Frequentist Confidence Intervals via optimization: Resolving the Burrus conjecture

Pau Batlle

Department of Computing and Mathematical Sciences
California Institute of Technology

SIAM UQ - February 28^{th} 2024

Warmup: inverse problems setup

$$\underbrace{y}_{\text{Data}} = \underbrace{f}_{\substack{\text{Forward model}\\\text{(known)}}}\Big(\underbrace{x^*}_{\substack{\text{True parameter}\\\text{(unknown)}}}\Big) + \underbrace{\varepsilon}_{\substack{\text{Noise}\\\text{(known distr.)}}}$$

Goal

Recover an unknown x^* given observations y and give UQ about the estimate

Warmup: inverse problems setup

$$\underbrace{y}_{\text{Data}} = \underbrace{f}_{\substack{\text{Forward model}\\ \text{(known)}}} \Big(\underbrace{x^*}_{\substack{\text{True parameter}\\ \text{(unknown)}}} \Big) + \underbrace{\varepsilon}_{\substack{\text{Noise}\\ \text{(known distr.)}}}$$

Goal

Recover an unknown x^* given observations y and give UQ about the estimate

Today, we will focus on problems in which:

- The only known information on x^* is $x^* \in \mathcal{X}$
- The problem is possibly ill-posed (those that even without noise, admit multiple x^* compatible with an observation y)
- Very few observations are available (no asymptotic results)

The need for transparent and well-calibrated UQ

For certification in safety-based applications, we need UQ that is:

- **Transparent**: With a clear set of assumptions and meaningful outputs
- **Well-calibrated**: Precisely adjustable to the desired level of coverage
  - If it **undercovers**, we incur **unnecessary risk**
  - If it **overcovers**, we incur **economic cost**

Today's talk

Batlle P., Pratik P., Stanley M. Kuusala M. and Owhadi H., "Optimization-based frequentist confidence intervals for constrained inverse problems: Resolving the Burrus conjecture", arXiv:2310.02461, 2023.

Collaborators:

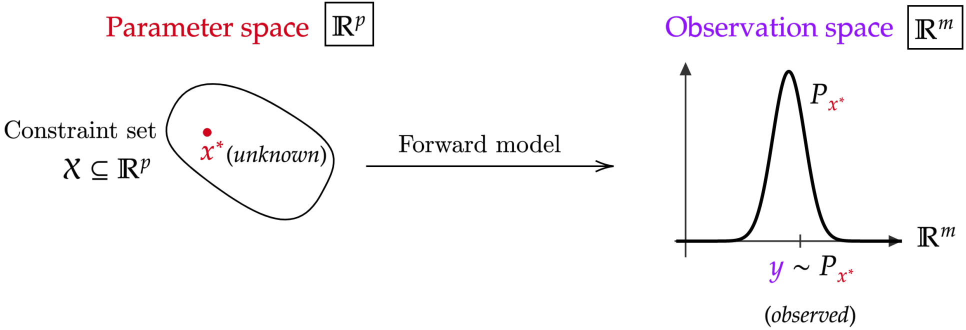
Mike Stanley
(CMU)

Pratik Patil
(Berkeley)

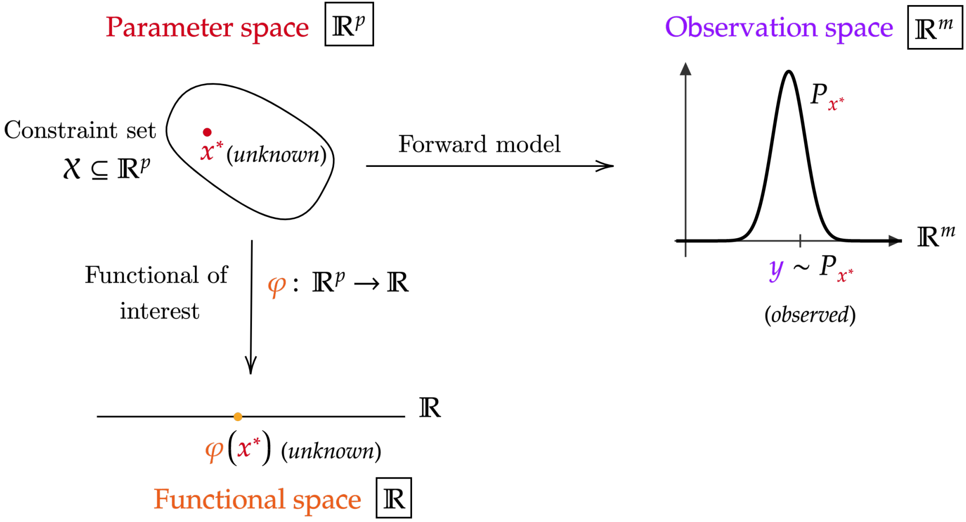
Houman Owhadi
(Caltech)

