrix $\boldsymbol{D} \in \mathbb{R}^{p \times p}$.
$\boldsymbol{I}$, $\mathbf{1}$, $\mathbf{0}$	The identity matrix, the all ones vector, the all zeros vector of appropriate dimensions
$\langle \mathbf{u}, \mathbf{v} \rangle$	The inner product of vectors $\mathbf{u}$ and $\mathbf{v}$.
$\|\boldsymbol{u}\|_{\boldsymbol{D}}$	The induced ℓ_2 norm of a vector $\boldsymbol{u}$ with respect to a positive semidefinite matrix $\boldsymbol{D}$
$\|\mathbf{u}\|_q$	The ℓ_q norm of a vector $\mathbf{u}$ for $q \geq 1$.
$\|f\|_{L_q}$	The L_q norm of a function f for $q \geq 1$.
$\boldsymbol{u} \leq \boldsymbol{v}$	Lexicographic ordering for vectors $\boldsymbol{u}$ and $\boldsymbol{v}$.
$\boldsymbol{A} \preceq \boldsymbol{B}$	Loewner ordering for symmetric matrices $\boldsymbol{A}$ and $\boldsymbol{B}$.
$X \preceq Y$	Stochastic dominance order for random variables X and Y (see Section 3.2 for details).
$Y = \mathcal{O}_\alpha(X)$	Deterministic big-O notation, indicating that Y is bounded by $|Y| \leq C_\alpha X$ for some constant C_α that may depend on the ambient parameter α.
$\mathcal{O}_p$	Probabilistic big-O notation.
$\xrightarrow{\mathrm{d}}$	Convergence in distribution.
$\xrightarrow{\mathrm{p}}$	Convergence in probability.

Table 2: Summary of general notation used throughout the paper and the supplement.

A Proofs in Section 2

A.1 Proof of Lemma 2.1

To prove the lemma, we need to show that the probability of μ^* being in the confidence set $\mathcal{C}_\alpha(\boldsymbol{y})$ is at least $1 - \alpha$. Towards this end, observe that

$$
\begin{aligned}
\mathbb{P}_{\boldsymbol{y} \sim P_{\boldsymbol{x}^*}}(\mu^* \in \mathcal{C}_\alpha(\boldsymbol{y})) &= \mathbb{P}_{\boldsymbol{y} \sim P_{\boldsymbol{x}^*}}(\boldsymbol{y} \in A_\alpha(\mu^*)) \\
&= 1 - \mathbb{P}_{\boldsymbol{y} \sim P_{\boldsymbol{x}^*}}(\boldsymbol{y} \notin A_\alpha(\mu^*)) \\
&\geq 1 - \sup_{\boldsymbol{x} \in \Phi_{\mu^*} \cap \mathcal{X}} \mathbb{P}_{\boldsymbol{y} \sim P_{\boldsymbol{x}}}(\boldsymbol{y} \notin A_\alpha(\mu^*)) \\
&\geq 1 - \alpha,
\end{aligned}
$$

as desired. This completes the proof.

A.2 Proof of Lemma 2.2

The value $Q_\mu^{\max}$ is a valid decision value for any given μ, since for all $\boldsymbol{x} \in \Phi_\mu \cap \mathcal{X}$ it holds that

$$
\mathbb{P}_{\lambda \sim F_{\boldsymbol{x}}}\left(\lambda > \sup_{\boldsymbol{x}' \in \Phi_\mu \cap \mathcal{X}} Q_{F_{\boldsymbol{x}'}}(1 - \alpha)\right) \leq \mathbb{P}_{\lambda \sim F_{\boldsymbol{x}}}(\lambda > Q_{F_{\boldsymbol{x}}}(1 - \alpha)) = \alpha. \tag{A.1}
$$

For any $v < Q_\mu^{\max}$ that one could use as a decision value, there exists $\widetilde{\boldsymbol{x}} \in \Phi_\mu \cap \mathcal{X}$ such that $Q_{F_{\widetilde{\boldsymbol{x}}}}(1 - \alpha) > v$. Therefore, $\mathbb{P}_{\lambda \sim F_{\widetilde{\boldsymbol{x}}}}(\lambda > v) > \alpha$, and thus v is not a valid decision value. $Q^{\max}$ is clearly valid for all μ, and a similar argument shows that choosing any smaller v would make it not valid, as there exists a $\widetilde{\boldsymbol{x}} \in \mathcal{X}$ with a larger quantile than v, making v invalid as a decision value for $\mu = \varphi(\widetilde{\boldsymbol{x}})$. The equality between the two formulations comes from the same argument that shows (2.10) is equivalent to (2.11).

A.3 Proof of Theorem 2.4

Assume $\bar{\mathcal{X}}_\alpha(\boldsymbol{y})$ is nonempty and write as shorthand $\inf_{\boldsymbol{x} \in \mathcal{X}} / \sup_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{x})$ for the interval

$$
\left[\inf_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{x}), \sup_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{x})\right].
$$

Observe that

$$
\mathcal{C}_\alpha(\boldsymbol{y}) \subseteq \inf_{\mu \in \mathcal{C}_\alpha(\boldsymbol{y})} / \sup_{\mu \in \mathcal{C}_\alpha(\boldsymbol{y})} \mu. \tag{A.2}
$$

From Lemma 2.1, $\mathcal{C}_\alpha(\boldsymbol{y}) \subseteq \mathcal{I}_\alpha(\boldsymbol{y})$ implies that $\mathcal{I}_\alpha(\boldsymbol{y})$ is also a $1 - \alpha$ confidence interval. We prove this interval exactly equals the defined $\mathcal{I}_\alpha(\boldsymbol{y})$ in (2.17). Unpacking the definition of $\mathcal{C}_\alpha(\boldsymbol{y})$, we write the interval

$$
\begin{aligned}
&\inf_\mu / \sup_\mu \quad \mu \\
&\text{subject to} \quad \mu \in \mathbb{R} \\
&\qquad\qquad\quad - 2 \log \Lambda(\mu, \boldsymbol{y}) \leq q_\alpha(\mu).
\end{aligned} \tag{A.3}
$$

We can write different optimization problems which are equivalent to the optimization problem (A.3). First, we use the definition of Λ to write:

$$
\begin{aligned}
&\inf_{\mu} / \sup_{\mu} \quad \mu \\
&\text{subject to} \quad \mu \in \mathbb{R} \\
&\qquad\qquad \inf_{\varphi(\boldsymbol{x})=\mu, \boldsymbol{x} \in \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}) - \inf_{\boldsymbol{x} \in \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}) \leq q_{\alpha}(\mu).
\end{aligned} \tag{A.4}
$$

Notice that we can rewrite the feasibility condition of μ as follows:

$$
\inf_{\varphi(\boldsymbol{x})=\mu, \boldsymbol{x} \in \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}) \leq q_{\alpha}(\mu) + \inf_{\boldsymbol{x} \in \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y})
$$

as there exists $\boldsymbol{x} \in \mathcal{X}$ such that $\varphi(\boldsymbol{x}) = \mu$ and

$$
-2\ell_{\boldsymbol{x}}(\boldsymbol{y}) \leq q_{\alpha}(\mu) + \inf_{\boldsymbol{x} \in \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}).
$$

Therefore, the optimization problem can be rewritten with $\boldsymbol{x}$ and μ as the optimization variables:

$$
\begin{aligned}
&\inf_{\mu, \boldsymbol{x}} / \sup_{\mu, \boldsymbol{x}} \quad \mu \\
&\text{subject to} \quad \boldsymbol{x} \in \mathcal{X}, \ \mu \in \mathbb{R} \\
&\qquad\qquad \varphi(\boldsymbol{x}) = \mu \\
&\qquad\qquad -2\ell_{\boldsymbol{x}}(\boldsymbol{y}) \leq q_{\alpha}(\mu) + \inf_{\boldsymbol{x} \in \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}).
\end{aligned} \tag{A.5}
$$

And μ can be eliminated using the constraint, yielding

$$
\begin{aligned}
&\inf_{\boldsymbol{x}} / \sup_{\boldsymbol{x}} \quad \varphi(\boldsymbol{x}) \\
&\text{subject to} \quad \boldsymbol{x} \in \mathcal{X} \\
&\qquad\qquad -2\ell_{\boldsymbol{x}}(\boldsymbol{y}) \leq q_{\alpha}(\varphi(\boldsymbol{x})) + \inf_{\boldsymbol{x} \in \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}),
\end{aligned} \tag{A.6}
$$

that is, $\inf_{\boldsymbol{x} \in \bar{\mathcal{X}}_{\alpha}(\boldsymbol{y})} / \sup_{\boldsymbol{x} \in \bar{\mathcal{X}}_{\alpha}(\boldsymbol{y})} \varphi(\boldsymbol{x})$. The choice when $\bar{\mathcal{X}}_{\alpha}(\boldsymbol{y})$ is empty does not affect coverage properties. An alternative proof comes by observing the set equality $\varphi(\{\boldsymbol{x} \in \mathcal{X} : -2\ell_{\boldsymbol{x}}(\boldsymbol{y}) \leq q_{\alpha}(\varphi(\boldsymbol{x})) + \inf_{\boldsymbol{x} \in \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y})\}) = \{\mu \in \varphi(\mathcal{X}) : \inf_{\varphi(\boldsymbol{x})=\mu, \boldsymbol{x} \in \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}) - \inf_{\boldsymbol{x} \in \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}) \leq q_{\alpha}(\mu)\}$, and the result follows. This finishes the proof.

