Confidence intervals for functionals in constrained inverse problems via data-adaptive sampling-based calibration

Michael Stanley[†]
michael.c.stanley@ama-inc.com

Pau Batlle[‡]
pbatllef@caltech.edu

Pratik Patil[§]
pratikpatil@berkeley.edu

Houman Owhadi[‡]
owhadi@caltech.edu

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

Abstract

We address the problem of functional uncertainty quantification for ill-posed inverse problems where it is possible to evaluate a possibly rank-deficient forward model, the observation noise distribution is known, and there are known parameter constraints. We present four constraint-aware confidence interval constructions extending the theoretical test inversion framework in Batlle et al. (2023) by making the intervals both computationally feasible and less conservative. Our approach makes the potentially unbounded constraint set compact in a data-adaptive way, samples from this set and relevant test statistics to estimate a quantile function, and then uses these computed quantities to produce the constraint-aware confidence intervals. Our approach to bounding the constraint set in a data-adaptive way is based on the approach by Berger and Boos (1994), and involves defining a subset of the constraint set where the true parameter is guaranteed to exist with high probability. The probabilistic guarantee of this compact subset is then incorporated into the final coverage guarantee in the form of an uncertainty budget. We then propose custom sampling algorithms to efficiently sample from this subset, even when the parameter space is high-dimensional. Optimization-based interval methods formulate confidence interval computation as two endpoint optimizations, where the optimization constraints can be set to achieve different types of interval calibration while seamlessly incorporating parameter constraints. However, choosing optimization constraints to obtain coverage for a single functional has been elusive. We show that all four proposed intervals achieve nominal coverage for a particular functional both theoretically and in practice, with several numerical examples demonstrating superior performance of our intervals over the OSB interval in terms of both coverage and expected interval length. In particular, we show the superior performance of our intervals in a realistic unfolding simulation from high-energy physics that is severely ill-posed and involves a rank-deficient forward model.

[†]Analytical Mechanics Associates, Hampton, VA 23666, USA.
[‡]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, Carnegie Mellon University, Pittsburgh, PA 15213, USA.

1 Introduction

This paper proposes a novel uncertainty quantification (UQ) approach for ill-posed inverse problems characterized by a known parametric forward model[1] $f\colon \mathbb{R}^p \to \mathbb{R}^n$ mapping a parameter $\boldsymbol{x} \in \mathbb{R}^p$ to an observation in $\mathbb{R}^n$, additive noise with a known distribution $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, and constraints on the forward model parameters denoted by $\boldsymbol{x} \in \mathcal{X}$, where $\mathcal{X}$ is primarily assumed to be of the form $\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}$. We consider scenarios where the parameter space dimension p may exceed the observation space dimension n, which often results in an ill-posed problem. Additionally, we do not require the forward model f to be injective, allowing for the possibility that f is many-to-one. We assume observations are generated by $\boldsymbol{y} = f(\boldsymbol{x}^*) + \boldsymbol{\varepsilon}$, where $\boldsymbol{x}^* \in \mathcal{X}$ is the true but unknown model parameter. Our UQ object of interest is a confidence interval, $\mathcal{I}_\alpha(\boldsymbol{y})$, on a one-dimensional quantity of interest (QoI) derived from x^* by $\varphi : \mathcal{X} \to \mathbb{R}$. Our approach produces constraint-aware intervals with a finite-sample *frequentist coverage guarantee*, namely, for each $\alpha \in (0, 1)$

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

The probability in the coverage guarantee (1) is taken over the distribution of the observation $\boldsymbol{y} = f(\boldsymbol{x}^*) + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ is the additive noise. The level $\alpha \in (0, 1)$ corresponds to the desired coverage probability, meaning that the interval $\mathcal{I}_\alpha(\boldsymbol{y})$ will contain the true value $\varphi(\boldsymbol{x}^)$ with probability at least $1 - \alpha$ for any realization of the noise $\boldsymbol{\varepsilon}$.

This paper builds on the confidence interval framework of Batlle et al. (2023), who provided a unified perspective for many previous interval constructions by introducing a particular hypothesis test framework which, when inverted, yields confidence intervals that can be computed using two endpoint optimizations. By characterizing the quantile function of a particular log-likelihood ratio test statistic, they produced two interval constructions of increasing complexity and statistical power: one calibrates the interval to the largest value of the quantile function over the entire constrained parameter space, while the other calibrates the interval to the largest values of the quantile function as sliced along level sets of the constrained parameter space defined by the QoI. As such, one of the primary computational challenges of computing the intervals described therein is the potentially unbounded constraint set over which to search for these worst-case parameter settings. A significant challenge also arises from the fact that we do not know the exact distribution of the test statistic or the precise location where the maximum quantile is achieved, making it difficult to determine the worst-case parameter settings directly. This paper solves this unboundedness challenge by shrinking the constraint set in a data-adaptive way to produce the compact "Berger-Boos" set, inspired by Berger and Boos (1994). We then introduce a custom sampling algorithm to draw a set of design points in this set such that each respective thetest statistic can be sampled, thereby producing a dataset on which to train a quantile regressor to learn the quantile function. We finally use the learned quantile functions to define four interval constructions, all satisfying (1) with finite-data guarantees as the number of samples of our algorithm approaches infinity.

With advances in data collection and computational processing, high-dimensional ill-posed inverse problems have become more prevalent, especially in fields like remote sensing and data assimilation. This setting includes a wide array of physical science applications, spanning Earth science (Rodgers, 2000), atmospheric science and remote sensing (Liu et al., 2016; Patil et al., 2022), and high energy physics (Kuusela, 2016; Stanley et al., 2022), among many others. Guaranteed UQ in parameter inference from indirect observations is essential for assessing the precision of scientific

[1]Even though we write the forward model as a known function f, we only require evaluation access to the map throughout the paper.

inferences made in these contexts. However, the inherent ill-posed nature of these problems often leads to inferences that are highly sensitive to noise, posing significant challenges for UQ. Namely, the ill-posedness leads to an identifiability issue in which statistical inference is impossible without providing some form of regularization, which usually takes either a deterministic (e.g., SVD truncation (Höcker and Kartvelishvili, 1996) and Tikhonov regularization (Schmitt, 2012)) or probabilistic (e.g., priors and Bayesian inference) form. Under some assumptions, these two approaches are mathematically equivalent and are therefore subject to the same pitfalls, namely a disruption of the coverage guarantees of downstream intervals as a result of the incurred regularization bias as thoroughly discussed in Kuusela (2016). Our method's focus on one-dimensional QoI's and incorporation of parameter constraints allows for implicit regularization and therefore produces intervals with the promised coverage guarantee while avoiding the problem of regularization bias altogether.

Although including parameter constraints provides implicit regularization and enables handling non-trivial null spaces, it shifts complexity to statistical inference under constraints, a non-trivial problem in even elementary settings (Gouriéroux et al., 1982; Wolak, 1987; Robertson et al., 1988; Shapiro, 1988; Wolak, 1989; Molenberghs and Verbeke, 2007). One elegant solution to this problem originates in what we refer to as *optimization-based UQ*, originating in the work of Burrus et al. (1965) and Rust and Burrus (1972). Stark (1992a) extended and generalized their approach, calling this method *strict bounds*, since it produced guaranteed simultaneous coverage interval estimators complying with known physical constraints on the model parameters. This collection of work defines endpoint optimization problems over the physically constrained parameter space to directly compute confidence intervals for one-dimensional QoI's. Not only does the optimization form of the confidence interval computation shift statistical inference complexity to numerical optimization, but it also allows known physical constraints to be directly included in the endpoint optimizations. This practical advantage is dulled by the difficulty of proving the coverage properties of the intervals resulting from the defined endpoint optimizations in the one-at-a-time setting. Even under the relatively strong assumptions in Rust and Burrus (1972) of a linear forward model, non-negativity parameter constraints and linear QoI, the authors were only able to conjecture the coverage of their interval (known as the Burrus Conjecture). The interval coverage was unsuccessfully proven in Rust and O'Leary (1994) (the error pointed out in Tenorio et al. (2007)) and finally generally refuted in Batlle et al. (2023). These optimized-based intervals that have gained recent attention take the form:

$$\mathcal{I}(\psi_\alpha^2, \boldsymbol{y}) := \left[\varphi^l(\psi_\alpha^2, \boldsymbol{y}), \varphi^u(\psi_\alpha^2, \boldsymbol{y})\right] = \left[\min_{\boldsymbol{x}\in D(\psi_\alpha^2, \boldsymbol{y})} \varphi(\boldsymbol{x}), \max_{\boldsymbol{x}\in D(\psi_\alpha^2, \boldsymbol{y})} \varphi(\boldsymbol{x})\right], \tag{2}$$

where

$$D(\psi_\alpha^2, \boldsymbol{y}) := \left\{\boldsymbol{x} \in \mathcal{X} : \|\boldsymbol{y} - f(\boldsymbol{x})\|_2^2 \leq \psi_\alpha^2\right\}. \tag{3}$$

The statistical challenge for intervals of this form is then to choose ψ_α^2 such that $\mathcal{I}(\psi_\alpha^2, \boldsymbol{y})$ has the desired coverage guarantee. It is important to note that the definition of $D(\psi_\alpha^2, \boldsymbol{y})$ assumes that the noise has been standardized; in other words, the covariance matrix of the noise has been transformed to the identity matrix.

The literature proposes two settings to guarantee coverage. One approach is to set ψ_α^2 such that $D(\psi_\alpha^2, \boldsymbol{y})$ is itself a confidence set for $\boldsymbol{x}^*$ in the parameter space. Since this choice would then automatically guarantee coverage for $\mathcal{I}(\psi_\alpha^2, \boldsymbol{y})$ regardless of the chosen QoI, this setting has been called "simultaneous" in Rust and O'Leary (1994), and "Simultaneous Strict Bounds" (SSB) in Stanley et al. (2022); Batlle et al. (2023) since it also aligns with the setting of the strict bounds construction in Stark (1992a). Under the Gaussian assumption, this can be achieved by setting

$\psi^2_{SSB,\alpha} := \chi^2_{n,\alpha}$, where $\chi^2_{n,\alpha}$ is the upper α-quantile for a chi-squared distribution with n degrees of freedom. Although this approach is relatively simple and obtains the desired coverage guarantee, it is conservative since the guarantee holds for all possible QoI choices simultaneously. To tailor the interval to one particular QoI, there is the "one-at-a-time" setting as described in Rust and Burrus (1972); Rust and O'Leary (1994), or "one-at-a-time strict bounds" (OSB) as called in Stanley et al. (2022); Batlle et al. (2023). Under the Gaussian assumption this proposed setting was $\psi^2_{OSB,\alpha} = \chi^2_{1,\alpha} + s(\boldsymbol{y})^2$, where $s(\boldsymbol{y})^2 := \min_{\boldsymbol{x}\in\mathcal{X}} \|\boldsymbol{y} - f(\boldsymbol{x})\|_2^2$. Unlike the simultaneous setting, $D(\psi^2_{OSB,\alpha}, \boldsymbol{y})$ is not a $1-\alpha$ confidence set for $\boldsymbol{x}^*$, which makes proving its coverage guarantee difficult. The validity of this claim was proposed by the Burrus conjecture (Rust and Burrus, 1972) and generally disproven by Batlle et al. (2023).

To address the challenge of calibrating strict bounds and optimization-based confidence intervals, Batlle et al. (2023) approached these intervals as an inverted likelihood ratio test. The distribution of the log-likelihood ratio (LLR) statistic is non-standard due to the presence of the constraints, adding to the complexity of finding these max quantiles. This framework development allowed for the general disproving of the long-standing Burrus conjecture, invalidating the general calibration of the intervals proposed by Rust and Burrus (1972). However, the framework reveals that the characterization of a particular quantile function allows for proper interval calibration in much more general settings than those in which the Burrus conjecture applied (linear forward model, Gaussian noise, and linear functional). Batlle et al. (2023) introduces the concept of "Maximum Quantile" (MQ) in two forms: one based on a global quantile, q_α, and one based on a level-set quantile, $q_\alpha(\mu)$. While these quantiles provide valid one-at-a-time confidence intervals, they are challenging to obtain and tend to be conservative, protecting against the worst-case scenario. As mentioned in Batlle et al. (2023), finding these max quantiles can be formulated as chance-constrained optimization problems which are known to be strongly NP-hard (Geng and Xie, 2019). In lower-dimensional examples, gradient-free optimizers can be used to find these max quantiles, but this approach becomes intractable in higher dimensions. Although these quantiles offer coverage even in the worst-case scenario, they can be much larger than the quantile generated by the true parameter if the constraint set is unbounded (e.g., with non-negativity constraints). Therefore, we aim to find an alternative approach that is more computationally feasible and less conservative than these max quantiles.

We propose constructing confidence intervals in the present setting in the following way. We first use the observed data to define a $1-\eta$ confidence set in the parameter space, referred to as the "Berger-Boos" set following Berger and Boos (1994), which intersects with known parameter constraints to create a smaller bounded region of the constrained parameter space. A custom sampling algorithm is then developed to draw design points from the Berger-Boos set followed by sampling their respective test statistics. These sampled design points and test statistics constitute a data set on which a quantile regressor is trained to estimate the underlying quantile function. With this estimated quantile function, we construct confidence intervals along two axes; Global versus Sliced and Inverted versus Optimized. The Global intervals are computed using an estimate of the largest quantile in the Berger-Boos set whereas the Sliced intervals rely upon estimating the max quantile along level sets of the quantity of interest. These two options are analogous to the global and level-set quantiles defined in Batlle et al. (2023). On the other axis, the Inverted interval constructions directly accept or reject design points sampled in the Berger-Boos set according to their test statistic evaluation and the estimated quantile surface, producing an interval in the QoI space. The Optimized intervals use the estimated quantile function to define ψ^2_α in (3) and proceed to compute the interval endpoints via the corresponding optimizations. By balancing the level of the

Berger-Boos confidence set $(1-\eta)$ and the computed quantile (γ), we ensure the final confidence intervals achieve the desired coverage level $(1-\alpha)$. This approach is flexible and does not rely on assumptions about the linearity of the forward model, Gaussianity of the noise, non-negativity of the parameter constraints, or linearity of the functional. Although implementing this method becomes more challenging with relaxed assumptions, it can accommodate such violations.

To validate our method, we conduct a series of numerical experiments to demonstrate the effectiveness of the intervals in various scenarios. Since the OSB interval described in Patil et al. (2022) and Stanley et al. (2022) is the current standard approach in this setting, it is our primary point of comparison. First, we present low-dimensional examples. The first example is a two-dimensional constrained Gaussian noise model, a well-studied case in the literature (see Tenorio et al. (2007), Batlle et al. (2023)) and illustrates how our intervals are competitive with the OSB interval in a scenario where the OSB interval is known to achieve nominal coverage. The second example is a three-dimensional constrained Gaussian noise model, which illustrates the advantage of our method in achieving nominal coverage where the OSB interval fails. Second, we apply our method to a simulated version of particle unfolding from high-energy physics (see Stanley et al. (2022)), a binned deconvolution problem, in an 80-dimensional parameter space with a rank-deficient linear forward model. We use the previously studied setting realistic setting from Stanley et al. (2022) where the OSB interval is known to provide coverage to show that our Sliced intervals substantially outperform OSB in terms of expected length. Then, we use a parameter setting where the OSB interval does not achieve nominal coverage, but our intervals do, with the Sliced intervals still outperforming OSB in terms of expected length. This application highlights the practical significance of our approach, with our confidence intervals consistently achieving nominal coverage and often outperforming existing methods in terms of expected interval length.

The contributions of this work are both methodological and computational. Methodologically, we propose new confidence interval constructions for the setting described at the beginning of the section. Although the use of the Berger-Boos set is inspired by Berger and Boos (1994), they originally applied it in a hypothesis testing setting for handling nuisance parameters. Masserano et al. (2024) applied a similar idea to nuisance parameters in a classification setting, but to the best of our knowledge, our paper is the first to apply the idea in the ill-posed inverse problem UQ setting. Inspired by the methods used in simulation-based inference (Dalmasso et al. (2020, 2021); Masserano et al. (2023)) we apply the approach of estimating the quantile function of a test statistic using simulated data. However, this work differs in the underlying model assumptions and composite nature of the null hypotheses. In addition, the prior work assumed bounded and relatively low-dimensioanl parameter spaces. The search for confidence interval endpoints is sometimes done via the Robbins-Monro (RM) procedure, which iteratively refines estimates to achieve desired coverage probabilities (Garthwaite and Buckland, 1992; Carpenter, 1999). Methods based on quantile regression (Koenker, 2005) have been proposed to improve accuracy. For instance, Fisher et al. (2020) introduced a technique that inverts estimated quantiles to determine the endpoints of the confidence intervals. This work is the first to combine test inversion, sampling and quantile regression to estimate calibrated constraint-aware confidence intervals.

Computationally, we propose two custom sampling methods designed specifically to draw design points within the defined Berger-Boos set. This set can be particularly challenging to sample since it is the intersection between a pre-image based on the observed data and linear parameter constraints, which can produce sharp edges and corners. This set is also quite elongated due to the ill-posedness of the underlying inverse problem. While there are some approaches to handle similar scenarios arising in more traditional MCMC sampling (e.g., nested sampling (Skilling, 2004; Buchner, 2023)

and Hamiltonian Monte Carlo sampling with constraints (Pakman and Paninski, 2014)), these algorithms can be intricate, so we develop the following alternatives. In the low-dimensional setting ($p < 10$) where the forward model is linear and Gaussian noise distribution, we design an accept-reject sampler based on sampling from p-balls as described in Voelker et al. (2017). Although the approach in Voelker et al. (2017) can rapidly sample from the pre-image ellipsoids as defined by the data-generating model assumptions, with even a simple non-negativity constraint on the parameters, the proportion of rejected points quickly get impractically large for even moderate dimensions. For moderate and higher dimensional settings ($p \geq 10$), we leverage the Vaidya walk MCMC algorithm from Chen et al. (2018) to generate random walks about a polytope enclosing the Berger-Boos set. Using the Chebyshev ball center to define a notion of polytope centrality along with the extreme points of the Berger-Boos set with respect to the QoI, we create two lines along which starting positions are defined to generate a collection of parallel Markov chains. The design points resulting from these chains are then combined to create the complete set of sampled design points, which crucially span the full range of the QoI values over the Berger-Boos set.

The rest of the paper proceeds as follows. In Section 2, we recapitulate the framework and theoretical components of Batlle et al. (2023) upon which this paper's work is built. Section 3 then presents the theoretical Global and Sliced intervals under the Berger-Boos set formulation, along with the four interval constructions and theoretical justifications to guarantee that the constructed sampling-based intervals converge in probability to their theoretical counterparts. In Section 4, we give theoretical results regarding our proposed methods. Section 5 presents the custom sampling algorithms along with a brief description of how quantile regression works in our intervals. Section 6 presents four numerical experiments to demonstrate the coverage and length advantages of our intervals over the OSB interval. Finally, Section 7 provides some discussion and conclusions.

2 Background

A key part of the approach in Batlle et al. (2023) was to view the original optimization-based intervals as inverted hypothesis tests. More specifically, a particular inverted likelihood ratio test. This change in perspective allowed for interval analysis through the lens of the testing properties of a particular log-likelihood ratio (LLR) test statistic. We summarize this connection here, as it is critical to our method development carried out in Section 3. Although the numerical examples in Section 6 focus on the linear-Gaussian version of the data-generating process, we follow the more general exposition of Batlle et al. (2023) to indicate how this framework is not limited to that scenario.

Suppose $\boldsymbol{y} \sim P_{\boldsymbol{x}^*}$, where $P_{\boldsymbol{x}^*}$ is a distribution parameterized by a fixed but unknown $\boldsymbol{x}^* \in \mathbb{R}^p$. Let $L_{\boldsymbol{x}}(\boldsymbol{y})$ denote the likelihood of $P_{\boldsymbol{x}}$ evaluated at $\boldsymbol{y}$ and similarly $\ell_{\boldsymbol{x}}(\boldsymbol{y})$ the log-likelihood evaluated at $\boldsymbol{y}$. We furthermore suppose that we know $\boldsymbol{x}^* \in \mathcal{X}$.

