

A tutorial on the Bayesian statistical approach to inverse problems

Cite as: APL Mach. Learn. 1, 041101 (2023); doi: 10.1063/5.0154773
Submitted: 15 April 2023 • Accepted: 29 September 2023 •
Published Online: 6 November 2023



View Online



Export Citation



CrossMark

Faaq G. Waqar¹  Swati Patel,^{2,a)}  and Cory M. Simon^{3,a)} 

AFFILIATIONS

¹School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, Oregon 97331, USA

²Department of Mathematics, Oregon State University, Corvallis, Oregon 97331, USA

³School of Chemical, Biological, and Environmental Engineering, Oregon State University, Corvallis, Oregon 97331, USA

^{a)}Authors to whom correspondence should be addressed: patelswa@oregonstate.edu and cory.simon@oregonstate.edu

ABSTRACT

Inverse problems are ubiquitous in science and engineering. Two categories of inverse problems concerning a physical system are (1) estimate parameters in a model of the system from observed input–output pairs and (2) given a model of the system, reconstruct the input to it that caused some observed output. Applied inverse problems are challenging because a solution may (i) not exist, (ii) not be unique, or (iii) be sensitive to measurement noise contaminating the data. Bayesian statistical inversion (BSI) is an approach to tackle ill-posed and/or ill-conditioned inverse problems. Advantageously, BSI provides a “solution” that (i) quantifies uncertainty by assigning a probability to each possible value of the unknown parameter/input and (ii) incorporates prior information and beliefs about the parameter/input. Herein, we provide a tutorial of BSI for inverse problems by way of illustrative examples dealing with heat transfer from ambient air to a cold lime fruit. First, we use BSI to infer a parameter in a dynamic model of the lime temperature from measurements of the lime temperature over time. Second, we use BSI to reconstruct the initial condition of the lime from a measurement of its temperature later in time. We demonstrate the incorporation of prior information, visualize the posterior distributions of the parameter/initial condition, and show posterior samples of lime temperature trajectories from the model. Our Tutorial aims to reach a wide range of scientists and engineers.

© 2023 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/5.0154773>

I. INTRODUCTION

A. Mathematical models of physical systems

In science and engineering, we often apply physical principles to develop a mathematical model of a physical system in order to capture some phenomenon of interest. Physics-based models are useful for acquiring insights and understanding about a system, explaining observations, guiding future experiments, discovering new scientific questions, and making predictions.¹

Abstractly, a physics-based model of a system is typically an operator $f_\beta : x \mapsto y$ that (i) predicts the output y in response to any given input x and (ii) contains a vector β of physical parameters characterizing the system. See Fig. 1.

The input and output could be scalars, vectors, or functions. Evaluating the operator f_β at an input x could constitute evaluating an analytic expression, solving a system of algebraic equations,

evaluating an integral, solving a differential equation, or running a computer simulation.^{2,3} Note that the abstraction in Fig. 1 is not predicated on the system having mass and/or energy input/output; broadly, by definition, the input is merely a causal factor set in the beginning of an experiment on the system, and the output is an effect/result/consequence of the input.⁴

B. The forward problem

The *forward problem* is, given the model structure $f_\beta(x)$ and its parameters β , predict the output y of the physical system in response to a given input x —i.e., predict the effect of a cause.

The forward problem is solved by evaluating $f_\beta(x)$.

Forward problems tend to be (but are not always, e.g., chaotic dynamical systems⁵) well-posed and well-conditioned (i.e., they have a unique solution that is insensitive to error in the measurement of x).⁶

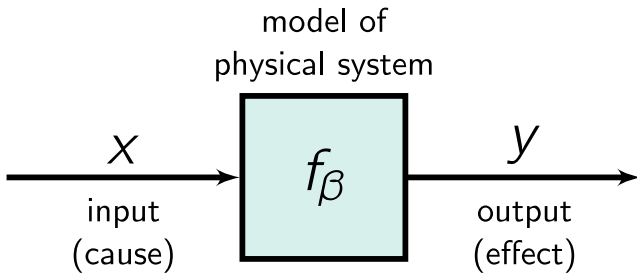


FIG. 1. A mathematical model $y = f_\beta(x)$, parameterized by β , predicts the output y of a physical system in response to an input x .

C. Inverse problems

Mathematicians find it difficult to define “inverse problem,” yet most recognize one when they see it.

Charles W. Groetsch³

Two categories of inverse problems^{2,3,7-12} are (see Fig. 2)

1. *parameter identification*: determine the parameters β characterizing a system that produced a set of observed input–output pairs $\{(x_i, y_i)\}_{i=1}^N$;
2. *reconstruction*: determine/reconstruct the input x' that produced an observed system output y' , given the model parameters β ; that is, predict the cause of an observed effect.

Inverse problems are ubiquitous in science and engineering. The development of a quantitatively predictive model often involves parameter identification. Reconstruction problems emerge when measurements we ideally wish to make in order to probe a system are infeasible, inaccessible, invasive, dangerous, and/or destructive, e.g., (i) sensing or imaging using indirect measurements (an elementary example: exploiting a model of the thermal expansion of mercury to infer the temperature of the air from a measurement of the volume of mercury in a mercury-in-glass thermometer^{13,14}), (ii) making inferences about the interior of a domain from measurements on its boundary, and (iii) time reversal: reconstructing

inverse problem	unknown	data	model structure	model parameter
parameter identification	β	$\{(x_i, y_i)\}_{i=1}^N$	$f_\beta(x)$	β
reconstruction	x'	y'	$f_\beta(x)$	β

(a)

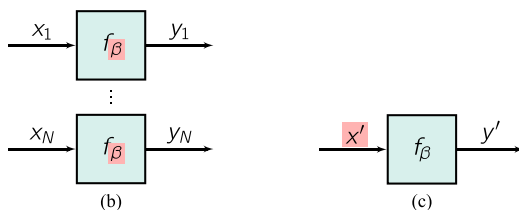


FIG. 2. Two categories of inverse problems: (b) parameter identification and (c) reconstruction. The unknown is highlighted red.

the past state of a system from a measurement of its current state.^{2,3}

A famous inverse problem, one of reconstruction, is “can one hear the shape of a drum?”¹⁵⁻¹⁸

• Can one hear the shape of a drum?

Suppose we strike the head of a drum with a stick at $t = 0$. The force induces transverse waves in the drumhead, which transmit to the air, producing longitudinal waves (sound).

Treating the drumhead as a homogeneous, elastic, two-dimensional membrane under tension, the wave equation is a mathematical model for the vertical displacement $u(x, t)$ of the vibrating membrane at a point $x \in \Omega$ and time $t \geq 0$,

$$\frac{\partial^2 u}{\partial t^2} = c \Delta_x u, \quad x \in \Omega, t \geq 0, \quad \text{wave equation,}$$

$$u|_{\partial\Omega} = 0, \quad \text{clamped boundary conditions.} \quad (1)$$

The parameter $c > 0$ is the ratio of the spatially uniform tension in the membrane to its density, Δ_x is the Laplace operator, and $\Omega \subset \mathbb{R}^2$ defines the geometry of the membrane ($\partial\Omega \subset \mathbb{R}^2$ is the boundary of Ω). Damping is neglected. Equation (1) is subject to an initial position and velocity set by the striking stick.¹⁹

The pure tones the drumhead can produce are dictated by standing wave solutions of the wave equation, $u(x, t) = e^{i\omega t} \phi(x)$. Substituting into Eq. (1), we find that the frequency ω of a pure tone of the drumhead is related to an eigenvalue of the Laplace operator on the domain Ω ,

$$c \Delta_x \phi = -\omega^2 \phi, \quad x \in \Omega, \quad \text{an eigenvalue problem,}$$

$$\phi|_{\partial\Omega} = 0, \quad \text{clamped boundary conditions.} \quad (2)$$

According to Eq. (2), the membrane, characterized by its physical property c and its geometry Ω , can produce a discrete spectrum of vibration frequencies $\omega_1 \geq \omega_2 \geq \dots$.¹⁵

The conventional *forward problem* is, given the physical property c and geometry Ω of the membrane, use Eq. (2) to determine the spectrum of vibration frequencies $\omega_1 \geq \omega_2 \geq \dots$ it can produce. In undergraduate mathematics courses, students solve the forward problem analytically for a rectangular or circular drum head.²⁰

An *inverse problem* is, given the physical property c of the membrane and its spectrum of vibration frequencies $\omega_1 \geq \omega_2 \geq \dots$, use Eq. (2) to determine its geometry Ω . Certainly, the set of pure tones a drum can produce contains information about the geometry of its membrane, but it is not obvious if these pure tones (and c) *uniquely determine* its geometry. This inverse problem, “can one hear the shape of a drum?,” was popularized by Kac in 1966.¹⁵ It was not until 1992 that Gordon, Webb, and Wolpert¹⁷ found two non-isometric domains with an identical spectrum of vibration frequencies, shown in Fig. 3. That is, one cannot necessarily hear the shape of a drum.¹⁸

Classically, a solution to an inverse problem is (conceptually) sought by adjusting the input x (for a reconstruction problem) or parameter β (for a parameter identification problem) until the model

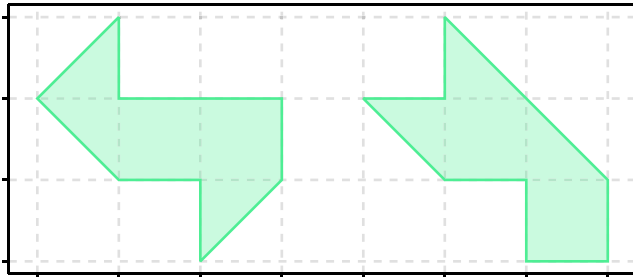


FIG. 3. Two non-isometric, but isospectral domains. Thus, the inverse problem of reconstructing the geometry of a membrane from its pure tones may not have a unique solution.

output(s) match(es) or (approximately) fit(s) the observed system output(s) to achieve (maximal) consistency between the model and the data,²¹ e.g., least squares.^{22,23}

1. Challenges in applied inverse problems

As opposed to theoretical, *exact* inverse problems, in *applied* inverse problems,²⁴ we must cope with noise contaminating the measurements of the system output. This noise may originate from (i) an imperfect measurement instrument and/or (ii) variance in the system output in response to a repeated input due to (a) inherent stochasticity and/or (b) unrecognized or poorly controlled, and thus unaccounted-for in the model, inputs/conditions that influence the output. In addition, the model $f_{\beta}(x)$ may inadequately approximate physical reality—perhaps, in part, owing to (ii-b).²¹

Applied inverse problems are particularly interesting and challenging because, in contrast to forward problems, they are often ill-posed and/or ill-conditioned.^{6,25}

- A solution to an applied inverse problem may not exist owing to noise in the measured output and model inadequacy:
 - in reconstruction problems, the observed output may fall outside the image of f_{β} so that no physically viable input is consistent with it;
 - in parameter identification, a parameter β giving a model $f_{\beta}(x)$ that exactly reproduces all observed input–output pairs may not exist.
- A solution to an applied inverse problem may not be unique:
 - in reconstruction problems, many different inputs may be consistent with the observed output owing to a many-to-one f_{β} ;
 - in parameter identification, (i) the number of observed input–output pairs may be insufficient to fully constrain the parameters or (ii) the model may be inherently unidentifiable (i.e., no amount of data can fully constrain the parameters).^{25–27}
- A solution to an applied inverse problem may discontinuously depend on or be sensitive to measurement noise contaminating the data. Even if the solution to the inverse

problem exists and is unique, noise contaminating measurements of the system output propagates onto and corrupts the solution to the inverse problem. In ill-conditioned inverse problems,¹⁴ small realizations of noise in the data produce large changes in the solution.

To tackle applied inverse problems, these challenges implore us to account for the noise contaminating measurements of the output, reframe our concept of a “solution,” and quantify our uncertainty in the solution.^{28,29}

D. Bayesian statistical inversion (BSI)

Bayesian statistical inversion (BSI)^{13,29–37} is a versatile framework to tackle (possibly) ill-posed and/or ill-conditioned applied inverse problems. BSI has two key advantages. First, BSI allows us to incorporate prior (i.e., before data are collected) information and/or beliefs about the input/parameter into our solution to the inverse problem. Second, BSI yields a probabilistic solution to inverse problems that quantifies uncertainty about the input/parameter.

Key references for BSI:^{13,38,39}

Modeling uncertainty. To model uncertainty about the unknown input/parameter, we treat it as a random variable and model its probability distribution. In this Bayesian view, the probability assigned to each region of input/parameter space reflects our degree of belief that the input/parameter falls in that region. This belief is based on some combination of subjective belief and objective information.^{40,41} Loosely speaking, the spread of the probability density over the input/parameter space reflects our uncertainty about its value, whereas sharp concentration of the density reflects certainty.

Prior vs posterior distributions. Our uncertainty about the input/parameter is different before and after we conduct an experiment, collect data, and compare these data with our model. Consequently, the input/parameter has a *prior* density before the data are considered and then an updated, *posterior* density after we are enlightened by the data.

Modeling the data-generating process. To acknowledge the unobservable noise that contaminates our measurements, we treat the outcomes of our measurements of the system output (i.e., the data) as realizations of random variables. A model of the stochastic data-generating process follows from combining (i) the mathematical model $f_{\beta}(x)$ of the system and (ii) a model of the probability distribution of the noise.