A.4 Proof of Proposition 2.5

We have, by definition and test inversion, that $q_{\alpha}(\mu)$ are valid if and only if

$$
\mathcal{C}_{\alpha}(\boldsymbol{y}) := \{\mu : \lambda(\mu, \boldsymbol{y}) \leq q_{\alpha}(\mu)\}
$$

is a valid $1 - \alpha$ confidence interval for any $\boldsymbol{x} \in \mathcal{X}$. Since $\mathcal{I}_{\alpha}(\boldsymbol{y})$ is the smallest interval that contains $\mathcal{C}_{\alpha}(\boldsymbol{y})$, we aim to prove that $\mathcal{C}_{\alpha}(\boldsymbol{y})$ is already an interval (including singletons or empty sets), so that $\mathcal{C}_{\alpha}(\boldsymbol{y}) = \mathcal{I}_{\alpha}(\boldsymbol{y})$ and the result holds. Define the function:

$$
\mu \mapsto \mathcal{F}(\mu) = \inf_{\substack{\varphi(\boldsymbol{x})=\mu \\ \boldsymbol{x} \in \mathcal{X}}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}) \tag{A.7}
$$

for a given $\boldsymbol{y}$, supported in all μ such that $\Phi_\mu \cap \mathcal{X} \neq \emptyset$. Write $\mathcal{C}_\alpha(\boldsymbol{y})$ explicitly using (2.6), we get

$$\mathcal{C}_\alpha(\boldsymbol{y}) := \left\{ \mu \colon \inf_{\substack{\varphi(\boldsymbol{x})=\mu \\ \boldsymbol{x} \in \mathcal{X}}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}) - \inf_{\boldsymbol{x} \in \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}) \leq q_\alpha \right\}. \tag{A.8}$$

The second term on the left-hand side does not depend on μ, so it is enough to prove that any set of the form $\{\mu \colon \mathcal{F}(\mu) \leq z\}$ is an interval, which is implied by the function $\mathcal{F}(\mu)$ being convex in μ (for a fixed $\boldsymbol{y}$). Indeed, if the set is not an interval, we have $\mu^- < \mu < \mu^+$ with $\mu^-, \mu^+ \in \mathcal{C}_\alpha(\boldsymbol{y})$ and $\mu \notin \mathcal{C}_\alpha(\boldsymbol{y})$ which contradicts convexity, since

$$\mathcal{F}(\mu) \geq z > \gamma \mathcal{F}(\mu^-) + (1-\gamma)\mathcal{F}(\mu^+).$$

To see convexity and finish the proof, let $\mu_1 \neq \mu_2$ and let $\mathcal{G}(\boldsymbol{x}) := -2\ell_{\boldsymbol{x}}(\boldsymbol{y})$, a convex function by assumption. Write for $i = 1, 2$:

$$x_i \in \operatorname*{argmin}_{\substack{\varphi(\boldsymbol{x})=\mu \\ \boldsymbol{x} \in \mathcal{X}}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}),$$

with x_i being any possible element in the set of minimizers, so that $\mathcal{F}(\mu_i) = \mathcal{G}(\boldsymbol{x}_i)$. For any $0 < \gamma < 1$, $\gamma \boldsymbol{x}_1 + (1-\gamma)\boldsymbol{x}_2 \in \mathcal{X}$ since $\mathcal{X}$ is a convex cone and

$$\varphi(\gamma \boldsymbol{x}_1 + (1-\gamma)\boldsymbol{x}_2) = \gamma \mu_1 + (1-\gamma)\mu_2,$$

since φ is linear, so $\gamma \boldsymbol{x}_1 + (1-\gamma)\boldsymbol{x}_2$ is a feasible point of the optimization problem

$$\inf_{\substack{\varphi(\boldsymbol{x})=\gamma\mu_1+(1-\gamma)\mu_2 \\ \boldsymbol{x} \in \mathcal{X}}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}),$$

that has optimal value $\mathcal{F}(\gamma \mu_1 + (1-\gamma)\mu_2)$. Therefore, by convexity of $\mathcal{G}$ and definition of the $\boldsymbol{x}_i$, we have that:

$$\mathcal{F}(\gamma\mu_1 + (1-\gamma)\mu_2) \leq \mathcal{G}(\gamma\boldsymbol{x}_1 + (1-\gamma)\boldsymbol{x}_2) \leq \gamma\mathcal{G}(\boldsymbol{x}_1) + (1-\gamma)\mathcal{G}(\boldsymbol{x}_2) = \gamma\mathcal{F}(\mu_1) + (1-\gamma)\mathcal{F}(\mu_2). \tag{A.9}$$

This completes the proof.

A.5 Proof of Example 2.6

Since the case when $\mu^* = 0$ is of particular interest, we show the result in this specific case and then generalize to the case of $\mu^* > 0$.

Case of $\mu^* = 0$. When $\mu^* = 0$, we can argue from symmetry of the standard Gaussian about the origin to write down the CDF in closed form. For $c \geq 0$, we have

$$\begin{aligned}
\mathbb{P}_{\mu_0}(\ell_0 \leq c) &= \mathbb{P}_{\mu_0}(\ell_0 \leq c, y < 0) + \mathbb{P}_{\mu_0}(\ell_0 \leq c, y \geq 0) \\
&= \mathbb{P}_{\mu_0}(\ell_0 \leq c \mid y < 0)\, \mathbb{P}_{\mu_0}(y < 0) + \mathbb{P}_{\mu_0}(\ell_0 \leq c \mid y \geq 0)\, \mathbb{P}_{\mu_0}(y \geq 0).
\end{aligned} \tag{A.10}$$

By definition, $\mathbb{P}_{\mu_0}(y < 0) = \mathbb{P}_{\mu_0}(y \geq 0) = \frac{1}{2}$, so only the conditional probabilities remain. By (2.21), we have

$$\begin{aligned}
\mathbb{P}_{\mu_0}(\ell_0 \leq c \mid y < 0) &= \mathbb{P}_{\mu_0}(0 \leq c \mid y < 0) = 1 \\
\mathbb{P}_{\mu_0}(\ell_0 \leq c \mid y \geq 0) &= \mathbb{P}_{\mu_0}(y^2 \leq c \mid y \geq 0).
\end{aligned} \tag{A.11}$$

In (A.11), we immediately observe that

$$\mathbb{P}_{\mu_0}\left(y^2 \leq c \mid y \geq 0\right) = \mathbb{P}_{\mu_0}\left(y^2 \leq c, y \geq 0\right)\mathbb{P}_{\mu_0}\left(y \geq 0\right)^{-1} = 2\mathbb{P}_{\mu_0}\left(0 \leq y \leq \sqrt{c}\right) = 2\Phi(\sqrt{c}) - 1.$$
(A.12)

But we also have that

$$\mathbb{P}_{\mu_0}\left(y^2 \leq c\right) = \mathbb{P}_{\mu_0}\left(-\sqrt{c} \leq y \leq \sqrt{c}\right) = 2\Phi(\sqrt{c}) - 1.$$

So we have

$$\mathbb{P}_{\mu_0}\left(y^2 \leq c \mid y \geq 0\right) = \mathbb{P}_{\mu_0}\left(y^2 \leq c\right).$$

Hence, we obtain

$$\mathbb{P}_{\mu_0}\left(\ell_0 \leq c \mid y \geq 0\right) = \chi_1^2(c).$$

Note this independence on the sign of y means that the magnitude of y is statistically independent of its direction. Thus, when $\mu_0 = 0$, the log-likelihood ratio has the following distribution:

$$\ell_0 \sim \frac{1}{2} + \frac{1}{2}\chi_1^2.$$
(A.13)

This completes the case when $\mu^* = 0$.