The duality between hypothesis tests and confidence sets is a well-studied connection in statistics (see, e.g., Chapter 7 of Casella and Berger (2002) or Chapter 5 of Panaretos (2016)). We invert the following hypothesis test:

$$H_0 : \boldsymbol{x}^* \in \Phi_\mu \cap \mathcal{X} \quad \text{versus} \quad H_1 : \boldsymbol{x}^* \in \mathcal{X} - \Phi_\mu, \tag{4}$$

where $\Phi_\mu := \{\boldsymbol{x} : \varphi(\boldsymbol{x}) = \mu\}$. Since we are looking to obtain a confidence interval in $\mathbb{R}$, it makes sense that each hypothesis test is defined by $\mu \in \mathbb{R}$. Notice that the composite structure of the null hypothesis includes all parameter settings within the μ-level set of the functional of interest.

We define the following test statistic to evaluate Test (4):

$$\lambda(\mu, \boldsymbol{y}; \mathcal{X}) := \inf_{\boldsymbol{x} \in \Phi_\mu \cap \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}) - \inf_{\boldsymbol{x} \in \mathcal{X}} -2\ell_{\boldsymbol{x}}(\boldsymbol{y}), \tag{5}$$

where μ denotes the level set of the null hypothesis, $\boldsymbol{y}$ is the observed data, and $\mathcal{X}$ is the constraint set in which we know the parameter exists. We control the behavior of this test statistic, and therefore the test, by bounding above the probability of erroneously rejecting the null hypothesis (i.e., type-1 error control). As such, we consider the distribution of $\lambda(\mu, \boldsymbol{y}; \mathcal{X})$ under the null. For each $\boldsymbol{x} \in \mathcal{X}$, let $\mu = \varphi(\boldsymbol{x})$ and suppose $\boldsymbol{y} \sim P_{\boldsymbol{x}}$. Define $Q_{P_{\boldsymbol{x}}} : (0, 1) \to \mathbb{R}$ such that, for all $\alpha \in (0, 1)$,

$$\mathbb{P}\left(\lambda(\mu, \boldsymbol{y}; \mathcal{X}) \leq Q_{P_{\boldsymbol{x}}}(1 - \alpha)\right) = 1 - \alpha. \tag{6}$$

We refer to $Q_{P_{\boldsymbol{x}}}$ as the *quantile function* of the LLR under the null hypothesis at $(\boldsymbol{x}, \varphi(\boldsymbol{x}) = \mu)$. Since the null hypothesis is composite, using this quantile function to define a cutoff is not enough to control type-1 error. Thus, we use the *sliced* maximum quantile function over the level set under consideration:

$$Q^{\max}_{\mu, 1-\alpha} := \sup_{\boldsymbol{x} \in \Phi_\mu \cap \mathcal{X}} Q_{P_{\boldsymbol{x}}}(1 - \alpha). \tag{7}$$

Note, we refer to this version as "sliced" since we are considering the maximum of the quantile function defined by the level-set of the functional defining a slice through the constrained parameter space. By Lemma 2.1 in Batlle et al. (2023), the set $\mathcal{C}_\alpha(\boldsymbol{y}; \mathcal{X}) := \{\mu : \lambda(\mu, \boldsymbol{y}; \mathcal{X}) \leq Q^{\max}_{\mu, 1-\alpha}\} \subset \mathbb{R}$ defines a $1 - \alpha$ confidence set for the true functional value, $\mu^* = \varphi(\boldsymbol{x}^*)$. $Q^{\max}_{\mu, 1-\alpha}$ as defined above controls type-1 error for a value of μ specified in advance. We can define a more conservative confidence set using the following *global* maximum quantile:

$$Q^{\max}_{1-\alpha} := \sup_{\mu \in \varphi(\mathcal{X})} Q^{\max}_{\mu, 1-\alpha}. \tag{8}$$

(8) is more conservative than (7) since it holds for all null hypotheses, and therefore also produces a $1 - \alpha$ confidence set.

Connecting this framework back to Interval (2), suppose the data generating process is of the form, $P_{\boldsymbol{x}} = \mathcal{N}(\boldsymbol{K}\boldsymbol{x}, \boldsymbol{I})$. The LLR test statistic is then:

$$\lambda(\mu, \boldsymbol{y}; \mathcal{X}) = \min_{\boldsymbol{x} \in \Phi_\mu \cap \mathcal{X}} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2 - \min_{\boldsymbol{x} \in \mathcal{X}} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2 \tag{9}$$

Batlle et al. (2023) (Theorem 2.4) proved that

$$\left[\inf_{\mu : \lambda(\mu, \boldsymbol{y}) \leq Q^{\max}_{1-\alpha}} \mu, \sup_{\mu : \lambda(\mu, \boldsymbol{y}) \leq Q^{\max}_{1-\alpha}} \mu\right] = \left[\inf_{\boldsymbol{x} \in D(Q^{\max}_{1-\alpha} + s(\boldsymbol{y})^2, \boldsymbol{y})} \varphi(\boldsymbol{x}), \sup_{\boldsymbol{x} \in D(Q^{\max}_{1-\alpha} + s(\boldsymbol{y})^2, \boldsymbol{y})} \varphi(\boldsymbol{x})\right]. \tag{10}$$

This equivalence asserts that $\psi^2_\alpha := Q^{\max}_{1-\alpha} + s(\boldsymbol{y})^2$ guarantees coverage for Interval (2). Furthermore, it asserts that the original OSB interval formulation is only valid if and only if $\chi^2_{1,\alpha} \geq Q^{\max}_{1-\alpha}$. Batlle et al. (2023) showed that this inequality does not hold in general.

As explored in Batlle et al. (2023) and reiterated above, Interval (2) can be calibrated by computing $Q^{\max}_{\mu, 1-\alpha}$ or $Q^{\max}_{1-\alpha}$. However, pursuing calibration in this way is computationally challenging and statistically conservative. Both of these values require the ability to evaluate $Q_{P_{\boldsymbol{x}}}$ either directly if this function is explicitly known, or indirectly through sampling. Without parameter constraints, this quantile function can be constant under Gaussian-linear assumptions (Batlle et al., 2023), i.e.,

the test statistic is pivotal. But even in a relatively simple two-dimensional Gaussian noise model example with non-negativity constraints, this quantile function becomes non-trivial (e.g., see Figure 5.3 in Batlle et al. (2023)). Beyond the practical difficulty of dealing with the underlying quantile function, both (7) and (8) can be expressed as chance-constrained optimization problems, which are known to be NP-hard in general (Geng and Xie, 2019; Pena-Ordieres et al., 2020; Batlle et al., 2023). Statistically, since both (7) and (8) are the largest quantile function values subject to their respective constraints, they are by definition conservative, especially in scenarios where the true $\boldsymbol{x}^*$ is far from these most conservative points. Since we do not know $\boldsymbol{x}^*$, this conservatism is potentially warranted.

As we see in Section 3, our method addresses both of these challenges. In scenarios where sampling the LLR under the null hypothesis is difficult, we sample a collection of design points in a bounded subset of the constraint set, sample the LLR under the null at each design point and use quantile regression to estimate the quantile surface. Furthermore, we can potentially remove these most conservative points from consideration by only considering parameter values that are *not unlikely* given the observed data.

3 Adaptive confidence interval: constructions and theoretical justifications

Having reiterated the theoretical framework originally presented in Batlle et al. (2023), in this section, we present four related interval constructions to implement this theory. The first implementation challenge is to handle the potentially unbounded constraint set, for which we define and apply the "Berger-Boos" set to create a data-dependent subset of the original constraint set. This set leads to our four interval definitions, which follow a two-stage taxonomy; Global versus Sliced and Inverted versus Optimized. The second implementation challenge is computing these interval constructions in practice which we achieve using a combination of novel sampling algorithms and quantile regression. We will cover this in the following section (Section 5).

We rewrite the data-generating process articulated in the introduction:

$$\boldsymbol{y} = f(\boldsymbol{x}^*) + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}), \quad \boldsymbol{x}^* \in \mathcal{X} \tag{11}$$

where $f : \mathbb{R}^p \to \mathbb{R}^n$ is a known forward model and $\text{Cov}(\boldsymbol{\varepsilon}) = \boldsymbol{I}$, without loss of generality. Let $f^{-1}(A) := \{\boldsymbol{x} : f(\boldsymbol{x}) \in A\}$ be the pre-image of a set $A \subset \mathbb{R}^n$ under the forward map f.

3.1 Global and sliced confidence sets using the Berger-Boos set

Both $Q^{\max}_{\mu,1-\alpha}$ and $Q^{\max}_{1-\alpha}$ in Section 2 suffer from the same theoretical and practical concerns. Theoretically, they are both *conservative* in the sense that they must control type-1 error probability under the worst case for the truth (i.e., the parameter setting with the largest quantile). Practically, not only can $Q_{P_{\boldsymbol{x}}}$ be difficult to optimize but if the constraint set $\mathcal{X}$ is unbounded, computing the aforementioned quantiles becomes even more difficult. These challenges both stem from the composite nature of the null hypothesis. Berger and Boos (1994) introduce one compelling solution in the context of hypothesis testing with nuisance parameters, which is to control type-1 error only over a data-informed region of the parameter space. A similar approach can be applied to the aforementioned sliced and global max-quantiles, where we limit the maximization to the intersection of the original sets and a data-informed set. This solution addresses both the potential theoretical

conservatism of controlling type-1 error over the entire parameter space and the practical difficulty of dealing with unboundedness.

For $\eta \in (0,1)$, under the additive Gaussian noise assumption, we can easily find a $1-\eta$ confidence set, $\Gamma_\eta(\boldsymbol{y})$, such that $\mathbb{P}\left(f(\boldsymbol{x}^*) \in \Gamma_\eta(\boldsymbol{y})\right) = 1-\eta$, namely $\Gamma_\eta(\boldsymbol{y}) := \{\boldsymbol{y}' \in \mathbb{R}^n : \|\boldsymbol{y}-\boldsymbol{y}'\|_2^2 \leq \chi^2_{n,\eta}\}$. Define $\mathcal{B}_\eta := f^{-1}(\Gamma_\eta(\boldsymbol{y})) \cap \mathcal{X}$ and $\mathcal{B}_\eta^\mu := \mathcal{B}_\eta \cap \Phi_\mu$. We refer to this pre-image confidence set, $\mathcal{B}_\eta$, as the "Berger-Boos" set. The pre-image of $\Gamma_\eta(\boldsymbol{y})$ in the observation space with either $\mathcal{B}_\eta^{\mu^*}$ or $\mathcal{B}_\eta$ provides a confidence set in the parameter space such $\mathbb{P}(\boldsymbol{x}^* \in \mathcal{B}_\eta^{\mu^*}) \geq 1-\eta$ and $\mathbb{P}(\boldsymbol{x}^* \in \mathcal{B}_\eta) \geq 1-\eta$. Notice that intersecting the pre-image $f^{-1}(\Gamma_\eta(\boldsymbol{y}))$ with the constraint set $\mathcal{X}$ or $\mathcal{X} \cap \Phi_{\mu^*}$ does not change the coverage probability since we know $\boldsymbol{x}^* \in \mathcal{X}$ and $\boldsymbol{x}^* \in \mathcal{X} \cap \Phi_{\mu^*}$.

Analogous to the presentation in Batlle et al. (2023), we define *sliced* and *global* max-quantiles that include the Berger-Boos set to control the type-1 error probability of Test (4). Respectively, these are defined as follows:

$$\bar{q}_{\gamma,\eta}(\mu) := \max_{\boldsymbol{x} \in \mathcal{B}_\eta^\mu} Q_{P_{\boldsymbol{x}}}(1-\gamma), \tag{12}$$

$$\bar{q}_{\gamma,\eta} := \max_{\boldsymbol{x} \in \mathcal{B}_\eta} Q_{P_{\boldsymbol{x}}}(1-\gamma), \tag{13}$$

which define the following sliced (sl) and global (gl) confidence sets:

$$C_\alpha^{\mathrm{sl}}(\boldsymbol{y}; \mathcal{B}_\eta) := \{\mu : \lambda(\mu, \boldsymbol{y}; \mathcal{B}_\eta) \leq \bar{q}_{\gamma,\eta}(\mu)\} \tag{14}$$

$$C_\alpha^{\mathrm{gl}}(\boldsymbol{y}; \mathcal{B}_\eta) := \{\mu : \lambda(\mu, \boldsymbol{y}; \mathcal{B}_\eta) \leq \bar{q}_{\gamma,\eta}\}. \tag{15}$$

Both of these max-quantiles maximize a larger γ-quantile over the set of interest instead of the α-quantile as shown in (7) and (8). The following Lemma (analogous to Lemma 2.1 in Batlle et al. (2023)) gives a sufficient condition to select the values γ and η to ensure a $1-\alpha$ confidence set.

Lemma 3.1 (Setting η and γ to Guarantee $1-\alpha$ Coverage)**.** *Let $\alpha \in (0,1)$ and define $\bar{q}_{\gamma,\eta}(\mu)$ according (12) and $C_\alpha^{\mathrm{sl}}(\boldsymbol{y}; \mathcal{B}_\eta)$ according to (14). For $\eta \in (0, \alpha)$, $C_\alpha^{\mathrm{sl}}(\boldsymbol{y}; \mathcal{B}_\eta)$ is a $1-\alpha$ confidence set as long as $(1-\gamma)(1-\eta) = 1-\alpha$.*

Proof. See Appendix A.1. □

The coverage guarantee implied by Lemma 3.1 also implies coverage for $C_\alpha^{\mathrm{gl}}(\boldsymbol{y}; \mathcal{B}_\eta)$, as shown in the following Corollary.

Corollary 3.2. *Assume the same setting as Lemma 3.1. $C_\alpha^{\mathrm{gl}}(\boldsymbol{y}; \mathcal{B}_\eta)$ is a $1-\alpha$ confidence set as long as $(1-\gamma)(1-\eta) = 1-\alpha$.*

Proof. See Appendix A.2. □

Navigating the Berger-Boos parameter choices and trade-offs. Setting $\eta = 0$ in $C_\alpha^{\mathrm{gl}}(\boldsymbol{y}; \mathcal{B}_\eta)$ we maximize $Q_{P_{\boldsymbol{x}}}(1-\gamma) = Q_{P_{\boldsymbol{x}}}(1-\alpha)$ over $\mathcal{X}$, which returns $Q_{1-\alpha}^{\max}$ from Equation (8). For $\eta > 0$, the Berger-Boos construction restricts the quantile maximization to a smaller, bounded subset of the parameter space while maximizing a larger quantile to maintain the desired $1-\alpha$ confidence level of the final interval. Fundamentally, the choice of η reflects a trade-off between these two opposing effects: As we increase η away from 0, the set over which the quantile function is optimized shrinks, but the quantile level (which is equal to $1-\gamma = \frac{1-\alpha}{1-\eta}$) is increased. In Section 6.2, we perform a numerical experiment comparing average interval length for different values of η, showing that a

small $\eta > 0$ can be beneficial not only numerically (since it permits sampling in a bounded space) but also in terms of average length.

3.2 LLR invariance under the BB set

The argument proving Lemma 3.1 relies on guaranteeing $\mathbb{P}\big(\mu^* \in C_\alpha^{\text{sl}}(\boldsymbol{y}; \mathcal{B}_\eta) \mid \boldsymbol{x}^* \in \mathcal{B}_\eta\big) \geq 1 - \gamma$. We emphasize that the introduction of the Berger-Boos set implies a change to the test being inverted. Namely, instead of working with the assumption $\boldsymbol{x}^* \in \mathcal{X}$, we assume $\boldsymbol{x}^* \in \mathcal{B}_\eta = f^{-1}(\Gamma_\eta(\boldsymbol{y})) \cap \mathcal{X}$. This refined assumption propagates to the test statistic definition such that we compute $\lambda(\mu, \boldsymbol{y}; \mathcal{B}_\eta)$ instead of $\lambda(\mu, \boldsymbol{y}; \mathcal{X})$. Because both the Berger-Boos set and the LLR depend on the term $\|f(\boldsymbol{x}) - \boldsymbol{y}\|_2^2$, we have $\lambda(\mu, \boldsymbol{y}; \mathcal{B}_\eta) = \lambda(\mu, \boldsymbol{y}; \mathcal{X})$, which we state formally in the following lemma.

Lemma 3.3 (LLR Invariance)**.** *Suppose $\boldsymbol{x}^* \in \mathcal{B}_\eta$. For all $\mu \in \mathbb{R}$ and $\boldsymbol{y} \in \mathbb{R}^m$, $\lambda(\mu, \boldsymbol{y}; \mathcal{B}_\eta) = \lambda(\mu, \boldsymbol{y}; \mathcal{X})$.*

Proof. See Appendix A.3. □

Fundamentally, there are two ways to consider the incorporation of the Berger-Boos set into the confidence set definitions with respect to the underlying quantile function of the LLR. One way is to see the Berger-Boos set as a way to "cut out" a section of the quantile function defined under the original constraint set. Another way is that for any instantiation of the Berger-Boos set, there is a unique quantile function defined only on that set that is not necessarily equal to the original quantile function defined over the entire constraint set when limited to the Berger-Boos set. Lemma 3.3 states that these two ways are the same when the true parameter is contained in the Berger-Boos set. Note that the proof relies upon the particular ellipsoidal form of the Berger-Boos set constraint which allows us to directly compare quantities with the LLR objective functions under the Gaussian assumption. As such, this invariance depends upon selecting the Berger-Boos set in accordance with the noise distribution. Although this choice makes intuitive sense, the point is worth making since the geometry of the Berger-Boos set is potentially something the statistician can choose.

Given this invariance, in the subsequent theoretical and computational development involving the LLR within the Berger-Boos set, we use $\lambda(\mu, \boldsymbol{y}; \mathcal{X})$ since it is more streamlined.

3.3 Interval constructions

There are possibly many ways to compute $C_\alpha^{\text{gl}}(\boldsymbol{y}; \mathcal{B}_\eta)$ and $C_\alpha^{\text{sl}}(\boldsymbol{y}; \mathcal{B}_\eta)$ in practice. Obtaining either set comes down to the computation of $\bar{q}_{\gamma,\eta}(\mu)$ and $\bar{q}_{\gamma,\eta}$. If the quantile function $Q_{P_{\boldsymbol{x}}}$ and its gradient $\nabla_{\boldsymbol{x}} Q_{P_{\boldsymbol{x}}}$ could be evaluated, computing these quantities could potentially be achieved using a first-order numerical optimizer. However, we emphasize that since the sliced max-quantile, $\bar{q}_{\gamma,\eta}(\mu)$, is a function of the level-set parameter, μ, such an optimization would have to be done for each possible functional value. Unfortunately, such easy function and gradient evaluations rarely exist. As such, this paper develops a *sampling*-based approach to estimate these quantities to construct Confidence Sets (14) and (15). As we demonstrate below, once we estimate $\bar{q}_{\gamma,\eta}(\mu)$ and $\bar{q}_{\gamma,\eta}$, we can either use the output from the sampling algorithm to compute the Global or Sliced interval via classical test inversion, or we can use the estimated max-quantiles in Optimizations (10). As such, we introduce four interval constructions; Global Inverted, Global Optimized, Sliced Inverted, and Sliced Optimized.

Since we can typically sample $\lambda(\mu, \boldsymbol{y}; \mathcal{X})$ by sampling $\boldsymbol{y} \sim P_{\boldsymbol{x}}$, our approach leverages this ability to estimate the desired max-quantiles. We present two algorithms, both of which first generate a random set of design points from the Berger-Boos set defined by the observation. If we can directly *efficiently* compute the desired quantile at each design point, this capability is leveraged in Algorithm 1. If we cannot afford such a computation, Algorithm 2 presents an alternative, which first samples one realization of the LLR test statistic at each sampled design point, and then performs quantile regression with the generated pairs of design points and LLR values to estimate the underlying quantile surface. The sampled design points and either true or estimated quantiles at the design points are then used to produce the final intervals.