The two ingredients of BSI: the prior and likelihood. The two key ingredients for tackling an inverse problem with BSI are as follows:

1. The *prior distribution* of the input/parameter, expressing our beliefs about it before the data are collected/observed. The prior we construct depends on the amount of information we have about the input/parameter before the experiment, and it involves a degree of subjectivity and judgment.^{36,40} A prior may be roughly categorized as diffuse, weakly informative, or informative³⁹ based on the amount of uncertainty it admits about the input/parameter. An informative prior, e.g., a Gaussian distribution with a small variance, expresses a high degree of certainty about the input/parameter. On

the other hand, a diffuse prior, e.g., a uniform distribution, is more appropriate when we lack any prior information about the input/parameter. A diffuse prior adheres to the principle of indifference and spreads its density widely over input/parameter space to express a very high degree of uncertainty.³⁹ An informative prior influences the posterior more than a diffuse prior, which “lets the data speak for itself.”⁴² Generally, as we gather more data, the influence of the prior on the posterior tends to (but does not for non-identifiable systems²⁶) weaken/diminish as the data override/overwhelm the prior.

2. The *likelihood function* of the input/parameter, giving the probability density of the data conditioned on each value of the input/parameter. The likelihood function is constructed from (i) the data [our observation(s) of the system output(s), paired with inputs for the parameter identification problem] and (ii) our model of the data-generating process, which constitutes (a) the forward model f_β and (b) the probability distribution that models the noise contaminating the data. The likelihood quantifies the support that the data lend to each value of the unknown input/parameter.³⁹
 - From these two ingredients, we next “turn the Bayesian crank”⁴³ to obtain the posterior distribution of the unknown input/parameter.

The BSI solution to an inverse problem: the posterior. The BSI “solution” to the inverse problem, the *posterior distribution* of the unknown input/parameter, follows from the prior density and likelihood function of the input/parameter via Bayes’ theorem. The posterior density of the unknown input/parameter gives the probability that the input/parameter falls in any given region of input/parameter space, conditioned on the data. The posterior updates the prior in light of the data, offers a compromise between the prior and likelihood, and constitutes the raw solution to the inverse problem that quantifies uncertainty about the input/parameter through its degree of spread.²⁸

For some combinations of families of likelihood and prior probability distributions [e.g., a Gaussian (conjugate) prior for a Gaussian likelihood density], we obtain a closed-form expression for the posterior distribution. Otherwise, we resort to approximating the posterior distribution (i) as a tractable analytical distribution [e.g., as a Gaussian via the Laplace (quadratic) approximation or some other parameterized distribution via variational inference] or (ii) as an empirical distribution constructed using samples from the posterior obtained from a Markov Chain Monte Carlo (MCMC) method.^{43,44}

We may summarize the posterior distribution of the input/parameter with a *credible region* (e.g., a highest-density region⁴⁵) that contains some large fraction α of the posterior density and thus credibly contains—given the assumptions in the likelihood and prior hold—the input/parameter (precisely, with probability α).⁴⁶ Note that a Bayesian credible interval for the parameter/input is distinct from and arguably more intuitive and natural than a frequentist confidence interval.⁴⁷

We next present an optional warm-up problem that gives an interpretable, closed-form posterior to illustrate and gain intuition about Bayesian inference.

• **Bayesian inference of the pH of a vat of wine**

Suppose we work at a winery and wish to estimate the average pH, x , of Pinot Noir in a large, stirred vat. We possess a noisy pH-meter for the task.⁴⁸ We adopt the Bayesian approach and view the pH as a random variable X to model our uncertainty about it.

Prior. First, we impose a Gaussian prior distribution on X based on the history of pH measurements from previous batches of Pinot Noir produced by our winery over many years,

$$\pi_{\text{prior}}(x) = \frac{1}{\sigma_{\text{pr}}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x - x_{\text{pr}}}{\sigma_{\text{pr}}}\right)^2\right], \quad (3)$$

with x_{pr} and σ_{pr}^2 being the mean and variance of average pH among previous batches. This prior reflects our assumption that we followed closely the fermentation protocols from previous years, sourced our grapes from the same vineyard, employed the same yeast strain for fermentation, etc.

Probabilistic data-generating model. We treat the extraction of a sample of wine from the vat and subsequent measurement of its pH, giving an observed pH x_{obs} , as a stochastic process. Particularly, we treat x_{obs} as a realization of an random variable,

$$X_{\text{obs}} \mid (X = x) = x + E, \quad (4)$$

where $E \sim \mathcal{N}(0, \sigma^2)$ is a normally distributed random variable whose variance σ^2 models the imprecision of the pH-meter and variation in pH among samples from the vat. Assume that σ is known from previously observed variance in pH measurements of samples from the same vat (but different batch).

Data. To gather information about the pH X of the vat of the wine, we take n samples of wine from the vat and measure the pH of each sample with our noisy pH-meter, giving data $\{x_{\text{obs},1}, \dots, x_{\text{obs},n}\}$.

Likelihood function. Under our probabilistic data-generating model in Eq. (4), the *likelihood density*—the probability density of the data given the pH of the wine in the vat is x (and, implicitly, given σ)—is, assuming independent measurements,

$$\begin{aligned} \pi_{\text{likelihood}}(x_{\text{obs},1}, \dots, x_{\text{obs},n} \mid x) &= \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x - x_{\text{obs},i}}{\sigma}\right)^2\right]. \end{aligned} \quad (5)$$

Since we collected the data, we now see $\pi_{\text{likelihood}}$ as a function of x , indicating the support the data lend for each value of x . Then, with some algebra (including completing the square), the likelihood function

$$\pi_{\text{likelihood}}(x_{\text{obs},1}, \dots, x_{\text{obs},n} \mid x) \propto \exp\left[-\frac{1}{2}\left(\frac{x - x_{\text{obs}}}{\sigma/\sqrt{n}}\right)^2\right], \quad (6)$$

where $\bar{x}_{\text{obs}} := n^{-1} \sum_{i=1}^n x_{\text{obs},i}$ is the mean pH among the samples. The likelihood function diminishes as x deviates from the sample mean \bar{x}_{obs} , and it diminishes more rapidly as n increases.

Posterior. The posterior, the Bayesian solution to this task at the winery, is the probability density of the pH of Pinot Noir in the vat X , given the data $\{x_{\text{obs},1}, \dots, x_{\text{obs},n}\}$ (and, implicitly, σ). It follows from Bayes theorem that

$$\pi_{\text{posterior}}(x \mid x_{\text{obs},1}, \dots, x_{\text{obs},n}) = \frac{\pi_{\text{likelihood}}(x_{\text{obs},1}, \dots, x_{\text{obs},n} \mid x) \pi_{\text{prior}}(x)}{\int_{-\infty}^{\infty} \pi_{\text{likelihood}}(x_{\text{obs},1}, \dots, x_{\text{obs},n} \mid x') \pi_{\text{prior}}(x') dx'}. \quad (7)$$

After some algebra (including completing the square; the denominator is a constant we do not need to handle), we find the posterior is also Gaussian,

$$\pi_{\text{posterior}}(x \mid x_{\text{obs},1}, \dots, x_{\text{obs},n}) = \frac{1}{\sigma_{\text{post}} \sqrt{2\pi}} \times \exp \left[-\frac{1}{2} \left(\frac{x - x_{\text{post}}}{\sigma_{\text{post}}} \right)^2 \right], \quad (8)$$

with an inverse variance, a measure of *concentration* of the Gaussian, of

$$\frac{1}{\sigma_{\text{post}}^2} := \frac{1}{\sigma_{\text{pr}}^2} + n \frac{1}{\sigma^2} \quad (9)$$

and mean

$$x_{\text{post}} := \left(\frac{n/\sigma^2}{n/\sigma^2 + 1/\sigma_{\text{pr}}^2} \right) \bar{x}_{\text{obs}} + \left(1 - \frac{n/\sigma^2}{n/\sigma^2 + 1/\sigma_{\text{pr}}^2} \right) x_{\text{pr}}. \quad (10)$$

Intuitively: (1) The concentration of the posterior Gaussian increases (i) as we collect more data (n increases) and (ii) if we use a more precise pH-meter (small σ), reflecting a reduction in our uncertainty about the pH. (2) The mean of the posterior is a weighted average of the sample mean pH \bar{x}_{post} and the prior mean pH x_{pr} . With more data and a more precise pH-meter, we weigh the sample mean more than the prior mean. A more informative prior (smaller σ_{pr}) means we need more data (larger n) to override the prior.⁴⁹⁻⁵¹

Example. Figure 4 illustrates how (top) the diffuseness of the prior influences the posterior and (bottom) how more data override the prior and reduce posterior uncertainty.

Conjugate priors. Note that the posterior distribution in Eq. (10) belongs to the same family (Gaussian) as the prior distribution. Hence, the Gaussian prior is a *conjugate prior* for the Gaussian likelihood distribution in Eq. (5). Generally, a conjugate prior can simplify Bayesian inference by giving a closed-form expression for the posterior.⁴⁶

E. Our contribution: a tutorial of BSI, illustrated on inverse problems of heat transfer

We provide a tutorial of BSI to solve inverse problems of both parameter identification and reconstruction while incorporating

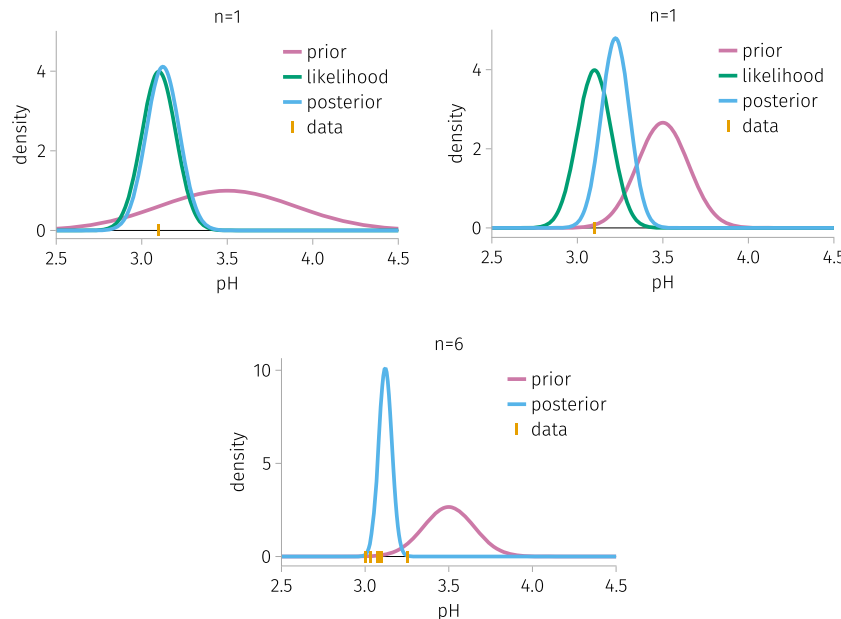


FIG. 4. Illustrating Bayesian inference of the pH of a vat of wine, X . Suppose $\sigma = 0.1$ and $x_{\text{pr}} = 3.5$. (top) Furthermore, suppose $n = 1$ and $x_{\text{obs}} = 3.1$. The prior distribution, likelihood function, and posterior distribution of the pH are shown for (left) a more diffuse prior ($\sigma_{\text{pr}} = 0.4$) and (right) a more concentrated prior ($\sigma_{\text{pr}} = 0.15$). The more concentrated prior pulls the posterior mean toward it more than the more diffuse prior. (bottom) For $n = 6$ measurements, the posterior becomes more concentrated as we reduce our uncertainty about the pH, and the more concentrated prior has a less dramatic effect on the posterior mean (likelihood omitted due to its scale).

prior information and quantifying uncertainty. Our Tutorial is by way of examples regarding a simple, intuitive physical process: convective heat transfer from ambient air to a cold lime fruit in our kitchen. We aim to engage a wide range of scientists and engineers.

In Sec. II A, we describe our experimental setup: (i) we take a cold lime fruit out of a refrigerator and allow it to exchange heat with indoor air via natural convection; (ii) a probe inserted into the lime measures its internal temperature. In Sec. II B, we pose a mathematical model governing the time-evolution of the lime temperature based on Newton's "law" of cooling. The model contains a single parameter. To describe the data-generating process, in Sec. II C, we augment this model with a probabilistic model of taking a noisy measurement of the lime temperature with the imperfect temperature probe. Next, in Sec. II D, we employ BSI to infer the parameter in the dynamic model of the lime temperature using time series data of the lime temperature (an overdetermined inverse problem). We then employ the model of the lime temperature with the inferred parameter to tackle two time reversal problems: reconstruct the initial temperature of the lime, given a measurement of its current temperature and the time it has been outside of the refrigerator (Sec. II E, ill-conditioned), and reconstruct the initial temperature of the lime and the duration it has been out of the refrigerator, given a measurement of its current temperature (Sec. II F, underdetermined). The solution to each inverse problem under BSI expresses uncertainty through the posterior probability density of the parameter/initial condition. We assess the fidelity of the solution to the reconstruction problems by comparing with the measured initial condition held out from the BSI procedure. From these concrete, instructive examples, we intend for readers to recognize inverse problems in their research domain where BSI may be employed to incorporate prior information and quantify uncertainty.

II. A TUTORIAL OF BAYESIAN STATISTICAL INVERSION (BSI)

As a tutorial of BSI to solve inverse problems while incorporating prior information and quantifying uncertainty, we tackle a variety of inverse problems pertaining to heat exchange between ambient air and a lime fruit via natural convection.

A. Experimental setup: heat exchange between a lime and the air

We allowed a lime fruit (~5 cm diameter) to reside in a refrigerator for several days. Then, at time $t := t_0$ (min), we removed the lime from the refrigerator, placed it on a thin slab of insulating styrofoam, and allowed it to exchange heat with the indoor air via natural convection. An electrical-resistance-based temperature sensor inserted into the lime allows us to measure its internal temperature at any given time t to generate a data point (t, θ_{obs}) ("obs" = observed). [To avoid the early-time data reflecting the (short) time dynamics of the temperature probe coming into thermal equilibrium with the lime, owing to its nonzero thermal mass, we inserted the probe into the lime *before* we placed it in the refrigerator, so it begins thermally equilibrated with the lime.] Figure 5 shows our experimental setup, and Sec. S1 explains our temperature sensor and Arduino microcontroller setup.