Case of $\mu^* > 0$. When $\mu > 0$, the closed-form solution to the CDF of ℓ_0 becomes more complicated, as we can no longer use symmetry around the origin. Picking up at (A.10), we first note that when $y \sim \mathcal{N}(\mu_0, 1)$, we have

$$\mathbb{P}_{\mu_0}\left(y < 0\right) = \Phi(-\mu_0) \quad \text{and} \quad \mathbb{P}_{\mu_0}\left(y \geq 0\right) = \Phi(\mu_0).$$

Next, we must find the conditional probabilities. Starting with the case when $\{y < 0\}$, we obtain

$$\mathbb{P}_{\mu_0}\left((y - \mu_0)^2 - y^2 \leq c \mid y < 0\right) \tag{A.14}$$
$$= \mathbb{P}_{\mu_0}\left(-2y\mu_0 + \mu_0^2 \leq c \mid y < 0\right)$$
$$= \mathbb{P}_{\mu_0}\left(y \geq \frac{\mu_0^2 - c}{2\mu_0} \mid y < 0\right)$$
$$= \Phi(-\mu_0)^{-1}\mathbb{P}_{\mu_0}\left(y \geq \frac{\mu_0^2 - c}{2\mu_0}, y < 0\right)$$
$$= \Phi(-\mu_0)^{-1}\left\{0 \cdot \mathbf{1}\{c \leq \mu_0^2\} + \mathbb{P}_{\mu_0}\left(\frac{\mu_0^2 - c}{2\mu_0} \leq y \leq 0\right)\mathbf{1}\{c > \mu_0^2\}\right\}$$
$$= \Phi(-\mu_0)^{-1}\mathbb{P}_{\mu_0}\left(\frac{-\mu_0^2 - c}{2\mu_0} \leq y - \mu_0 \leq -\mu_0\right)\mathbf{1}\{c > \mu_0^2\}$$
$$= \Phi(-\mu_0)^{-1}\left\{\Phi(-\mu_0) - \Phi\left(\frac{-\mu_0^2 - c}{2\mu_0}\right)\right\}\mathbf{1}\{c > \mu_0^2\}. \tag{A.15}$$

Then, when $\{y \geq 0\}$, we have

$$\mathbb{P}_{\mu_0}\left((y - \mu_0)^2 \leq c \mid y \geq 0\right) \tag{A.16}$$
$$= \mathbb{P}_{\mu_0}\left(-\sqrt{c} \leq y - \mu_0 \leq \sqrt{c} \mid y \geq 0\right)$$
$$= \Phi(\mu_0)^{-1}\mathbb{P}_{\mu_0}\left(-\sqrt{c} \leq y - \mu_0 \leq \sqrt{c}, y \geq 0\right)$$
$$= \Phi(\mu_0)^{-1}\mathbb{P}_{\mu_0}\left(0 \leq y \leq \sqrt{c} + \mu_0\right)\mathbf{1}\{-\sqrt{c} + \mu_0 \leq 0\}$$

$$+ \Phi(\mu_0)^{-1} \mathbb{P}_{\mu_0} \left(-\sqrt{c} + \mu_0 \leq y \leq \sqrt{c} + \mu_0 \right) \mathbf{1}\{-\sqrt{c} + \mu_0 > 0\}$$
$$= \Phi(\mu_0)^{-1} \left\{ \left(\Phi(\sqrt{c}) - \Phi(-\mu_0) \right) \mathbf{1}\{c \geq \mu_0^2\} + (2\Phi(\sqrt{c}) - 1)\mathbf{1}\{c < \mu_0^2\} \right\}. \tag{A.17}$$

Putting together (A.15) and (A.17), we obtain the following CDF:

$$\mathbb{P}_{\mu_0} \left(\ell_0 \leq c \right) = \chi_1^2(c) \cdot \mathbf{1}\{c < \mu_0^2\} + \left\{ \Phi(\sqrt{c}) - \Phi\left(\frac{-\mu_0^2 - c}{2\mu_0} \right) \right\} \cdot \mathbf{1}\{c \geq \mu_0^2\}. \tag{A.18}$$

This completes the case of $\mu^* > 0$.

A.6   Proof of Example 2.7

We derive this result using a duality argument inspired by [18]. By definition, we have

$$\lambda(\mu^*, \boldsymbol{y}) = \min_{\boldsymbol{x}:\varphi(\boldsymbol{x})=\mu^*} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2 - \min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2. \tag{A.19}$$

For ease of notation, let $\widehat{\boldsymbol{x}}^* = \underset{\boldsymbol{x}:\,\boldsymbol{h}^\top \boldsymbol{x}=\mu^*}{\mathrm{argmin}} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2$. Consider the Lagrangian for the first optimization in (A.19):

$$L(\boldsymbol{x}, \lambda) = \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2 + \lambda(\boldsymbol{h}^\top \boldsymbol{x} - \mu^*). \tag{A.20}$$

First-order optimality allows solving for $\widehat{\boldsymbol{x}}^*$ as a function of the dual variable λ:

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}, \lambda) = -2\boldsymbol{K}^\top(\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}) + \lambda\boldsymbol{h} = 0$$
$$\implies -2\boldsymbol{K}^\top \boldsymbol{y} + 2\boldsymbol{K}^\top \boldsymbol{K}\boldsymbol{x} + \lambda\boldsymbol{h} = 0$$
$$\implies \widehat{\boldsymbol{x}}^* = (\boldsymbol{K}^\top \boldsymbol{K})^{-1}\boldsymbol{K}^\top \boldsymbol{y} - \frac{1}{2}\lambda(\boldsymbol{K}^\top \boldsymbol{K})^{-1}\boldsymbol{h}$$
$$\implies \widehat{\boldsymbol{x}}^* = \widehat{\boldsymbol{x}} - \frac{1}{2}\lambda(\boldsymbol{K}^\top \boldsymbol{K})^{-1}\boldsymbol{h}.$$

Substituting back into the LLR, we obtain

$$\lambda(\mu^*, \boldsymbol{y}) = \|\boldsymbol{y} - \boldsymbol{K}\widehat{\boldsymbol{x}}^*\|_2^2 - \|\boldsymbol{y} - \boldsymbol{K}\widehat{\boldsymbol{x}}\|_2^2$$
$$= \|\boldsymbol{y} - \boldsymbol{K}\widehat{\boldsymbol{x}} + \frac{1}{2}\lambda\boldsymbol{K}(\boldsymbol{K}^\top \boldsymbol{K})^{-1}\boldsymbol{h}\|_2^2 - \|\boldsymbol{y} - \boldsymbol{K}\widehat{\boldsymbol{x}}\|_2^2. \tag{A.21}$$

Performing some algebra, we note that

$$\|\boldsymbol{y} - \boldsymbol{K}\widehat{\boldsymbol{x}} + \frac{1}{2}\lambda\boldsymbol{K}(\boldsymbol{K}^\top \boldsymbol{K})^{-1}\boldsymbol{h}\|_2^2 = \|\boldsymbol{y} - \boldsymbol{K}\widehat{\boldsymbol{x}}\|_2^2$$
$$+ \lambda(\boldsymbol{y} - \boldsymbol{K}\widehat{\boldsymbol{x}})^\top \boldsymbol{K}(\boldsymbol{K}^\top \boldsymbol{K})^{-1}\boldsymbol{h} + \frac{1}{4}\lambda^2 \boldsymbol{h}^\top(\boldsymbol{K}^\top \boldsymbol{K})^{-1}\boldsymbol{h}.$$

Thus, we have

$$\lambda(\boldsymbol{y} - \boldsymbol{K}\widehat{\boldsymbol{x}})^\top \boldsymbol{K}(\boldsymbol{K}^\top \boldsymbol{K})^{-1}\boldsymbol{h} = \lambda\boldsymbol{y}^\top \boldsymbol{K}(\boldsymbol{K}^\top \boldsymbol{K})^{-1}\boldsymbol{h} - \lambda\widehat{\boldsymbol{x}}^\top \boldsymbol{K}^\top \boldsymbol{K}(\boldsymbol{K}^\top \boldsymbol{K})^{-1}\boldsymbol{h}$$
$$= \lambda\widehat{\boldsymbol{x}}^\top \boldsymbol{h} - \lambda\widehat{\boldsymbol{x}}^\top \boldsymbol{h}$$
$$= 0.$$

So the substitution in (A.21) can be further simplified such that:

$$\lambda(\mu^*, \boldsymbol{y}) = \frac{1}{4}\lambda^2 \boldsymbol{h}^\top(\boldsymbol{K}^\top \boldsymbol{K})^{-1}\boldsymbol{h}. \tag{A.22}$$