Both algorithms start by sampling $\mathcal{B}_\eta = f^{-1}(\Gamma_\eta(\boldsymbol{y})) \cap \mathcal{X}$ uniformly at random to generate a collection of M design points across the Berger-Boos set, i.e., $\bar{\boldsymbol{x}}_1, \bar{\boldsymbol{x}}_2, \ldots, \bar{\boldsymbol{x}}_M \sim \mathcal{U}(\mathcal{B}_\eta)$. This step is non-trivial, even under an additive Gaussian noise assumption where the pre-image is an ellipsoid since the ill-posedness can produce an extremely narrow or unbounded ellipsoid, resulting in sharp boundaries when intersected with the known constraints. We address these sampling challenges in Section 5.1. If computing $\lambda(\mu, \boldsymbol{y}; \mathcal{X})$ for a given μ and $\boldsymbol{y}$ is inexpensive, Algorithm 1 directly estimates the quantile function at each design point. This computation is most likely inexpensive when there are closed-form solutions to the LLR's subordinate optimizations. As shown in Algorithm 1, an easy way to estimate each design point's quantile is to sample its test statistic N times and take the appropriate percentile. More often, computing $\lambda(\mu, \boldsymbol{y}; \mathcal{X})$ is expensive since it involves two constrained optimizations which are possibly non-convex, or just numerically challenging due to the ill-posedness of the problem. In this scenario, one can use Algorithm 2 to sample one realization of the test statistic at each design point and estimate the quantile function over the Berger-Boos set using quantile regression. While Algorithm 1 relies upon computational strength to compute the LLR $N \times M$ times, Algorithm 2 shifts complexity to the quantile regression and relieves the computational burden by assuming that there is information to be shared about the quantile surface between design points. The emphasis on the quantile regression further necessitates that the quantile regression be performed well. We discuss some considerations to this end in Section 5.2. We note that although Algorithm 1 and Algorithm 2 make use of the additive Gaussian noise assumption, they are not limited to this assumption. The consistency of the output for Algorithm 1 is proven by Lemma 4.1, while the consistency of the output of Algorithm 2 is proven by Lemma 4.2.

The approach of Algorithm 2 is inspired by recent uncertainty quantification approaches in likelihood-free inference where one can sample from a likelihood but cannot easily compute it Dalmasso et al. (2020, 2021); Masserano et al. (2023, 2024). Although these approaches differ in detail and implementation from our approach (e.g., they typically focus on low-dimensional settings), they overlap in the sampling and quantile regression perspectives, which effectively allow Machine Learning to act as a bridge over the chasm caused by computationally intensive or intractable forward models. Namely, rather than assuming a stochastic forward model and therefore only being able to sample from the likelihood, we assume f is a deterministic function involved in $\lambda(\mu, \boldsymbol{y}; \mathcal{X})$, which is a random quantity due to the additive noise. This test statistic is likely difficult to compute for the reasons mentioned above, and therefore sampling once from its distribution is computationally expedient.

With the generated pairs from Algorithm 1 and Algorithm 2, we present two strategies to estimate $C_\alpha^{\text{gl}}(\boldsymbol{y}; \mathcal{B}_\eta)$ and two strategies to estimate $C_\alpha^{\text{sl}}(\boldsymbol{y}; \mathcal{B}_\eta)$. To streamline notation, let $\left(\boldsymbol{x}_k, q_\gamma^k\right)$ denote the k-th pair from either algorithm. This is notationally helpful since Algorithm 1 only generates one set of parameter samples, whereas Algorithm 2 generates two. That is, using Algorithm 1, $q_\gamma^k := \widehat{q}_\gamma^k$ and using Algorithm 2, $q_\gamma^k := \widehat{q}_\gamma(\widetilde{\boldsymbol{x}}_k)$.

Algorithm 1 Direct estimation of quantile surface

Input: $\alpha, \gamma, \eta \in (0,1)$, $M, N \in \mathbb{N}$.

1: **Construct Berger-Boos confidence set**: Create $\Gamma_\eta(\boldsymbol{y}) \subseteq \mathbb{R}^m$ such that $\mathbb{P}\left(f(\boldsymbol{x}^*) \in \Gamma(\boldsymbol{y})\right) \geq 1-\eta$. The pre-image $f^{-1}\left(\Gamma(\boldsymbol{y})\right)$ is also a $1-\eta$ confidence set for $\boldsymbol{x}^*$, as is $\mathcal{B}_\eta = f^{-1}(\Gamma_\eta(\boldsymbol{y})) \cap \mathcal{X}$.

2: **Sample from the Berger-Boos confidence set** $\mathcal{B}_\eta$: Sample $\overline{\boldsymbol{x}}_1, \ldots, \overline{\boldsymbol{x}}_M \sim \mathcal{U}\left(\mathcal{B}_\eta\right)$.

3: **for** $k = 1, 2, \ldots, M$ **do**

4: **Sample noise realizations**: Sample N noise realizations from the known noise distribution: $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_N \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{I}\right)$.

5: **Sample from the LLR distribution**: Create an ensemble of LLR draws under $\overline{\boldsymbol{x}}_k$: $\{\lambda_i\}_{i=1}^N$, where $\lambda_i := \lambda\left(\varphi(\overline{\boldsymbol{x}}_k), f(\overline{\boldsymbol{x}}_k) + \boldsymbol{\varepsilon}_i\right)$.

6: **Compute percentile estimate of the γ-quantile**: Compute the $(1-\gamma) \times 100$ percentile of the LLR samples for the data generating process under $\overline{\boldsymbol{x}}_k$, i.e., $\widehat{q}_\gamma^k := \lambda_{(\{(1-\gamma)N\})}$, where $\{\cdot\}$ denotes the nearest whole number and $\lambda_{(i)}$ denotes the i-th order statistic.

7: **end for**

Output: Pairs of sampled design points and their respective γ-quantiles, i.e., $\{(\overline{\boldsymbol{x}}_k, \widehat{q}_\gamma^k)\}_{k=1}^M$.

Algorithm 2 Quantile regression estimate of quantile surface

Input: $\alpha, \gamma, \eta \in (0,1)$; $M_{\text{tr}}, M \in \mathbb{N}$.

1: **Construct Berger-Boos confidence set**: Create $\Gamma_\eta(\boldsymbol{y}) \subseteq \mathbb{R}^m$ such that $\mathbb{P}\left(f(\boldsymbol{x}^*) \in \Gamma(\boldsymbol{y})\right) \geq 1-\eta$. The pre-image $f^{-1}\left(\Gamma(\boldsymbol{y})\right)$ is also a $1-\eta$ confidence set for $\boldsymbol{x}^*$, as is $\mathcal{B}_\eta = f^{-1}(\Gamma_\eta(\boldsymbol{y})) \cap \mathcal{X}$.

2: **Sample from the Berger-Boos confidence set** $\mathcal{B}_\eta$: Sample $\overline{\boldsymbol{x}}_1, \ldots, \overline{\boldsymbol{x}}_{M_{\text{tr}}} \sim \mathcal{U}\left(\mathcal{B}_\eta\right)$ design points to train the quantile regressor and $\widetilde{\boldsymbol{x}}_1, \ldots, \widetilde{\boldsymbol{x}}_M \sim \mathcal{U}\left(\mathcal{B}_\eta\right)$ test points to invert the interval, generating $M_{\text{tr}} + M$ total samples. Since the test points are used for the interval inversion, they are used as out-of-sample points for the quantile regressor.

3: **for** $k = 1, 2, \ldots, M_{\text{tr}}$ **do**

4: **Sample a noise realization:** Sample a noise realization from the known noise distribution: $\boldsymbol{\varepsilon}_k \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{I}\right)$.

5: **Sample from the LLR distribution:** Compute the LLR under $\overline{\boldsymbol{x}}_k$ with sampled noise $\boldsymbol{\varepsilon}_k$: $\lambda_k := \lambda\left(\varphi(\overline{\boldsymbol{x}}_k), f(\overline{\boldsymbol{x}}_k) + \boldsymbol{\varepsilon}_k\right)$.

6: **end for**

7: **Estimate the quantile function using quantile regression:** Using the generated pairs $\{(\overline{\boldsymbol{x}}_k, \lambda_k)\}_{k=1}^{M_{\text{tr}}}$, estimate the upper γ-conditional quantile function, $\widehat{q}_\gamma(\boldsymbol{x})$, using quantile regression.

Output: Generate γ-quantile predictions at out-of-sample test points, $\{(\widetilde{\boldsymbol{x}}_k, \widehat{q}_\gamma(\widetilde{\boldsymbol{x}}_k))\}_{k=1}^M$.

To estimate the global confidence set, we estimate $\bar{q}_{\gamma,\eta}$ using the empirical max $\widehat{q} := \max_k q_\gamma^k$. This results in the following two interval constructions:

$$C_{\text{inv}}^{\text{gl}}(\boldsymbol{y}) := \left[\min_{k \in I^{\text{gl}}(\boldsymbol{y})} \varphi(\boldsymbol{x}_k), \max_{k \in I^{\text{gl}}(\boldsymbol{y})} \varphi(\boldsymbol{x}_k)\right], \tag{16}$$

$$C_{\text{opt}}^{\text{gl}}(\boldsymbol{y}) := \left[\min_{\boldsymbol{x} \in D(\widehat{q}+s(\boldsymbol{y})^2, \boldsymbol{y})} \varphi(\boldsymbol{x}), \max_{\boldsymbol{x} \in D(\widehat{q}+s(\boldsymbol{y})^2, \boldsymbol{y})} \varphi(\boldsymbol{x})\right], \tag{17}$$

where $I^{\text{gl}}(\boldsymbol{y}) = \{k : \lambda(\varphi(\boldsymbol{x}_k), \boldsymbol{y}; \mathcal{X}) \leq \widehat{q}\} \subseteq [M]$. We refer to $I^{\text{gl}}(\boldsymbol{y})$ as the "global" index set, i.e.,

those indices for which the LLR at a particular functional value is less than the maximum estimated quantile. We refer to Interval (16) as the Global Inverted interval construction since it is endpoints defined by only those sampled parameter values that comport with the LLR cutoff. We refer to Interval (17) as the Global Optimized interval construction since its endpoints are defined by the extreme functional values of a feasible region defined by $\widehat{q}$. Although, $C^{\text{gl}}_{\text{inv}}(\boldsymbol{y}) \neq C^{\text{gl}}_{\text{opt}}(\boldsymbol{y})$ due to finite sample, they are asymptotically equal and in practice show similar performance in terms of coverage and expected length, as shown in Section 6. We prove the consistency of the interval constructed via inversion in Lemma 4.3 and via optimization in Lemma 4.5. Since the endpoints of both the inverted and optimized intervals both converge in probability to the endpoints of $C^{\text{gl}}_{\alpha}(\boldsymbol{y}; \mathcal{B}_\eta)$, the two interval constructions are asymptotically equivalent.

To estimate the sliced set, we estimate $\bar{q}_{\gamma,\eta}(\mu)$ using a rolling maximum of the sampled q^k_γ values, as ordered by the sampled functional values. This results in two interval constructions:

$$C^{\text{sl}}_{\text{inv}}(\boldsymbol{y}) := \left[\min_{k \in I^{\text{sl}}(\boldsymbol{y})} \varphi(\boldsymbol{x}_k), \max_{k \in I^{\text{sl}}(\boldsymbol{y})} \varphi(\boldsymbol{x}_k) \right], \tag{18}$$

$$C^{\text{sl}}_{\text{opt}}(\boldsymbol{y}) := \left[\min_{\mu \in L(\boldsymbol{y})} \mu, \max_{\mu \in L(\boldsymbol{y})} \mu \right], \tag{19}$$

where $I^{\text{sl}}(\boldsymbol{y}) = \{k : \lambda(\varphi(\boldsymbol{x}_k), \boldsymbol{y}; \mathcal{X}) \leq q^k_\gamma\} \subseteq [M]$, $L(\boldsymbol{y}) := \{\mu : \lambda(\mu, \boldsymbol{y}; \mathcal{X}) \leq \widehat{m}_\gamma(\mu)\}$ and $\widehat{m}_\gamma(\mu)$ denotes rolling estimate of $\bar{q}_{\gamma,\eta}(\mu)$ defined as follows. We refer to $I^{\text{sl}}(\boldsymbol{y})$ as the "sliced" index set, i.e., those indices for which the LLR at a particular functional value is less than the estimated quantile at a design point generating that functional value.

The rolling maximum quantile is defined using estimated quantiles ordered by the sampled functional values. Choose "rolling" parameter, $T \in \mathbb{N}$. Namely, let $\sigma(1), \sigma(2), \ldots, \sigma(M)$ define an ordering such that $\mu_{\sigma(k)} \leq \mu_{\sigma(k+1)}$ for all $k = 1, \ldots, M-1$. Define $Q_k := \{q^{\sigma(k)}_\gamma, q^{\sigma(k-1)}_\gamma, \ldots, q^{\sigma(k-T)}_\gamma\}$ Then, for a given $\mu \in \left[\min_{\boldsymbol{x} \in \mathcal{B}_\eta} \varphi(\boldsymbol{x}), \max_{\boldsymbol{x} \in \mathcal{B}_\eta} \varphi(\boldsymbol{x})\right]$, define

$$k^*(\mu) := \operatorname*{argmin}_k |\mu - \mu_{\sigma(k)}|. \tag{20}$$

and $\widehat{m}_\gamma(\mu) := \max\{q \in Q_{k^*(\mu)}\}$. Using the rolling maximum quantile is one possible way to estimate $\bar{q}_{\gamma,\eta}(\mu)$. One could also bin the quantiles by functional value, compute the maximum predicted quantile in each bin, and then fit a nonparametric regression to fit the maximum binned quantiles to the functional values.

This estimator choice for $\widehat{m}_\gamma(\mu)$ affects how one computes $C^{\text{sl}}_{\text{opt}}(\boldsymbol{y})$. To see why, note that we can re-express $C^{\text{sl}}_{\text{opt}}(\boldsymbol{y})$ as follows:

$$C^{\text{sl}}_{\text{opt}}(\boldsymbol{y}) := \left[\min_{\boldsymbol{x} \in D(\widehat{m}_\gamma(\varphi(\boldsymbol{x})) + s(\boldsymbol{y})^2, \boldsymbol{y})} \varphi(\boldsymbol{x}), \max_{\boldsymbol{x} \in D(\widehat{m}_\gamma(\varphi(\boldsymbol{x})) + s(\boldsymbol{y})^2, \boldsymbol{y})} \varphi(\boldsymbol{x}) \right]. \tag{21}$$

As such, one could substitute the estimated $\widehat{m}_\gamma(\cdot)$ into each endpoint optimization. However, this estimated curve is likely not convex in $\boldsymbol{x}$ and therefore complicates the optimizations. One could also pursue a root-finding approach to find the set of μ where $\lambda(\mu, \boldsymbol{x}; \mathcal{X}) = \widehat{m}_\gamma(\mu)$. This approach can also be complex in proportion to the complexity of the estimated curve. In our view, the most pragmatic approach is to simply determine all $\mu_k = \varphi(\widetilde{\boldsymbol{x}}_k)$ such that $\lambda(\mu_k; \boldsymbol{y}; \mathcal{X}) \leq \widehat{m}_\gamma(\mu_k)$, and then define the endpoints of $C^{\text{sl}}_{\text{opt}}(\boldsymbol{y})$ to be the minimum and maximum of those accepted sampled values. The consistency of the computed endpoints for both the inverted and optimized sliced intervals is proven in Lemmas 4.4 and 4.6, respectively. Similar to the global interval constructions,

Table 1: Summary of methods based on global and sliced max quantile, and whether they are optimization-based or inversion-based.

Category	Global max quantile	Sliced max quantile
Inversion-based methods	Interval (16)	Interval (18)
Optimization-based methods	Interval (17)	Interval (19)

the consistency of both constructions to $C_\alpha^{\mathrm{sl}}(\boldsymbol{y}; \mathcal{B}_\eta)$ also establishes the asymptotic equivalent of the inverted and optimized approaches.

The statement and result of Lemma 3.1 show that given $\alpha \in (0,1)$, the statistician has the flexibility to select (η, γ) such that $(1-\eta)(1-\gamma) = 1-\alpha$. In practice, one selects $\alpha \in (0,1)$ followed by $\eta \in (0, \alpha)$, implying a unique value for γ. As mentioned in the discussion following Lemma 3.1, larger values of η mean smaller Berger-Boos sets and smaller upper-quantiles. Ideally, one wants to pick $\eta \in (0, \alpha)$ that minimizes the maximum γ-quantile over the Berger-Boos set, but this is difficult to do in practice.

4 Theoretical results

Although both Algorithm 1 and Algorithm 2 can produce an estimate $\bar{q}_\gamma$, it is worth noting a few key differences between the two. First, Algorithm 1 involves a nested sampling loop. Thus, any statistical guarantee regarding the validity of its output has the two moving parts of the accuracy of $\widehat{q}_\gamma^k$ as N gets large and the proximity of $\bar{q}_\gamma$ to the true maximum quantile function value over $\mathcal{B}_\eta$ as M gets large. By contrast, Algorithm 2 only has one sampling loop and estimates the full quantile function surface over $\mathcal{B}_\eta$, producing an estimate of the maximum as a consequence.

In this subsection, we provide the necessary mathematical results supporting the use of Algorithm 1 and Algorithm 2 as estimators for $\bar{q}_{\gamma,\eta}$ in addition to proving that extremes of the estimated intervals defined in Section 3.3 converge in probability to their respective theoretical targets, $C_\alpha^{\mathrm{gl}}(\boldsymbol{y}; \mathcal{B}_\eta)$ and $C_\alpha^{\mathrm{sl}}(\boldsymbol{y}; \mathcal{B}_\eta)$. Let $\bar{q}_{\gamma,\eta}^{\mathrm{de}}$ denote the empirical max-quantile from Algorithm 1 and $\bar{q}_{\gamma,\eta}^{\mathrm{qr}}$ denote the empirical max-quantile from Algorithm 2. Lemma 4.1 proves the consistency of $\bar{q}_{\gamma,\eta}^{\mathrm{de}}$ as an estimator for $\bar{q}_{\gamma,\eta}$ and Lemma 4.2 proves the consistency of $\bar{q}_{\gamma,\eta}^{\mathrm{qr}}$ as an estimator for $\bar{q}_{\gamma,\eta}$. Note that by changing the set that is sampled in the algorithms, the results hold true both for estimating the global and sliced maximum quantiles. Lemma 4.5 and Lemma 4.3 then establishes that the Global interval constructions using either $\bar{q}_{\gamma,\eta}^{\mathrm{de}}$ or $\bar{q}_{\gamma,\eta}^{\mathrm{qr}}$ to estimate the max-quantile converge in probability to the true Global interval, $C_\alpha^{\mathrm{gl}}(\boldsymbol{y}; \mathcal{B}_\eta)$. Finally, Lemma 4.4 and Lemma 4.6 establish that the Sliced interval constructions using the estimated quantiles from Algorithm 2 converge in probability to $C_\alpha^{\mathrm{sl}}(\boldsymbol{y}; \mathcal{B}_\eta)$.

Throughout the following results, we assume that for fixed $\gamma \in (0,1)$, $Q_{P_{\boldsymbol{x}}}(1-\gamma)$ is a continuous function of $\boldsymbol{x}$. We refer the reader to Kibzun and Kan (1997) for a full analysis of the properties of parametrized quantile functions. We furthermore assume that $\mathcal{X} \cap \mathcal{B}_\eta$ does not have isolated points, so that the sampling method eventually samples close to every point. A key lemma supporting the following results and pertaining to these two assumptions is stated and proven in Lemma B.1.