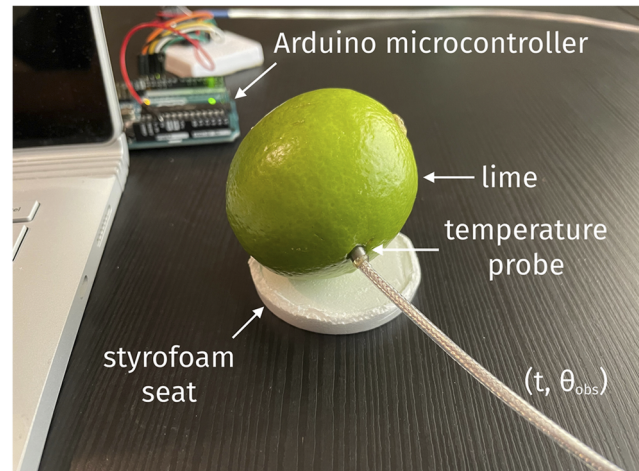


FIG. 5. Experimental setup. An initially cold lime fruit rests on a styrofoam slab and exchanges heat with the indoor air via natural convection. A temperature sensor inserted into the lime can measure its internal temperature at any given time t , producing a data point (t, θ_{obs}) .

B. The mathematical model of the lime temperature

We develop a mathematical model governing $\theta = \theta(t)$ ($^{\circ}\text{C}$), the temperature of the lime as a function of time $t \geq t_0$ (min), as it exchanges heat with the bath of ambient air at bulk (i.e., far from the lime) temperature θ^{air} ($^{\circ}\text{C}$). The mechanism of lime-air heat transfer is natural convection, which is the combined effect⁵² of both heat (i) conduction and (ii) advection, i.e., via motion of the air adjacent to the lime, driven by buoyant forces arising from spatial density-gradients in the air, caused by temperature-gradients.⁵³ The initial temperature of the lime is $\theta(t_0) =: \theta_0$.

Assumptions. We make several simplifying assumptions:

- The temperature of the lime is spatially uniform. [We estimate the Biot number⁵² for this system $\text{Bi} := hr/k \approx 0.6$ based on (i) our measurement of the radius of the lime $r \approx 2.5$ cm, (ii) the reported thermal conductivity of a lime $k \approx 0.595$ J/(s m $^{\circ}\text{C}$),⁵⁴ and (iii) an estimated heat transfer coefficient for natural convection through a gas, $h \approx 15$ J/(s m² $^{\circ}\text{C}$).^{52,55}]
- The bulk temperature of the air (a bath) θ^{air} is constant (sufficiently far from the lime).
- Heat conduction between the lime and the styrofoam surface on which it sits is negligible.
- The mass of the lime is constant (e.g., a negligible loss of moisture over time).
- The heat capacity of the lime is constant with temperature.
- Heat released/consumed due to chemical reactions (e.g., oxidation) is negligible.
- The temperature probe inserted into the lime has a negligible thermal mass.
- The rate of heat exchange between the air and the lime is governed by Newton's "law" of cooling.

Newton’s “law” of cooling^{56–59} prescribes the rate of heat transfer (J/min) from the air to the lime at time $t \geq t_0$ as proportional to the difference in temperature between the lime and the air (the thermodynamic driving force for heat transfer) at that time,

$$\text{rate of heat transfer from air to lime} \propto \theta^{\text{air}} - \theta(t). \quad (11)$$

The precise rate, $hA[\theta^{\text{air}} - \theta(t)]$, depends on two (assumed, constant) parameters: (i) the surface area of the lime in contact with the air A (cm²) and (ii) the (natural) convective heat transfer coefficient, h [J/(°C min cm²)], between the air and the surface of the lime.⁵²

Differential equation model. Conservation of energy applied to the lime gives its change in temperature $d\theta = d\theta(t)$ over a differential change in time dt ,

$$Cd\theta(t) = hA[\theta^{\text{air}} - \theta(t)]dt \quad \text{for } t \geq t_0, \quad (12)$$

with C (J/°C) the thermal mass of the lime. Equation (12) balances the amount (J) of sensible heat stored in the lime (left) and amount (J) of heat transferred to the lime (right) over the time interval $[t, t + dt)$. Thus, a first-order differential equation describes the time-evolution of the lime temperature,^{56,57}

$$\begin{aligned} \lambda \frac{d\theta}{dt} &= \theta^{\text{air}} - \theta(t) \quad \text{for } t \geq t_0, \\ \theta(t = t_0) &= \theta_0. \end{aligned} \quad (13)$$

The single, lumped parameter $\lambda := C/(hA)$ (min) of the model is a time constant governing the dynamics of heat transfer between the lime and the air.

Analytical solution to the model. Equation (13) admits an analytical solution through a variable transformation and then integration,

$$\theta(t) = \theta^{\text{air}} + (\theta_0 - \theta^{\text{air}})e^{-(t-t_0)/\lambda} \quad \text{for } t \geq t_0. \quad (14)$$

Starting at θ_0 , the lime temperature $\theta(t)$ monotonically approaches the air temperature $[\lim_{t \rightarrow \infty} \theta(t) = \theta^{\text{air}}]$ as the temperature

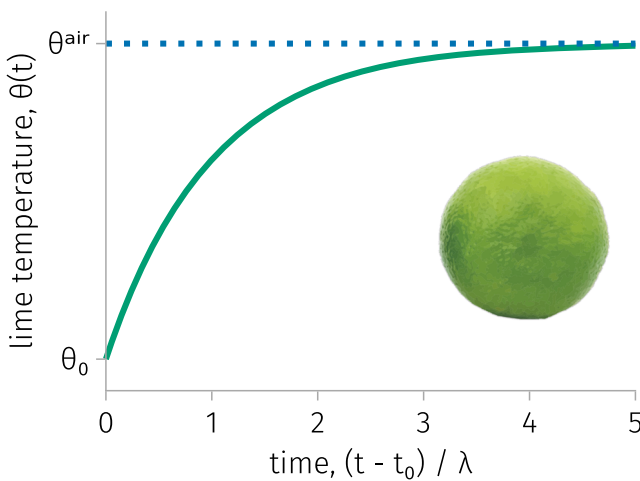


FIG. 6. The solution to the mathematical model of the lime temperature, $\theta(t; \lambda, t_0, \theta_0, \theta^{\text{air}})$.

difference between the lime and air decays exponentially. The parameter λ is a time scale for the lime to reach thermal equilibrium with the air. Specifically, after a duration $t - t_0 = \lambda$ out of the refrigerator, the difference between the air temperature and lime temperature is $e^{-1} \approx 37\%$ of the initial difference.

Figure 6 shows the solution to the model of the lime temperature, which we write as $\theta(t; \lambda, t_0, \theta_0, \theta^{\text{air}})$ to emphasize its dependence on the parameter λ , the initial condition (t_0, θ_0) , and the air temperature θ^{air} .

C. The probabilistic model of the data-generating process

Consider the process of employing our temperature probe, an imperfect measurement instrument, to measure the lime temperature at time t , giving a data point (t, θ_{obs}) .

We treat the measured temperature θ_{obs} as a realization of a random variable Θ_{obs} owing to two sources of stochasticity. First, *measurement noise*: unobservable noise originating from the temperature probe corrupts the measurement. Second, *residual variability*: under repeated experiments with identical conditions $(t_0, \theta_0, \theta^{\text{air}})$, the observed lime temperature at time t may exhibit variance due to variable conditions/inputs that are poorly controlled or unrecognized and thus unaccounted for in our model for $\theta(t)$.²¹ For example, (i) the air temperature θ^{air} is not perfectly controlled and may fluctuate over time and (ii) the opening and closing of doors in the building may introduce fluctuating air currents in the room, making the heat transfer coefficient h fluctuate over time.

We model the noise in the observed lime temperature as a random variable E_σ additive to the model prediction, independent among repeated measurements, and having an identical distribution over time. Then, our probabilistic model of the data-generating process is

$$\Theta_{\text{obs}} = \theta(t; \lambda, t_0, \theta_0, \theta^{\text{air}}) + E_\sigma, \quad (15)$$

where E_σ is a zero-centered Gaussian with variance σ^2 ,

$$E_\sigma \sim \mathcal{N}(0, \sigma^2). \quad (16)$$

Equation (15) assumes the time scale for the temperature probe to thermally equilibrate with the lime is negligibly small, avoiding a time delay.

The random variable E_σ can capture noise emanating from both the measurement instrument and residual variability. However, Eq. (15) assumes that the mean measured lime temperature at time t over repeated experiments with the same conditions $(t_0, \theta_0, \theta^{\text{air}})$ is given by the model $\theta(t; \lambda, t_0, \theta_0, \theta^{\text{air}})$. That is, our data-generating model neglects the possibility of *model discrepancy* (with physical reality), a nonzero difference between (a) the expected measured lime temperature at time t over multiple experiments with the same conditions $(t_0, \theta_0, \theta^{\text{air}})$ and (b) the model prediction of the lime temperature, $\theta(t; \lambda, t_0, \theta_0, \theta^{\text{air}})$.²¹

Under Eq. (15), the probability density function governing the distribution of the measured lime temperature Θ_{obs} at time t , given (i.e., conditioned on) $\lambda, t_0, \theta_0, \theta^{\text{air}}, \sigma$, is obtained by translating the density of the noise in Eq. (16) to center it at the model prediction,

$$\begin{aligned} \pi_{\text{likelihood}}(\theta_{\text{obs}} | \lambda, t_0, \theta_0, \theta^{\text{air}}, \sigma) \\ = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\theta_{\text{obs}} - \theta(t; \lambda, t_0, \theta_0, \theta^{\text{air}})}{\sigma}\right)^2\right]. \end{aligned} \quad (17)$$

The likelihood regards a temperature measurement far from the model prediction, conditioned on knowing the parameters/conditions, as unlikely.

D. Inverse problem I: Parameter identification

Overview of problem and approach

Task: employ BSI to infer the parameter Λ (i.e., λ treated as a random variable) characterizing the lime, appearing in the model of the lime temperature in Eq. (14), using data from a heat transfer experiment.

First, we estimate λ with a back-of-the-envelope calculation and use this estimate to construct a weakly informative prior distribution of Λ .

To admit our uncertainty about the variance σ^2 of the measurement noise in our model of the data-generating process in Eq. (15), we also treat it as a random variable Σ^2 to also be inferred from the data. We impose a diffuse prior distribution on Σ^2 , granted support based on the noise characteristics of the temperature probe.

Next, we set up a heat transfer experiment with the lime (see Sec. II A). We take two measurements of the conditions of the experiment: (i) the initial temperature of the lime, giving datum $(t_0 = 0, \theta_{0,\text{obs}})$, and (ii) the air temperature, giving datum $\theta_{\text{obs}}^{\text{air}}$. To admit these are noisy measurements, we treat the initial temperature of the lime Θ_0 and air temperature Θ^{air} as random variables to be inferred from the data, but impose informative prior distributions on them based on these measurements.

We then measure the lime temperature at different times over the course of the heat transfer experiment, giving time series data $\{(t_i, \theta_{i,\text{obs}})\}_{i=1}^N$ that provide information about Λ .

Finally, we use Bayes' theorem to construct the posterior distribution of $(\Lambda, \Theta_0, \Theta^{\text{air}}, \Sigma)$ in light of the data $\{(t_i, \theta_{i,\text{obs}})\}_{i=1}^N$, sample from it using a Markov Chain Monte Carlo algorithm, and obtain a credible interval for the parameter Λ that quantifies posterior uncertainty about its value.

Quick overview:

- *measurements of the experimental conditions:* the measured initial ($t = t_0 = 0$) temperature of the lime $\theta_{0,\text{obs}}$ and air temperature $\theta_{\text{obs}}^{\text{air}}$;
- *data collected during the experiment:* the lime temperature time series data $\{(t_i, \theta_{i,\text{obs}})\}_{i=1}^N$;
- *random variables to infer from the data:* the parameter Λ , the initial lime temperature Θ_0 , the air temperature Θ^{air} , and the variance of the measurement noise Σ^2 ;
- *sources of priors:* Λ : back-of-the-envelope calculation, $\Theta_0, \Theta^{\text{air}}$: our noisy measurements of them, and Σ^2 : precision of temperature sensor.

Summary of results: See Fig. 7.

Classically, this inverse problem is overdetermined because a parameter λ giving a model $\theta(t; \lambda, t_0, \theta_0, \theta^{\text{air}})$ that passes through all data points $\{(t_i, \theta_{i,\text{obs}})\}_{i=1}^N$ does not exist. [Note that we do not

attempt to determine each C , h , and A in Eq. (12) from the data $\{(t_i, \theta_{i,\text{obs}})\}_{i=1}^N$ but rather the lumped parameter $\lambda := C/(hA)$. The former attempt would make the system unidentifiable because multiple distinct (C, h, A) give the same λ and hence produce the same model $\theta(t)$.]

1. The experimental setup

To characterize the setup of the lime heat transfer experiment, we use the temperature probe to measure the air temperature, giving datum $\theta_{\text{obs}}^{\text{air}}$, and the initial lime temperature at $t = t_0 := 0$, giving datum $\theta_{0,\text{obs}}$. These data are plotted in Fig. 7(a) as the horizontal dashed line and first point, respectively.

2. The prior distributions

Before the data $\{(t_i, \theta_{i,\text{obs}})\}_{i=1}^N$ are considered, in BSI, we must express our prior beliefs and information about the value of each variable via a prior probability density function.