We now turn our attention to finding λ. Note that this optimization defining the Lagrangian (A.20) is convex with an affine equality constraint. Therefore, strong duality holds. We then define the dual function as follows:

$$
\begin{aligned}
g(\lambda) = \min_{\boldsymbol{x}} L(\boldsymbol{x}, \lambda) &= L(\widehat{\boldsymbol{x}}^*, \lambda) \\
&= \|\boldsymbol{y} - \boldsymbol{K}\widehat{\boldsymbol{x}}^*\|_2^2 + \lambda(\boldsymbol{h}^\top \widehat{\boldsymbol{x}}^* - \mu^*) \\
&= \|\boldsymbol{y} - \boldsymbol{K}\widehat{\boldsymbol{x}} + \frac{1}{2}\lambda \boldsymbol{K}(\boldsymbol{K}^\top \boldsymbol{K})^{-1}\boldsymbol{h}\|_2^2 + \lambda \left(\boldsymbol{h}^\top \widehat{\boldsymbol{x}} - \frac{1}{2}\lambda \boldsymbol{h}^\top (\boldsymbol{K}^\top \boldsymbol{K})^{-1}\boldsymbol{h} - \mu^* \right).
\end{aligned} \tag{A.23}
$$

We note that we can make many of the same simplifications above to arrive at the simplified dual function:

$$
g(\lambda) = \|\boldsymbol{y} - \boldsymbol{K}\widehat{\boldsymbol{x}}\|_2^2 + \lambda \boldsymbol{h}^\top - \frac{1}{4}\lambda^2 \boldsymbol{h}^\top (\boldsymbol{K}^\top \boldsymbol{K})^{-1}\boldsymbol{h} + \lambda \boldsymbol{h}^\top \widehat{\boldsymbol{x}} - \lambda \mu^*. \tag{A.24}
$$

To maximize $g(\lambda)$, we again use the following first order optimality condition:

$$
\begin{aligned}
\frac{dg}{d\lambda} &= -\frac{1}{2}\lambda \boldsymbol{h}^\top (\boldsymbol{K}^\top \boldsymbol{K})^{-1}\boldsymbol{h} + \boldsymbol{h}^\top \widehat{\boldsymbol{x}} - \mu^* = 0 \\
\implies \widehat{\lambda} &= \frac{2\left(\boldsymbol{h}^\top \widehat{\boldsymbol{x}} - \mu^*\right)}{\boldsymbol{h}^\top (\boldsymbol{K}^\top \boldsymbol{K})^{-1}\boldsymbol{h}}.
\end{aligned} \tag{A.25}
$$

Substituting (A.25) back into (A.22), we obtain

$$
\begin{aligned}
\lambda(\mu^*, \boldsymbol{y}) &= \frac{1}{4}\left(\frac{2\left(\boldsymbol{h}^\top \widehat{\boldsymbol{x}} - \mu^*\right)}{\boldsymbol{h}^\top (\boldsymbol{K}^\top \boldsymbol{K})^{-1}\boldsymbol{h}} \right)^2 \boldsymbol{h}^\top (\boldsymbol{K}^\top \boldsymbol{K})^{-1}\boldsymbol{h} \\
&= \frac{(\boldsymbol{h}^\top \widehat{\boldsymbol{x}} - \mu^*)^2}{\boldsymbol{h}^\top (\boldsymbol{K}^\top \boldsymbol{K})^{-1}\boldsymbol{h}}.
\end{aligned}
$$

For the second part, observe that when $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{K}\boldsymbol{x}^*, \boldsymbol{I}_m)$, we have

$$
\boldsymbol{h}^\top (\boldsymbol{K}^\top \boldsymbol{K})^{-1}\boldsymbol{K}^\top \boldsymbol{y} \sim \mathcal{N}(\boldsymbol{h}^\top \boldsymbol{x}^*, \boldsymbol{h}^\top (\boldsymbol{K}^\top \boldsymbol{K})^{-1}\boldsymbol{h}),
$$

hence (2.26) is the square of a one-dimensional standard Gaussian distribution. This finishes the proof.

B Proofs in Section 3

B.1 Proof of Lemma 3.2

Let $Y := \lambda(\boldsymbol{y}, \mu^*)$. Recall the validity of q_α can be written as $\mathbb{P}(Y \leq q_\alpha) \geq 1 - \alpha$ from (2.7) as:

$$
\begin{aligned}
X \succeq Y &\iff \mathbb{P}(X \geq \gamma) \geq \mathbb{P}(Y \geq \gamma), \text{ for all } \gamma \\
&\iff \alpha = \mathbb{P}(X \geq Q_X(1-\alpha)) \geq \mathbb{P}(Y \geq Q_X(1-\alpha)), \text{ for all } \alpha \\
&\iff 1 - \alpha = \mathbb{P}(X \leq Q_X(1-\alpha)) \leq \mathbb{P}(Y \leq Q_X(1-\alpha)), \text{ for all } \alpha \\
&\iff Q_X(1-\alpha) \text{ is a valid } q_\alpha \text{for all } \alpha.
\end{aligned}
$$

This finishes the proof.

B.2 Joint formulation of change-constrained optimization problem (3.1)

In this section, we provide details on formulating the optimization problem (3.1) and the interval optimization problem (2.17) as a single chance-constrained optimization problem. By joining (3.1) and (2.17) as a single optimization problem, we can use a similar argument as in the proof of Theorem 2.4. Starting with problem (A.4):

$$
\begin{aligned}
&\inf_{\mu} / \sup_{\mu} \quad \mu \\
&\text{subject to} \quad \mu \in \mathbb{R} \\
&\qquad \inf_{\varphi(\boldsymbol{x})=\mu, \boldsymbol{x} \in \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}) - \inf_{\boldsymbol{x} \in \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}) \le q_\alpha(\mu),
\end{aligned}
\tag{B.1}
$$

and substituting $q_\alpha(\mu) = \sup_{\boldsymbol{x} \in \Phi_\mu \cap \mathcal{X}} Q_{F_{\boldsymbol{x}}}(1-\alpha) = -\inf_{\boldsymbol{x} \in \Phi_\mu \cap \mathcal{X}} -Q_{F_{\boldsymbol{x}}}(1-\alpha)$, we obtain:

$$
\begin{aligned}
&\inf_{\mu} / \sup_{\mu} \quad \mu \\
&\text{subject to} \quad \mu \in \mathbb{R} \\
&\qquad \inf_{\substack{\varphi(\boldsymbol{x}_1)=\mu, \boldsymbol{x}_1 \in \mathcal{X}, \\ \varphi(\boldsymbol{x}_2)=\mu, \boldsymbol{x}_2 \in \mathcal{X}}} \left[-2\ell_{\boldsymbol{x}_1}(\boldsymbol{y}) - Q_{F_{\boldsymbol{x}_2}}(1-\alpha)\right] \le \inf_{\boldsymbol{x} \in \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}),
\end{aligned}
\tag{B.2}
$$

which can be transformed to parameter space as:

$$
\begin{aligned}
&\inf_{\boldsymbol{x}_1, \boldsymbol{x}_2} / \sup_{\boldsymbol{x}_1, \boldsymbol{x}_2} \quad \varphi(\boldsymbol{x}_1) \\
&\text{subject to} \quad \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{X} \\
&\qquad \varphi(\boldsymbol{x}_1) = \varphi(\boldsymbol{x}_2) \\
&\qquad -2\ell_{\boldsymbol{x}_1}(\boldsymbol{y}) - Q_{F_{\boldsymbol{x}_2}}(1-\alpha) \le \inf_{\boldsymbol{x} \in \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}).
\end{aligned}
\tag{B.3}
$$

Further unpacking $Q_{F_{\boldsymbol{x}_2}}(1-\alpha)$ as in Lemma 3.1, we obtain the chance-constrained optimization problem:

$$
\begin{aligned}
&\inf_{\boldsymbol{x}_1, \boldsymbol{x}_2, q} / \sup_{\boldsymbol{x}_1, \boldsymbol{x}_2, q} \quad \varphi(\boldsymbol{x}_1) \\
&\text{subject to} \quad \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{X} \\
&\qquad \varphi(\boldsymbol{x}_1) = \varphi(\boldsymbol{x}_2) \\
&\qquad -2\ell_{\boldsymbol{x}_1}(\boldsymbol{y}) \le \inf_{\boldsymbol{x} \in \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}) + q \\
&\qquad \mathbb{P}_{u \sim \mathcal{U}([0,1])}(\mathcal{F}(\boldsymbol{x}_2, u) \le q) \le 1-\alpha,
\end{aligned}
\tag{B.4}
$$

where $\mathcal{F}(\boldsymbol{x}, u) = F_{\boldsymbol{x}}^{-1}(u)$.

B.3 Proof of Example 3.3

Similar to Example 2.6 in Appendix A.5, this proof is divided into two cases.