Lemma 4.1 (Consistency of $\bar{q}_{\gamma,\eta}^{\mathrm{de}}$)**.** *Let $\bar{q}_{\gamma,\eta}^{\mathrm{de}}$ be generated from Algorithm 1. We have $\bar{q}_{\gamma,\eta}^{\mathrm{de}} \xrightarrow{\mathrm{p}} \bar{q}_{\gamma,\eta}$ as $N, M \to \infty$.*

Proof. See Appendix B.1. □

To establish a similar consistency result for Algorithm 2, it is necessary to assume the consistency of quantile regression estimator used to estimate $Q_{P_{\boldsymbol{x}}}(1-\gamma)$. Consistency is true for a variety of nonparametric quantile regression options (e.g., Meinshausen (2006); Takeuchi et al. (2006)). We define consistency as follows:

Definition 1 (Consistency of a quantile estimator)**.** Let $\mathcal{C} \subset \mathcal{X}$ be compact, $\epsilon > 0$, and $\gamma \in (0,1)$. For each $\boldsymbol{x} \in \mathcal{C}$, the quantile estimator $\widehat{q}_\gamma$ is *consistent* if

$$\mathbb{P}(|\widehat{q}_\gamma(\boldsymbol{x}) - Q_{P_{\boldsymbol{x}}}(1-\gamma)| > \epsilon) \to 0 \text{ as } M_{tr} \to \infty. \tag{22}$$

We also write $\widehat{q}_\gamma(\boldsymbol{x}) \xrightarrow{\text{p}} Q_{P_{\boldsymbol{x}}}(1-\gamma)$.

Using the consistency of a quantile estimator, Lemma 4.2 shows the implied consistency of the maximum quantile.

Lemma 4.2 (Consistency of $\bar{q}^{\text{qr}}_{\gamma,\eta}$)**.** *Let $\bar{q}^{\text{qr}}_{\gamma,\eta}$ be the maximum estimated quantile output of Algorithm 2. Suppose $\widehat{q}_\gamma$ is a consistent estimator per Definition 1. Then, $\bar{q}^{\text{qr}}_{\gamma,\eta}$ is a consistent estimate of $\bar{q}_{\gamma,\eta}$.*

Proof. See Appendix B.2. □

By Lemmas 4.1 and 4.2, it is guaranteed that as the number of samples gets large, the resulting maximum quantile estimate approaches the true maximum quantile over $f^{-1}(\Gamma_\eta(\boldsymbol{y})) \cap \mathcal{X}$ with high probability. We henceforth assume we have access to an algorithm with these guarantees to estimate quantiles (and in particular, the maximum quantiles) and prove that it can be used to asymptotically approximate the optimal confidence intervals.

We prove the asymptotic equality of the four total interval constructions: The inversion and the optimization approach to both $C^{\text{gl}}_\alpha(\boldsymbol{y}; \mathcal{B}_\eta)$ and $C^{\text{sl}}_\alpha(\boldsymbol{y}; \mathcal{B}_\eta)$. In particular, we focus on convergence in probability to the extremes of the interval. We assume henceforth continuity of the quantity of interest function φ The first two results analyze the inverted constructions.

Lemma 4.3. *The extremes of $C^{\text{gl}}_{\text{inv}}(\boldsymbol{y})$ converge in probability to the extremes of $C^{\text{gl}}_\alpha(\boldsymbol{y}; \mathcal{B}_\eta)$ as $M \to \infty$*

Proof. See Appendix B.3. □

Lemma 4.4. *The extremes of $C^{\text{sl}}_{\text{inv}}(\boldsymbol{y})$ converge in probability to the extremes of $C^{\text{sl}}_\alpha(\boldsymbol{y}; \mathcal{B}_\eta)$ as $M_{\text{tr}}, M \to \infty$*

Proof. See Appendix B.4. □

Convergence of the optimized constructions requires extra assumptions. For the global version, we assume the existence of a strictly feasible point in the optimization problems in (21). If $\lambda(\mu, \boldsymbol{y}; \mathcal{X})$ is convex in μ, as it is for linear forward models and quantities of interest, see (Batlle et al., 2023, Proposition 2.5), this condition is equivalent to the interval having positive Lebesgue measure.

Lemma 4.5. *For all $\boldsymbol{y} \in \mathbb{R}^m$ such that there exists a point $\bar{\mu} \in \varphi(\mathcal{X} \cap \mathcal{B}_\eta)$ satisfying $\lambda(\bar{\mu}, \boldsymbol{y}; \mathcal{X}) < \bar{q}_{\gamma,\eta}$, $C^{\text{gl}}_{\text{opt}}(\boldsymbol{y})$ converges in probability to $C^{\text{gl}}_\alpha(\boldsymbol{y}; \mathcal{B}_\eta)$ as $M_{\text{tr}}, M \to \infty$.*

Proof. See Appendix B.5. □

For the convergence of the local optimized version, we aim to show $\inf_{\mu:\lambda(\mu,y)\leq\widehat{m}_\gamma(\mu)}\mu$ to $\inf_{\mu:\lambda(\mu,y)\leq m_\gamma(\mu)}\mu$ as $\widehat{m}_\gamma(\mu)$ converges to $m_\gamma(\mu)$. We study convergence of optimization problems of the form $\inf_{\mu:\widehat{f}(\mu)\geq 0}\mu$ to $\inf_{\mu:f(\mu)\geq 0}\mu$ as $\widehat{f}$ converges to f. While the result is not true in general, we provide sufficient technical conditions about f and the uniformness of convergence of $\widehat{f}$ to f for the result to hold. We discuss the details in appendix B.

Lemma 4.6. *(Informal). Assume $\widehat{m}_\gamma(\mu) \xrightarrow{\text{p}} m_\gamma(\mu)$ uniformly as $M_{\text{tr}} \to \infty$ for all μ and technical conditions of smoothness and lack of oscillation of $m_\gamma(\mu) - \lambda(\mu, y)$. Then, $C^{\text{sl}}_{\text{opt}}(\boldsymbol{y})$ converges in probability to $C^{\text{sl}}_{\alpha}(\boldsymbol{y}; \mathcal{B}_\eta)$ as $M \to \infty$.*

Proof. See Appendix B.6. □

5 Adaptive confidence interval: implementation methodology

5.1 Sampling the pre-image Berger-Boos set

The viability of this method directly relies upon our ability to sample from the Berger-Boos set, $\mathcal{B}_\eta = f^{-1}(\Gamma_\eta(\boldsymbol{y}))\cap\mathcal{X}$. Under the assumed data generating process in Equation (11), the Berger-Boos set is defined as follows:

$$\mathcal{B}_\eta = f^{-1}(\Gamma_\eta(\boldsymbol{y})) \cap \mathcal{X} = \left\{\boldsymbol{x} \in \mathcal{X} : (\boldsymbol{y} - f(\boldsymbol{x}))^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - f(\boldsymbol{x})) \leq \chi^2_{n,\gamma}\right\}, \tag{23}$$

where we have generalized to a non-identity covariance matrix, $\boldsymbol{\Sigma}$. This set is equivalent to the set over which the strict bounds intervals are optimized in Stark (1992b), also called "SSB" intervals in Stanley et al. (2022); Batlle et al. (2023). Note however, that one would use $\chi^2_{n,\alpha}$ instead of $\chi^2_{n,\gamma}$ for interval computation in that scenario.

We discuss sampling $\mathcal{B}_\eta$ in the linear-Gaussian case, where $f(\boldsymbol{x}) = \boldsymbol{K}\boldsymbol{x}$, as this scenario aligns to the numerical experiments presented in Section 6. We present two sampling approaches; the "Voelker-Gossman-Stewart" (VGS) algorithm based on Voelker et al. (2017) for efficiently sampling ellipsoids uniformly at random in low-dimensional scenarios and an MCMC-based algorithm we call the Polytope sampler based on the Vaidya walk presented in Chen et al. (2018). When f is linear and full-column rank, $\mathcal{B}_\eta$ is an ellipsoid intersected with the constraint set, making the VGS algorithm an effective option (see Section 5.1.1). In all other scenarios, especially higher dimensional ones, the Polytope sampler is a better option, as naive accept-reject algorithms become intractable (see Section 5.1.2).

5.1.1 VGS Sampler for low-dimensional and full column rank settings

Under the linear-Gaussian assumptions, when $\boldsymbol{K}$ is full column rank, it can be shown that

$$\mathcal{B}_\eta = \left\{\boldsymbol{x} \in \mathcal{X} : (\boldsymbol{x} - \widehat{\boldsymbol{x}})^\top \boldsymbol{K}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{K}(\boldsymbol{x} - \widehat{\boldsymbol{x}}) \leq \chi^2_{n,\gamma}\right\}, \tag{24}$$

where $\widehat{\boldsymbol{x}}$ is the Generalized Least-Squares (GLS) estimator. Equation (24) describes an ellipsoid in $\mathbb{R}^p$ intersected with $\mathcal{X}$ with axis directions and lengths determined by $\chi^2_{n,\gamma}$ and the eigenvectors and eigenvalues of $\boldsymbol{K}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{K}$, respectively (see Rust and Burrus (1972) for more details). An ellipsoid is nothing but a deformed ball. As such, we can sample uniformly at random from the ellipsoid in $\mathbb{R}^p$ using an algorithm sampling uniformly at random from a p-ball, followed by the appropriate linear transformation and translation. We can then include an additional accept-reject step to account for the constraint set, $\mathcal{X}$. Note, all subsequent discussions of spheres and balls assume unit radii

and centering at the origin. Also note that because the ellipsoid in (24) is centered around the GLS estimator, we can sample points first from within an ellipsoid of the same shape centered at the origin, and then translate those points by $\widehat{\boldsymbol{x}}$.

Voelker et al. (2017) propose a particularly efficient and clever algorithm to sample uniformly at random from the p-ball by proving a connection between uniform sampling on the $(p+1)$-sphere (i.e., the surface of the ball in $\mathbb{R}^{p+2}$) and uniform sampling within the p-ball (i.e., *within* the unit ball in $\mathbb{R}^p$). Namely, one can sample a Gaussian $\widetilde{\boldsymbol{z}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{p+2})$ and normalize $\boldsymbol{z} := \widetilde{\boldsymbol{z}}/\|\widetilde{\boldsymbol{z}}\|_2$ to sample a point uniformly at random from the $(p+1)$-sphere. Next, one simply drops the last two elements of $\boldsymbol{z}$ to obtain a point that is uniformly sampled from within a p-ball. The validity of this approach is substantiated by Lemma 1 and Theorem 1 in Voelker et al. (2017) and relies upon a distribution result about the ratio of chi-squared distributions and preservation of distribution under orthogonal transformation. To denote sampling a point from this procedure, we use the notation $\boldsymbol{x} \sim \text{VGS}(p)$.

To sample from our desired ellipsoid in (24), first consider the eigendecomposition of $\boldsymbol{K}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{K} = \boldsymbol{P} \boldsymbol{\Omega} \boldsymbol{P}^\top$, where $\boldsymbol{\Omega} = \text{diag}\left(\omega_1^2, \omega_2^2, \ldots, \omega_p^2\right)$, ω_i is the i-th eigenvalue of $\boldsymbol{K}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{K}$ and $\boldsymbol{P}$ is an orthonormal matrix where the columns vectors are the eigenvectors of $\boldsymbol{K}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{K}$. Denote $\boldsymbol{\Omega}^{1/2} = \text{diag}\left(\omega_1, \omega_2, \ldots, \omega_p\right)$ and by extension, $\boldsymbol{\Omega}^{-1/2} = \text{diag}\left(\omega_1^{-1}, \omega_2^{-1}, \ldots, \omega_p^{-1}\right)$ for when $\omega_i > 0$ for all i. These decompositions imply the correct transformation to apply to points sampled from the p-ball via $\boldsymbol{x} \sim \text{VGS}(p)$. Namely, define $\boldsymbol{w} := \sqrt{\chi^2_{n,\gamma}} \boldsymbol{P} \boldsymbol{\Omega}^{-1/2} \boldsymbol{x}$. To know that $\boldsymbol{w}$ is sampled from the correct ellipsoid, it should be the case that $\boldsymbol{w}^\top \boldsymbol{K}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{K} \boldsymbol{w} \leq \chi^2_{n,\gamma}$, as then we can simply make the update $\boldsymbol{w} := \boldsymbol{w} + \widehat{\boldsymbol{x}}$ to ensure that we have a point sampled from (24). This guarantee is verified as follows:

$$\begin{aligned} \boldsymbol{w}^\top \boldsymbol{K}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{K} \boldsymbol{w} &= \boldsymbol{w}^\top \boldsymbol{P} \boldsymbol{\Omega}^{1/2} \boldsymbol{\Omega}^{1/2} \boldsymbol{P}^\top \boldsymbol{w} \\ &= \chi^2_{n,\gamma} \cdot \boldsymbol{x}^\top \boldsymbol{\Omega}^{-1/2} \boldsymbol{P}^\top \boldsymbol{P} \boldsymbol{\Omega}^{1/2} \boldsymbol{\Omega}^{1/2} \boldsymbol{P}^\top \boldsymbol{P} \boldsymbol{\Omega}^{-1/2} \boldsymbol{x} \\ &= \chi^2_{n,\gamma} \cdot \boldsymbol{x}^\top \boldsymbol{x} \leq \chi^2_{n,\gamma}, \end{aligned} \tag{25}$$

where $\boldsymbol{P}^\top \boldsymbol{P} = \boldsymbol{I}$ by definition and the last line follows since we know $\boldsymbol{x}$ is sampled from the p-ball and therefore $\boldsymbol{x}^\top \boldsymbol{x} \leq 1$. Finally, we accept $\boldsymbol{w}$ if $\boldsymbol{w} \in \mathcal{X}$ and reject if $\boldsymbol{w} \notin \mathcal{X}$. This procedure for sampling at random from $\mathcal{B}_\eta$ is summarized in Algorithm 3.

Algorithm 3 VGS Sampler

Input: $p \in \mathbb{N}$, $M \in \mathbb{N}$, $\boldsymbol{P}$, $\boldsymbol{\Omega}^{-1/2}$, $\widehat{\boldsymbol{x}}$, $\chi^2_{n,\gamma}$.

1: Define $\mathcal{S} := \{\}$ to be the initialized set in which the sampled points are to be placed.
2: **for** $k = 1, 2, \ldots, M$ **do**
3: **Sample from the p-ball using VGS:** $\boldsymbol{x}_k := \boldsymbol{z}_{1:p}/\|\boldsymbol{z}\|_2$, where $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{p+2})$ and $\boldsymbol{z}_{1:p}$ denotes taking the first through p-th indices (inclusive).
4: **Transform VGS output:** $\boldsymbol{w}_k := \sqrt{\chi^2_{n,\gamma}} \boldsymbol{P} \boldsymbol{\Omega}^{-1/2} \boldsymbol{x}_k$.
5: **Translate $\boldsymbol{w}_k$ by the GLS estimator:** $\boldsymbol{w}_k \leftarrow \boldsymbol{w}_k + \widehat{\boldsymbol{x}}$.
6: **Accept-Reject to incorporate constraints**: If $\boldsymbol{w}_k \in \mathcal{X}$, then include $\mathcal{S}[k] \leftarrow \boldsymbol{w}_k$, else start loop iteration k again.
7: **end for**

Output: $\mathcal{S}$ containing uniformly sampled points over $\mathcal{B}_\eta$ (as defined in (24)).

Generating samples from the VGS Sampler is efficient, but its feasibility diminishes in high dimensions because of the accept-reject step. To illustrate this point, consider the simple scenario where

$\mathcal{X} = \mathbb{R}^p_+$, i.e., the non-negative orthant of $\mathbb{R}^p$ and suppose we sample from the p-ball intersected with $\mathbb{R}^p_+$. Let $\boldsymbol{x} \sim \text{VGS}(p)$. Then, $\mathbb{P}(\boldsymbol{x} \in \mathbb{R}^p_+) = 2^{-p}$ and the acceptance probability of each sample goes to zero exponentially in p. To more concretely make this point, we generate data $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{x}^*_p, \boldsymbol{I}_p)$ where $\boldsymbol{x}^*_p \in \mathbb{R}^p_+$ and is a vector of ones. For a collection of dimensions $p \in [2, 30]$, we estimate the probability that $\widetilde{\boldsymbol{x}} \sim \text{VGS}(p)$ is in $\mathbb{R}^p_+$ and plot the results in the left panel of Figure 5.1. Since the acceptance probability decays exponentially (under 10^{-4} for 30 dimensions), it becomes clear that this algorithm is efficient in dimensions larger than 10, since the acceptance probability quickly becomes prohibitively small in regimes where larger sample sizes are even more important.

Dimension is only one of two primary complicating factors. Not only is less of a (potentially) shifted p-ball's volume in the non-negative orthant as p grows but if our data are generated via $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{K}\boldsymbol{x}^*, \boldsymbol{I}_p)$ with $\boldsymbol{x}^* \in \mathbb{R}^p_+$ where the conditioning number of $\boldsymbol{K}$ is large, even less of the ellipsoid from which the VGS Sampler draws points intersects with the parameter constraint. To illustrate this point, we generate a single observation from the aforementioned model using a $\boldsymbol{K} \in \mathbb{R}^{40\times 40}$ as described in Section 6.3. The $\boldsymbol{x}^* \in \mathbb{R}^{40}$ is created in the same way as that of Section 6.3.1. The right panel of Figure 5.1 shows a computed probability mass function for the number of coordinates in a VGS Sampler draw complying with the non-negativity parameter constraints. For this particular setup, we critically note that out of 5×10^4 draws from the VGS Sampler, none of the draws had all coordinates comply with the non-negativity constraint. By contrast, for the aforementioned noise model, the computed acceptance probability is approximately 3.35×10^{-6}, in both cases emphasizing the VGS Sampler's poor performance in high-dimensional and non-identity forward model regimes.

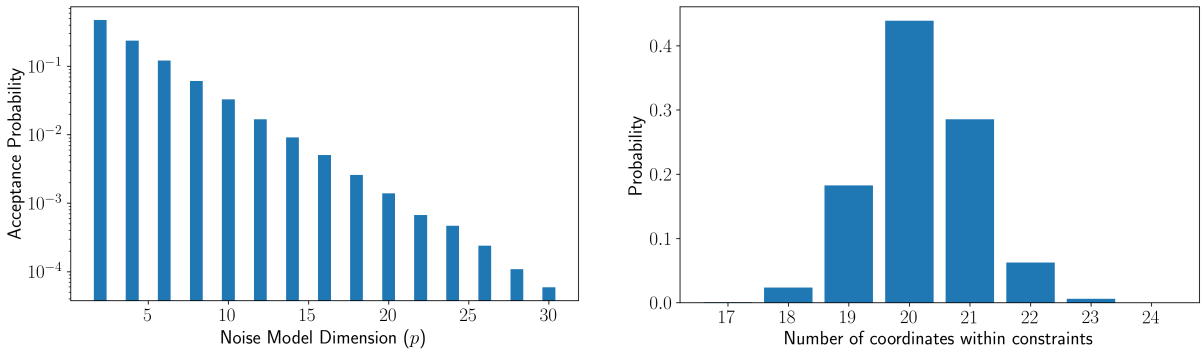

Figure 5.1: Numerical illustrations of the VGS Sampler's infeasibility in high dimensional regimes. The **left** panel shows the computed acceptance probability of a point drawn by the VGS Sampler with data generated from a non-negatively constrained Gaussian noise model. Crucially, at only 30 dimensions, the acceptance probability is already less than 10^{-4} for this particular setup. The **right** panel shows the computed probability mass function for the number of non-negative constraint complying coordinates of a VGS Sample with data generated from a non-negatively constrained linear Gaussian model in 40 dimensions with a non-identity forward model. Since this is an example using a forward model with a high conditioning number ($\approx 1.6 \times 10^4$), we critically note that there is empirically zero probability of generating a sample within the non-negativity constraints.