The parameter, Λ . A back-of-the-envelope estimate of $\lambda = C/(hA)$ is ≈ 1 h based on the following. The diameter of the lime, approximated as a sphere, is ~ 5 cm. The mass of the lime is ~ 100 g. The specific heat of a lime is reported as ~ 4.0 kJ/(kg °C).^{54,60} A typical heat transfer coefficient h for natural convection via gas is 15 J/(s m² °C).^{52,55}

We specify a weakly informative prior density $\pi_{\text{prior}}(\lambda)$ as that of a Gaussian distribution (i) centered at our back-of-the-envelope estimate of λ , (ii) with a high variance to reflect our low confidence in this rough estimate, and (iii) truncated below zero to enforce positivity,

$$\Lambda \sim \mathcal{N}_{>0}(1\text{h}, (0.3\text{h})^2). \quad (18)$$

See Fig. 7(b).

The experimental conditions, Θ_0 and Θ^{air} . We impose informative prior distributions on the initial lime temperature Θ_0 and air temperature Θ^{air} based on our (noisy) measurements of them,

$$\Theta_0 \sim \mathcal{N}(\theta_{0,\text{obs}} = 6.54^\circ\text{C}, \sigma^2), \quad (19)$$

$$\Theta^{\text{air}} \sim \mathcal{N}(\theta_{\text{obs}}^{\text{air}} = 18.47^\circ\text{C}, \sigma^2). \quad (20)$$

The variance of the measurement noise, Σ^2 . Our prior distribution of the standard deviation of the measurement noise, reflecting our beliefs about the precision of the temperature probe, is

$$\Sigma \sim \mathcal{U}([0^\circ\text{C}, 1^\circ\text{C}]), \quad (21)$$

where $\mathcal{U}(\cdot)$ is a uniform distribution over the interval \cdot .

The joint prior distribution. The joint prior distribution of all of the unknowns for this inverse problem factorizes since we imposed independent priors, corresponding to plausible assumptions, including that the parameter λ of the lime has no causal link with the air temperature,

$$\pi_{\text{prior}}(\lambda, \theta_0, \theta^{\text{air}}, \sigma) = \pi_{\text{prior}}(\lambda)\pi_{\text{prior}}(\theta_0)\pi_{\text{prior}}(\theta^{\text{air}})\pi_{\text{prior}}(\sigma). \quad (22)$$

The prior distribution $\pi_{\text{prior}}(\lambda, \theta_0, \theta^{\text{air}}, \sigma)$ summarizes the information and beliefs we have about the unknowns $(\lambda, \theta_0, \theta^{\text{air}}, \sigma)$ at this stage, before the data $\{(t_i, \theta_{i,\text{obs}})\}_{i=1}^N$ are considered.

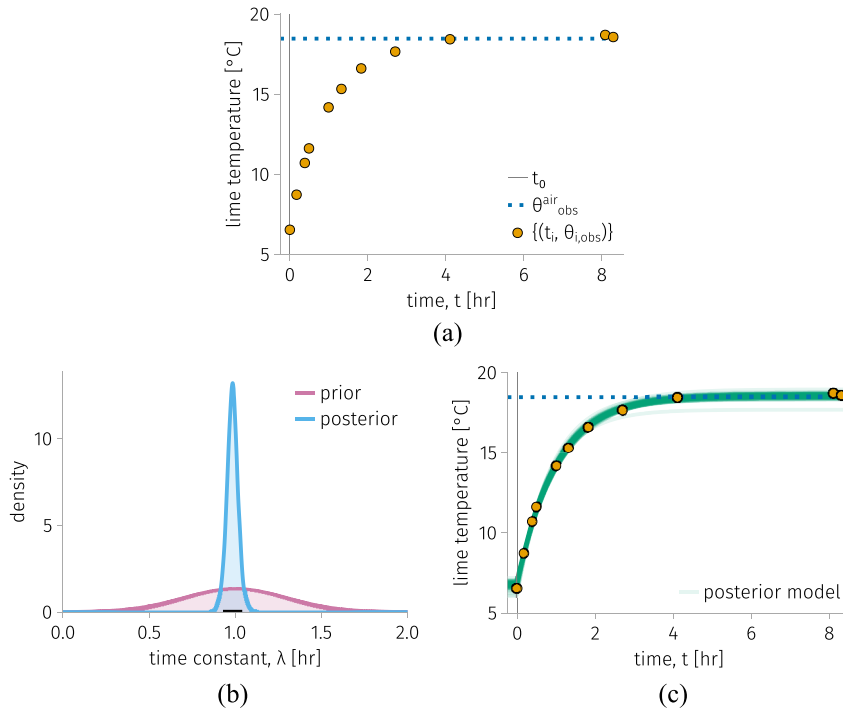


FIG. 7. Inverse problem I: parameter identification: infer the parameter Λ in the model of the lime temperature, $\theta(t; \lambda, t_0, \theta_0, \theta^{\text{air}})$, given measurements of the conditions of the experiment, $(t_0 = 0, \theta_{0,\text{obs}})$ and $\theta^{\text{air}}_{\text{obs}}$, and time series data of the lime temperature $\{(t_i, \theta_{i,\text{obs}})\}_{i=1}^N$. (a) The measurements/data. (b) The prior and posterior distributions of Λ . The black bar shows an equal-tailed, 90% posterior credible interval for Λ . (c) A sample of 100 model trajectories $\theta(t; \lambda, t_0 = 0, \theta_0, \theta^{\text{air}})$, with $(\lambda, \theta_0, \theta^{\text{air}})$ sampled from the posterior.

3. The data and likelihood function

The data. We measure and consider the lime temperature over the course of the heat transfer experiment, giving the time series data $\{(t_i, \theta_{i,\text{obs}})\}_{i=1}^N$ displayed in Fig. 7(a).

The likelihood function. The likelihood function gives the probability density of the data $\{(t_i, \theta_{i,\text{obs}})\}_{i=1}^N$ conditioned on each possible value of the parameters $\Lambda = \lambda$ and $\Sigma = \sigma$ and experimental conditions $\Theta_0 = \theta_0$ and $\Theta^{\text{air}} = \theta^{\text{air}}$. We construct the likelihood from the (i) data $\{(t_i, \theta_{i,\text{obs}})\}_{i=1}^N$ and (ii) model of the data-generating process in Eq. (17). The likelihood function is

$$\begin{aligned} \pi_{\text{likelihood}}(\theta_{1,\text{obs}}, \dots, \theta_{N,\text{obs}} \mid \lambda, \theta_0, \theta^{\text{air}}, \sigma) \\ = \prod_{i=1}^N \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\theta_{i,\text{obs}} - \theta(t_i; \lambda, t_0 = 0, \theta_0, \theta^{\text{air}})}{\sigma}\right)^2\right]. \end{aligned} \tag{23}$$

The likelihood factorizes because we model the measurement noise E_σ as an independent random variable. Note, inherently, that the likelihood is conditioned on the model structure [i.e., functional form/shape for $\theta(t)$ in Eq. (14)] as well.

Since we possess the data $\{(t_i, \theta_{i,\text{obs}})\}_{i=1}^N$ at this stage, we (i) view the likelihood as a function of the unknowns $(\lambda, \theta_0, \theta^{\text{air}}, \sigma)$ and (ii) interpret it as a measure of the support the data $\{(t_i, \theta_{i,\text{obs}})\}_{i=1}^N$ lend to each value of the unknowns, $(\lambda, \theta_0, \theta^{\text{air}}, \sigma)$.³⁹

4. The posterior distribution

The (joint) *posterior density* governs the probability distribution of the unknowns $(\Lambda, \Theta_0, \Theta^{\text{air}}, \Sigma)$ conditioned on the time series data $\{(t_i, \theta_{i,\text{obs}})\}_{i=1}^N$. By Bayes’ theorem,⁴⁶ the posterior density is proportional to the product of the likelihood function and prior density,

$$\begin{aligned} \pi_{\text{posterior}}(\lambda, \theta_0, \theta^{\text{air}}, \sigma \mid \theta_{1,\text{obs}}, \dots, \theta_{N,\text{obs}}) \\ \propto \pi_{\text{likelihood}}(\theta_{1,\text{obs}}, \dots, \theta_{N,\text{obs}} \mid \lambda, \theta_0, \theta^{\text{air}}, \sigma) \\ \times \pi_{\text{prior}}(\lambda, \theta_0, \theta^{\text{air}}, \sigma). \end{aligned} \tag{24}$$

Note, because we will employ a Markov chain Monte Carlo algorithm to sample from the posterior and approximate it empirically, we do not need to know the normalizing factor; these samplers only require *ratios* of posterior densities.

We are particularly interested in the posterior distribution of the parameter Λ , with $(\Theta_0, \Theta^{\text{air}}, \Sigma)$ marginalized out.

The posterior density of the unknowns $\pi_{\text{posterior}}(\lambda, \theta_0, \theta^{\text{air}}, \sigma \mid \theta_{1,\text{obs}}, \dots, \theta_{N,\text{obs}})$

- is the raw solution to this parameter identification problem as it assigns a probability to each region of $(\lambda, \theta_0, \theta^{\text{air}}, \sigma)$ -space to reflect our posterior degree of belief that the unknowns $(\lambda, \theta_0, \theta^{\text{air}}, \sigma)$ fall in that region;

07 May 2026 18:12:23

- represents an *update* to our prior density $\pi_{\text{prior}}(\lambda, \theta_0, \theta^{\text{air}}, \sigma)$, in light of the data $\{(t_i, \theta_{i,\text{obs}})\}_{i=1}^N$; and
- reflects a compromise between (i) our prior knowledge and beliefs about $(\lambda, \theta_0, \theta^{\text{air}}, \sigma)$ and (ii) the support the data $\{(t_i, \theta_{i,\text{obs}})\}_{i=1}^N$ lend to $(\lambda, \theta_0, \theta^{\text{air}}, \sigma)$, according to our model of the data-generating process.

Remark on sources of posterior uncertainty about Λ . Uncertainty about Λ remains even after the data $\{(t_i, \theta_{i,\text{obs}})\}_{i=1}^N$ are considered. Sources of this posterior uncertainty are a lack of data (small N) coupled with two sources of noise captured through our model of the data-generating process in Eq. (15):²¹ (i) measurements of the lime temperature being corrupted by (unobservable) noise from using an imperfect temperature sensor and (ii) unrecognized and/or poorly controlled inputs/conditions that influence the lime temperature *over the course of the experiment* and result in white noise. However, our posterior distribution does *not* capture uncertainty due to (i) residual variability: unrecognized and/or poorly controlled inputs/conditions that influence the lime temperature and vary *from experiment-to-experiment* (because we infer Λ using data from only a single experiment), or (ii) model inadequacy: when $\theta(t)$ does not faithfully predict the expected [over many experiments under the same conditions $(t_0, \theta_0, \theta^{\text{air}})$] measured lime temperature [perhaps, in part, owing to (i)], which would introduce bias and violate the white noise assumption in Eq. (15).

5. Sampling from the posterior

We employ a *Markov Chain Monte Carlo* (MCMC) algorithm, the No-U-Turn Sampler (NUTS)⁶¹ implemented in `Turing.jl`⁶² in Julia,⁶³ to obtain samples from the joint posterior distribution in Eq. (24) in order to (i) approximate it with an empirical distribution using kernel density estimation and (ii) compute means and credible intervals of the unknowns. Over four independent chains, we collect 2500 samples/chain, with the first half discarded for warm-up.

- **Markov chain Monte Carlo (MCMC) sampling from the posterior distribution**

The computational challenge of computing the posterior. In a general setting for parameter inference via BSI, let

- $U \in \mathbb{R}^K$ be the random vector of the K unknown parameters in the inverse problem;
- $d \in \mathbb{R}^N$ be the data vector of noisy measurements/observations.

The posterior density of U follows from Bayes' theorem,⁴⁶

$$\pi(u) := \pi_{\text{posterior}}(u | d) = \frac{\pi_{\text{likelihood}}(d | u)\pi_{\text{prior}}(u)}{\pi_{\text{evidence}}(d)}. \quad (25)$$

The denominator, *the evidence*, is the probability density of the data—a constant that serves as a normalizing factor for the numerator,

$$\pi_{\text{evidence}}(d) = \int \pi_{\text{likelihood}}(d | u)\pi_{\text{prior}}(u)du. \quad (26)$$

Typically, we cannot analytically evaluate this K -dimensional integral. If K is large, it may be intractable to use numerical cubature to approximate $\pi_{\text{evidence}}(d)$ as well.