Case of $\mu^* = 0$. Since the case when $\mu^* = 0$ is of particular interest, we show the result in this specific case and then generalize. Thus, when $\mu_0 = 0$, the log-likelihood ratio has the following distribution:

$$
\ell_0 \sim \frac{1}{2} + \frac{1}{2}\chi_1^2.
\tag{B.5}
$$

Additionally, this distribution implies the following stochastic dominance:

$$\mathbb{P}_{\mu_0}\left(\ell_0 \leq c\right) = \frac{1}{2}\left(1 + \chi_1^2(c)\right) \geq \chi_1^2(c), \tag{B.6}$$

that is, the log-likelihood ratio CDF is stochastically dominated by the chi-squared with one degree of freedom distribution. This means that the type-I error of the test can be controlled at the α level.

When $\mu > 0$, the closed-form solution to the CDF of ℓ_0 becomes more complicated, as we can no longer use symmetry around the origin. From the result of Example 2.6, we have the following CDF:

$$\mathbb{P}_{\mu_0}\left(\ell_0 \leq c\right) = \chi_1^2(c) \cdot \mathbf{1}\{c < \mu_0^2\} + \left\{\Phi(\sqrt{c}) - \Phi\left(\frac{-\mu_0^2 - c}{2\mu_0}\right)\right\} \cdot \mathbf{1}\{c \geq \mu_0^2\}. \tag{B.7}$$

Note, a quick check of (B.7) when $\mu_0 = 0$ reveals agreement with (B.6) such that

$$\mathbb{P}_{\mu_0}\left(\ell_0 \leq c\right) = \Phi(\sqrt{c}) = \Phi(\sqrt{c}) - \frac{1}{2} + \frac{1}{2} = \frac{1}{2}\left(2\Phi(\sqrt{c}) - 1\right) + \frac{1}{2} = \frac{1}{2}\chi_1^2(c) + \frac{1}{2}. \tag{B.8}$$

This completes the case of $\mu^* = 0$.

Case of $\mu^* > 0$. We already demonstrated above the chi-squared with one degree of freedom dominates the log-likelihood ratio when $\mu_0 = 0$. We now show that the dominance holds when $\mu_0 > 0$. Clearly, when $c < \mu_0^2$, $\mathbb{P}_{\mu_0}\left(\ell_0 \leq 0\right) = \chi_1^2(c)$, making it in fact equal to the chi-squared with one degree of freedom. Suppose $c \geq \mu_0^2$. Define

$$h(c) := \Phi(\sqrt{c}) - \Phi\left(\frac{-\mu_0^2 - c}{2\mu_0}\right) - \chi_1^2(c).$$

The stochastic dominance occurs if and only if $h(c) \geq 0$ for all $c \geq \mu_0^2$.

Note first that $\chi_1^2(c) = \Phi(\sqrt{c}) - \Phi(-\sqrt{c})$ and therefore $h(c) = \Phi(-\sqrt{c}) - \Phi\left(\frac{-\mu_0^2 - c}{2\mu_0}\right)$. Since $\Phi(\cdot)$ is a monotonically increasing function, it is sufficient to show that $-\sqrt{c} - \frac{-\mu_0^2 - c}{2\mu_0} \geq 0$ for all $c \geq \mu_0^2$. We do so below.

Define a function f as follows:

$$f(c) = -\sqrt{c} - \frac{-\mu_0^2 - c}{2\mu_0}.$$

Observe that when $c = \mu_0^2$, $f(c) = 0$. Consider when $c > \mu_0^2$. We obtain the following first and second derivatives:

$$f'(c) = \frac{-\mu_0 + \sqrt{c}}{2\mu_0\sqrt{c}} \quad \text{and} \quad f''(c) = \frac{1}{4}c^{-3/2}.$$

By the constraint $c > \mu_0^2$, it follows that $-\mu_0 + \sqrt{c} > 0$, and therefore, $f'(c) > 0$ for all $c > \mu_0^2$. Additionally, $f''(c) > 0$ for all $c > \mu_0^2$, so f is convex. Hence, we conclude that f is a monotonically increasing function for $c > \mu_0^2$, which starts at 0 when $c = \mu_0^2$, and thus $f(c) \geq 0$ for all $c \geq \mu_0^2$. It therefore follows that

$$\Phi(-\sqrt{c}) \geq \Phi\left(\frac{-\mu_0^2 - c}{2\mu_0}\right),$$

and hence $h(c) \geq 0$ for all $c \geq \mu_0^2$. As such, we conclude that $\mathbb{P}_{\mu_0}\left(\ell_0 \leq c\right) \geq \chi_1^2(c)$ for all $c \geq 0$. In other words, that the sampling distribution for the log-likelihood ratio is stochastically dominated by a chi-squared distribution with one degree of freedom. This completes the case of $\mu^* > 0$.

C Proofs in Section 4

C.1 Proof of Theorem 4.1

The proof follows by combining Lemmas 4.2 and 4.3.

C.2 Proof of Lemma 4.2

The proof follows by direct inspection and substitution of q_α and $-2\ell_{\boldsymbol{x}}(\boldsymbol{y}) = \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2$. The interval has the coverage if the q_α is valid by Theorem 2.4 and only if by Proposition 2.5.

C.3 Proof of Lemma 4.3

The proof follows by observing $z_{\alpha/2}^2 = Q_{\chi_1^2}(1 - \alpha)$ and applying Lemma 3.2

C.4 Proof of Lemma 4.4

We argue by coupling. Note that $\frac{1}{2}(y_1 - y_2)^2 \sim \chi_1^2$, so that it suffices to show $\lambda \leq \frac{1}{2}(y_1 - y_2)^2$ for every y to constitute a valid coupling that proves stochastic dominance. This is clearly true when $y_1 + y_2 \geq 0$, since $\lambda - \frac{1}{2}(y_1 - y_2)^2$ is equal to non-positive terms only. When $y_1 + y_2 < 0$, if both are strictly negative, then $\lambda = 0 \leq \frac{1}{2}(y_1 - y_2)^2$. Then assume without loss of generality that y_1 is non-negative, then y_2 has to be negative. Then $\lambda = y_1^2$, but $y_1 \geq 0$, $y_2 < 0$ and $y_1 < -y_2$ imply that $|y_1 - y_2| = y_1 - y_2 \geq 2y_1 \geq \sqrt{2}y_1$, squaring both sides gives $\frac{1}{2}(y_1 - y_2)^2 < y_1^2 = \lambda$. This finishes the proof.

C.5 Proof of Lemma 4.5

Consider the LLR

$$\lambda(\mu^* = -1, \boldsymbol{y}) = \min_{\substack{x_1 + x_2 - x_3 = -1 \\ \boldsymbol{x} \geq \boldsymbol{0}}} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 - \min_{\boldsymbol{x} \geq \boldsymbol{0}} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2. \tag{C.1}$$

The goal of this proof is to show that χ_1^2 does not stochastically dominate (C.1) when $\boldsymbol{y} \sim \mathcal{N}(x^* = (0, 0, 1), \boldsymbol{I}_3)$. By Corollary 4.26 in [45], $X \succeq Y$ implies $\mathbb{E}[x] > \mathbb{E}[y]$, so it suffices to show that

$$\mathbb{E}[\lambda(\mu^* = -1, \boldsymbol{y})] > \mathbb{E}[\chi_1^2] = 1$$

to complete the proof.

Observe that

$$\mathbb{E}[\lambda(\mu^* = -1, \boldsymbol{y})] = \mathbb{E}\left[\min_{\substack{\boldsymbol{h}^\top \boldsymbol{x} = -1 \\ \boldsymbol{x} \geq \boldsymbol{0}}} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2\right] - \mathbb{E}\left[\min_{\boldsymbol{x} \geq \boldsymbol{0}} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2\right]. \tag{C.2}$$

We begin by computing the second term. Since

$$\min_{\boldsymbol{x} \geq \boldsymbol{0}} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 = \sum_{i=1}^{3}(y_i - \max\{y_i, 0\})^2,$$

we have

$$\mathbb{E}\left[\min_{\boldsymbol{x} \geq \boldsymbol{0}} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2\right] = \sum_{i=1}^{3}\mathbb{E}\left[(y_i - \max\{y_i, 0\})^2\right]$$

$$= 2\mathbb{E}_{z\sim\mathcal{N}(0,1)}\left[(z - \max\{z,0\})^2\right] + \mathbb{E}_{z\sim\mathcal{N}(1,1)}\left[(z - \max\{z,0\})^2\right]$$

Let $g(z) := (z - \max\{z,0\})^2$. Using in both cases, we obtain

$$\mathbb{E}[g(z)] = \underbrace{\mathbb{E}[g(z) \mid z \geq 0]}_{0} \cdot \mathbb{P}(z \geq 0) + \mathbb{E}[g(z) \mid z < 0] \cdot \mathbb{P}(z < 0) = \mathbb{E}[z^2 \mid z < 0] \cdot \mathbb{P}(z < 0).$$