5.1.2 Polytope sampler for general settings

In settings where the forward model is not linear and full-column rank, Algorithm 3 fails. The ineffectiveness of this algorithm expands if the conditioning number of the linear forward model is large such that most of the pre-image ellipsoid lies outside of the constraint set. Although these scenarios induce particular geometric challenges, we are still fundamentally sampling a convex set for which

there is a vast literature. In the statistics literature, there is a vast sampling literature for computing Bayesian posteriors in high dimensions via nested sampling (Skilling, 2004; Ashton et al., 2022; Buchner, 2023). In particular, nested sampling has been successfully applied in high-dimensional cosmology settings using sophisticated approaches to strategically sample the parameter space by restricting prior sampling in various ways (Buchner, 2023; Montel et al., 2023). Although these approaches provide tools addressing a sampling setting similar to ours (i.e., the Berger-Boos set can be viewed as a portion of the parameter space defined by a cutoff on the likelihood) they are ultimately aimed at sampling from a particular distribution (i.e., the posterior), which is a stronger criterion than required here. For sampling general convex sets, simple algorithms like Hit-and-Run are available (Smith, 1984; Lovasz, 1999; Lovasz and Vempala, 2006). However, with more assumptions, more sophisticated and efficient algorithms can be devised. In particular, there exists a deep literature on random walks over polytopes such that the asymptotic stationary distribution of the chain is a uniform distribution over the polytope of interest (Kannan and Narayanan, 2012; Narayanan, 2016; Chen et al., 2018). As such, we propose to first construct a bounding polytope, $\mathcal{P}^d$ composed of d halfspaces around $\mathcal{B}_\eta$, sample C random walks within the Berger-Boos set using the Vaidya walk as described in Chen et al. (2018) each starting at a point from a collection of strategically chosen positions, and then combine the parallel chains to create the final sample set. The precisely written algorithm can be seen in Algorithm 4.

Although MCMC algorithms are typically evaluated using trace plots on individual dimensions of the parameter space, given the final step combining several MCMC chains in the parameter space started from different positions, it is more meaningful to evaluate this algorithm's ability to sample fully from the functional space. Namely, we can solve for the largest and smallest values the functional can take within the Berger-Boos set as follows:

$$I_{\mathsf{BB}}(\boldsymbol{y}) := \left[\mu^l_{\mathsf{BB}}, \mu^u_{\mathsf{BB}}\right] = \left[\min_{\boldsymbol{x}\in\mathcal{B}_\eta} \varphi(\boldsymbol{x}), \max_{\boldsymbol{x}\in\mathcal{B}_\eta} \varphi(\boldsymbol{x})\right]. \tag{26}$$

When $\varphi(\boldsymbol{x}) = \boldsymbol{h}^\top \boldsymbol{x}$ for some $\boldsymbol{h} \in \mathbb{R}^p$, $I_{\mathsf{BB}}(\boldsymbol{y})$ corresponds to the SSB interval in Stanley et al. (2022). Computing the sets $C^{\text{sl}}_\alpha(\boldsymbol{y}; \mathcal{B}_\eta)$ and $C^{\text{gl}}_\alpha(\boldsymbol{y}; \mathcal{B}_\eta)$ well is then contingent upon sampling functional values within $I_{\mathsf{BB}}(\boldsymbol{y})$ well since a functional value can only be included in the inverted set if the sampler has a non-zero probability of sampling arbitrarily near it. By "well", we informally mean that the sampled functional values range at least between the endpoints of $I_{\mathsf{BB}}(\boldsymbol{y})$ and that if we partition $I_{\mathsf{BB}}(\boldsymbol{y})$ into n pieces, that most if not all of the pieces contain at least one sample. In practice, it will often be the case that the sampled functional values can lie outside the endpoints of $I_{\mathsf{BB}}(\boldsymbol{y})$ since the MCMC chains are sampled in a bounding polytope of the Berger-Boos set.

Choosing the bounding polytope. Any bounded polytope of $\mathcal{B}_\eta$ is the intersection of a finite number of half-spaces defined in $\mathbb{R}^p$. The task of constructing such a bounding polytope is then equivalent to choosing a collection of d hyperplanes in $\mathbb{R}^p$ to construct a set $\mathcal{P}^d := \{\boldsymbol{x} \in \mathbb{R}^p : \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\}$ such that $\mathcal{B}_\eta \subseteq \mathcal{P}^d$, where $\boldsymbol{A} \in \mathbb{R}^{d\times p}$ and $\boldsymbol{b} \in \mathbb{R}^d$. Let $\boldsymbol{a}_i \in \mathbb{R}^p$ denote the i-th row vector of $\boldsymbol{A}$. We compute b_i (the i-th element of $\boldsymbol{b}$) as

$$b_i = \max_{\boldsymbol{x}\in\mathcal{B}_\eta} \boldsymbol{a}_i^\top \boldsymbol{x}. \tag{27}$$

This construction ensures the necessary inclusion. We consider three approaches to pick the vectors, $\boldsymbol{a}_i$; from the physical constraints, from the known eigenvectors defining the bounded directions of the pre-image ellipsoid, and randomly. In practice, we combine these approaches to ensure that we consider only parameter settings in agreement with our physical constraints, and to tighten as much

as possible the bounding-Berger-Boos set polytope as much as possible. There is a tradeoff with respect to the latter consideration since the mixing time and computational costs for the Vaidya walk increase with the number of hyperplanes (Chen et al., 2018).

To incorporate the physical constraints, consider the non-negativity constraint used in Section 6, i.e., $\boldsymbol{x} \in \mathbb{R}^p_+$. To enforce non-negativity, we set $\boldsymbol{a}_i := -\boldsymbol{e}_i$ for $i = 1, \ldots, p$, where $\boldsymbol{e}_i$ is defined by its i-th element set to one and the rest of its elements set to zero and $b_i := 0$. These choices produce p rows in $\boldsymbol{A}$ corresponding to the desired lower bounds (i.e., $x_i \geq 0$ for all i), but we can compute an additional p constraints using (27) with $\boldsymbol{a}_i := \boldsymbol{e}_i$ for $i = p+1, \ldots 2p$. These $2p$ constraints define a hyperrectangle enclosing the Berger-Boos set in the parameter space. To incorporate polytope constraints based upon the forward model, we use the ellipsoidal definition of the pre-image as shown in (24) and eigendecomposition of $\boldsymbol{K}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{K} = \boldsymbol{P} \boldsymbol{\Omega} \boldsymbol{P}^\top$ shown in (25). Note that this ellipsoid form is valid under the Gaussian noise assumption. The column vectors of $\boldsymbol{P}$ corresponding to the non-zero entries on the diagonal of $\boldsymbol{\Omega}$ are the eigenvectors corresponding to the bounded principal axes. As such, both $\boldsymbol{p}_i$ and $-\boldsymbol{p}_i$ can be used as rows of $\boldsymbol{A}$ with their corresponding bounds defined by (27). Finally, to further tighten the polytope around the Berger-Boos set, we sample a multivariate Gaussian, i.e., $\boldsymbol{a}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_p)$ to include random hyperplanes.

Once the components $(\boldsymbol{A}, \boldsymbol{b})$ have been defined using some or all of the above hyperplane generation strategies, the Vaidya walk can immediately be employed to perform a random walk around the space. The primary intuitive requirement we wish to satisfy with any sampling scheme in this context is that every region of the Berger-Boos set has a non-zero probability of being sampled. Asymptotically, the Vaidya walk samples the desired polytope uniformly at random, which satisfies a stronger requirement. In practice, the need to sample non-uniformly can arise if there are particularly meaningful parameter settings in regions that are difficult for the random walk to reach. We explore such a case in Section 6.2. Additionally, although the asymptotic distribution of the Vaidya sampler is theoretically sufficient, in practice it often has difficulty reaching the corners of the generated polytope. Although MCMC chain mixing is typically evaluated by looking at trace plots, this diagnostic is insufficient here because the dimension is high and the defined polytope can make different dimensions difficult to compare. Instead, we consider how well the sampler samples the functional space. This motivates taking a collection of starting points between the SSB endpoints and the Chebyshev center, as described in the following section.

Constructing the parallel chain starting points. Although the random walks defined in Chen et al. (2018) asymptotically sample uniformly over $\mathcal{P}^d$, given the long and thin nature of the pre-image, running the Vaidya walk from even a "good" starting position does not consistently sample the functional space well. Instead, we use the following heuristic to construct several parallel chains that constitute a complete sample when combined. We define an even number of starting points, $C \in 2\mathbb{N}$, to roughly span the Berger-Boos set, run the Vaidya walk for M_p steps from each starting point to collect parameter settings $\{\widetilde{\boldsymbol{x}}_1, \ldots, \widetilde{\boldsymbol{x}}_{M_p}\}$ for each starting point, and combine the samples from each walk, resulting in $C \times M_p$ total samples. In practice, C and M_p are chosen to yield a total of M samples as desired for Algorithm 1 or Algorithm 2. We define the collection of starting points as convex combinations of the parameter settings generating the endpoints of $I_{\mathsf{BB}}(\boldsymbol{y})$ and the Chebyshev center of $\mathcal{P}^d$ as defined in Boyd and Vandenberghe (2004), and denoted by $\boldsymbol{x}_c$. This point is the center of the largest ball contained within $\mathcal{P}^d$ and therefore acts as a center of $\mathcal{P}^d$. We denote the parameter settings generating the endpoints of $I_{\mathsf{BB}}(\boldsymbol{y})$ by $\widehat{\boldsymbol{x}}^l$ for the lower endpoint and $\widehat{\boldsymbol{x}}^u$ for the upper endpoint. We then define a grid of values $\{\tau_l\}_{l=1}^{C/2}$ such that $\tau_l \in (0, 1)$ and $\tau_l < \tau_{l+1}$ for all l. For each $k \in [C]$, define $k' := \lceil k/2 \rceil$ and set $\boldsymbol{x}_k^{start} := \tau_{k'} \widehat{\boldsymbol{x}}^l + (1 - \tau_{k'}) \boldsymbol{x}_c$ if k is odd

and $\boldsymbol{x}_k^{start} := \tau_{k'}\widehat{\boldsymbol{x}}^u + (1-\tau_{k'})\boldsymbol{x}_c$ if k is even. Creating starting positions along the lines connecting these endpoints and the Chebyshev center accommodates the chosen polytope while helping ensure that samples are chosen spanning the range of possible functional values over the Berger-Boos set.

Empirical performance for the Polytope sampler can be seen in Figure 5.2, showing a histogram of the sampled functional values in the left panel and a trace plot for the sampled Vaidya walks starting from points constructed as described above in the right panel. These plots are generated for one observation from the 80-dimensional ill-posed inverse problem in the coverage study performed in Section 6.3. Critically, the histogram shows that the Polytope sampler samples the functional space well and the trace plot shows that our starting point construction heuristic performs well in practice.

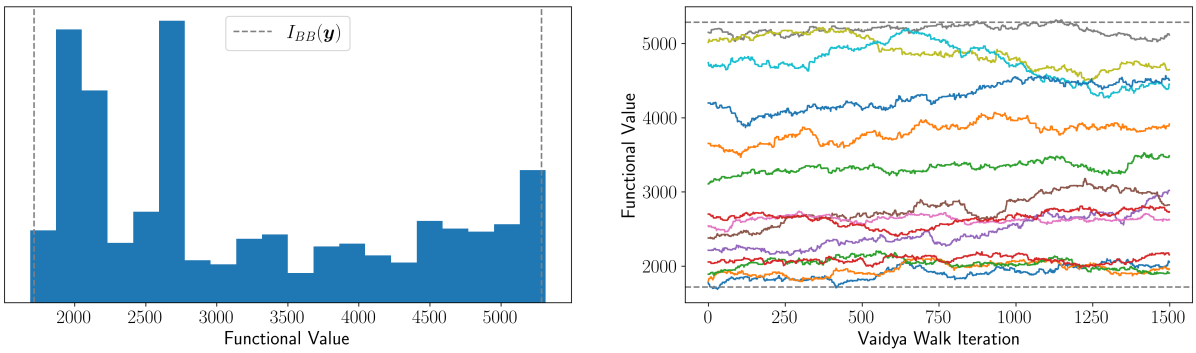

Figure 5.2: Polytope sampler output for a realization of the 80-dimensional ill-posed inverse problem studied in Section 6.3. The **left** panel contains a histogram of sampled functional values which both span and well cover the range of $I_{\mathsf{BB}}(\boldsymbol{y})$ (shown in by the dashed gray lines in both plots). The **right** panel contains trace plots of the 14 Vaidya walks (each indicated by a different color line) which together constitute the full sample. Our heuristic for choosing starting points along the lines connecting the parameter settings generating the endpoints of $I_{\mathsf{BB}}(\boldsymbol{y})$ and the Chebyshev center of the polytope provides a good initial spread of starting functional values.

Sampling from the Berger-Boos set. With $\mathcal{P}^d$ and starting positions defined, we construct a sample $\mathcal{S} := \{\widetilde{\boldsymbol{x}}_1, \ldots, \widetilde{\boldsymbol{x}}_{C\times M_p}\}$. We leave the details of the Vaidya walk to the original paper (Chen et al., 2018), but note an essential radius hyperparameter of the algorithm that must be chosen. This radius impacts the covariance of a Gaussian proposal distribution for the walk and thus affects a new proposed point's acceptance probability. Choosing a radius too large results in a low acceptance rate of proposed steps, thus creating a relatively sedentary walk. In contrast, choosing a radius too small results in a high acceptance rate with relatively small step sizes. In practice, we find a radius setting of 0.5 works well.

5.2 Quantile regression

Algorithm 2 explained in Section 3.3 involves using quantile regression to learn a quantile surface from a collection of pairs of design points and samples from LLR test statistics sampled at each design point. As previously mentioned, similar approaches have been taken in Dalmasso et al. (2020, 2021); Masserano et al. (2023, 2024), and since quantile regression is a technique facilitating our interval constructions, we wish only to give a brief overview of quantile regression and some different ways to implement it. Fundamentally, given a one-dimension random variable $z \sim P_{\boldsymbol{x}}$ that depends on the parameter $\boldsymbol{x} \in \mathbb{R}^p$, we are interested in the upper γ-quantile at every parameter

Algorithm 4 Polytope sampler

Input: $M_p \in \mathbb{N}$ and $C \in 2\mathbb{N}$. $\boldsymbol{A} \in \mathbb{R}^{d\times p}$, $\boldsymbol{b} \in \mathbb{R}^d$. $\{\tau_l\}_{l=1}^{C/2}$, where $\tau_l \in (0,1)$ and $\tau_l < \tau_{l+1}$ for all l.

1: Let $\mathcal{S} := \{\}$ be the set in which we store all sampled parameter settings.

2: **Construct Chebyshev center:** Solve for $\boldsymbol{x}_c$ using the following optimization.

$$\begin{aligned} &\underset{\boldsymbol{x}_c, r}{\text{maximize}} \quad r \\ &\text{subject to} \quad \boldsymbol{a}_i^\top \boldsymbol{x}_c + r\|\boldsymbol{a}_i\|_2 \le b_i, \quad i = 1, \ldots, d. \end{aligned}$$

3: **Compute φ extremes of Berger-Boos set:** Extreme points are computed with respect to the functional of interest:

$$\begin{aligned} \widehat{\boldsymbol{x}}^l &:= \text{argmin} \;\; \varphi(\boldsymbol{x}) \quad \text{subject to} \;\; \boldsymbol{x} \in \mathcal{B}_\eta, \\ \widehat{\boldsymbol{x}}^u &:= \text{argmax} \;\; \varphi(\boldsymbol{x}) \quad \text{subject to} \;\; \boldsymbol{x} \in \mathcal{B}_\eta. \end{aligned} \tag{28}$$

4: **for** $k = 1, 2, \ldots, C$ **do**

5: **Construct starting point:** Define $k' := \lceil k/2 \rceil$. Define $\boldsymbol{x}_k^{start} = \tau_{k'}\widehat{\boldsymbol{x}}^l + (1-\tau_{k'})\boldsymbol{x}_c$ if k is odd and $\boldsymbol{x}_k^{start} = \tau_{k'}\widehat{\boldsymbol{x}}^u + (1-\tau_{k'})\boldsymbol{x}_c$ if k is even.

6: **Run the Vaidya walk for M steps:** Collect samples $\{\widetilde{\boldsymbol{x}}_1, \ldots, \widetilde{\boldsymbol{x}}_{M_p}\}$ and add them to $\mathcal{S}$.

7: **end for**

Output: $\mathcal{S}$ containing sampled points over $\mathcal{B}_\eta$.

setting, i.e.,

$$\mathbb{P}_{\boldsymbol{x}}(z > Q_{P_{\boldsymbol{x}}}(1-\gamma)) = \gamma. \tag{29}$$

We note that the quantile surface itself is not random, so we can use draws from the distribution $P_{\boldsymbol{x}}$ to estimate $Q_{P_{\boldsymbol{x}}}$ at a given parameter setting, $\boldsymbol{x}$. However, as noted in Section 3.3, performing such an estimate is not always computationally feasible, and intuitively, we might expect quantiles to change smoothly over the parameter space, which would imply that information about a quantile at $\boldsymbol{x}_1$ should be related to a quantile at $\boldsymbol{x}_2$ if these points are close. In statistics literature, estimating $Q_{P_{\boldsymbol{x}}}$ is framed as estimating a quantile function conditional on known covariates and is often thought of as a generalization of estimating the conditional median (Koenker and Hallock, 2001). Just as conditional mean and conditional median estimation can be accomplished by using an appropriate loss function (sum of squares and absolute differences, respectively), estimating conditional quantiles can be accomplished by minimizing the pinball loss defined as follows:

$$L_\gamma(z, q) := \begin{cases} (1-\gamma)(q-z) & z < q \\ \gamma(z-q) & z \ge q. \end{cases} \tag{30}$$

(Koenker, 2005; Steinwart and Christmann, 2011). As such, estimating the quantile surface can be framed as a risk-minimization problem, leaving only standard modeling choices to fill in for q in (30). Although initial efforts were focused on linear parametric quantile regressors (Koenker and Bassett Jr, 1978), in recent years, modeling efforts have focused on nonparametric varieties. Meinshausen (2006) adapted random forests to quantile regression. Takeuchi et al. (2006) leveraged Reproducing Kernel Hilbert Spaces to construct smooth quantile regressors. Closer to our applica-

tion, Masserano et al. (2023) used neural networks to optimize the pinball loss to learn the quantile surface for their application.

6 Numerical experiments

For scenarios within the data-generating process (11), the OSB interval was the state-of-the-art option for computing constraint-aware confidence intervals. Although these intervals have empirically achieved nominal coverage in applications (Patil et al., 2022; Stanley et al., 2022), they generally do not Batlle et al. (2023). As such, we use the OSB interval in the following numerical experiments as a main comparison point to the intervals defined in this paper. In scenarios where the OSB interval achieves at least nominal coverage, we show that our intervals are either competitive or better in terms of estimated expected interval length. In scenarios where the OSB interval does not achieve nominal coverage, our intervals do achieve nominal coverage and in one instance, have a shorter estimated expected length. We provide four numerical experiments to make these points. The first set of two uses a constrained Gaussian noise model setup in two and three dimensions, respectively. These two experiments illustrate the aforementioned points in addition to constituent parts of the interval computation process due to the relatively low dimensions. The second set of two considers a wide-bin deconvolution setup inspired by particle unfolding in high-energy physics (Stanley et al., 2022). This setup features an 80-dimensional parameter space with a rank-deficient forward model and is thus a substantially more complicated computational scenario compared to the two and three-dimensional Gaussian noise models. These examples demonstrate the superior performance of our intervals over OSB in terms of both coverage and expected length.

In the following experiments, we estimate 68% confidence intervals, set $\eta = 0.01$ and compute γ according to Lemma 3.1. We draw 10^3 observations from each data-generating process to estimate both coverage and expected interval length for our four interval constructions and the OSB interval. We additionally provide 95% confidence intervals in the form of orange line segments to characterize statistical error for both coverage and expected length estimates. The coverage confidence intervals are Clopper-Pearson intervals for the probability of success parameter of a binomial distribution, while the expected length confidence intervals are the average length plus/minus the appropriately scaled standard error.