Mikael Kuusela
(CMU)

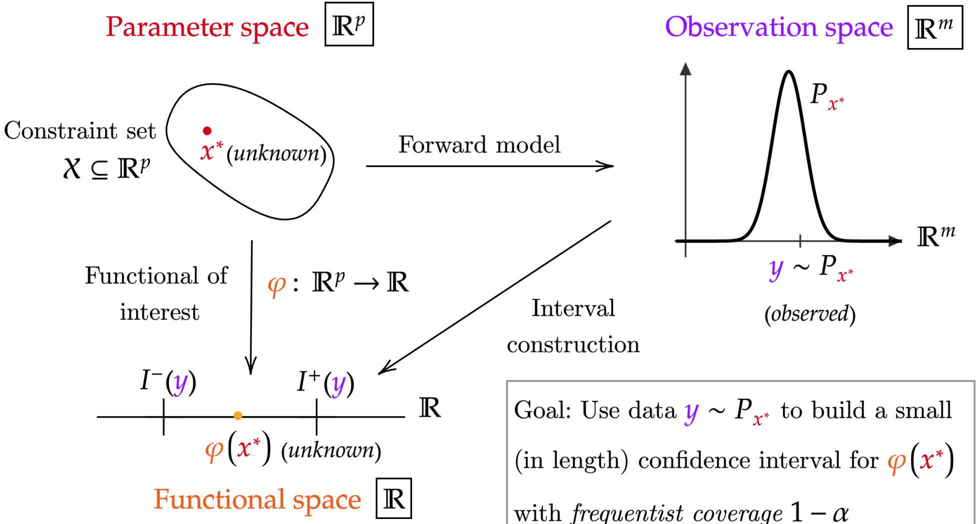
Constrained inverse problems setup



Constrained inverse problems setup



Constrained inverse problems setup



Goal: Use data $y \sim P_{x^*}$ to build a small (in length) confidence interval for $\varphi(x^*)$ with *frequentist coverage* $1 - \alpha$

Frequentist, well-calibrated intervals

Frequentist guarantees

Find an interval that contains $\varphi(x^*)$ with probability $1 - \alpha$:

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

with finite sample guarantees and that is as small as possible

Frequentist, well-calibrated intervals

Frequentist guarantees

Find an interval that contains $\varphi(x^*)$ with probability $1-\alpha$:

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

with finite sample guarantees and that is as small as possible

- Stronger notion than Bayesian credible intervals
- Only assumes likelihood model and constraints
- Well-calibration: Slack in the inequality as small as possible

Baseline: simultaneous approach

A general method to build intervals with correct coverage is the **simultaneous approach** [Stark 1992, 1994]:

1. Find a $1 - \alpha$ (frequentist) confidence set $\mathcal{C}(y)$ for x^*
2. Intersect it with the constraint set $\mathcal{X}$
3. Map through $\varphi(x)$

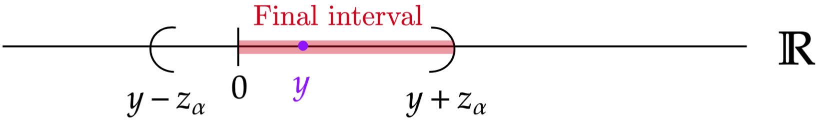
Baseline: simultaneous approach

A general method to build intervals with correct coverage is the **simultaneous approach** [Stark 1992, 1994]:

1. Find a $1-\alpha$ (frequentist) confidence set $\mathcal{C}(y)$ for x^*
2. Intersect it with the constraint set $\mathcal{X}$
3. Map through $\varphi(x)$

Example:

$$y = x^* + \varepsilon, \quad \varphi(x) = x, \quad \varepsilon \sim \mathcal{N}(0,1), \quad x^* \geq 0$$



Baseline: simultaneous approach

A general to build intervals with correct coverage is the **simultaneous approach** [Stark 1992, 1994]:

1. Find a $1-\alpha$ confidence set $\mathcal{C}(y)$ for x^*
2. Intersect it with the constraint set $\mathcal{X}$
3. Map through $\varphi(x)$

- Can be written as "refined worst-case" optimization problems ($\mathcal{X} \to \mathcal{X} \cap \mathcal{C}(y)$)

$$\left[\min_{\boldsymbol{x} \in \mathcal{X} \cap \mathcal{C}(y)} \varphi(x), \max_{x \in \mathcal{X} \cap \mathcal{C}(y)} \varphi(x)\right] =: \begin{array}{ll} \min_x / \max_x & \varphi(x) \\ \text{s.t.} & x \in \mathcal{X} \cap \mathcal{C}(y) \end{array}$$

- It overcovers, since $\mathcal{C}(y)$ does not need to be $1-\alpha$ for the interval to be $1-\alpha$

The linear Gaussian model with linear constraints and FoI

Linear Gaussian model with linear constraints and functional of interest

$$y = Kx^* + \varepsilon, \quad \mathcal{X} = \{x : Ax \leq b\}$$

with $y \in \mathbb{R}^m$, $x^* \in \mathbb{R}^p$, $\varepsilon \sim \mathcal{N}(0, I_m)$, and $\varphi(x) = h^T x$

- Generally ill-posed for rank-deficient K, incorporating the constraint allows the computation of finite-length intervals without priors

The linear Gaussian model with linear constraints and FoI

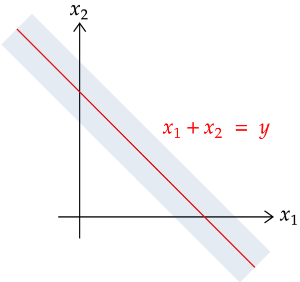
Linear Gaussian model with linear constraints and functional of interest

$$y = Kx^* + \varepsilon, \quad \mathcal{X} = \{x : Ax \leq b\}$$

with $y \in \mathbb{R}^m$, $x^* \in \mathbb{R}^p$, $\varepsilon \sim \mathcal{N}(0, I_m)$, and $\varphi(x) = h^T x$

- Generally ill-posed for rank-deficient K, incorporating the constraint allows the computation of finite-length intervals without priors

Example:



$x = (x_1, x_2)$
$y = x_1 + x_2 + \varepsilon \in \mathbb{R}$
Confidence interval for $h^t x := x_1 - x_2$?

Ill-posed inverse problem, can not give finite intervals with frequentist guarantees

The linear Gaussian model with linear constraints and FoI

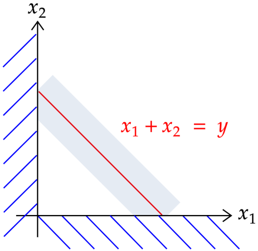
Linear Gaussian model with linear constraints and functional of interest

$$y = Kx^* + \varepsilon, \quad \mathcal{X} = \{x : Ax \leq b\}$$

with $y \in \mathbb{R}^m$, $x^* \in \mathbb{R}^p$, $\varepsilon \sim \mathcal{N}(0, I_m)$, and $\varphi(x) = h^T x$

- Generally ill-posed for rank-deficient K, incorporating the constraint allows the computation of finite-length intervals without priors

Example:



$x = (x_1, x_2) \geq 0$
$y = x_1 + x_2 + \varepsilon \in \mathbb{R}$
Confidence interval for $h^t x := x_1 - x_2$?

The known constraint allows finite intervals with frequentist guarantees without extra assumptions!

The linear Gaussian model with linear constraints and FoI

Linear Gaussian model with linear constraints and functional of interest

$$y = Kx^* + \varepsilon, \quad \mathcal{X} = \{x : Ax \leq b\}$$

with $y \in \mathbb{R}^m$, $x^* \in \mathbb{R}^p$, $\varepsilon \sim \mathcal{N}(0, I_m)$, and $\varphi(x) = h^T x$

- Generally ill-posed for rank-deficient K, incorporating the constraint allows the computation of finite-length intervals without priors
- Originally studied in the context of unfolding gamma-ray and neutron spectra from pulse-height distributions under rank-deficient linear systems [Burrus 1965]
- Has since then become the most studied/fundamental constrained inference problem [Rust and Burrus 1972, O'Leary and Rust 1986, 1994, Tenorio 2007]

The Burrus conjecture

Burrus conjecture (informal) [Burrus 1965, Rust and Burrus 1972]