The same difficulty may arise for the integral to (i) compute the mean of the posterior or (ii) marginalize out a subset of the unknowns we are less concerned with.⁴³

Markov chain Monte Carlo (MCMC) simulation. MCMC methods⁴⁴ permit us to obtain samples u_1, \dots, u_n from the posterior density $\pi(u)$ with only access to evaluations of a function to which the posterior density is *proportional* [as in Eq. (24)]. This circumvents the need to compute $\pi_{\text{evidence}}(d)$. From the samples u_1, \dots, u_n , we can (i) construct an empirical posterior distribution using kernel density estimation⁶⁴ and (ii) approximate (a) the mean of the posterior from the mean of the samples and (b) an equal-tailed credible interval of any unknown from the percentiles of its samples. The approximation to the posterior improves as more samples are collected. Note that we can construct the empirical *marginal* posterior distribution, of a subset of unknowns, trivially by ignoring the remaining dimensions of the sampled vectors u_1, \dots, u_n . The idea behind an MCMC method is to (i) construct a Markov chain U_1, U_2, \dots whose (a) state space is the parameter space and (b) transition kernel $\pi_t(u'|u)$ governing the probability of transitioning from one state u to another u' endows the chain with a stationary distribution equal to the posterior distribution $\pi(u)$ and then (ii) simulate the Markov chain to obtain a realization u_1, \dots, u_n , regarded as (autocorrelated) samples from $\pi(u)$.^{39,43,65,66}

For example, random walk Metropolis. Perhaps the simplest MCMC simulation algorithm to understand is random walk Metropolis.^{43,65} Here, a realization u_1, \dots, u_n of a Markov chain U_1, \dots, U_n is obtained by iterating a stochastic process of “propose then accept-or-reject” n times. We initiate the walk at some state u_1 in a region with reasonably high posterior density. Within the iterations, suppose u is the current state in the chain. We *propose* to move to a new state u' , chosen randomly according to an isotropic random walk starting at u . That is, the proposed new state is a random variable $U' = u + \Delta U$, where $\Delta U \sim \mathcal{N}(0, \sigma^2 I)$. We *accept* this proposed state transition with probability

$$\min \left[1, \frac{\pi(u')}{\pi(u)} \right] \quad (27)$$

and *reject* it otherwise, staying at u . This rule (i) always accepts proposed moves “uphill” to a state u' with higher density than u and (ii) occasionally accepts moves “downhill.” Note that the rule only requires the ratio of the densities of the two states. Hence, the normalization factor $\pi_{\text{evidence}}(d)$ cancels and is not needed. Together, this proposal distribution and acceptance rule specifies a transition kernel $\pi_t(u' | u)$ that grants the Markov chain U_1, U_2, \dots a stationary density equal to $\pi(u)$. Consequently, u_1, \dots, u_n are autocorrelated samples from the posterior $\pi(u)$.⁴³ The scale parameter σ in the random walk proposal distribution dictates the efficiency of the sampling, in terms of the amount of serial correlation in the samples. If σ is too small, too many proposed random walk steps are required to explore the state space. If σ is too large, too many

proposals will be to visit regions with low density, which will be rejected, making the walker stay in place. Both extremes make the sampler inefficient.^{65,67} NUTS,⁶¹ an extension of Hamiltonian Monte Carlo (HMC),⁶⁸ tends to be a more efficient MCMC sampler than random walk Metropolis owing to its more judicious state transition proposal scheme than a random walk in state space.

Warm-up, stopping rules, convergence diagnostics. If it is too difficult to find a high-density region of the posterior in which to initiate the Markov chain with state u_1 , some first fraction of the n MCMC samples are typically be discarded (i.e., not counted as samples from the posterior) as “warm-up” to allow the walker to find a region of state space with high posterior density.^{39,66}

Stopping rules and empirical convergence diagnostic tools may ascertain, or at least give us confidence, that we have run the Markov chain sufficiently long (i.e., that n is large enough) for the sampled chain u_1, \dots, u_n to give a good approximation of the posterior. More, they allow us to avoid wastefully running the chain for longer than needed for our desired accuracy. An example of a stopping rule is to terminate when the *effective* sample size exceeds a pre-specified number. The effective sample size is the number of *independent* samples with the same estimation power as the *autocorrelated* samples from the chain, u_1, \dots, u_n . An example of an empirical convergence diagnostic tool is a *trace plot* of a coordinate of the sampled chain u_1, \dots, u_n vs the iteration number. The trace plot shows the efficiency of the Markov chain. It should resemble a horizontal, approximately uniform-thickness “hairy caterpillar” to reflect low autocorrelation between the samples, thorough exploration of the support of the posterior, and sufficient warm-up. Another convergence check is for consistency between the empirical posterior distributions from independent chains (perhaps) run in parallel.^{66,69,70}

6. Summary of results

Figure 7(a) shows all data from the heat transfer experiment that we use to infer the parameter Λ of the lime with BSI.

Figure 7(b) compares (i) the prior distribution of the parameter Λ and (ii) its updated, (marginal) empirical posterior distribution constructed via kernel density estimation. The bar shows the *equal-tailed 90% posterior credible interval* for Λ , [0.94 h, 1.04 h]. By definition, the true parameter λ of the lime is situated in this interval with 90% probability, falls below it with 5% probability, and falls above it with 5% probability. The width of the interval, then, reflects our posterior uncertainty about Λ . This interpretation is predicated upon our model of the data-generating process and prior assumptions holding.

Figure 7(c) illustrates the posterior distribution over functions $\theta(t)$ modeling the lime temperature by showing a random sample of 100 realizations of models for the lime temperature, $\theta(t; \lambda, t_0 = 0, \theta_0, \theta^{\text{air}})$, with $(\lambda, \theta_0, \theta^{\text{air}})$ a sample from the posterior distribution. The models fit the data $\{(t_i, \theta_{i,\text{obs}})\}_{i=1}^N$ well and exhibit little variance, reflecting low uncertainty about the parameter λ in light of the data.

We found little correlation (magnitude of Pearson correlation < 0.02) between Λ and Σ in the joint posterior distribution. The (marginal) posterior distributions of Λ and Σ are well-approximated as independent Gaussian distributions $\mathcal{N}(0.98 \text{ h}, (0.03 \text{ h})^2)$ and $\mathcal{N}(0.18^\circ \text{ C}, (0.05^\circ \text{ C})^2)$.

Checks. To assess the efficiency of an MCMC sampler and give confidence the MCMC samples reliably approximate the posterior distribution,⁶⁶ we drew trace plots and visualized the empirical posterior distribution of λ over four independent chains in Fig. S2. The trace plot indicates that each sampled chain exhibits little autocorrelation and thoroughly explores an interval of λ , and the empirical posterior distributions are consistent with each other.

To visually inspect (i) the amount of prior information we include about λ in the prior distribution and (ii) the consistency of the observed data $\{(t_i, \theta_{i,\text{obs}})\}_{i=1}^N$ with the prior distribution, we show a prior predictive check^{39,71} in Fig. S3, where we generated synthetic data obtained by simulating the data-generating process with parameters sampled from the prior.

To visually assess the fit of the posterior model of the lime temperature, we, in a posterior predictive check,^{39,71} plot the distribution of residuals between our observed data and synthetic data obtained by simulating the data-generating process with parameters sampled from the posterior in Fig. S4. The mean residual is less than $\pm 0.25^\circ \text{ C}$, an excellent fit.

E. Inverse problem IIa: time reversal

Overview of problem and approach

Task: employ BSI to infer the initial temperature of the lime Θ_0 (i.e., θ_0 treated as a random variable) based on a single measurement of its temperature at a later time $t' > t_0 = 0$, θ'_{obs} , and a measurement of the air temperature, $\theta^{\text{air}}_{\text{obs}}$.

First, we impose a diffuse prior distribution on Θ_0 based on the range of temperatures encountered in refrigerators.

Second, we impose informative prior distributions on the parameter Λ characterizing the lime and the variance of the measurement noise Σ^2 based on the posterior distributions from our parameter identification phase in Sec. II D. “Yesterday’s posterior is today’s prior.”⁷²

Next, we set up another heat transfer experiment on the same lime. To (partially) characterize the conditions of the experiment, we measure the air temperature, giving datum $\theta^{\text{air}}_{\text{obs}}$, which we use to construct an informative prior distribution on Θ^{air} . [Note that we also measured the initial temperature of the lime, but we hold the data ($t_0 = 0, \theta_{0,\text{obs}}$) out from the BSI procedure until the end, to test the fidelity of the posterior distribution of Θ_0 .]

Then, during the experiment, we (indirectly) gather information about Θ_0 by measuring the temperature of the lime at time $t' > t_0 = 0$, i.e., a duration t' after the lime was taken out of the refrigerator, giving datum $(t, \theta'_{\text{obs}})$.

Finally, we use Bayes’ theorem to construct the posterior distribution of $(\Theta_0, \Lambda, \Theta^{\text{air}}, \Sigma)$ in light of the datum $(t, \theta'_{\text{obs}})$, sample from the posterior using an MCMC algorithm, and obtain a credible interval for the initial lime temperature Θ_0 that quantifies posterior uncertainty about its value.

Quick overview:

- *measurement of an experimental condition:* the measured air temperature $\theta^{\text{air}}_{\text{obs}}$;

- *datum collected during the experiment*: the measured lime temperature point $(t', \theta'_{\text{obs}})$ with $t' > t_0$ and $t_0 = 0$;
- *random variables to infer from the datum*: the initial lime temperature Θ_0 , the parameter Λ , the air temperature Θ^{air} , and the variance of the measurement noise Σ^2 ; and
- *sources of priors*: Θ_0 : range of temperatures encountered in refrigerators; Λ, Σ^2 : the posterior from our parameter identification phase in Sec. II D; Θ^{air} : our noisy measurement of the air temperature.

Summary of results: See Figs. 8 and 9.

Classically, this is a determined time-reversal problem that becomes ill-conditioned for large t' . To see the ill-conditioning, let $\delta\theta'$ be the error in our measurement of the true lime temperature $\theta(t')$ and $\delta\theta_0$ be the resulting error in our prediction of the initial temperature, $\hat{\theta}_0 := \theta_0 + \delta\theta_0$, with θ_0 being the true initial lime temperature. The errors $\delta\theta_0$ and $\delta\theta'$ are related through a perturbed version of Eq. (14) (given it holds),

$$\theta(t') + \delta\theta' = \theta^{\text{air}} + (\theta_0 + \delta\theta_0 - \theta^{\text{air}})e^{-(t'-t_0)/\lambda}, \quad (28)$$

implying that the error in the predicted initial temperature $\delta\theta_0 = \delta\theta' e^{(t'-t_0)/\lambda}$ grows exponentially with the time $t' > t_0$ at which we take the measurement of the lime temperature. This ill-conditioning is apparent from the graphical solution to this time reversal problem, of tracing the model trajectory of the lime temperature backward in time, starting at the point $(t', \theta(t') + \delta\theta')$, back to $(t_0 = 0, \hat{\theta}_0)$. A small error $\delta\theta'$ in the measured lime temperature at t' results in a large change in the trajectory traced backward, if t' is large enough to place the measured lime temperature close to the air temperature. See Fig. S5.

1. The experimental setup

To characterize the setup of the additional lime heat transfer experiment (different from the one for parameter identification in Sec. II D, but with the same lime), we use the temperature probe to measure the air temperature, giving datum $\theta^{\text{air}}_{\text{obs}}$, plotted in Fig. 8(a) as the horizontal dashed line. (Note, we actually also measured the initial lime temperature at $t = t_0 := 0$, giving datum $\theta_{0,\text{obs}} = 6.3^\circ\text{C}$, but let us pretend we did not for now.)

2. The prior distributions

The initial temperature, Θ_0 . We impose a diffuse prior distribution on the initial temperature of the lime based on a (generous) range of temperatures encountered in refrigerators,

$$\Theta_0 \sim \mathcal{U}([0^\circ\text{C}, 20^\circ\text{C}]). \quad (29)$$

The air temperature, Θ^{air} . We impose an informative prior distribution on the air temperature Θ^{air} based on our (noisy) measurement of it,

$$\Theta^{\text{air}} \sim \mathcal{N}(\theta^{\text{air}}_{\text{obs}} = 18.64^\circ\text{C}, \sigma^2). \quad (30)$$

The parameter Λ and variance of the measurement noise Σ^2 . We exploit the information we obtained about Λ and Σ during our parameter identification phase in Sec. II D to construct informative prior distributions on the parameter Λ characterizing the

(same) lime and the variance in the measurement noise Σ emanating from the (same) temperature sensor. Particularly, we use the posterior distributions from Sec. II D. “Yesterday’s posterior is today’s prior,”⁷²

$$\Lambda \sim \mathcal{N}(0.98 \text{ h}, (0.03 \text{ h})^2), \quad (31a)$$

$$\Sigma \sim \mathcal{N}(0.18^\circ\text{C}, (0.05^\circ\text{C})^2). \quad (31b)$$

The joint prior distribution. Again, the joint prior distribution of all of the unknowns $\pi_{\text{prior}}(\theta_0, \lambda, \theta^{\text{air}}, \sigma)$ for this inverse problem factorizes since we impose independent priors.

3. The datum and likelihood function

The datum. During the second heat transfer experiment, we take a single measurement of the lime temperature at time $t' > t_0 = 0$, giving the datum point $(t', \theta'_{\text{obs}})$ displayed in Fig. 8(a). Note that the time the lime was taken out of the refrigerator $t_0 = 0$ is known.

The likelihood function. The likelihood function gives the probability density of the datum θ'_{obs} conditioned on each possible value of the parameters $\Lambda = \lambda$ and $\Sigma = \sigma$ and experimental conditions $\Theta_0 = \theta_0$ and $\Theta^{\text{air}} = \theta^{\text{air}}$. We construct the likelihood from the (i) datum $(t', \theta'_{\text{obs}})$ and (ii) model of the data-generating process in Eq. (15). The likelihood function

$$\begin{aligned} \pi_{\text{likelihood}}(\theta'_{\text{obs}} | \theta_0, \lambda, \theta^{\text{air}}, \sigma) \\ = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\theta'_{\text{obs}} - \theta(t'; \lambda, t_0 = 0, \theta_0, \theta^{\text{air}})}{\sigma}\right)^2\right] \end{aligned} \quad (32)$$

quantifies the support the datum $(t', \theta'_{\text{obs}})$ lends to each value of the unknowns $(\theta_0, \lambda, \theta^{\text{air}}, \sigma)$.

4. The posterior distribution

The (joint) *posterior density* governs the probability distribution of the unknowns $(\Theta_0, \Lambda, \Theta^{\text{air}}, \Sigma)$ conditioned on the data $(t', \theta'_{\text{obs}})$. By Bayes’ theorem, the posterior density is proportional to the product of the likelihood function and prior density,

$$\begin{aligned} \pi_{\text{posterior}}(\theta_0, \lambda, \theta^{\text{air}}, \sigma | \theta'_{\text{obs}}) \propto \pi_{\text{likelihood}}(\theta'_{\text{obs}} | \theta_0, \lambda, \theta^{\text{air}}, \sigma) \\ \times \pi_{\text{prior}}(\theta_0, \lambda, \theta^{\text{air}}, \sigma). \end{aligned} \quad (33)$$

We are particularly interested in the posterior distribution of the initial lime temperature Θ_0 , with $(\Lambda, \Theta^{\text{air}}, \Sigma)$ marginalized out.