6.1 Constrained Gaussian in two dimensions

The two-dimensional Gaussian noise model is defined as follows:

$$\boldsymbol{y} = \boldsymbol{x}^* + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_2), \quad \boldsymbol{x}^* \in \mathbb{R}^2_+, \tag{31}$$

where $\varphi(\boldsymbol{x}) = x_1 - x_2$ and $\boldsymbol{x}^* = \begin{pmatrix} 0.5 & 0.5 \end{pmatrix}^\top$. The LLR is then defined as follows:

$$\lambda(\mu, \boldsymbol{y}; \mathbb{R}^2_+) = \min_{\substack{x_1 - x_2 = \mu \\ \boldsymbol{x} \in \mathbb{R}^2_+}} \|\boldsymbol{y} - \boldsymbol{x}\|_2^2 - \min_{\boldsymbol{x} \in \mathbb{R}^2_+} \|\boldsymbol{y} - \boldsymbol{x}\|_2^2. \tag{32}$$

This example first appeared in Tenorio et al. (2007) as an example where the OSB interval failed to achieve nominal coverage when the true parameter $\boldsymbol{x}^*$ is such that $\varphi(\boldsymbol{x}^*) = 0$. However, Batlle et al. (2023) overturned this result by proving OSB validity in this case. We note that the OSB interval does not generally provide coverage for two-dimensional cases when $\varphi(\boldsymbol{x}^*) \neq 0$, but we leave this analysis to future work. As such, this example is important to include because of its methodological historical context and OSB interval validity. The proof that OSB intervals cover in

this particular example relies upon showing that $Q_{1-\alpha}^{\max} = \chi^2_{1,\alpha}$ for all $\alpha \in (0,1)$, where $\chi^2_{1,\alpha}$ is the upper α-quantile of a chi-squared distribution with one degree of freedom. Alternatively stated, it must be the case that $\lambda(\varphi(\boldsymbol{x}^*), \boldsymbol{y}; \mathbb{R}^2_+)$ is stochastically dominated by χ^2_1. This result is shown in Lemma 4.4 of Batlle et al. (2023).

Estimated coverage and length results are shown in Figure 6.1. We note that all four of our interval constructions are competitive with OSB in terms of coverage, while all but our Sliced Inverted interval have higher estimated expected length. Since the OSB interval is defined using $Q_{1-\alpha}^{\max} = \chi^2_{1,\alpha}$ and the α-quantile surface of the LLR rapidly approaches this global max-quantile as one moves away from the origin (see Figure 5.3 in Batlle et al. (2023)), the OSB interval lengths are difficult to beat in practice with intervals based on the Berger-Boos sets since these sets likely contain parameter settings with quantiles near $\chi^2_{1,\alpha}$. The left panel of Figure 6.2 shows four realizations of the data-generating process with the observations shown as red points. For each observation, the blue points show uniformly distributed draws within its Berger-Boos set, sampled using the VGS sampler. Cross-referencing the spread of the Berger-Boos set samples in Figure 6.2 with Figure 5.3 in Batlle et al. (2023), it is clear that there are always samples in the parameter space where the quantile surface is nearly the same as the $\chi^2_{1,\alpha}$ quantile. Further, when including the Berger-Boos set in the interval construction, we instead construct our intervals using the γ-quantile, where $\gamma < \alpha$, as the LLR cutoff, resulting in a more relaxed constraint. This fact can be clearly observed in the central panel of Figure 6.2, showing the sampled γ-quantiles within the Berger-Boos set of one observation from the data generating process. Since a non-trivial portion of this distribution is above $\chi^2_{1,\alpha}$, the longer average length of the Global intervals is explained. In the right panel of Figure 6.2, for the same observation, we show the estimated sliced max-quantile function, $\widehat{m}_\gamma(\mu)$, in orange alongside $\chi^2_{1,\alpha}$. Since this estimated function is above $\chi^2_{1,\alpha}$ at their intersection points with the underlying LLR function shown in the solid blue line, it further makes sense that the Sliced interval constructions provide no additional length improvement compared to the OSB interval.

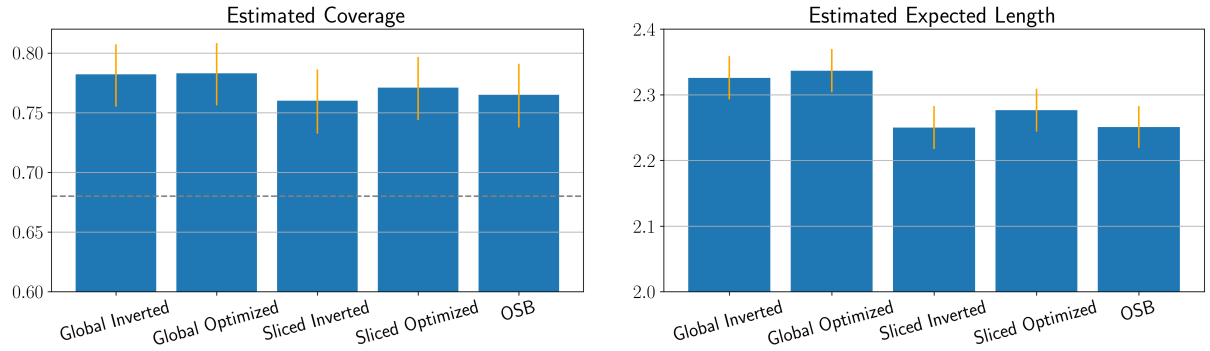

Figure 6.1: Estimated coverages and expected lengths across all four interval constructions and OSB for comparison at the 68% level for the two-dimensional constrained Gaussian setting. All four of our interval constructions are competitive with OSB with respect to coverage, but OSB shows better expected length performance, aside from our two sliced interval constructions. Although the OSB intervals are defined using the global max-quantile ($Q_{1-\alpha}^{\max}$) and therefore can potentially be improved upon by limiting the considered parameter space via the Berger-Boos set, due to the rapidity with which the α-quantile surface meets the $\chi^2_{1,\alpha}$ quantile (see Figure 5.3 in Batlle et al. (2023)), the OSB interval lengths are difficult to beat in practice.

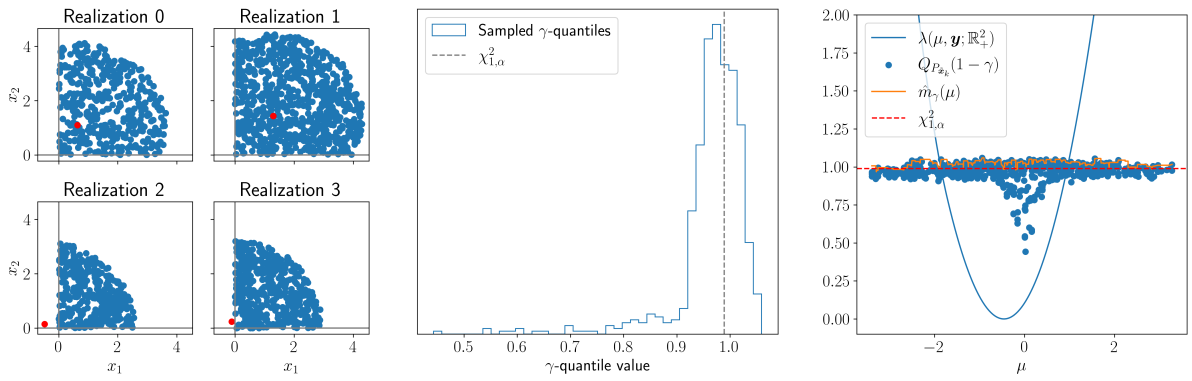

Figure 6.2: (**Left**) Four realizations of the data-generating process where the observations are shown in red. For each realization, the blue points are uniformly distributed samples from its Berger-Boos set, sampled using the VGS sampler. (**Center**) For a realization of the data-generating process, we plot the distribution of γ-quantiles for the points sampled by the VGS sampler. Notably, a non-trivial percent of these are above $\chi^2_{1,\alpha}$ defining the OSB interval. (**Right**) For the same realization, we plot the estimated sliced max-quantile function, $\widehat{m}_\gamma(\mu)$ in orange alongside $\chi^2_{1,\alpha}$ in red. The blue points correspond to samples in the center histogram, while the solid blue line shows the LLR over the functional space. All intervals can be read immediately from this image.

6.2 Constrained Gaussian in three dimensions

As seen in the previous example, in a case where the OSB interval is known to achieve nominal coverage, its expected length can be difficult to beat. However, an OSB interval coverage guarantee is difficult to prove, since it amounts to proving stochastic dominance on the non-trivial LLR object. As such, in situations where the OSB interval does not provide nominal coverage, our intervals immediately provide a clear theoretical advantage. One such three-dimensional constrained Gaussian case was explored in Batlle et al. (2023), to which we apply our four interval constructions. The three-dimensional Gaussian noise model is defined as follows:

$$\boldsymbol{y} = \boldsymbol{x}^* + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_3), \quad \boldsymbol{x}^* \in \mathbb{R}^3_+, \tag{33}$$

where $\varphi(\boldsymbol{x}) = x_1 + x_2 - x_3$ and $\boldsymbol{x}^* = \begin{pmatrix} 0.03 & 0.03 & 1 \end{pmatrix}^\top$. The LLR is then defined as follows:

$$\lambda(\mu, \boldsymbol{y}; \mathbb{R}^3_+) = \min_{\substack{x_1+x_2-x_3=\mu \\ \boldsymbol{x}\in\mathbb{R}^3_+}} \|\boldsymbol{y} - \boldsymbol{x}\|_2^2 - \min_{\boldsymbol{x}\in\mathbb{R}^3_+} \|\boldsymbol{y} - \boldsymbol{x}\|_2^2. \tag{34}$$

Figure 6.3 shows estimated coverage and expected length across all four interval constructions and the OSB interval. While the OSB interval fails to attain nominal coverage in this example, all four of our interval constructions do, with the Sliced constructions only slightly over-covering. While the Global constructions pay a fairly steep price for coverage in expected interval length, the Sliced constructions navigate well the tradeoff, paying for coverage with only slightly longer intervals compared to the OSB interval.

The setting of $\boldsymbol{x}^*$ used here is slightly different from that of Batlle et al. (2023), where they used $\boldsymbol{x}^* = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}^\top$. In Batlle et al. (2023), this setting was used as a counter-example for OSB coverage, since the $Q_{P_{\boldsymbol{x}^*}}(1-\alpha) > \chi^2_{1,\alpha}$ at least some $\alpha \in (0,1)$. However, as shown in Figure 5.5 of Batlle et al. (2023), when $\alpha = 0.05$, $Q_{P_{\boldsymbol{x}^*}}(1-\alpha) \leq \chi^2_{1,\alpha}$ for $\boldsymbol{x}^* = \begin{pmatrix} t & t & 1 \end{pmatrix}^\top$ when t is approximately greater than e^{-2}, which indicates that parameter settings violating stochastic dominance by χ^2_1 exist

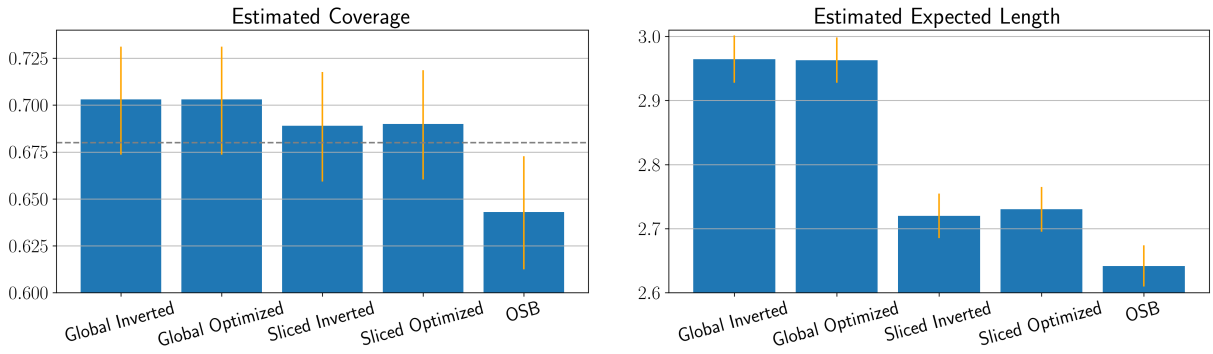

Figure 6.3: Estimated coverage and expected length across all four interval constructions and OSB for comparison at the 68% level for the three-dimensional constrained Gaussian example. All four of our interval constructions achieve nominal coverage while the OSB interval does not. While the Global interval constructions pay a steep price in expected length compared to OSB, the Sliced constructions are only slightly longer than OSB.

close to the parameter constraint boundary. The location of these key parameter settings presents a challenge for the Polytope sampler described by Algorithm 4. In Appendix C.1, Algorithm 5 presents a modified version of Algorithm 4 that better handles sampling in this example.

Berger-Boos set experiment. For this model, we investigate the effect of changing the parameter η that controls the Berger-Boos construction. We compute intervals for different y and fixed $x^* = (5, 5, 0)$ as η ranges between 0 and $\alpha = 0.32$. The length of such intervals is shown in Figure 6.4. We observe that the optimal value of η is around 0.01. Following these numerical results in the low-dimensional example, we choose $\eta = 0.01$ for $\alpha = 0.32$ (so that $\gamma = 0.313$) as a reasonable (but not necessarily optimal) default value throughout the examples, achieving competitive performance with this choice. Notably, this keeps the γ-quantile close to the original α-quantile of interest.

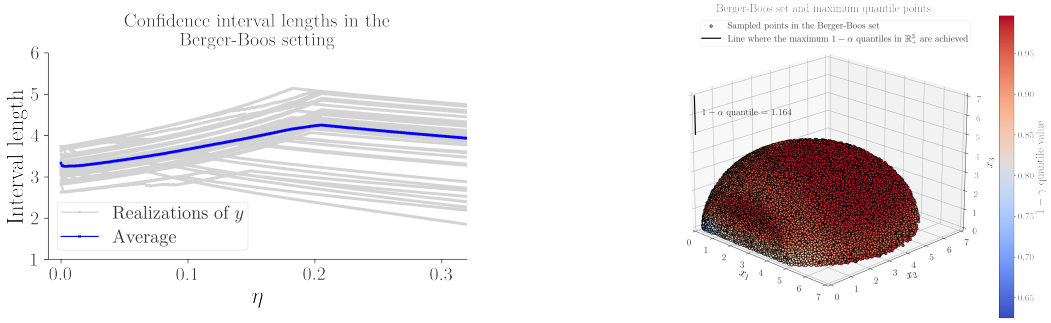

Figure 6.4: Confidence interval lengths in the Berger-Boos setting for different values of y and varying η (left) and representation of the Berger-Boos set in relation to the maximum quantile point (right). The minimum average length occurs at $\eta > 0$, showing that the construction is beneficial. This occurs because even for moderately small η, the Berger-Boos set, which is a three-dimensional sphere intersected with the non-negative orthant, avoids the line with the $1 - \alpha$ highest quantile, and the $1 - \gamma$ quantile in the set is smaller than the $1 - \alpha$ quantile in that line.

6.3 Wide-bin deconvolution

While the numerical experiments in Section 6.1 and Section 6.2 show that our interval constructions are competitive with the OSB interval in scenarios where it is known to provide coverage and superior to the OSB interval by achieving nominal coverage when OSB does not, this section shows the superior performance of our interval constructions relative to OSB in a more complex setting. We consider the problem of computing a confidence interval for the sum of adjacent bins of a deconvolved histogram as described by Stanley et al. (2022). This problem is a core statistical problem of particle unfolding in high-energy physics. For more detailed information, we refer the reader to Kuusela and Panaretos (2015); Kuusela (2016); Kuusela and Stark (2017); CMS Collaboration (2016, 2019). The data-generating process is linear with Gaussian noise,

$$\boldsymbol{y} = \boldsymbol{K}\boldsymbol{x}^* + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}), \quad \boldsymbol{x}^* \geq \boldsymbol{0}, \tag{35}$$

where $\boldsymbol{K} \in \mathbb{R}^{40\times 80}$ and $\varphi(\boldsymbol{x}) = \boldsymbol{h}^\top \boldsymbol{x}$. The vector $\boldsymbol{h}$ defines the bin-adjacent aggregation. The LLR is then defined as follows:

$$\lambda(\mu, \boldsymbol{y}; \mathbb{R}^{80}_+) = \min_{\substack{\boldsymbol{h}^\top \boldsymbol{x} = \mu \\ \boldsymbol{x} \in \mathbb{R}^{80}_+}} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2 - \min_{\boldsymbol{x} \in \mathbb{R}^{80}_+} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2. \tag{36}$$

We emphasize two features of this setup that complicate the task of computing confidence intervals for $\varphi(\boldsymbol{x})$. First, the forward model, $\boldsymbol{K}$, has a non-trivial null space, making any inverse problem point estimation and UQ markedly challenging. Typically, this sort of ill-posedness is handled with regularization of some kind, but as is well known in the inverse problem literature and specifically shown in Kuusela (2016), including such regularization induces a bias, which can undercut desired statistical guarantees (e.g., coverage) of the inference object of interest. Including constraints and focusing on a particular functional of the parameter vector *implicitly* regularizes the problem, but shifts the problem difficulty to inference with constraints. Although Batlle et al. (2023) and this paper proposes a theoretical framework to do inference with constraints, the second challenge is in the implementation due to the high-dimensional parameter space of scenarios like this example. As we show in the following sections, the Polytope sampler described by Algorithm 4 and quantile regression does an adequate job producing samples and fitting quantile surfaces in this high-dimensional space to ensure the desired coverage guarantee of the interval constructions.

As extensively discussed in Stanley et al. (2022), the OSB interval (i.e., using $\chi^2_{1,0.05}$ in the optimization-based interval construction) produces empirically valid confidence intervals in all tested scenarios, albeit typically with over-coverage. Since the scenarios considered in Stanley et al. (2022) were designed to be realistic particle unfolding simulations, the underlying intensity function generating the true histogram means ($\boldsymbol{x}^* \in \mathbb{R}^{80}_+$) was relatively smooth, likely contributing to the over-coverage for the internal bin aggregations. As such, we present two true parameter settings for $\boldsymbol{x}^*$ in (35) to highlight two advantages of our interval constructions over the OSB interval. First, we use the original smooth "realistic" parameter setting from Stanley et al. (2022) to show how our intervals improve over-coverage relative to the OSB interval by reducing the expected interval length. Second, we present an "adversarial" setting where our interval constructions achieve nominal coverage while the OSB interval does not. Figure 6.5 shows the realistic and adversarial settings for $\boldsymbol{x}^*$. We constructed the adversarial setting by first computing our interval constructions on the realistic setting and then looking at the maximum out-of-sample predicted quantile for a particular generated observation, $\boldsymbol{y}$. For each observation drawn within both settings, we draw 2.1×10^4 samples using the Polytope sampler as described by Algorithm 4.

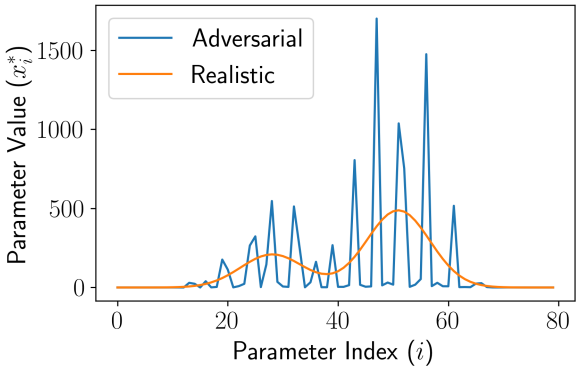

Figure 6.5: The figure shows the parameter values for the realistic and adversarial settings for $\boldsymbol{x}^*$ used to illustrate our interval construction performances versus the OSB interval. The realistic setting for which the OSB interval over-covers is relatively smooth compared to the adversarial setting.

6.3.1 Realistic setting

Using the realistic $\boldsymbol{x}^*$ shown in Figure 6.5, Stanley et al. (2022) showed that the OSB interval over-covers at the 95% level. Furthermore, it was shown that the OSB interval was the shortest across a range of other interval options, including SSB, prior optimized, and minimax. As such, for this setting, we show that our interval constructions not only achieve nominal coverage, but the sliced constructions dramatically reduce over-coverage compared to OSB by producing substantially shorter intervals on average. Estimated coverage and expected interval lengths are shown in Figure 6.6.

Both Global interval constructions and OSB dramatically over-cover which highlights the conservatism of the Global constructions. These estimated coverage values indicate that within each observation's Berger-Boos set, there is a parameter setting against which the method protects that is substantially more difficult to cover than the true realistic parameter setting. Interestingly, both Global constructions produce markedly longer intervals on average compared with the OSB interval. Aligning with the intuition from the Global and Sliced construction definitions, the Sliced intervals are less conservative as seen by their lower over-coverage and dramatically lower average lengths. Interestingly, the Sliced Inverted intervals have higher coverage and lower average length compared to the Sliced Optimized. Importantly, both Sliced constructions are shorter on average compared to the OSB interval, with the Sliced Inverted showing an 18.8% reduction in average length and the Sliced Optimized showing a 17.0% reduction in average length.