Consider the model $y = Kx^* + \varepsilon, \quad \varphi(x) = h^T x, \quad Ax^* \leq b, \quad \varepsilon \sim \mathcal{N}(0, I_m)$
A valid $1-\alpha$ confidence interval for $\varphi(x^*)$ has its extremes given by:

$$\begin{aligned} \min_x / \max_x \quad & h^T x \\ \text{s.t.} \quad & \|y - Kx\|_2^2 \leq \psi_\alpha^2(y) \\ & Ax \leq b \end{aligned}$$

With $\psi_\alpha^2(y)$ much smaller than the simultaneous approach equivalent constant $Q_{\chi_m^2}(1-\alpha)$

- The set $\{x : \|y - Kx\|_2^2 \leq \psi_\alpha^2(y)\}$ plays the role of $\mathcal{C}(y)$ in the simultaneous approach and is **not** necessarily a $1-\alpha$ set
- If true, provides short intervals for a range of fundamental problems

The Burrus conjecture

Burrus conjecture [Burrus 1965, Rust and Burrus 1972]

Consider the model

$$y = Kx^* + \varepsilon, \quad \varphi(x) = h^T x, \quad Ax^* \leq b, \quad \varepsilon \sim \mathcal{N}(0, I_m)$$

A valid $1-\alpha$ confidence interval for $\varphi(x^*)$ has its extremes given by:

$$\begin{aligned} \min_x / \max_x \quad & h^T x \\ \text{s.t.} \quad & \|y - Kx\|_2^2 \leq \psi_\alpha^2(y) \\ & Ax \leq b \end{aligned}$$

where $\psi_\alpha^2(y) = (c_{\alpha/2})^2 + s^2$, where $\mathbb{P}(Z > c_\alpha) = \alpha$ for $Z \sim \mathcal{N}(0,1)$ and

$$\begin{aligned} s^2 := \min_z \quad & \|y - Kz\|_2^2 \\ \text{s.t.} \quad & z \geq 0 \end{aligned}$$

The Burrus conjecture: more than 50 years of history

- Conjecture proposed [Burrus 1965, Rust and Burrus 1972]
- Theoretical analysis and "proof" [O'Leary and Rust 1986, 1994]
- Mistake pointed out in proof and counterexample proposal [Tenorio et al. 2007]

The Burrus conjecture: more than 50 years of history

- Conjecture proposed [Burrus 1965, Rust and Burrus 1972]
- Theoretical analysis and "proof" [O'Leary and Rust 1986, 1994]
- Mistake pointed out in proof and counterexample proposal [Tenorio et al. 2007]

Theorem [Batlle, Patil, Stanley, Owhadi, Kuusela 2023]

- The counterexample provided in Tenorio et al. is not valid
- There exist other counterexamples, so **the Burrus conjecture is false**

The Burrus conjecture: more than 50 years of history

- Conjecture proposed [Burrus 1965, Rust and Burrus 1972]
- Theoretical analysis and "proof" [O'Leary and Rust 1986, 1994]
- Mistake pointed out in proof and counterexample proposal [Tenorio et al. 2007]

Theorem [Batlle, Patil, Stanley, Owhadi, Kuusela 2023]

- The counterexample provided in Tenorio et al. is not valid
- There exist other counterexamples, so **the Burrus conjecture is false**

Our proof technique is novel in analyzing this problem and provides:

- A test to identify when the Rust-Burrus approach works and when it does not
- An algorithm to fix the faulty examples
- A generalization of the Rust-Burrus methodology beyond Gaussian-linear settings

Generalized Burrus conjecture intervals

We develop theoretical tools to study a large class of confidence intervals that contains the ones posited in the conjecture

Generalization to non-linear QoI
and non-gaussian likelihoods

$$\ell_x(y) := \log p(y|x)$$

Rust-Burrus intervals

$$\begin{aligned} &\min_x / \max_x \; h^t x \\ &\text{s. t.} \quad \|y - Kx\|_2^2 - \min_{z\geq 0}\|y - Kz\|_2^2 \leq q_\alpha \\ &\quad x \geq 0 \end{aligned}$$

$$\begin{aligned} &\min_x / \max_x \; \varphi(x) \\ &\text{s. t.} \quad -2\ell_x(y) - \inf_{z\in\mathcal{X}} -2\ell_z(y) \leq q_\alpha \\ &\quad x \in \mathcal{X} \end{aligned}$$