Note that

$$\mathbb{E}[z^2 \mid z < 0] = (\mathbb{E}[z \mid z < 0])^2 + \mathrm{Var}[z \mid z < 0]$$

$$= \begin{cases} \left(-\dfrac{\phi(0)}{\Phi(0)}\right)^2 + \left(1 - \left(\dfrac{\phi(0)}{\Phi(0)}\right)^2\right), & z \sim \mathcal{N}(0,1) \\ \left(1 - \dfrac{\phi(-1)}{\Phi(-1)}\right)^2 + 1 + \dfrac{\phi(-1)}{\Phi(-1)} - \left(\dfrac{\phi(-1)}{\Phi(-1)}\right)^2, & z \sim \mathcal{N}(1,1) \end{cases}$$

$$= \begin{cases} 1, & z \sim \mathcal{N}(0,1) \\ 2 - \dfrac{\phi(-1)}{\Phi(-1)}, & z \sim \mathcal{N}(1,1), \end{cases}$$

where we used the formulas for mean and variance of a truncated Gaussian. Finally,

$$\mathbb{E}\left[\min_{\boldsymbol{x}\geq\boldsymbol{0}} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2\right] = 2 \cdot 1/2 \cdot 1 + (2 - \phi(-1)/\Phi(-1)) \cdot (\Phi(-1))$$

$$= 1 + 2\Phi(-1) - \phi(-1)$$

$$\approx 1.0753.$$

It suffices to prove that

$$\mathbb{E}\left[\min_{\substack{\boldsymbol{h}^\top\boldsymbol{x}=-1 \\ \boldsymbol{x}\geq\boldsymbol{0}}} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2\right] > 2 + 2\Phi(-1) - \phi(-1) \approx 2.0753.$$

We will prove that

$$\mathbb{E}\left[\min_{\substack{\boldsymbol{h}^\top\boldsymbol{x}=-1 \\ \boldsymbol{x}\geq\boldsymbol{0}}} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2\right] = 13/6 \approx 2.166.$$

Note that the intersection of the plane $\boldsymbol{h}^\top\boldsymbol{x} = x_1 + x_2 - x_3 = -1$ and $\boldsymbol{x} \geq \boldsymbol{0}$ is the parametric surface $\mathcal{S} = \{(u, v, u + v + 1), u \geq 0, v \geq 0\}$, so we can write

$$\min_{\boldsymbol{x}\in\mathcal{S}} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 = \min_{u\geq 0, v\geq 0} (y_1 - u)^2 + (y_2 - v)^2 + (y_3 - u - v - 1)^2. \tag{C.3}$$

It is convenient to define a new variable $z_3 = 1 - y_3 \sim \mathcal{N}(0,1)$, so that (y_1, y_2, z_3) is sampled from a standard three dimensional Gaussian. Abusing notation we will still write y_3 for z_3 and then $\boldsymbol{y} \sim \mathcal{N}((0,0,0), \boldsymbol{I})$. The optimization problem becomes:

$$\min_{u\geq 0, v\geq 0} (y_1 - u)^2 + (y_2 - v)^2 + (-y_3 - u - v)^2. \tag{C.4}$$

This can be explicitly solved to yield:

$$\min_{\substack{h^\top x = -1 \\ x \geq 0}} \|x - y\|_2^2 = \begin{cases} y_1^2 + y_2^2 + y_3^2 & y_1 - y_3 \leq 0 \text{ and } y_2 - y_3 \leq 0 \\ \frac{1}{2}\left(y_1^2 + 2y_1 y_3 + 2y_2^2 + y_3^2\right) & y_1 - y_3 \geq 0 \text{ and } y_1 - 2y_2 + y_3 \geq 0 \\ \frac{1}{2}\left(2y_1^2 + y_2^2 + 2y_2 y_3 + y_3^2\right) & y_2 - y_3 \geq 0 \text{ and } 2y_1 - y_2 - y_3 \leq 0 \\ \frac{1}{3}\left(y_1 + y_2 + y_3\right)^2 & \begin{cases} 2y_1 - y_3 \geq y_2 \geq \max\{y_1, y_3\} \\ 2y_2 - y_3 \geq y_1 \geq \max\{y_2, y_3\} \end{cases} \end{cases} . \tag{C.5}$$

We split

$$\int_{\mathbb{R}^3} \min_{\substack{h^\top x = -1 \\ x \geq 0}} \|x - y\|_2^2 \, \phi(y_1)\phi(y_2)\phi(y_3) \, \mathrm{d}y$$

into the different domains given by (C.5), with the value of the expectation being equal to the sum of the different integrals, which we proceed to compute.

Region 1:

$$I_1 = \int_{y_3 \geq y_1, y_3 \geq y_2} \frac{e^{-\frac{1}{2}(y_1^2 + y_2^2 + y_3^2)}}{2\sqrt{2}\pi^{3/2}} (y_1^2 + y_2^2 + y_3^2) \, \mathrm{d}y.$$

Note that by symmetry of the variables in the integrand, we have

$$I_1 = \int_{y_2 \geq y_1, y_2 \geq y_3} \frac{e^{-\frac{1}{2}(y_1^2 + y_2^2 + y_3^2)}}{2\sqrt{2}\pi^{3/2}} (y_1^2 + y_2^2 + y_3^2) \, \mathrm{d}y$$

$$= \int_{y_1 \geq y_3, y_1 \geq y_2} \frac{e^{-\frac{1}{2}(y_1^2 + y_2^2 + y_3^2)}}{2\sqrt{2}\pi^{3/2}} (y_1^2 + y_2^2 + y_3^2) \, \mathrm{d}y.$$

And since one of the y_i will always be the largest one, the sum of the domains is $\mathbb{R}^3$ (modulo measure zero intersections that do not affect integration) and we can write

$$I_1 = \frac{1}{3} \int_{\mathbb{R}^3} \frac{e^{-\frac{1}{2}(y_1^2 + y_2^2 + y_3^2)}}{2\sqrt{2}\pi^{3/2}} (y_1^2 + y_2^2 + y_3^2) \, \mathrm{d}y = \frac{1}{3} \cdot 3 = 1.$$

Here we used that the integral is the expected value of $y_1^2 + y_2^2 + y_3^2$, which is 3 since the y_i are centered with unit variance.

Region 2:

$$I_2 = \int_{y_1 - y_3 \geq 0, y_1 - 2y_2 + y_3 \geq 0} \frac{e^{-\frac{1}{2}(y_1^2 + y_2^2 + y_3^2)}}{4\sqrt{2}\pi^{3/2}} \left(y_1^2 + 2y_1 y_3 + 2y_2^2 + y_3^2\right) \, \mathrm{d}y. \tag{C.6}$$

Partition $\mathbb{R}^3$ in four spaces with measure zero intersection, and we aim to argue that the integral of the integrand in (C.6) has the same value when integrating over any of them:

$$A := \left\{ y : y_1 \geq y_3, y_2 \geq \frac{y_1 + y_3}{2} \right\}$$

$$B := \left\{ \boldsymbol{y} : y_1 \geq y_3, y_2 \leq \frac{y_1 + y_3}{2} \right\}$$

$$C := \left\{ \boldsymbol{y} : y_1 \leq y_3, y_2 \geq \frac{y_1 + y_3}{2} \right\}$$

$$D := \left\{ \boldsymbol{y} : y_1 \leq y_3, y_2 \leq \frac{y_1 + y_3}{2} \right\}.$$

Clearly $I_2 = I_B = \int_A \boldsymbol{h}(y_1, y_2, y_3) \, \mathrm{d}y$. Since $\boldsymbol{h}$ satisfies $\boldsymbol{h}(x_1, x_2, x_3) = \boldsymbol{h}(x_3, x_2, x_1)$, we can exchange y_1 and y_3 in the definitions of the sets, so $I_A = I_C$ and $I_B = I_D$. And since $\boldsymbol{h}$ is even with respect to x_2 and odd with respect to x_1, x_3 we can exchange y_i to $-y_i$ for $i = 1, 2, 3$ without the result changing. This flips both inequalities, proving $I_A = I_D$ and $I_B = I_C$. We therefore have

$$I_2 = \frac{1}{4} \int_{\mathbb{R}^3} \frac{e^{-\frac{1}{2}(y_1^2 + y_2^2 + y_3^2)}}{4\sqrt{2}\pi^{3/2}} \left(y_1^2 + 2y_1 y_3 + 2y_2^2 + y_3^2 \right) \mathrm{d}y = \frac{1}{4} \cdot 2 = \frac{1}{2}.$$

Here, in the integral, we factor out the sum and using that, the expected value of $y_i y_j$ is δ_{ij}.