6.3.2 Adversarial setting

The critical result in Section 6.3.1 is that the Sliced interval constructions both reduce over-coverage and expected interval length compared with the OSB interval. In this section, we show that for a true parameter setting where the OSB interval does not achieve nominal coverage, all four of our interval constructions do achieve nominal coverage while still reducing the expected interval length in the case of the Sliced interval constructions compared to the OSB interval. The estimated coverage and expected length results are shown in Figure 6.7.

Both Global interval constructions and the Sliced Optimized interval over-cover, with the Sliced Optimized over-covering to a lesser extent than the Global intervals. The Sliced Inverted interval achieves nominal coverage within statistical error. The estimated expected lengths tell a story similar to that of the realistic example, with the Global intervals showing the longest average interval lengths, the Sliced showing the shortest, and the OSB interval being right in the middle.

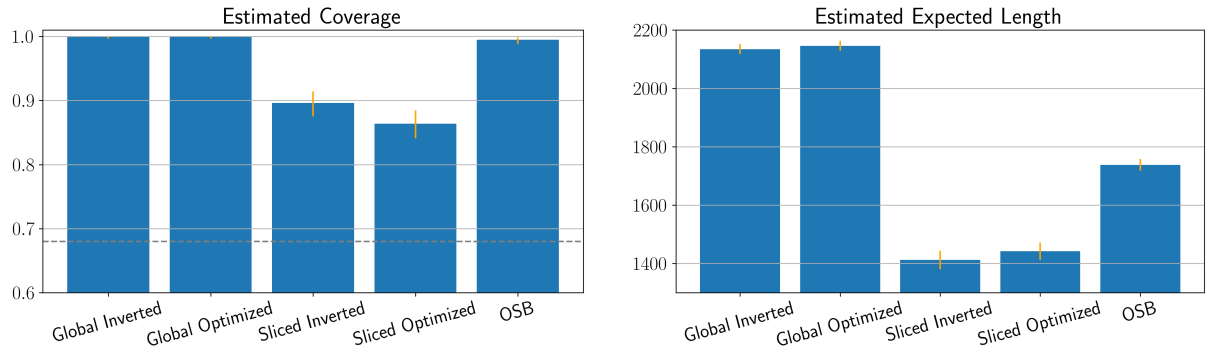

Figure 6.6: Estimated coverage and expected length across all four interval constructions and OSB for comparison at the 68% level for the **realistic** wide-bin deconvolution experiment. While the Global interval constructions over-cover like the OSB interval, the Sliced interval constructions both reduce over-coverage and expected interval length.

Importantly, the Sliced intervals are again dramatically shorter than the OSB interval, even though the OSB interval does not achieve nominal coverage. The Sliced Inverted interval shows a 19.5% average interval length reduction over OSB while the Sliced Optimized shows an 11.7% average interval length reduction.

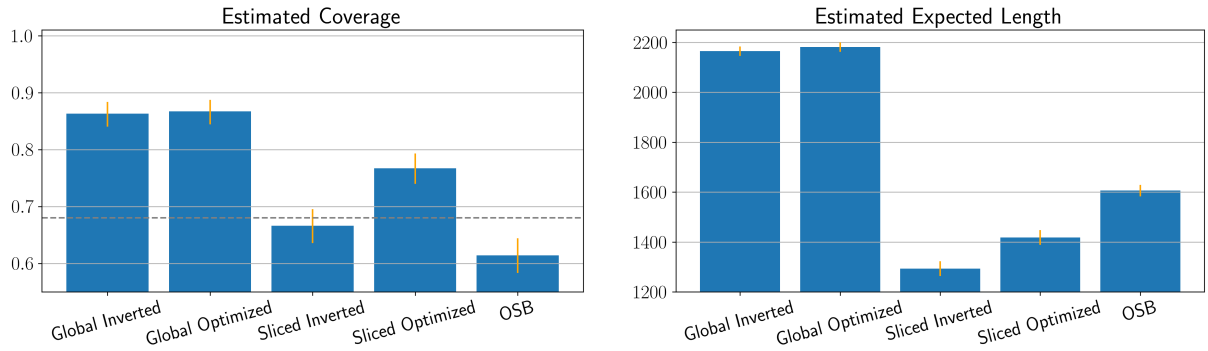

Figure 6.7: Estimated coverage and expected length across all four interval constructions and OSB for comparison at the 68% level for the **adversarial** wide-bin deconvolution experiment. While the OSB interval fails to achieve nominal coverage, all four of our interval constructions do. Interestingly, the Sliced interval constructions are meaningfully shorter than the OSB interval on average.

7 Conclusion

This paper proposes several confidence interval constructions for functionals in constrained ill-posed inverse problems. Our approach is based on two key ideas: data-adaptive constraints using a Berger-Boos construction and sampling-based inversion. Two independent decisions when constructing intervals provide four different valid intervals: Global/Sliced, depending on the optimization domain of the quantile function and Inverted/Optimized, depending on the construction after the sampling process. All of the constructions are built upon the preliminary constraint by the data-informed Berger-Boos set, followed by a sampling procedure to estimate a quantile function that can be used to invert or optimize the intervals. We have validated the method (including all four aforementioned interval constructions) through several numerical examples, demonstrating its ability to provide

correct coverage and better calibration compared to the OSB interval baseline approach. Overall, our approach offers a flexible framework that can incorporate constraints directly and can be tailored to various types of inverse problems. The main takeaway is that data-adaptive constraining helps improve the length of the resulting confidence intervals, while sampling makes it feasible to carry out the test inversion needed to construct confidence intervals with the desired nominal coverage.

There are several promising directions for future work. One direction is to extend our method to even higher dimensional problems, which are more challenging. This would involve developing techniques to handle the curse of dimensionality and exploring the trade-off between accuracy and computational complexity. For the approaches in this paper in particular, this extension would require a more tailored sampling approach. Another direction is to leverage more sophisticated machine learning algorithms (deep learning models or ensemble methods) to improve the estimate of the quantile function and thus improve the accuracy and efficiency of our confidence intervals. Additionally, applying our approach to other applications of ill-posed inverse problems, such as medical imaging or geophysics, would provide further validation of the effectiveness of our approach. Finally, it is of interest to conduct a further theoretical analysis of our approach under different constraints and noise conditions to better understand its limitations and strengths. This would involve studying the statistical properties of our confidence intervals and investigating the impact of various assumptions on their performance. Overall, these future research directions have the potential to enhance the applicability and robustness of our approach in a wide range of domains.

Acknowledgements

We thank members of the STAMPS (Statistical Methods in the Physical Sciences) working group at Carnegie Mellon University for fruitful discussions on this work.

References

Ashton, G. et al. (2022). Nested sampling for physical scientists. *Nature Reviews Methods Primers*, 2.

Batlle, P., Stanley, M., Patil, P., Kuusela, M., and Owhadi, H. (2023). Optimization-based frequentist confidence intervals for functionals in constrained inverse problems: Resolving the burrus conjecture. *arXiv preprint arXiv:2310.02461*.

Berger, R. L. and Boos, D. D. (1994). P values maximized over a confidence set for the nuisance parameter. *Journal of the American Statistical Association*, 89:1012–1016.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

Buchner, J. (2023). Nested sampling methods. *arXiv preprint arXiv:2101.09675v4*.

Burrus, W., Commission, U. A. E., and Laboratory, O. R. N. (1965). *Utilization of a Priori Information by Means of Mathematical Programming in the Statistical Interpretation of Measured Distributions*. Oak Ridge National Laboratory.

Carpenter, J. (1999). Test inversion bootstrap confidence intervals. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(1):159–172.

Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Duxbery, 2 edition.

Chen, Y., Dwivedi, R., Wainwright, M. J., and Yu, B. (2018). Fast mcmc sampling algorithms on polytopes. *Journal of Machine Learning Research*, 19:1–86.

CMS Collaboration (2016). Measurement of differential cross sections for Higgs boson production in the diphoton decay channel in pp collisions at $\sqrt{s} = 8$ TeV. *The European Physical Journal C*, 76(13).

CMS Collaboration (2019). Measurement of inclusive and differential Higgs boson production cross sections in the diphoton decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Journal of High Energy Physics*, 2019(183).

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

Dalmasso, N., Masserano, L., Zhao, D., Izbicki, R., and Lee, A. B. (2021). Likelihood-free frequentist inference: Bridging classical statistics and machine learning for reliable simulator-based inference. *arXiv preprint arXiv:2107.03920*.

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.

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.

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.

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.

Höcker, A. and Kartvelishvili, V. (1996). Svd approach to data unfolding. *Nuclear Instruments and Methods in Physics Research A*, 372:469–481.

Kannan, R. and Narayanan, H. (2012). Random walks on polytopes and an affine interior point method for linear programming. *Mathematics of Operations Research*, 37:1 – 20.

Kibzun, A. I. and Kan, Y. S. (1997). Stochastic programming problems with probability and quantile functions. *Journal of the Operational Research Society*, 48(8):849–849.

Koenker, R. (2005). *Quantile regression*, volume 38. Cambridge university press.

Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50.

Koenker, R. and Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, 15(4):143–156.

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

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.

Kuusela, M. and Stark, P. B. (2017). Shape-constrained uncertainty quantification in unfolding steeply falling elementary particle spectra. *The Annals of Applied Statistics*, 11(3):1671–1710.

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.

Lovasz, L. (1999). Hit-and-run mixes fast. *Mathematical Programming*, 86:443 – 461.

Lovasz, L. and Vempala, S. (2006). Hit-and-run from a corner. *SIAM Journal on Computing*, 35:985 – 1005.

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. *AISTATS*.

Masserano, L., Shen, A., Doro, M., Dorigo, T., Izbiki, R., and Lee, A. (2024). Classification under nuisance parameters and generalized label shift in likelihood-free inference. *arXiv preprint arXiv:2402.05330*.

Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research*, 7:983–999.

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

Montel, N. A., Alvey, J., and Weniger, C. (2023). Scalable inference with autoregressive neural ratio estimation. *arXiv preprint arXiv:2308:08597v1*.

Narayanan, H. (2016). Ranomized interior point methods for sampling and optimization. *The Annals of Applied Probability*, 26:597–641.

Ok, E. A. (2007). *Real Analysis with Economic Applications*. Econometric Society Monographs. Princeton University Press, Princeton, NJ, 1st edition. Includes bibliographical references and index.

Pakman, A. and Paninski, L. (2014). Exact hamiltonian monte carlo for truncated multivariate gaussians. *Journal of Computational and Graphical Statistics*, 23:518 – 542.

Panaretos, V. M. (2016). *Statistics for Mathematicians*. Springer International Publishing.

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

Pena-Ordieres, A., Luedtke, J. R., and Wachter, A. (2020). Solving chance-constrained problems via a smooth sample-based nonlinear approximation. *SIAM Journal on Optimization*, 30(3):2221–2250.

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

Rodgers, C. D. (2000). *Inverse Methods for Atmospheric Sounding*. World Scientific Publishing.

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

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.

Schmitt, S. (2012). Tunfold, an algorithm for correcting migration effects in high energy physics. *Journal of Instrumentation*, 7:T10003.

Shapiro, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *International Statistical Review/Revue Internationale de Statistique*, pages 49–62.

Skilling, J. (2004). Nested sampling. *AIP Conference Proceedings*, 735:295.

Smith, R. L. (1984). Efficient monte carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research*, 32:1296–1308.

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.

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

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

Steinwart, I. and Christmann, A. (2011). Estimating conditional quantiles with the help of the pinball loss. *Bernoulli*, 117(1):211–225.

Takeuchi, I., Le, Q. V., Sears, T. D., and Smola, A. J. (2006). Nonparametric quantile estimation. *Journal of Machine Learning*, 7:1231–1264.

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.

Voelker, A. R., Gosmann, J., and Stewart, T. C. (2017). Simultaneous confidence intervals for linear estimates of linear functionals. Technical report, Centre for Theoretical Neuroscience.

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.

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

Appendix

This document serves as an appendix to the paper " Confidence intervals for functionals in constrained inverse problems via data-adaptive sampling-based calibration." Below we provide an organization for the appendix, followed by a summary of the main notation used in both the paper and the appendix. The equation and figure numbers in this appendix begin with the letter "S" to differentiate them from those appearing in the main paper.

Organization

The content of this Appendix is organized as follows.

Appendix	Description
Appendix A	Proofs in Section 3 (Proofs of Lemmas 3.1, 3.3, 4.1 and 4.2 and Corollary 3.2)
Appendix C	Additional details and illustrations in Section 6 (importance-like sampler)

Table 2: Roadmap of the supplement.

Notation

An overview of the main notation used in this paper is as follows.

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.
$\mathbb{R}_+$	Set of non-negative real numbers.
$[n]$	Set $\{1, \ldots, n\}$ for a positive integer n.
$\|\boldsymbol{u}\|_2$	The ℓ_2 norm of vector $\boldsymbol{u}$.
$\|f\|_{L_2}$	The L_2 norm of function f.
$\boldsymbol{K}^\top$	The transpose of a matrix $\boldsymbol{K} \in \mathbb{R}^{m \times p}$.
$\boldsymbol{I}_m$ or $\boldsymbol{I}$	The $m \times m$ identity matrix.
$\boldsymbol{v} \preceq \boldsymbol{u}$	Lexicographic ordering for vectors $\boldsymbol{v}$ and $\boldsymbol{u}$.
$\boldsymbol{A} \preceq \boldsymbol{B}$	The Loewner ordering for symmetric matrices $\boldsymbol{A}$ and $\boldsymbol{B}$.
$\mathbf{1}\{A\}$	Indicator random variable associated with event A.
$Y = \mathcal{O}_\alpha(X)$	Deterministic big-O notation, indicating that Y is bounded by $\|Y\| \leq C_\alpha X$.
C_α	A numerical constant that may depend on the parameter α in context.
$\mathcal{O}_p$	Probabilistic big-O notation.
$\xrightarrow{\mathrm{p}}$	Convergence in probability.
$X \succeq Y$	Stochastic dominance of X by Y, indicating that $\mathbb{P}(X \geq z) \geq \mathbb{P}(Y \geq z)$ for all $z \in \mathbb{R}$.

Table 3: Summary of main notation used in the paper and the appendix.

A Proofs in Section 3

A.1 Proof of Lemma 3.1

Let $\alpha \in (0,1)$ and $\eta \in (0,\alpha)$. Fix the true parameter value, $\boldsymbol{x}^* \in \mathcal{X}$, and let $\mu^* = \varphi(\boldsymbol{x}^*)$. We use the notation $\mathbb{P}_{\boldsymbol{x}^*}(A)$ to denote the probability of event A when the true parameter is $\boldsymbol{x}^*$. We first show that $\mathbb{P}_{\boldsymbol{x}^*}\big(\mu^* \in C_\alpha^{\mathrm{sl}}(\boldsymbol{y};\mathcal{B}_\eta) \mid \boldsymbol{x}^* \in \mathcal{B}_\eta\big) \geq 1-\gamma$, from which the result follows by the law of total probability.

Following from the definitions of the constituent objects, we have:

$$\mathbb{P}_{\boldsymbol{x}^*}\big(\mu^* \in C_\alpha^{\mathrm{sl}}(\boldsymbol{y};\mathcal{B}_\eta) \mid \boldsymbol{x}^* \in \mathcal{B}_\eta\big) = \mathbb{P}_{\boldsymbol{x}^*}\big(\lambda(\mu^*,\boldsymbol{y};\mathcal{B}_\eta) \leq \bar{q}_{\gamma,\eta}^{\mu^*} \mid \boldsymbol{x}^* \in \mathcal{B}_\eta\big) \tag{S.1}$$

$$\geq \mathbb{P}_{\boldsymbol{x}^*}\big(\lambda(\mu^*,\boldsymbol{y};\mathcal{B}_\eta) \leq Q_{P_{\boldsymbol{x}^*}}(1-\gamma) \mid \boldsymbol{x}^* \in \mathcal{B}_\eta\big) = 1-\gamma. \tag{S.2}$$

Since the above steps hold for any chosen $\boldsymbol{x}^* \in \mathcal{X}$, we have $\mathbb{P}_{\boldsymbol{x}^*}\big(\mu^* \in C_\alpha^{\mathrm{sl}}(\boldsymbol{y};\mathcal{B}_\eta) \mid \boldsymbol{x}^* \in \mathcal{B}_\eta\big) \geq 1-\gamma$, for all $\boldsymbol{x}^* \in \mathcal{X}$.

By definition, we have $\mathbb{P}_{\boldsymbol{x}^*}(\boldsymbol{x}^* \in \mathcal{B}_\eta) \geq 1-\eta$, for all $\boldsymbol{x}^* \in \mathcal{X}$. The law of total probability implies the following decomposition:

$$\mathbb{P}_{\boldsymbol{x}^*}\big(\mu^* \in C_\alpha^{\mathrm{sl}}(\boldsymbol{y};\mathcal{B}_\eta)\big) = \mathbb{P}_{\boldsymbol{x}^*}\big(\mu^* \in C_\alpha^{\mathrm{sl}}(\boldsymbol{y};\mathcal{B}_\eta), \boldsymbol{x}^* \in \mathcal{B}_\eta\big) + \mathbb{P}_{\boldsymbol{x}^*}\big(\mu^* \in C_\alpha^{\mathrm{sl}}(\boldsymbol{y};\mathcal{B}_\eta), \boldsymbol{x}^* \notin \mathcal{B}_\eta\big) \tag{S.3}$$

$$\geq \mathbb{P}_{\boldsymbol{x}^*}\big(\mu^* \in C_\alpha^{\mathrm{sl}}(\boldsymbol{y};\mathcal{B}_\eta) \mid \boldsymbol{x}^* \in \mathcal{B}_\eta\big)(1-\eta) + 0 \cdot \eta \tag{S.4}$$

$$\geq (1-\gamma)(1-\eta), \tag{S.5}$$

for all $\boldsymbol{x}^* \in \mathcal{X}$. Hence, $\mathbb{P}_{\boldsymbol{x}^*}\big(\mu^* \in C_\alpha^{\mathrm{sl}}(\boldsymbol{y};\mathcal{B}_\eta)\big) \geq 1-\alpha$ for all $\boldsymbol{x}^* \in \mathcal{X}$ if $(1-\gamma)(1-\eta) = 1-\alpha$.

A.2 Proof of Corollary 3.2

By definition, $\bar{q}_{\gamma,\eta}^{\mu} \leq \bar{q}_{\gamma,\eta}$ for all $\mu \in \mathbb{R}$. Therefore, $C_\alpha^{\mathrm{sl}}(\boldsymbol{y};\mathcal{B}_\eta) \subseteq C_\alpha^{\mathrm{gl}}(\boldsymbol{y};\mathcal{B}_\eta)$, and thus

$$\mathbb{P}_{\boldsymbol{x}^*}\big(\mu^* \in C_\alpha^{\mathrm{gl}}(\boldsymbol{y};\mathcal{B}_\eta)\big) \geq \mathbb{P}_{\boldsymbol{x}^*}\big(\mu^* \in C_\alpha^{\mathrm{sl}}(\boldsymbol{y};\mathcal{B}_\eta)\big) \geq 1-\alpha. \tag{S.6}$$

A.3 Proof of Lemma 3.3

We show the equivalence by showing how each constituent optimization of the LLR is unchanged by including the Berger-Boos set into its constraints. First, the assumption $\boldsymbol{x}^* \in \mathcal{B}_\eta$ implies $\mathcal{B}_\eta \neq \emptyset$. It also follows by definition that $\mathcal{B}_\eta \subseteq \mathcal{X}$ and further $\mathcal{B}_\eta \cap \Phi_\mu \subseteq \mathcal{X} \cap \Phi_\mu$, for all $\mu \in \mathbb{R}$. As such, it follows that

$$\min_{\boldsymbol{x} \in \mathcal{B}_\eta \cap \Phi_\mu} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2 \geq \min_{\boldsymbol{x} \in \mathcal{X} \cap \Phi_\mu} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2, \tag{S.7}$$

$$\min_{\boldsymbol{x} \in \mathcal{B}_\eta} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2 \geq \min_{\boldsymbol{x} \in \mathcal{X}} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2. \tag{S.8}$$

Suppose $\widehat{\boldsymbol{x}}$ minimizes $\|\boldsymbol{y}-\boldsymbol{K}\boldsymbol{x}\|_2^2$ over $\mathcal{X} \cap \Phi_\mu$ and that $\|\boldsymbol{y}-\boldsymbol{K}\widehat{\boldsymbol{x}}\|_2^2 > \chi^2_{m,\eta}$. This assumption violates (S.7), therefore $\|\boldsymbol{y}-\boldsymbol{K}\widehat{\boldsymbol{x}}\|_2^2 \leq \chi^2_{m,\eta}$. Then $\widehat{\boldsymbol{x}} \in \mathcal{B}_\eta \cap \Phi_\mu$ and since the minimum over $\mathcal{B}_\eta \cap \Phi_\mu$ must always be greater than or equal to the minimum over $\mathcal{X} \cap \Phi_\mu$, $\widehat{\boldsymbol{x}}$ also minimizes $\|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2$ over $\mathcal{B}_\eta \cap \Phi_\mu$. The same argument applies to minimizing over $\mathcal{B}_\eta$ and $\mathcal{X}$.