Generalized Burrus conjecture intervals

We develop theoretical tools to study a large class of confidence intervals that contains the ones posited in the conjecture

Generalization to non-linear QoI
and non-gaussian likelihoods

$$\ell_x(y) := \log p(y|x)$$

Rust-Burrus intervals

$$\begin{aligned} \min_x / \max_x \quad & h^t x \\ \text{s. t.} \quad & \|y - Kx\|_2^2 - \min_{z \geq 0} \|y - Kz\|_2^2 \leq q_\alpha \\ & x \geq 0 \end{aligned}$$

$$\begin{aligned} \min_x / \max_x \quad & \varphi(x) \\ \text{s. t.} \quad & -2\ell_x(y) - \inf_{z \in \mathcal{X}} -2\ell_z(y) \leq q_\alpha \\ & x \in \mathcal{X} \end{aligned}$$

Burrus conjecture is a particular choice of q_α: $q_\alpha = Q_{\chi^2_1}(1-\alpha)$, where Q is the quantile function

Finding valid q_α

Definition

Let $\lambda(\mu, y) := \inf_{\substack{\varphi(x)=\mu \\ x\in\mathcal{X}}} -2\ell_x(y) - \inf_{x\in\mathcal{X}} -2\ell_x(y)$ and, for $x \in \mathcal{X}$ let $Z_x := \lambda(\varphi(x), y)$ where $y \sim P_x$

Theorem [Batlle, Patil, Stanley, Owhadi, Kuusela 2023]

The Rust-Burrus interval covers $\varphi(x^*)$ at a level α iff $q_\alpha \geq \sup_{\substack{\varphi(x)=\varphi(x^*) \\ x\in\mathcal{X}}} Q_{Z_x}(1-\alpha)$

Very strong consequences:

- We know the theoretical optimal values to use in this construction
- We can use $q_\alpha = Q_X(1-\alpha)$ for a given r.v X at all levels α iff X stochastically dominates Z_x ($\mathbb{P}(X \geq \beta) \geq \mathbb{P}(Z_x \geq \beta)\, \forall \beta$, noted $X \succeq Z_x$) $\forall x \in \mathcal{X}$
- We can improve the intervals from $q_\alpha \to q_\alpha(\varphi(x))$ (more on that later)

A new provable counterexample

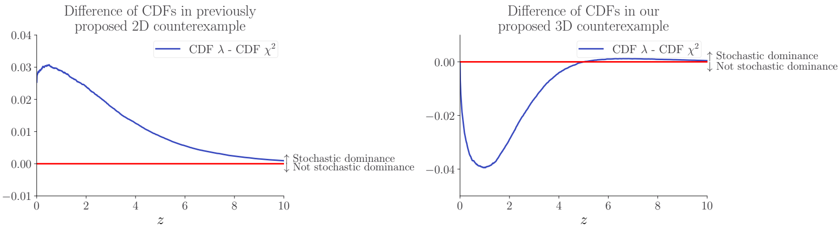
Our theoretical analysis shows that the Burrus conjecture is equivalent to:

$$Z_{x^*} = \min_{\substack{h^T x = h^T x^* \\ x \geq 0}} \|y - Kx\|_2^2 - \min_{x \geq 0} \|y - Kx\|_2^2 \preceq \chi_1^2 \text{ when } y \sim P_{x^*}$$

A new provable counterexample

Our theoretical analysis shows that the Burrus conjecture is equivalent to:

$$Z_{x^*} = \min_{\substack{h^T x = h^T x^* \\ x \geq 0}} \|y - Kx\|_2^2 - \min_{x \geq 0} \|y - Kx\|_2^2 \preceq \chi_1^2 \text{ when } y \sim P_{x^*}$$



Proof technique

The proof of stochastic dominance (left) is via a coupling of random variables argument. The disproof of stochastic dominance (right) is via computing $\mathbb{E}[Z]$

A general recipe for constructing provably correct intervals

The proof technique comes with a general recipe for constructing $1-\alpha$ intervals.