Region 3:

$$I_3 = \int_{y_2 - y_3 \geq 0, y_2 - 2y_1 + y_3 \geq 0} \frac{e^{-\frac{1}{2}(y_1^2 + y_2^2 + y_3^2)}}{4\sqrt{2}\pi^{3/2}} \left(2y_1^2 + 2y_2 y_3 + y_2^2 + y_3^2 \right) \mathrm{d}y.$$

This is exactly the same integral as I_2 by switching y_2 with y_1, so $I_3 = \frac{1}{2}$.

Region 4:

$$I_4 = \int_{2y_1 - y_3 \geq y_2 \geq \max(y_1, y_3)} \frac{e^{-\frac{1}{2}(y_1^2 + y_2^2 + y_3^2)}}{6\sqrt{2}\pi^{3/2}} (y_1 + y_2 + y_3)^2 \, \mathrm{d}y$$

$$+ \int_{2y_2 - y_3 \geq y_1 \geq \max(y_2, y_3)} \frac{e^{-\frac{1}{2}(y_1^2 + y_2^2 + y_3^2)}}{6\sqrt{2}\pi^{3/2}} (y_1 + y_2 + y_3)^2 \, \mathrm{d}y. \tag{C.7}$$

We partition $\mathbb{R}^3$ in 12 subspaces with measure 0 intersection and we aim to argue that the integral of the integrand in (C.7) (considering one of the integrals only) has the same value when integrating over any of them. For σ a permutation of (y_1, y_2, y_3), we define the first 6 subsets as:

$$\{\boldsymbol{y} : 2y_{\sigma(1)} - y_{\sigma(2)} \geq y_{\sigma(3)} \geq \max\{y_{\sigma(1)}, y_{\sigma(2)}\}\},$$

and the last 6 subsets as:

$$\{\boldsymbol{y} : 2y_{\sigma(1)} - y_{\sigma(2)} \leq y_{\sigma(3)} \leq \min\{y_{\sigma(1)}, y_{\sigma(2)}\}\}.$$

We need to prove that the integral has the same value in any of the 12 subsets. Since that the integrand

$$\boldsymbol{h}(y_1, y_2, y_3) := \frac{e^{-\frac{1}{2}(y_1^2 + y_2^2 + y_3^2)}}{6\sqrt{2}\pi^{3/2}} (y_1 + y_2 + y_3)^2$$

50

satisfies $\boldsymbol{h}(y_1, y_2, y_3) = \boldsymbol{h}(y_{\sigma(1)}, y_{\sigma(2)}, y_{\sigma(3)})$ for all permutations σ, the value of the integral in between the first and second groups of 6 subsets is the same. For a fixed σ (say, the identity), since $\boldsymbol{h}(y_1, y_2, y_3) = \boldsymbol{h}(-y_1, -y_2, -y_3)$, the value over

$$\{\boldsymbol{y} : 2y_1 - y_2 \geq y_3 \geq \max\{y_1, y_2\}\}$$

is the same as the value over

$$\{\boldsymbol{y} : -2y_1 + y_2 \geq -y_3 \geq \max\{-y_1, -y_2\}\} = \{\boldsymbol{y} : 2y_1 - y_2 \leq y_3 \leq \min\{y_1, y_2\}\},$$

so the value over the 12 sets is complete.

It remains to be seen that for a generic $\boldsymbol{y} = (y_1, y_2, y_3)$, $y_1 \neq y_2 \neq y_3 \neq y_1$ (which can be assumed with probability 1 without affecting the integral), the point belongs to one and just one of the sets. Assume without loss of generality that y_1 is the greater of the three and y_3 is the smallest. Then since $y_1 > \max\{y_2, y_3\}$ and $y_3 < \min\{y_1, y_2\}$ the only subsets that $\boldsymbol{y}$ can belong to are:

$$A := \{\boldsymbol{y} : 2y_2 - y_3 \geq y_1 \geq \max\{y_2, y_3\}\}$$
$$B := \{\boldsymbol{y} : 2y_3 - y_2 \geq y_1 \geq \max\{y_2, y_3\}\}$$
$$C := \{\boldsymbol{y} : 2y_1 - y_2 \leq y_3 \leq \min\{y_1, y_2\}\}$$
$$D := \{\boldsymbol{y} : 2y_2 - y_1 \leq y_3 \leq \min\{y_1, y_2\}\}.$$

But $\boldsymbol{y}$ is not in B because that would require $y_3 \geq \frac{y_1+y_2}{2}$ but $y_3 < y_1$ and $y_3 < y_2$, and it is also not in C because that would require $y_1 \leq \frac{y_2+y_3}{2}$ and $y_1 > y_2$ and $y_1 > y_3$. $\boldsymbol{y}$ will be in A if $y_2 > \frac{y_1+y_3}{2}$ and in D if, on the contrary, $y_2 < \frac{y_1+y_3}{2}$, both of which are possible, but not at the same time. We conclude by identifying I_4 as the sum of two integrals over subsets that we have defined, and therefore

$$I_4 = \frac{2}{12} \int_{\mathbb{R}^3} \frac{e^{-\frac{1}{2}(y_1^2 + y_2^2 + y_3^2)}}{6\sqrt{2}\pi^{3/2}}(y_1 + y_2 + y_3)^2 \, \mathrm{d}y = \frac{1}{6} \cdot 1 = \frac{1}{6}.$$

Here we expand the sum and use again that the expected value of $y_i y_j$ is δ_{ij}. The proof concludes by adding up

$$\mathbb{E}\left[\min_{\substack{\boldsymbol{h}^\top \boldsymbol{x} = -1 \\ \boldsymbol{x} \geq \boldsymbol{0}}} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2\right] = I_1 + I_2 + I_3 + I_4 = \frac{13}{6}.$$

C.6 Proof of Proposition 4.6

We construct a series of counterexamples, indexed by the dimension p, and prove that as $p \to \infty$, the expected value of the LLR diverges. Since stochastic dominance implies inequality of expectations (when expectations are finite), we conclude that the distribution can not be stochastically dominated. For all $p \in \mathbb{N}$, consider the example in $\mathbb{R}^p (= \mathbb{R}^m)$, $\boldsymbol{K} = \boldsymbol{I}_p$, $\boldsymbol{x}^* = (0, \ldots, 0, 1)$, $\boldsymbol{h} = (1, \ldots, 1, -1)$ (such that $\mu^* = -1$). Let

$$\lambda_n(\mu^* = -1, \boldsymbol{y}) = \min_{\substack{\sum_{i=1}^{p-1} x_i - x_p = -1 \\ \boldsymbol{x} \geq \boldsymbol{0}}} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 - \min_{\boldsymbol{x} \geq \boldsymbol{0}} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2. \tag{C.8}$$

And compute

$$\mathbb{E}_{\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{x}^*, \boldsymbol{I}_n)}[\lambda_n(-1, \boldsymbol{y})] = \mathbb{E}_{\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{x}^*, \boldsymbol{I}_n)}\left[\min_{\substack{\sum_{i=1}^{p-1} x_i - x_p = -1 \\ \boldsymbol{x} \geq \boldsymbol{0}}} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2\right] - \mathbb{E}_{\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{x}^*, \boldsymbol{I}_n)}\left[\min_{\boldsymbol{x} \geq \boldsymbol{0}} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2\right].$$
$$\tag{C.9}$$

For the second term, we have

$$
\mathbb{E}\left[\min_{\boldsymbol{x}\geq\boldsymbol{0}}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2\right] = \sum_{i=1}^{p}\mathbb{E}\left[(y_i-\max\{y_i,0\})^2\right]
$$

$$
= (p-1)\mathbb{E}_{z\sim\mathcal{N}(0,1)}\left[(z-\max\{z,0\})^2\right] + \mathbb{E}_{z\sim\mathcal{N}(1,1)}\left[(z-\max\{z,0\})^2\right]
$$

$$
= (p-1)\frac{1}{2} + (2-\phi(-1)/\Phi(-1))\cdot(\Phi(-1)),
$$

using similar arguments as the proof in Appendix C.5. We will lower bound the first term using duality. For simplicity, define $\boldsymbol{z} = (y_1,\ldots,y_{p-1},y_n-1)\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{I}_n)$, and equivalently optimize

$$
\min_{\substack{\sum_{i=1}^{p-1}x_i=x_p \\ \boldsymbol{x}\geq\boldsymbol{0}}} \|\boldsymbol{x}-\boldsymbol{z}\|_2^2, \tag{C.10}
$$

where we defined the feasible $\widetilde{\boldsymbol{x}} = (x_1,\ldots,x_n-1=\sum_{i=1}^{p-1}x_i)$ and replaced $\widetilde{\boldsymbol{x}}$ by $\boldsymbol{x}$, abusing notation. Using Fenchel duality, we have that

$$
\min_{\substack{\sum_{i=1}^{p-1}x_i=x_p \\ \boldsymbol{x}\geq\boldsymbol{0}}} \|\boldsymbol{x}-\boldsymbol{z}\|_2^2 \geq \sup_{\boldsymbol{\xi}\in\mathbb{R}^p}\left(-f^*(\boldsymbol{\xi})-g^*(-\boldsymbol{\xi})\right), \tag{C.11}
$$

where we have noted by f^* the convex conjugate of $f(\boldsymbol{x}) := \|\boldsymbol{x}-\boldsymbol{z}\|_2^2$ and, letting S be the feasible set, we denoted by g^* the convex conjugate of

$$
g(\boldsymbol{x}) = \begin{cases} 0 & \text{if } \boldsymbol{x}\in S \\ \infty & \text{if } \boldsymbol{x}\notin S. \end{cases} \tag{C.12}
$$