Therefore, since $\min_{\boldsymbol{x} \in \mathcal{B}_\eta \cap \Phi_\mu} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2 = \min_{\boldsymbol{x} \in \mathcal{X} \cap \Phi_\mu} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2$ and $\min_{\boldsymbol{x} \in \mathcal{B}_\eta} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2 = \min_{\boldsymbol{x} \in \mathcal{X}} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2$, we conclude that $\lambda(\mu, \boldsymbol{y}; \mathcal{B}_\eta) = \lambda(\mu, \boldsymbol{y}; \mathcal{X})$.

B Proofs in Section 4

Throughout the proofs, we make use of the following lemma:

Lemma B.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $\mathcal{X} \subset \mathbb{R}^n$ such that $\min_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x})$ is achieved, and at least one of the minimizers $\boldsymbol{x}^*$ satisfies:*

1. *f is continuous at $\boldsymbol{x}^*$*
2. *$\boldsymbol{x}^*$ is not an isolated point of $\mathcal{X}$ (i.e., $\forall \delta > 0, B_\delta(\boldsymbol{x}^*) \cap \mathcal{X} \neq \emptyset$)*

Let μ be a measure on $\mathcal{X}$ such that $\mu(B) > 0$ for all $B \subseteq \mathcal{X}$ such that $\lambda(B) > 0$. Let $Y_m = \min_{i=1,\dots,m} f(\boldsymbol{x}_i)$, where $\boldsymbol{x}_i$ are i.i.d samples from μ. Then $Y_m \xrightarrow{p} f(\boldsymbol{x}^*)$

Proof. Fix $\varepsilon > 0$ and let us show that $\mathbb{P}(|Y_m - f(\boldsymbol{x}^*)| > \varepsilon) \to 0$. Since f is continuous at $\boldsymbol{x}^*$ there exists a $\delta > 0$ such that $f(B_\delta(\boldsymbol{x}^*)) \subset B_\varepsilon(f(\boldsymbol{x}^*))$ so that

$$\mathbb{P}(|Y_m - f(\boldsymbol{x}^*)| \geq \varepsilon) \leq \mathbb{P}(\boldsymbol{x}_i \notin B_\delta(\boldsymbol{x}^*),\ \forall i = 1, \dots, m) = (\mathbb{P}_{\boldsymbol{x}\sim\mu}(\boldsymbol{x} \notin B_\delta(\boldsymbol{x}^*)))^m \tag{S.9}$$

Since $\boldsymbol{x}^*$ is not an isolated point, we have $\lambda(B_\delta(\boldsymbol{x}^*) \cap \mathcal{X}) > 0$ and therefore $\mathbb{P}_{\boldsymbol{x}\sim\mu}(\boldsymbol{x} \notin B_\delta(\boldsymbol{x}^*)) < 1$ and $\mathbb{P}(|Y_m - f(\boldsymbol{x}^*)| \geq \varepsilon) \to 0$ □

B.1 Proof of Lemma 4.1

By definition, $\bar{q}^{\mathrm{de}}_{\gamma,\eta} = \max_{i=1,\dots,M} \widehat{q}^i_\gamma(N)$, where we explicitly write the dependence with the number of samples and the index i refers to the quantile being estimated at i-th sampled point $\boldsymbol{x}_i$. We know that as $N \to \infty$, $\widehat{q}^i_\gamma(N)$ converges in probability to $Q_{P_{\boldsymbol{x}_i}}(1-\gamma)$. We have

$$\left|\bar{q}^{\mathrm{de}}_{\gamma,\eta} - \bar{q}_{\gamma,\eta}\right| \leq \left|\bar{q}^{\mathrm{de}}_{\gamma,\eta} - \max_{i=1,\dots,M} Q_{P_{\boldsymbol{x}_i}}(1-\gamma)\right| + \left|\max_{i=1,\dots,M} Q_{P_{\boldsymbol{x}_i}}(1-\gamma) - \bar{q}_{\gamma,\eta}\right| \tag{S.10}$$

The first term can be made smaller than $\varepsilon/2$ as $N \to \infty$ by convergence of the estimator, and the second term can be made smaller than $\varepsilon/2$ as $M \to \infty$ by application of Lemma B.1 to the quantile function (maximizing instead of minimizing).

B.2 Proof of Lemma 4.2

The proof is identical to that of Lemma 4.1, with the only difference of replacing the quantiles estimated via sampling to the quantile regressor, and N to M_{tr}, the number of samples needed to train the quantile regressor. Since the quantile regression is assumed to be consistent, the result follows.

B.3 Proof of Lemma 4.3

Recall,

$$C^{\mathrm{gl}}_{\mathrm{inv}}(\boldsymbol{y}) = \left[\min_{k:\lambda(\varphi(\boldsymbol{x}_k),\boldsymbol{y})\leq\widehat{q}} \varphi(\boldsymbol{x}_k),\ \max_{k:\lambda(\varphi(\boldsymbol{x}_k),\boldsymbol{y})\leq\widehat{q}} \varphi(\boldsymbol{x}_k)\right] \tag{S.11}$$

$$= \left[\min_{k:\lambda(\mu_k,\boldsymbol{y})\leq\widehat{q}} \mu_k, \max_{k:\lambda(\mu_k,\boldsymbol{y})\leq\widehat{q}} \mu_k \right] \tag{S.12}$$

where $\widehat{q} := \max_{i=1,\ldots,M} q(\boldsymbol{x}_i)$, the $\boldsymbol{x}_i$'s are samples from $\mathcal{X}\cap\mathcal{B}_\eta$ and $\mu_i := \varphi(\boldsymbol{x}_i)$. Note that μ_i are samples in $\varphi(\mathcal{X}\cap\mathcal{B}_\eta) \subset \mathbb{R}$. Also note that we have more explicitly written out the interval definition (i.e., Equation (16)) to emphasize clarity rather than presentation. Additionally, to make notion more compact, we have written the LLR as $\lambda(\varphi(\boldsymbol{x}), \boldsymbol{y})$ rather than $\lambda(\varphi(\boldsymbol{x}), \boldsymbol{y}; \mathcal{X})$ as done elsewhere.

Consider the left extreme of the interval without loss of generality, since a similar argument follows from the right extreme. Consider three quantities:

$$\mu_1 := \min \mu_k \quad \text{s.t.} \quad k = 1,\ldots,M \text{ and } \lambda(\mu_k, \boldsymbol{y}) \leq \widehat{q} \tag{S.13}$$

$$\mu_2 := \min \mu_k \quad \text{s.t.} \quad k = 1,\ldots,M \text{ and } \lambda(\mu_k, \boldsymbol{y}) \leq \bar{q}_{\gamma,\eta} \tag{S.14}$$

$$\mu_3 := \min \mu \quad \text{s.t.} \quad \mu \in \varphi(\mathcal{X}\cap\mathcal{B}_\eta) \text{ and } \lambda(\mu, \boldsymbol{y}) \leq \bar{q}_{\gamma,\eta} \tag{S.15}$$

Our goal is to show that as $M \to \infty$, $\mu_1 \xrightarrow{\text{p}} \mu_3$. Lemma B.1 shows that $\mu_2 \xrightarrow{\text{p}} \mu_3$. Indeed, the minimization over the indices k such that the condition is satisfied can be seen as a rejection sampling strategy in which all accepted samples are samples of the feasible region of the optimization in μ_3. As M grows, since the sampler eventually samples all areas of $\varphi(\mathcal{X}\cap\mathcal{B}_\eta)$, some samples are guaranteed to be close to the optimum with high probability. It remains to show that $\mu_1 \xrightarrow{\text{p}} \mu_2$. This follows from continuity of the optimization problem with respect to the right hand side of the constraint, and the fact that $\max_{i=1,\ldots,M} q(\boldsymbol{x}_i) \xrightarrow{\text{p}} \bar{q}_{\gamma,\eta}$.

B.4 Proof of Lemma 4.4

We have a memoryless sampler that first samples $\boldsymbol{x}_k$ and then accepts $\boldsymbol{h}^T\boldsymbol{x}_k$ as a sample if $\lambda(\boldsymbol{h}^T\boldsymbol{x}_k, \boldsymbol{y}) < q_\gamma^k$. If we consider the problem that we want to solve:

$$\max_{\mu\in\varphi(\mathcal{X}\cap\mathcal{B}_\eta)} \text{ s.t. } \lambda(\mu, y) \leq \max_{h^t x=\mu} q(x)$$

It then holds that all the sampled μ_i satisfy $\lambda(\mu_i, y) \leq \max_{h^t x=\mu_i} q(x)$ and are therefore feasible to the optimization problem. Furthermore, using an argument analogous to Lemma B.1, it can be shown that for every feasible point μ, there is eventually a sample close to it. This is because the quantity of interest is continuous, the measures of the samplers in x space and the sampler in μ space sample everywhere eventually, and the larger a quantile can be for an accepted μ_i is exactly $\max_{h^t x=\mu} q(x_i)$), by definition of our sample acceptance rule .

B.5 Proof of Lemma 4.5

We prove continuity of the optimization problem

$$\max_{\mu\in\varphi(\mathcal{X}\cap\mathcal{B}_\eta)} \text{ s.t. } \lambda(\mu, y) \leq q \tag{S.16}$$

as a function of q in the positive measure interval $(\lambda(\bar{\mu}, \boldsymbol{y}), \bar{q}_{\gamma,\eta}]$. Therefore, convergence in probability follows as we have convergence in probability to $\bar{q}_{\gamma,\eta}$, which in particular implies that the maximum quantile estimate is in the interval $(\lambda(\bar{\mu}, \boldsymbol{y}), \bar{q}_{\gamma,\eta}]$ almost surely. We do so by appealing to the maximum theorem (Ok, 2007, § E.3), which, in general, guarantees continuity of functions

of the form $f^*(\theta) = \sup\{f(x,\theta) : x \in C(\theta)\}$ as long as f is continuous, C is a continuous compact-valued correspondence and $C(\theta)$ is non-empty for all $\theta \in \Theta$. The continuity of C comes from the continuity of the LLR, f is equal to the identity and the strict feasibility condition ensures C is non-empty in $\Theta := (\lambda(\bar{\mu}, \boldsymbol{y}), \bar{q}_{\gamma,\eta}]$

B.6 Proof of Lemma 4.6

We will prove sufficient conditions for convergence of $\inf_{\mu:\widehat{f}_k(\mu)\geq 0} \mu$ to $\inf_{\mu:f(\mu)\geq 0} \mu$ as $\widehat{f}_k$ converges to f, and the result will follow by taking $\widehat{f}_k(\mu) = \widehat{m}_\gamma(\mu) - \lambda(\mu, \boldsymbol{y})$ and $f(\mu) = m_\gamma(\mu) - \lambda(\mu, \boldsymbol{y})$. A similar argument can be repeated for the supremum. Use use the notation $\widehat{f}_k$ to indicate that k sampled points are used to estimate this function via the definition of $\widehat{m}_\gamma(\mu)$ (see Section 3.3).

Lemma B.2. *Let $f : \mathbb{R} \to \mathbb{R}$ be a function, and let f_k be a sequence of functions $f_k : \mathbb{R} \to \mathbb{R}$. Let $\mu^* = \inf_{f(\mu)\geq 0} \mu$ and $\mu_k = \inf_{f_k(\mu)\geq 0} \mu$. Let the sequence of functions $\{f_k\}$ be such that for all $\delta > 0$,*

$$\mathbb{P}\left(\sup_{\mu} |f_k(\mu) - f(\mu)| > \delta\right) \to 0 \text{ as } k \to \infty, \tag{S.17}$$

namely, f_k converges in probability uniformly to f. Furthermore, let f be such that for all $\varepsilon > 0$, there exists $\delta' > 0$ such that $|f(\mu)| > \delta'$ if and only if $|\mu - \mu^*| < \varepsilon$. Then, we have $\mu_k \xrightarrow{\text{p}} \mu^*$.

We then have $\mu_k \xrightarrow{p} \mu^*$.

Proof. Assume for the sake of contradiction that there exists $\varepsilon > 0$ such that $\mathbb{P}(|\mu_k - \mu^*| \geq \varepsilon)$ does not go to 0. Then, by the condition on f, it must follow that $\mathbb{P}(|f(\mu_k)| > \delta)$ does not go to 0. But,

$$\mathbb{P}(|f(\mu_k)| > \delta) \leq \mathbb{P}(|f(\mu_k) - f_k(\mu_k)| > \delta) + \mathbb{P}(|f_k(\mu_k)| > \delta) \tag{S.18}$$

And the right-hand side goes to 0 by uniform convergence in probability (first term) and by feasibility of μ_k (second term)

□

C Additional details and illustrations in Section 6

C.1 Importance-like sampler for the three-dimensional example in Section 6.2

Since the parameter settings with quantiles meaningfully larger than $\chi^2_{1,\alpha}$ are located close to the constraint boundary, a sampling challenge is presented. Using the samplers as described in Section 5.1 results in under-sampling of this large-quantile region since both samplers provide uniform random samples over the Berger-Boos set. Algorithm 5 presents an importance-like sampler to increase the probability mass of samples close to the constraint boundary. Note, we say "importance-like" because we do not provide any theoretical guarantee regarding this sampler's ability to produce draws from a particular target distribution. We alternatively tailored Algorithm 5 to the particular three-dimensional example in Section 6.2.

The key to Algorithm 5 is the additional accept/reject step where the k-th sample is accepted with probability p_k. Additionally, by setting $q \in (0, 1)$, we prioritize retaining samples closer to the non-negativity boundary. This effect can be seen in Figure S.1, where the vast majority of samples are found closer to the constraint boundary. The ability of Algorithm 5 to better sample

Algorithm 5 Importance-like sampler for three-dimensional constrained Gaussian

Input: Number of samples: $M \in \mathbb{N}$, inverse length scale: γ_p, and order of norm: $q \in (0, 1)$.

Instantiate a list $\mathcal{S}$ of length M to store sampled points.
while $|\mathcal{S}| < M$ **do**
  Draw M realizations from Algorithm 4: $\widetilde{\boldsymbol{x}}_1, \ldots, \widetilde{\boldsymbol{x}}_M$
  for $k = 1, \ldots, M$: **do**
    Compute the probability of accepting the k-th draw: $p_k := \exp\left(-\gamma_p \|\widetilde{\boldsymbol{x}}_k\|_q\right)$.
    Draw $z_k \sim \text{Bernoulli}(p_k)$.
    if $z_k = 1$ **then**
      $\mathcal{S}[k] \leftarrow \widetilde{\boldsymbol{x}}_k$
    end if
  end for
end while

Output: Sampled parameters in Berger-Boos set $\mathcal{S}$.

the high-quantile regions of the Berger-Boos set can be seen in Figure S.2. In particular, for the functional values $\mu \in (-4, -2)$, the importance-like sampler is substantially more effective than the Polytope sampler at finding parameter setting with γ-quantiles greater than 1.1.

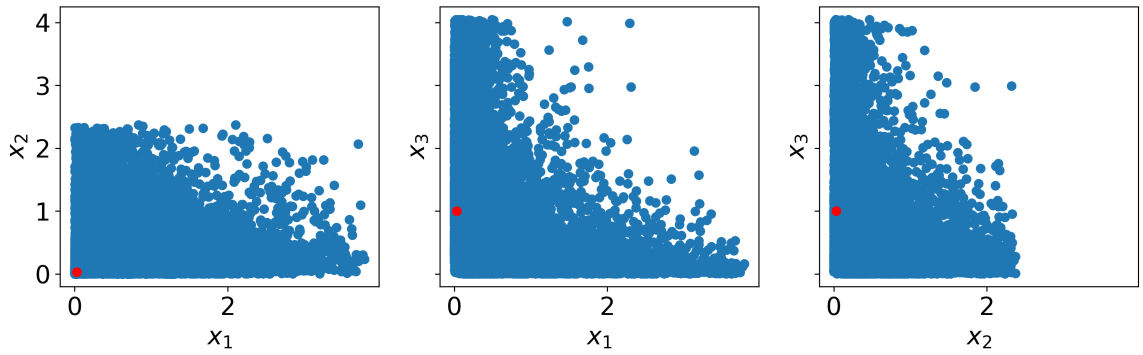

Figure S.1: When sampling points using Algorithm 5, the vast majority of samples are found closer to the non-negativity constraint boundary. The true parameter setting is shown by the red point, while the parameter settings sampled by Algorithm 5 are shown by the blue points. This sampling prioritization helps adequately sample the regions of the Berger-Boos set where the quantile surface is larger than $\chi^2_{1,\alpha}$.

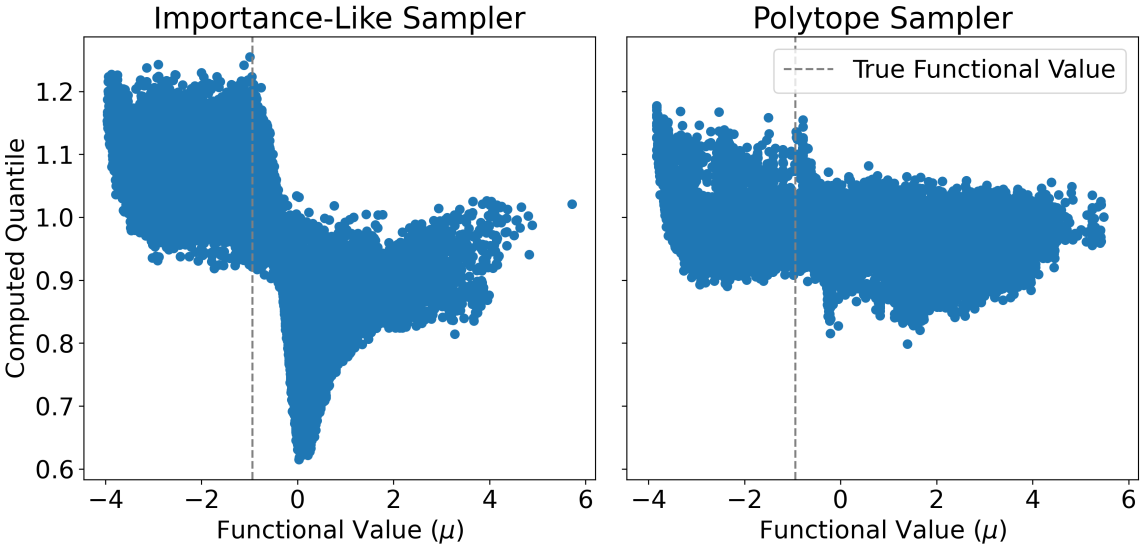

Figure S.2: The importance-like sampler described by Algorithm 5 is more effective than the Polytope sampler described by Algorithm 4 at sampling parameter settings with γ-quantile greater than 1.1. Each parameter setting sampled by Algorithm 5 is shown by a blue point. This improved ability helps ensure the coverage guarantee shown in the left panel of Figure 6.3.