1. Write down the random variable Z_x for all $x \in \mathcal{X}$
2. Obtain valid q_α by solving/bounding:
   - $\sup_{x\in\mathcal{X}} Q_{Z_x}(1-\alpha)$ (to set q_α)
   - $\sup_{\substack{\varphi(x)=\mu \\ x\in\mathcal{X}}} Q_{Z_x}(1-\alpha)$ (to set q_α depending on $\mu = \varphi(x)$)
3. Solve

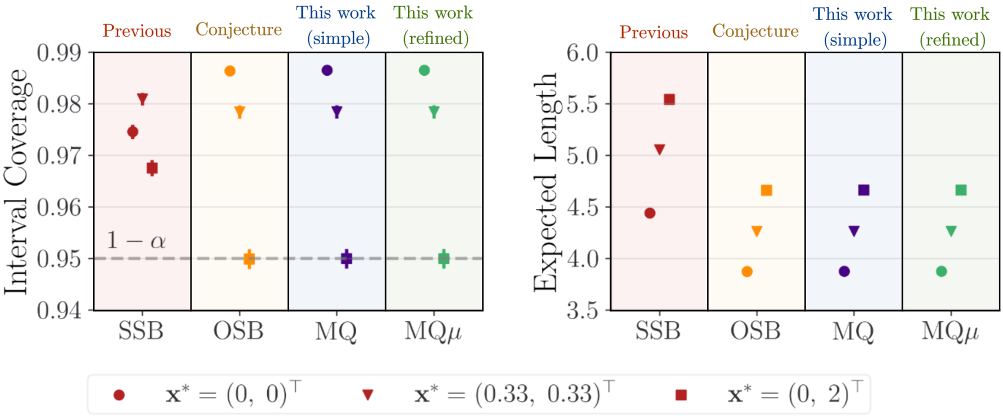
$$\begin{aligned} \min_x / \max_x \quad & \varphi(x) \\ \text{s.t.} \quad & x \in \mathcal{X} \\ & -2\ell_x(y) - \inf_{x'\in\mathcal{X}} -2\ell_{x'}(y) \leq q_\alpha(\varphi(x)) \end{aligned}$$

Remarks: The intervals with $q_\alpha(\varphi(x))$ are shorter, but harder to compute.

Numerical results

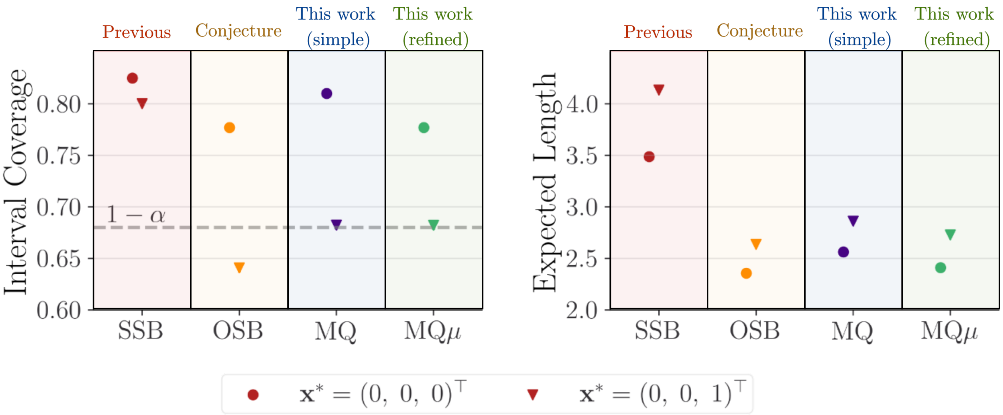
The Burrus conjecture intervals correctly cover in the previously proposed counterexample

$$x, y \in \mathbb{R}^2, \quad y = x + \varepsilon, \quad x^* \geq 0, \quad \varphi(x) = [1, -1]^T x, \quad \varepsilon \sim \mathcal{N}(0, I_2)$$



Numerical results

$$x, y \in \mathbb{R}^3, \quad y = x + \varepsilon, \quad x^* \geq 0, \quad \varphi(x) = [1, 1, -1]^T x, \quad \varepsilon \sim \mathcal{N}(0, I_3)$$