Note that with these definitions, $\min_{\substack{\sum_{i=1}^{p-1}x_i=x_p \\ \boldsymbol{x}\geq\boldsymbol{0}}}\|\boldsymbol{x}-\boldsymbol{z}\|_2^2 = \inf_{\boldsymbol{x}}(f(x)+g(x))$ so the weak Fenchel duality applies. We compute $f^*(\boldsymbol{\xi}) = \frac{1}{4}\|\boldsymbol{\xi}\|_2^2 + \boldsymbol{z}^\top\boldsymbol{\xi} - \boldsymbol{z}^\top\boldsymbol{z}$, and

$$
g^*(\boldsymbol{\xi}) = \begin{cases} 0 & \text{if } \xi_i+\xi_p\leq 0 \text{ for } i\in[p-1] \\ \infty & \text{otherwise,} \end{cases} \tag{C.13}
$$

so that

$$
\sup_{\boldsymbol{\xi}\in\mathbb{R}^p}\left(-f^*(\boldsymbol{\xi})-g^*(-\boldsymbol{\xi})\right) = \sup_{\xi_i+\xi_n\geq 0,\text{ for all } i\in[p-1]}\left[-\frac{1}{4}\|\boldsymbol{\xi}\|_2^2 - \boldsymbol{z}^\top\boldsymbol{\xi} + \boldsymbol{z}^\top\boldsymbol{z}\right] \tag{C.14}
$$

$$
\geq \sup_{\xi_i+\xi_n\geq 0,\text{ for all } i\in[p-1]}\left[-\frac{1}{4}\|\boldsymbol{\xi}\|_2^2 - \boldsymbol{z}^\top\boldsymbol{\xi}\right]. \tag{C.15}
$$

Since the supremum is lower bounded by any feasible point, we can further bound by picking a feasible $\boldsymbol{\xi}^*$ for each possible $\boldsymbol{z}$. We define the following:

$$
\boldsymbol{\xi}^*(\boldsymbol{z}) = \begin{cases} -\boldsymbol{z} & \text{if } -\boldsymbol{z} \text{ is feasible } (-z_i\geq z_n \text{ for all } i) \\ (-z_1,\ldots,-z_{p-1},\max_{i\in[p-1]}z_i) & \text{otherwise.} \end{cases} \tag{C.16}
$$

Observe that

$$
\min_{\substack{\sum_{i=1}^{p-1} x_i = x_p \\ \boldsymbol{x} \geq \boldsymbol{0}}} \|\boldsymbol{x} - \boldsymbol{z}\|_2^2
$$

$$
\geq -\frac{1}{4}\|\boldsymbol{\xi}^*(\boldsymbol{z})\|_2^2 - \boldsymbol{z}^\top \boldsymbol{\xi}^*(\boldsymbol{z})
$$

$$
= \begin{cases}
\dfrac{3}{4}\|\boldsymbol{z}\|^2 & \text{if } -\boldsymbol{z} \text{ is feasible } (-z_i \geq z_n \text{ for all } i) \\[2ex]
\dfrac{3}{4}\displaystyle\sum_{i=1}^{p-1} z_i^2 + z_n \max_{i \in [p-1]} z_i - \dfrac{1}{4}\left(\max_{i \in [p-1]} z_i\right)^2 & \text{otherwise.}
\end{cases}
$$

We note that $-\boldsymbol{z}$ is feasible with probability $1/p$, by symmetry. Taking expected value over the inequality and using the law of total expectation yields

$$
\mathbb{E}\left[\min_{\substack{\sum_{i=1}^{p-1} x_i = x_p \\ \boldsymbol{x} \geq \boldsymbol{0}}} \|\boldsymbol{x} - \boldsymbol{z}\|_2^2\right]
$$

$$
\geq \frac{1}{p} \times \frac{3}{4}\mathbb{E}\big[\|z\|_2^2\big] + \frac{p-1}{p}\left\{\mathbb{E}\left[\frac{3}{4}\sum_{i=1}^{p-1} z_i^2\right] + \mathbb{E}\left[z_n \max_{i \in [p-1])} z_i\right] - \frac{1}{4}\mathbb{E}\left[\left(\max_{i \in [p-1]} z_i\right)^2\right]\right\}
$$

$$
= \frac{3}{4} + \frac{p-1}{p}\left\{\frac{3(p-1)}{4} + 0 - \frac{1}{4}\mathbb{E}\left[\left(\max_{i \in [p-1]} z_i\right)^2\right]\right\}.
$$

To bound the last term, we use

$$
\mathbb{E}\left[\left(\max_{i \in [p-1]} z_i\right)^2\right] = \mathbb{E}\left[\max_{i \in [p-1]} z_i\right] + \mathrm{Var}\left[\max_{i \in [p-1]} z_i\right] \leq \sqrt{2\log(p-1)} + 1,
$$

where the moment bounds are standard results: the expectation bound can be found using Jensen's inequality on $\exp(\sqrt{2\log p} \max_i z_i)$ and then bounding $\max_i z_i \leq \sum_i z_i$, and the variance bound with Poincaré's inequality applied to a smooth maximum, even though it can be refined [3]. Putting everything together, we obtain

$$
\mathbb{E}_{\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{x}^*, \boldsymbol{I}_n)}[\lambda_n(-1, \boldsymbol{y})] \geq \frac{p-1}{p}\left\{\frac{3(p-1)}{4} - \frac{1}{4}\sqrt{2\log(p-1)}\right\} - \frac{p}{2} + \mathcal{O}(1),
$$

which is $\mathcal{O}(p)$ and therefore tends to ∞ as $p \to \infty$. This completes the proof.

D   Additional numerical illustrations in Section 5

D.1   Constrained Gaussian in three dimensions

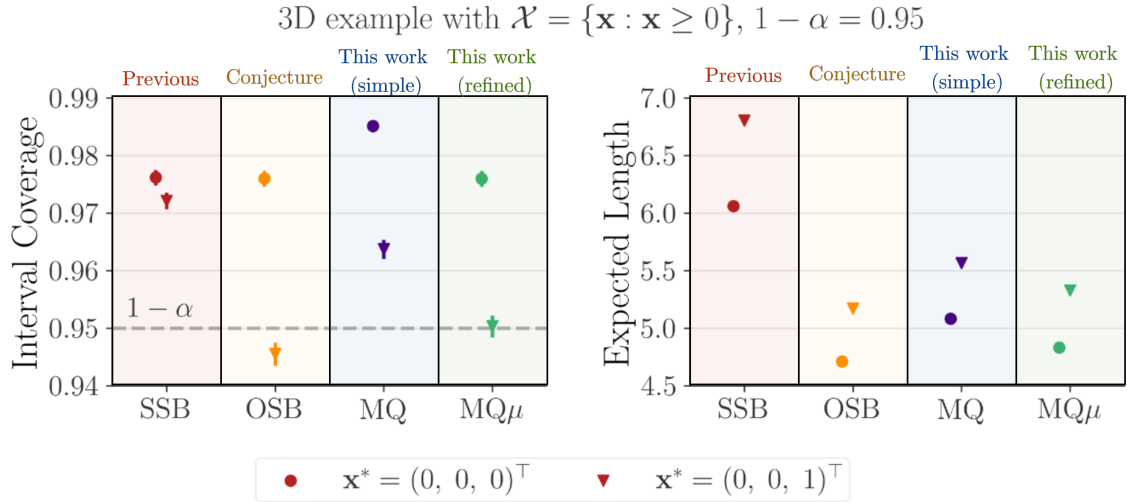We include the analog of Figure 5.4 with $1 - \alpha = 0.95$ in Figure D.1.

Figure D.1: Estimated interval coverage **(left)** and expected lengths **(right)** for 95% intervals resulting from the SSB, OSB, MQ and MQμ methods for the Gaussian linear model in (1.4) with $\boldsymbol{K} = \boldsymbol{I}_3$, $\varphi(\boldsymbol{x}) = \boldsymbol{h}^\top \boldsymbol{x} = x_1 + x_2 - x_3$ and $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^3 : \boldsymbol{x} \geq 0\}$.