Again, we employ NUTS to obtain samples from the posterior and then construct an empirical approximation of it.

5. Summary of results

Figure 8(a) shows the data from the heat transfer experiment that we employ to infer the initial lime temperature Θ_0 with BSI.

Figure 8(b) compares (i) the prior distribution of the initial lime temperature Θ_0 with (ii) its updated (marginal) empirical posterior distribution constructed via kernel density estimation. The bar shows the 90% equal-tailed posterior credible interval for Θ_0 . Notably, the hold-out test data, the measured initial lime temperature $\theta_{0,\text{obs}}$, fall in this credible interval.

Figure 8(c) illustrates the posterior distribution of backward trajectories of the lime temperature by showing a random sample of 100 realizations of models for the lime temperature, $\theta(t; \lambda,$

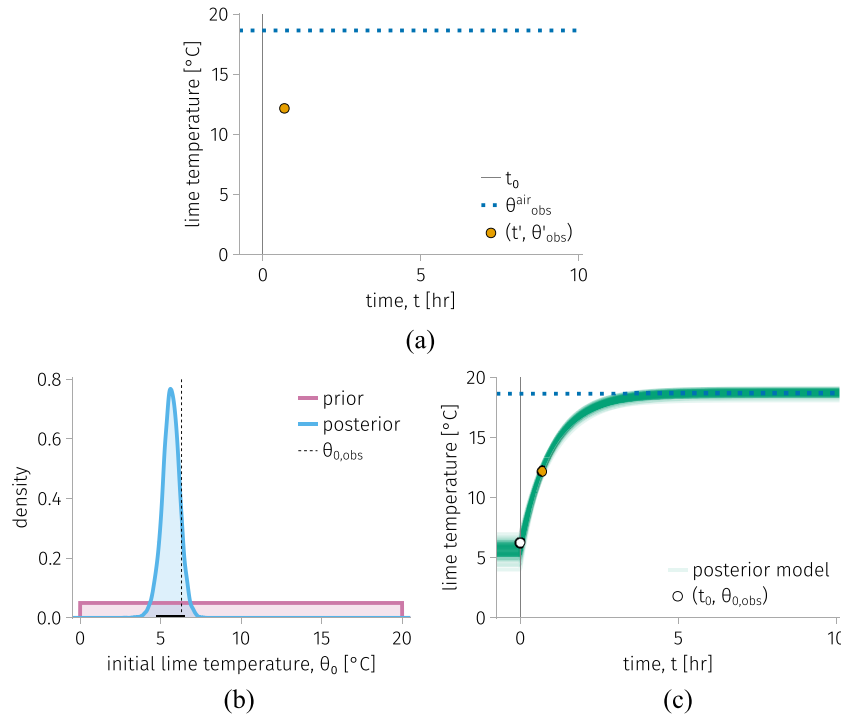


FIG. 8. Inverse problem IIa: time reversal: infer the initial (at time $t_0 = 0$) temperature Θ_0 of the lime from a measurement of the lime temperature later, (t', θ'_{obs}) , and the air temperature, θ_{obs}^{air} . (a) The data. (b) The prior and posterior distributions of Θ_0 . The black bar shows its equal-tailed 90% credible interval. The vertical line shows the held-out measurement $\theta_{0,obs}$. (c) A sample of 100 model trajectories $\theta(t; \lambda, t_0 = 0, \theta_0, \theta^{air})$ from the posterior compared to the true (held-out) initial condition.

$t_0 = 0, \theta_0, \theta^{air}$), with $(\theta_0, \lambda, \theta^{air})$ a sample from the posterior distribution. The true initial condition is covered by the ensemble of backward trajectories.

6. Ill-conditioning

Figure 9 shows the marginal posterior distribution of the initial lime temperature Θ_0 for various times t' at which we measure the lime temperature to obtain the datum (t', θ'_{obs}) . As t' becomes larger, the posterior distribution of Θ_0 spreads, reflecting higher uncertainty. For large t' , i.e., measurements after the lime has been outside of the refrigerator for a long time, we still have high posterior uncertainty about the initial lime temperature. That the datum (t', θ'_{obs}) has not much enlightened us much about the initial lime temperature Θ_0 for large t' owes to the ill-conditioned nature of the problem illustrated in Eq. (28). Hence, Fig. 9 demonstrates the ability of BSI to capture ill-conditioning in inverse problems of reconstruction.

F. Inverse problem IIb: Time reversal

Overview of problem and approach

Task: employ BSI to infer both the time T_0 (i.e., t_0 treated as a random variable) the lime was taken out of the refrigerator and the initial temperature Θ_0 of the lime based on a measurement of the lime temperature at time $t' > t_0, \theta'_{obs}$, and the measured air temperature θ_{obs}^{air} .

First, we impose a diffuse prior distribution on Θ_0 based on the range of temperatures encountered in refrigerators; a weakly informative prior distribution on T_0 based on our sense of time passing; and informative prior distributions on the parameter Λ characterizing the lime and the variance of the measurement noise Σ^2 based on the posterior distributions from our parameter identification phase in Sec. II D.

Next, we set up another heat transfer experiment (well, we use the same data from the second heat transfer experiment in Sec. II E; importantly, it is a different experiment from the one used for parameter identification in Sec. II D) on the same lime. To (partially) characterize the condition of the experiment, we measure the air temperature, giving datum θ_{obs}^{air} to construct an informative prior distribution for Θ^{air} . [Note that we also measure the initial temperature of the lime $\theta_{0,obs}$ and know the time t_0 it was taken out of the refrigerator, but we hold this datum $(t_0, \theta_{0,obs})$ out from the BSI procedure to test the fidelity of the posterior distribution of (T_0, Θ_0) .]

Then, during the experiment, to (indirectly) gather information about (T_0, Θ_0) , we measure the temperature of the lime at time $t' > t_0, \theta'_{obs}$.

Finally, we use Bayes' theorem to construct the posterior distribution of $(T_0, \Theta_0, \Lambda, \Theta^{air}, \Sigma)$ in light of the data, sample from it using a MCMC algorithm, and obtain an empirical marginal joint posterior distribution for the initial lime temperature Θ_0 and time it was taken out of the refrigerator T_0 that quantifies posterior uncertainty about their values.

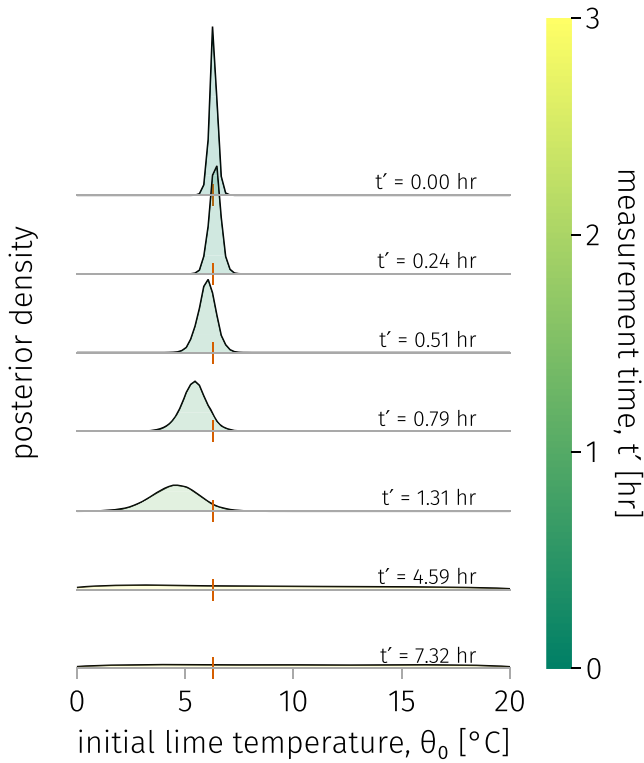


FIG. 9. Inverse problem IIa: time reversal: the (empirical, marginal) posterior distribution of Θ_0 as the time at which we measure the lime temperature, t' , increases. The true, held-out initial temperature is marked with vertical bars.

Quick overview:

- *measurement of an experimental condition:* the measured air temperature $\theta_{\text{obs}}^{\text{air}}$;
- *datum collected during the experiment:* the measured lime temperature point $(t', \theta'_{\text{obs}})$ with $t' > t_0$ and t_0 unknown;
- *random variables to infer from the datum:* the time the lime was taken out of the refrigerator T_0 , the initial lime temperature Θ_0 , the parameter Λ , the air temperature Θ^{air} , and the variance of the measurement noise Σ^2 ; and
- *sources of priors:* T_0 : our human judgment of the passing of time; Θ_0 : range of temperatures encountered in refrigerators; Λ, Σ^2 : the posterior from our parameter identification phase in Sec. II D; Θ^{air} : our noisy measurement of the air temperature.

Classically, this time reversal problem is underdetermined. Conceptually, the observed condition of the lime $(t', \theta'_{\text{obs}})$ is consistent with (1) “the lime was initially very cold and has been outside of the refrigerator for a long duration” and (2) “the lime was initially not very cold and has been outside of the refrigerator for a short duration.” Mathematically, there is a curve of infinite solutions in the (t_0, θ_0) plane (the two primary unknowns) that satisfy the model in Eq. (14) with known $(t', \theta(t'))$ and θ^{air} ,

$$\theta(t') = \theta^{\text{air}} + (\theta_0 - \theta^{\text{air}})e^{-(t'-t_0)/\lambda}. \quad (34)$$

Contrasting the classical curve of solutions in Eq. (34) with a solution via BSI, (i) depending on the prior distribution of T_0 and Θ_0 , the posterior distribution in BSI may assign different weights to each of the (inherently, equal-weighted) classical solutions, and (ii) by accounting for measurement noise, BSI entertains solutions off of the curve comprising the classical solutions. This time reversal problem still becomes ill-conditioned for large $t' - t_0$, as the curve described in Eq. (34) is sensitive to errors in the measurement of $\theta(t')$ when $t' - t_0$ is large.

1. The experimental setup

To (partially) characterize the experimentally setup of this lime heat transfer experiment, we use the temperature probe to measure the air temperature, giving datum $\theta_{\text{obs}}^{\text{air}}$ plotted as the horizontal dashed line in Fig. 10(a).

2. The prior distributions

We impose the same prior distributions of Θ_0 , Θ^{air} , Λ , and Σ^2 as in Sec. II E. Additionally, we now impose a prior distribution on the time T_0 at which the lime was taken out of the refrigerator based on our unreliable—truly, $t_0 = 0$, so the prior for T_0 is biased, as it is not centered at the true value—judgment of the passing of time,

$$T_0 \sim \mathcal{N}_{-1\text{h} \leq T_0 \leq 1\text{h}}(-0.1\text{h}, (0.25\text{h})^2). \quad (35)$$

This gives a joint prior distribution $\pi_{\text{prior}}(t_0, \theta_0, \lambda, \theta^{\text{air}}, \sigma)$ for this inverse problem. We visualize the marginal prior distribution of (T_0, Θ_0) in Fig. 10(b).

3. The datum and likelihood function

The datum. The datum from the heat transfer experiment is a single measurement of the lime temperature, $(t', \theta'_{\text{obs}})$ with $t' > t_0$, displayed in Fig. 10(a). The time the lime was taken out of the refrigerator t_0 is *not* known.

The likelihood function. The likelihood function gives the probability density of the datum θ'_{obs} conditioned on each possible value of the parameters $\Lambda = \lambda$ and $\Sigma = \sigma$ and experimental conditions $T_0 = t_0$, $\Theta_0 = \theta_0$, and $\Theta^{\text{air}} = \theta^{\text{air}}$,

$$\begin{aligned} \pi_{\text{likelihood}}(\theta'_{\text{obs}} | t_0, \theta_0, \lambda, \theta^{\text{air}}, \sigma) \\ = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\theta'_{\text{obs}} - \theta(t'; \lambda, t_0, \theta_0, \theta^{\text{air}})}{\sigma}\right)^2\right]. \end{aligned} \quad (36)$$

4. The posterior distribution

The (joint) *posterior density* governs the probability distribution of the unknowns $(T_0, \Theta_0, \Lambda, \Theta^{\text{air}}, \Sigma)$ conditioned on the data $(t', \theta'_{\text{obs}})$. By Bayes’ theorem, the posterior density is proportional to the product of the likelihood function and (joint) prior density,

$$\begin{aligned} \pi_{\text{posterior}}(t_0, \theta_0, \lambda, \theta^{\text{air}}, \sigma | \theta'_{\text{obs}}) \propto \pi_{\text{likelihood}}(\theta'_{\text{obs}} | t_0, \theta_0, \lambda, \theta^{\text{air}}, \sigma) \\ \times \pi_{\text{prior}}(t_0, \theta_0, \lambda, \theta^{\text{air}}, \sigma). \end{aligned} \quad (37)$$

We are particularly interested in the posterior distribution of the initial condition of the lime (T_0, Θ_0) , with $(\Lambda, \Theta^{\text{air}}, \Sigma)$ marginalized out.

Again, we employ NUTS to obtain samples from the posterior.