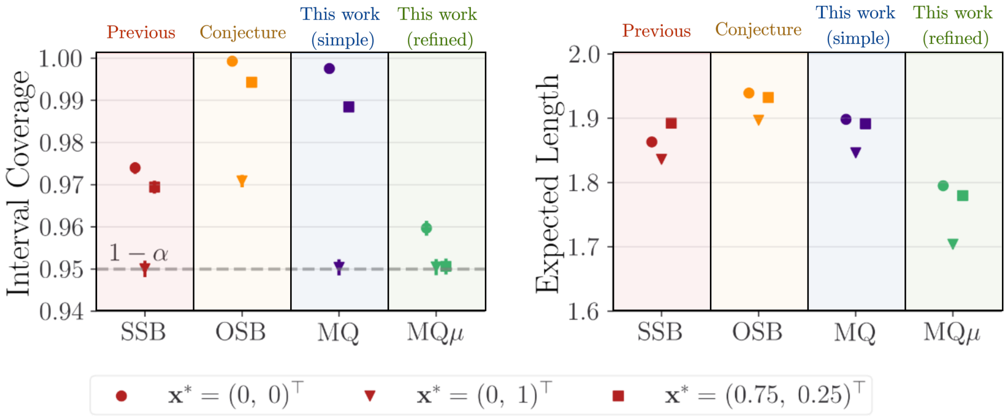


The Burrus conjecture intervals undercover for our proposed counterexample, and the described intervals fix the undercoverage

Numerical results

The Burrus conjecture intervals *can* overcover, and our described intervals can also fix the overcoverage

$$x, y \in \mathbb{R}^2, \quad y = x + \varepsilon, \quad 0 \leq x^* \leq 1, \quad \varphi(x) = [1, -1]^T x, \quad \varepsilon \sim \mathcal{N}(0, I_2)$$



Contributions

Our optimization-based framework to the problem of obtaining $1 - \alpha$ confidence intervals with frequentist guarantees in constrained inverse problems:

- Reinterprets previously proposed methods and disproves the long-standing Burrus conjecture (1965)
- Comes with an algorithm for constrained inverse problems that beats previous approaches in toy and real problems
- Explains the observed coverage (or lack thereof) throughout all numerical examples

Reach me at pau@caltech.edu

Backup slides

A high-level framework for certifiable UQ

Step 1: Write down all statistical model elements

Form of the likelihood/noise, forward model, underlying physics, prior distribution, data, parameter constraints...

A high-level framework for certifiable UQ

Step 1: Write down all statistical model elements

Form of the likelihood/noise, forward model, underlying physics, prior distribution, data, parameter constraints...

Step 2: Divide into knowns and unknowns

What you know/assume

What you do not know

A high-level framework for certifiable UQ

Step 1: Write down all statistical model elements

Form of the likelihood/noise, forward model, underlying physics, prior distribution, data, parameter constraints...

Step 2: Divide into knowns and unknowns

What you know/assume

What you do not know

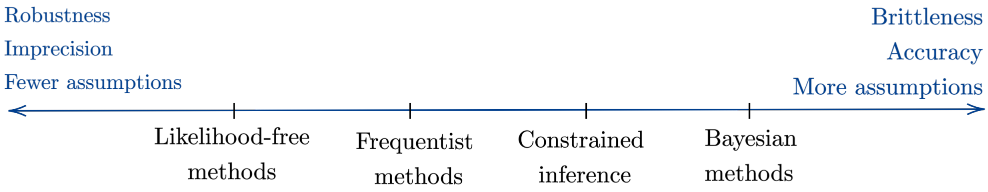
Use as information for more accurate UQ

Worst-case approach (min and max, minimax, ...)

Result: Certifiable UQ under the known knowns and the known unknowns

Can be used as an iterative process to falsify model assumptions

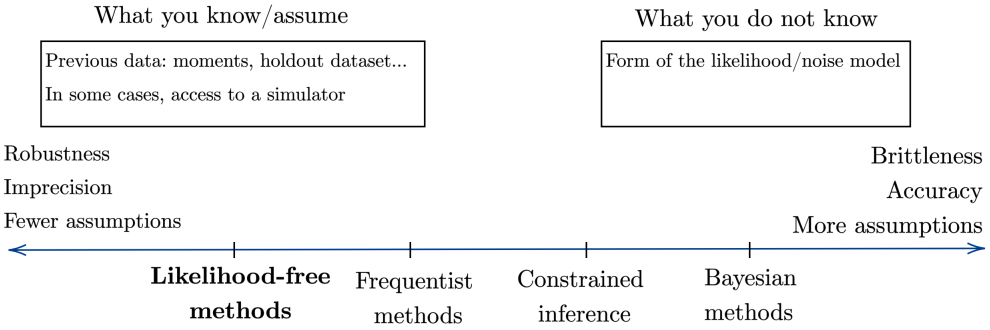
A scale of possible assumptions



We need theory and algorithms for all the scale, since:

- Reasonable assumptions vary by application
- There are (possibly steep) prices to pay for both underassuming and overassuming

A scale of possible assumptions



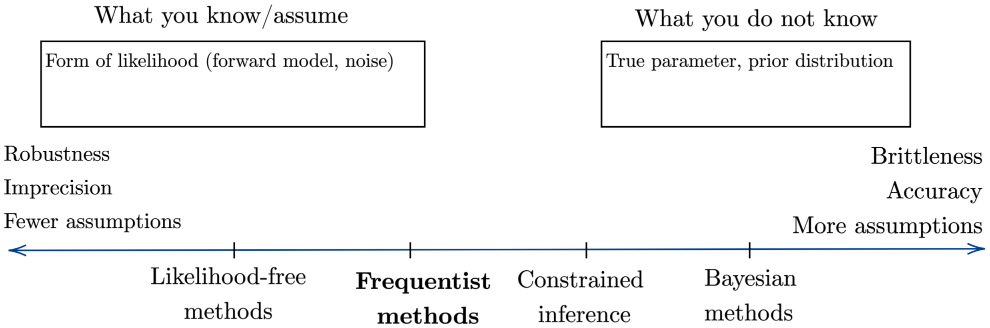
Optimal UQ: [Owhadi '13]

Universal Inference: [Wasserman '19]

Conformal Prediction: [Vapnik '98, Saunders '99, ...]

Likelihood-free inference: [Gutmann '16, Thomas '20...]

A scale of possible assumptions

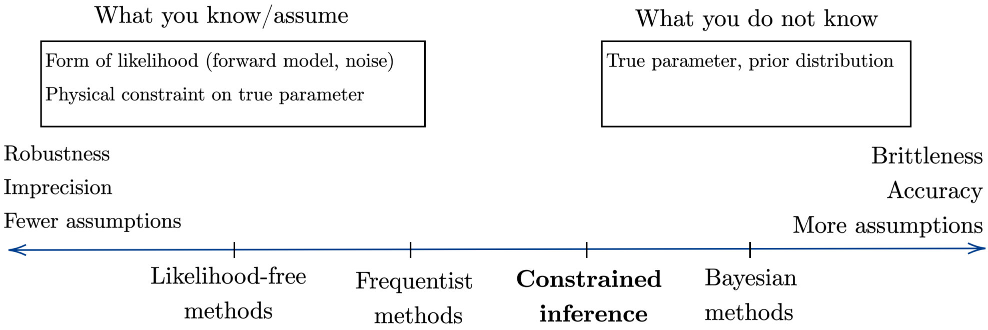


Frequentist statistics: [Neyman '37, ...]

Confidence intervals: [Neyman '37, ...]

Algorithms: [Dempster '97, ...]

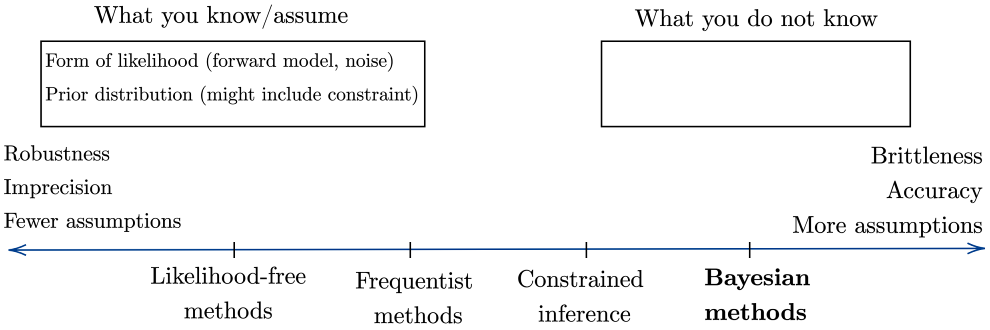
A scale of possible assumptions



Constrained tests: [Góurieroux '82, ...]

Constrained intervals: [Burrus '65, Rust-Burrus '72, ...]

A scale of possible assumptions



Theory: [Hjort '01, ...]

Algorithms: [Hastings '70, ...]

The regime of partial information

Today: Methods in the intermediate regime of partial information

This regime combines two challenging tasks:

(i) how to use the information we know

(ii) how to certify over what we do not know

We can borrow tools from worst-case and Bayesian methods