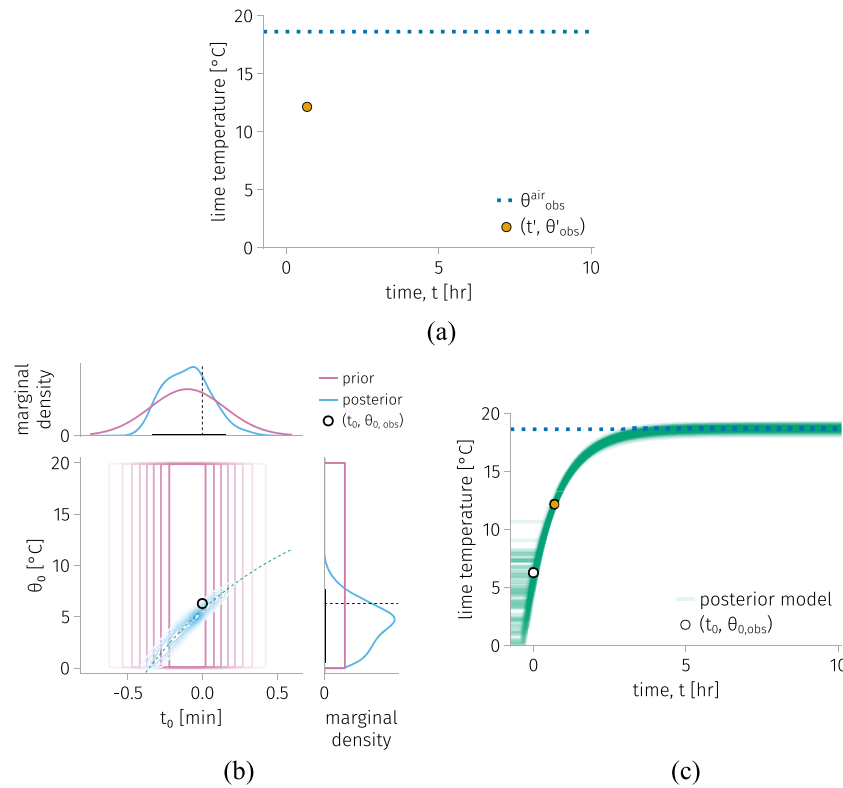


FIG. 10. Inverse problem IIb: time reversal: infer the initial temperature Θ_0 of the lime and the time T_0 it was taken out of the refrigerator from a measurement of the lime temperature later, $(t', \theta'_{\text{obs}})$, and the air temperature $\theta^{\text{air}}_{\text{obs}}$. (a) The data. (b) The joint and marginal prior and posterior distributions of (T_0, Θ_0) . The held-out measurement of the initial condition of the lime is indicated by the black point/dashed lines. The green dashed line shows the classical solution in Eq. (34) with λ set to be the mean of the posterior from Sec. II D. The black bars in the marginal plots show the 90% equal-tailed credible intervals. (c) A sample of 100 model trajectories $\theta(t; \lambda, t_0, \theta_0, \theta^{\text{air}})$, with $(T_0, \Theta_0, \Lambda, \Theta^{\text{air}})$ being a sample from the posterior.

5. Summary of results

Figure 10(a) shows the data from the heat transfer experiment that we employ to infer the initial condition of the lime (T_0, Θ_0) with BSI.

By showing contours, Fig. 10(b) compares (i) the joint prior distribution of the initial condition of the lime, (T_0, Θ_0) , with (ii) the updated, empirical, joint posterior distribution of (T_0, Θ_0) constructed via kernel density estimation. The curve drawn in the (t_0, θ_0) plane shows the classical solutions, given in Eq. (34) with $\theta^{\text{air}} := \theta^{\text{air}}_{\text{obs}}$, $\theta(t') := \theta'_{\text{obs}}$, and λ set to the mean of its posterior distribution from the parameter identification phase. The ridge of the posterior density follows the curve of classical solutions. However, owing to the non-uniform density and finite support of the prior distribution of (T_0, Θ_0) , the classical solutions are weighted differently and some are not entertained at all. The posterior density spreads orthogonal to (off) the curve of classical solutions owing to BSI accounting for noise corrupting our measurement of the air and lime temperature (unlike the classical approach); i.e., BSI entertains solutions off of the curve of classical solutions. The measured initial condition of the lime $(t_0 = 0, \theta_0, \text{obs})$, held out as test data, falls in a region of high posterior density. This result is, in part, a consequence of the mean of our prior for T_0 in Eq. (35) being close to the true $t_0 = 0$. The marginal posterior distributions of T_0 and Θ_0

are compared in Fig. 10(b) as well, including their 90%, equal-tailed posterior credible intervals. The credible interval for Θ_0 is much wider than in the inverse problem IIa in Fig. 8(b), despite using the same lime temperature measurement θ'_{obs} after the same duration t' the lime has been outside of the refrigerator, because t_0 now is not specified but rather T_0 is endowed a spread-out prior distribution, including density less than the true t_0 .

Figure 10(c) illustrates the posterior distribution of backward trajectories of the lime temperature by showing a random sample of 100 realizations of models for the lime temperature, $\theta(t; \lambda, t_0, \theta_0, \theta^{\text{air}})$, with $(t_0, \theta_0, \lambda, \theta^{\text{air}})$ being a sample from the posterior distribution. The intuition explaining the correlation of T_0 and Θ_0 in the joint posterior density in Fig. 10(b) is apparent in Fig. 10(c): the data $(t', \theta'_{\text{obs}})$ are consistent with both propositions: (i) “the lime was initially not very cold and taken out of the refrigerator recently” and (ii) “the lime was initially very cold and taken out of the refrigerator a while ago.”

G. Software for BSI

A variety of free and open-source probabilistic programming libraries are available, making Bayesian statistical inversion accessible to practitioners.³⁹ Broadly, the probabilistic (computer) programming paradigm^{73,74} allows practitioners to (i) easily specify

(a) the prior probability distributions and (b) the probabilistic forward model assumed to generate the data then (ii) call a Markov Chain Monte Carlo routine to draw samples from the posterior distribution that (implicitly) follows from the prior and likelihood. Popular Bayesian inference engines/probabilistic programming libraries include `Turing.jl`⁶² in Julia,⁶³ `PyMC3`,⁷⁵ `Pyro`,⁷⁶ and `Probability`⁷⁷ in Python, and `Stan`⁷⁸ that interfaces with several languages, including R.

BSI coding tutorial. We provide a minimal coding tutorial for BSI to tackle inverse problems I and IIa herein at https://simonensemble.github.io/pluto_nbs/bsi_tutorial.html using the probabilistic programming library `Turing.jl` in the Julia programming language.

H. Conclusions

By way of example, we provided a tutorial of Bayesian statistical inversion (BSI) to solve inverse problems while incorporating prior information and quantifying uncertainty. Our focus was a simple, intuitive physical process—heat transfer from ambient indoor air to a cold lime fruit via natural convection. First, we developed a simple mathematical model for the lime temperature, which contains a single parameter. Then, we used a time series dataset of the lime temperature to infer, via BSI, the posterior distribution of the model parameter. Next, we employed the model with the inferred parameter to tackle, via BSI, two reconstruction problems of time reversal. The first task, ill-conditioned, was to predict the initial temperature of the lime from a measurement of its temperature later in time. The second task, underdetermined, was to predict the initial temperature of the lime *and* the time it was taken out of the refrigerator from a measurement of its temperature later in time. We intend for our Tutorial to facilitate scientists and engineers to (i) recognize inverse problems in their research domain, (ii) question whether these problems are well-posed and well-conditioned, and (iii) leverage BSI to solve these problems while incorporating prior information and quantifying uncertainty.

Our BSI solutions to the inverse problems involving the lime are subject to limitations. First, our mathematical model of the lime temperature relies on several simplifying assumptions listed in Sec. II B. The model may be more accurate if we relaxed these assumptions and amended it to account for, e.g., the time-dependence of the bulk air temperature $\theta^{\text{air}}(t)$ and spatial temperature gradients in the lime in conjunction with its geometry and spatial heterogeneity (skin, flesh, seeds). Second, ideally, we would replicate the heat transfer experiment multiple times and use all of this time series data of the lime temperature for the inference of the parameter Λ in Sec. II D. This would allow the posterior distribution of Σ to capture the full residual variability of the lime temperature over repeated experiments owing to poorly controlled and/or unrecognized inputs/conditions that affect the lime temperature. Third, we neglected the possibility of model discrepancy,²¹ an unaccounted-for source of both systemic bias and uncertainty in the posterior of the parameter/initial condition of the lime.

III. DISCUSSION

We provided an introductory tutorial on Bayesian statistical inversion as a tool to tackle inverse problems, which could

be ill-posed, while (i) incorporating prior information and (ii) providing a solution in the form of a probability density function over the input/parameter space, which quantifies uncertainty via its spread. Inverse problems are pervasive throughout sciences and engineering, e.g., in heat or radiation transfer, gravitational fields, wave scattering, tomography, and electromagnetism;^{79–84} vibration of springs and beams;⁸⁵ imaging;^{7,86} fluid mechanics;⁸⁷ physiology;⁸⁸ epidemiology;^{89,90} ecology;⁹¹ geophysics;⁸⁴ environmental science;^{92,93} palaeoclimatology;⁹⁴ chemical/bio-chemical reactions;^{95–99} and adsorption.¹⁰⁰ Mathematically, reconstruction problems in these domains often reduce to using data to determine (1) a vector that was linearly transformed by a matrix;^{13,101,102} (2) the initial condition, boundary condition, or forcing term in an ordinary or partial differential equation;⁷ (3) an integrand,^{103,104} or (4) the geometry of a domain.⁹

A. Drawbacks of BSI

The Bayesian statistical approach to inverse problems has some drawbacks compared to simpler, classical methods, such as least squares.

A larger computational cost. The Bayesian approach to inverse problems typically demands a relatively large computational cost to approximate the posterior distribution via a Markov chain Monte Carlo algorithm—especially when the number of parameters/inputs is large and the likelihood is expensive to evaluate (owing to a large size of the data and/or a large computational expense to evaluate the forward model).^{35,49} To reduce the cost of Bayesian inference at the expense of accuracy/flexibility, we may resort to approximate Bayesian computation (ABC),¹⁰⁵ make a Laplace (quadratic) approximation of the posterior, or employ a conjugate prior for the likelihood that gives a closed-form expression for the posterior.³⁵

The subjectivity of the prior. A common objection to Bayesian inference is the subjectivity involved in constructing the prior distribution (which influences the posterior).¹⁰⁶ First, we can assess the impact of the prior on the posterior through a sensitivity analysis (i.e., examine how much the posterior changes when the prior changes). Second, through *prior predictive checking*, we may ascertain the consistency of the prior with our observed data by comparing (a) the observed data with (b) synthetic data generated via sampling inputs/parameters from the prior then simulating the probabilistic model of the data-generating process under those inputs/parameters.^{39,71} Third, in the absence of information, we may adopt the principle of indifference and impose a diffuse prior that avoids biasing the posterior. Finally, one intriguing approach to construct a prior is to elicit estimates of the unknown parameter/input from a panel of experts.¹⁰⁷ In defense of the prior distribution, it provides an *opportunity/benefit* to (i) potentially obtain better estimates of the inputs/parameters by exploiting relevant external information or constraints about the input/parameters (e.g., the physical constraint and back-of-the-envelope estimate of λ for the lime) and (ii) interpret new data from an experiment in the context of different beliefs held among different scientists (and encoded in their priors) before the data were collected.^{36,40} In addition, the choice for the forward model (its structure, e.g., functional form), upon which non-Bayesian methods also rely, requires subjectivity/judgment as well.³⁶

B. Model discrepancy

*Model discrepancy*²¹ refers to a nonzero difference between the (i) expected measurement of the system output over repeated experiments with the same input/conditions and (ii) the prediction of the system output by the model. If significantly present and not accounted for in the model of the data-generating process, model discrepancy corrupts the BSI solution to an inverse problem.¹⁰⁸ First, disregarded model discrepancy introduces bias, i.e., the posterior density of the parameter/input will not be centered at the true value even as more and more data are collected. Second, it leads to mis-calibrated uncertainty quantification, i.e., the posterior credible interval will not be likely to contain the true value of the parameter/input.

To account for model discrepancy, we can modify the model of the data-generating process to explicitly include a model discrepancy function $\Delta(x)$ (a random variable dependent on the input, x) as a Gaussian process¹⁰⁹ and infer it from data.^{21,108,110} That is, we model the measured system output Y_{obs} as

$$Y_{\text{obs}}(x) = f_{\beta}(x) + \Delta(x) + E, \quad (38)$$

with $f_{\beta}(x)$ the (inadequate) forward model and E the noise, which by definition has mean zero. The model discrepancy function $\Delta(x)$ aims to capture the input-dependent difference between reality (which we probe through sampling the output of the system Y_{obs} for a given input x) and the model predictions.

C. Model selection

We assumed that the model structure—in our study, the functional form/shape of the lime temperature as a function of time, $\theta(t)$ —is known (see Table 2a). In *model selection*, the task is to select the best model of a physical system among a set of candidate models, given data. Bayesian approaches to model selection include the Bayesian information criterion¹¹¹ and Bayes factors^{39,112} based on the evidence function.^{40,113}

D. Other methods to approximate the posterior distribution

Here, we employed the NUTS⁶¹ Markov chain Monte Carlo method to obtain samples from the posterior distribution to approximate it. Other methods to approximate a posterior distribution include (i) other Monte Carlo sampling methods, including the Gibbs sampler,¹¹⁴ adaptive Metropolis algorithm,¹¹⁵ sequential Monte Carlo,⁴³ and approximate Bayesian computation (ABC),¹¹⁶ and (ii) obtaining an analytical expression as an approximation to the posterior, including the Laplace approximation and variational Bayes.¹¹⁷ The BSI practitioner must consider accuracy, speed, and ease of implementation⁴³ when choosing a method to approximate their posterior at hand.

E. Classical approaches to inverse problems

A traditional approach to inverse problems is least squares to minimize the mismatch between the data and the model predictions by tuning the unknown input/parameter. For overdetermined inverse problems, bootstrapping,¹¹⁸ asymptotic theory,¹¹⁹ and the Hessian of the loss function¹²⁰ can provide uncertainty

estimates for the unknown input/parameters within the least squares framework—and for new predictions by the model. To handle ill-posed and/or ill-conditioned inverse problems, Tikhonov regularization^{121,122} provides a means to incorporate prior assumptions into their solutions by augmenting least squares. For example, the ill-posed problem of determining the initial temperature profile of a rod from a measurement of its temperature profile later in time may be tackled by augmenting the least-squares loss function with a regularization term to promote smooth solutions.¹²³ Tikhonov regularization is intimately connected with BSI.^{13,14,34,38} However, BSI is more versatile and better-grounded in theory; consequently, BSI provides a more interpretable means of incorporating prior assumptions/information, and BSI entertains and weighs multiple solutions for uncertainty quantification.

SUPPLEMENTARY MATERIAL

The supplementary material shows the Arduino setup for lime temperature measurements, convergence diagnostics for MCMC sampling, prior predictive check, posterior predictive check, and graphical solution to time reversal problem.

ACKNOWLEDGMENTS

F.G.W. acknowledges NSF under Award No. 1920945 for support. C.M.S. acknowledges support from the US Department of Homeland Security Countering Weapons of Mass Destruction under CWMD Academic Research Initiative Cooperative Agreement No. 21CWDARI00043. This support does not constitute an expressed or implied endorsement on the part of the Government. The authors thank Edward Celarier, Luther Mahoney, and the anonymous reviewers for feedback on the manuscript.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Faaig G. Waqar: Conceptualization (equal); Data curation (lead); Formal analysis (equal); Investigation (equal); Methodology (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Swati Patel:** Conceptualization (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Cory M. Simon:** Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Methodology (equal); Project administration (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

All data and Julia code to reproduce the results/plots in this article are openly available in Github at <https://github.com/faaiqgwaqar/Inverse-Problems>.¹²⁴

REFERENCES

- ¹J. M. Epstein, “Why model?,” *J. Artif. Soc. Soc. Simul.* **11**(4), 12 (2008).
- ²R. C. Aster, B. Borchers, and C. H. Thurber, *Parameter Estimation and Inverse Problems* (Elsevier, 2018).
- ³C. W. Groetsch, *Inverse Problems: Activities for Undergraduates* (Cambridge University Press, 1999), Vol. 12.
- ⁴M. Iglesias and A. M. Stuart, “Inverse problems and uncertainty quantification,” *SIAM News* **20**, 2–3 (2014).
- ⁵S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (CRC Press, 2018).
- ⁶S. I. Kabanikhin, “Definitions and examples of inverse and ill-posed problems,” *J. Inverse Ill-Posed Probl.* **16**(4), 317–357 (2008).
- ⁷F. Duarte Moura Neto and A. J. da Silva Neto, *An Introduction to Inverse Problems with Applications* (Springer Science & Business Media, 2012).
- ⁸J. L. Fernández-Martínez, Z. Fernández-Muñiz, J. L. G. Pallero, and L. M. Pedruelo-González, “From Bayes to Tarantola: New insights to understand uncertainty in inverse problems,” *J. Appl. Geophys.* **98**, 62–72 (2013).
- ⁹K. Ito and B. Jin, *Inverse Problems: Tikhonov Theory and Algorithms* (World Scientific, 2014), Vol. 22.
- ¹⁰A. M. Denisov, *Elements of the Theory of Inverse Problems* (VSP, 1999), Vol. 14.
- ¹¹J. L. Mueller and S. Siltanen, *Linear and Nonlinear Inverse Problems with Practical Applications* (SIAM, 2012).
- ¹²L. Tenorio, *An Introduction to Data Analysis and Uncertainty Quantification for Inverse Problems* (SIAM, 2017).
- ¹³J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems* (Springer Science & Business Media, 2006), Vol. 160.
- ¹⁴M. Vauhkonen, T. Tarvainen, and T. Lähivaara, “Inverse problems,” in *Mathematical Modelling* (Springer, 2016), Chap. 12.
- ¹⁵M. Kac, “Can one hear the shape of a drum?,” *Am. Math. Mon.* **73**(4P2), 1–23 (1966).
- ¹⁶M. H. Protter, “Can one hear the shape of a drum? Revisited,” *SIAM Rev.* **29**(2), 185–197 (1987).
- ¹⁷C. Gordon, D. L. Webb, and S. Wolpert, “One cannot hear the shape of a drum,” *Bull. Am. Math. Soc.* **27**(1), 134–138 (1992).
- ¹⁸C. Gordon and D. Webb, “You can’t hear the shape of a drum,” *Am. Sci.* **84**(1), 46–55 (1996).
- ¹⁹E. Kreyszig, *Advanced Engineering Mathematics*, 10th ed. (Wiley, 2009).
- ²⁰J. R. Kuttler and V. G. Sigillito, “Eigenvalues of the Laplacian in two dimensions,” *SIAM Rev.* **26**(2), 163–193 (1984).
- ²¹M. C. Kennedy and A. O’Hagan, “Bayesian calibration of computer models,” *J. R. Stat. Soc.: Ser. B* **63**(3), 425–464 (2001).
- ²²G. Chavent, *Nonlinear Least Squares for Inverse Problems: Theoretical Foundations and Step-by-Step Guide for Applications* (Springer Science & Business Media, 2010).
- ²³L. R. Lines and S. Treitel, “A review of least-squares inversion and its application to geophysical problems,” *Geophys. Prospect.* **32**(2), 159–186 (1984).
- ²⁴P. C. Sabatier, “Past and future of inverse problems,” *J. Math. Phys.* **41**(6), 4082–4124 (2000).
- ²⁵O. J. Maclaren and R. Nicholson, “What can be estimated? Identifiability, estimability, causal inference and ill-posed inverse problems,” *arXiv:1904.02826* (2019).
- ²⁶J. H. A. Guillaume, J. D. Jakeman, S. Marsili-Libelli, M. Asher, P. Brunner, B. Croke, M. C. Hill, A. J. Jakeman, K. J. Keesman, S. Razavi, and J. D. Stigter, “Introductory overview of identifiability analysis: A guide to evaluating whether you have the right type of data for your modeling purpose,” *Environ. Modell. Software* **119**, 418–432 (2019).
- ²⁷O.-T. Chis, J. R. Banga, and E. Balsa-Canto, “Structural identifiability of systems biology models: A critical comparison of methods,” *PLoS One* **6**(11), e27755 (2011).
- ²⁸M. Dashti and A. M. Stuart, “The Bayesian approach to inverse problems,” in *Handbook of Uncertainty Quantification* (Springer, 2017), pp. 311–428.
- ²⁹A. Tarantola, B. Valette *et al.*, “Inverse problems = Quest for information,” *J. Geophys.* **50**(1), 159–170 (1982).
- ³⁰A. M. Stuart, “Inverse problems: A Bayesian perspective,” *Acta Numer.* **19**, 451–559 (2010).
- ³¹A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation* (SIAM, 2005).
- ³²J. Idier, *Bayesian Approach to Inverse Problems* (John Wiley & Sons, 2013).
- ³³T. J. Ulrych, M. D. Sacchi, and A. Woodbury, “A Bayes tour of inversion: A tutorial,” *Geophysics* **66**(1), 55–69 (2001).
- ³⁴B. G. Fitzpatrick, “Bayesian analysis in inverse problems,” *Inverse Probl.* **7**(5), 675 (1991).
- ³⁵G. D’Agostini, “Bayesian inference in processing experimental data: Principles and basic applications,” *Rep. Prog. Phys.* **66**(9), 1383 (2003).
- ³⁶U. Von Toussaint, “Bayesian inference in physics,” *Rev. Mod. Phys.* **83**(3), 943 (2011).
- ³⁷S. Andreon and B. Weaver, *Bayesian Methods for the Physical Sciences* (Springer International Publishing, Switzerland, 2015), p. 52.
- ³⁸D. Calvetti and E. Somersalo, “Inverse problems: From regularization to Bayesian inference,” *Wiley Interdiscip. Rev.: Comput. Stat.* **10**(3), e1427 (2018).
- ³⁹R. van de Schoot, S. Depaoli, R. King, B. Kramer, K. Märtens, M. G. Tadesse, M. Vannucci, A. Gelman, D. Veen, J. Willemsen, and C. Yau, “Bayesian statistics and modelling,” *Nat. Rev. Methods Primers* **1**(1), 1–26 (2021).
- ⁴⁰R. Trotta, “Bayes in the sky: Bayesian inference and model selection in cosmology,” *Contemp. Phys.* **49**(2), 71–104 (2008).
- ⁴¹J. K. Ghosh, M. Delampady, and T. Samanta, *An Introduction to Bayesian Analysis: Theory and Methods* (Springer, 2006), Vol. 725.
- ⁴²A. B. Downey, “Think Bayes 2,” <https://alldowney.github.io/ThinkBayes2/index.html>, 2021.
- ⁴³K. P. Murphy, *Probabilistic Machine Learning: Advanced Topics* (MIT Press, 2023).
- ⁴⁴C. P. Robert, G. Casella, and G. Casella, *Monte Carlo Statistical Methods* (Springer, 1999), Vol. 2.
- ⁴⁵R. J. Hyndman, “Computing and graphing highest density regions,” *Am. Stat.* **50**(2), 120–126 (1996).
- ⁴⁶K.-R. Koch, *Introduction to Bayesian Statistics* (Springer Science & Business Media, 2007).
- ⁴⁷L. Hespanhol, C. S. Vallio, L. M. Costa, and B. T. Saragiotto, “Understanding and interpreting confidence and credible intervals around effect estimates,” *Braz. J. Phys. Ther.* **23**(4), 290–301 (2019).
- ⁴⁸Š. Kubínová and J. Šlégr, “ChemDuino: Adapting Arduino for low-cost chemical measurements in Lecture and Laboratory,” *J. Chem. Educ.* **92**(10), 1751–1753 (2015).
- ⁴⁹V. Dose, “Bayesian inference in physics: Case studies,” *Rep. Prog. Phys.* **66**(9), 1421 (2003).
- ⁵⁰K. P. Murphy, “Conjugate Bayesian analysis of the Gaussian distribution,” <https://www.cs.ubc.ca/murphyk/Papers/bayesGauss.pdf>, 2007.
- ⁵¹D. Fink, *A compendium of conjugate priors*, 1997.
- ⁵²Y. Cengel, J. Cimbala, and R. Turner, *Fundamentals of Thermal-Fluid Sciences* (McGraw Hill, 2017).
- ⁵³R. B. Bird, W. E. Stewart, and E. N. Lightfoot, *Transport Phenomena* (John Wiley and Sons, 2002).
- ⁵⁴O. J. Ikegwu and F. C. Ekwu, “Thermal and physical properties of some tropical fruits and their juices in Nigeria,” *J. Food Technol.* **7**(2), 38–42 (2009).
- ⁵⁵P. Kosky, R. Balmer, W. Keat, and G. Wise, “Mechanical engineering,” in *Exploring Engineering*, 3rd ed., edited by P. Kosky, R. Balmer, W. Keat, and G. Wise (Academic Press, Boston, 2013), pp. 259–281.
- ⁵⁶M. Vollmer, “Newton’s law of cooling revisited,” *Eur. J. Phys.* **30**(5), 1063 (2009).
- ⁵⁷W. G. Rees and C. Viney, “On cooling tea and coffee,” *Am. J. Phys.* **56**(5), 434–437 (1988).
- ⁵⁸C. T. O’Sullivan, “Newton’s law of cooling—A critical assessment,” *Am. J. Phys.* **58**(10), 956–960 (1990).
- ⁵⁹C. F. Bohren, “Comment on ‘Newton’s law of cooling—a critical assessment,’ by Colm T. O’Sullivan [Am. J. Phys. 58, 956–960 (1990)],” *Am. J. Phys.* **59**(11), 1044–1046 (1991).

- ⁶⁰M. Mukama, A. Ambaw, and U. L. Opara, “Thermophysical properties of fruit—A review with reference to postharvest handling,” *J. Food Meas. Charact.* **14**(5), 2917–2937 (2020).
- ⁶¹M. D. Hoffman, A. Gelman *et al.*, “The No-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo,” *J. Mach. Learn. Res.* **15**(1), 1593–1623 (2014); [arXiv:1111.4246](https://arxiv.org/abs/1111.4246)
- ⁶²H. Ge, K. Xu, and Z. Ghahramani, “Turing: A language for flexible probabilistic inference,” in *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics PMLR* (2018), pp. 1682–1690.
- ⁶³J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, “Julia: A fresh approach to numerical computing,” *SIAM Rev.* **59**(1), 65–98 (2017).
- ⁶⁴Y.-C. Chen, “A tutorial on kernel density estimation and recent advances,” *Biostat. Epidemiol.* **1**(1), 161–187 (2017).
- ⁶⁵C. Sherlock, P. Fearnhead, and G. O. Roberts, “The random walk Metropolis: Linking theory and practice through a case study,” *Stat. Sci.* **25**(2), 172–190 (2010).
- ⁶⁶V. Roy, “Convergence diagnostics for Markov chain Monte Carlo,” *Annu. Rev. Stat. Appl.* **7**, 387–412 (2020).
- ⁶⁷G. O. Roberts and J. S. Rosenthal, “Optimal scaling for various Metropolis-Hastings algorithms,” *Stat. Sci.* **16**(4), 351–367 (2001).
- ⁶⁸M. Betancourt, “A conceptual introduction to Hamiltonian Monte Carlo,” [arXiv:1701.02434](https://arxiv.org/abs/1701.02434) (2017).
- ⁶⁹M. Plummer, N. Best, K. Cowles, and K. Vines, “CODA: Convergence diagnosis and output analysis for MCMC,” *R News* **6**(1), 7–11 (2006).
- ⁷⁰M. K. Cowles and B. P. Carlin, “Markov chain Monte Carlo convergence diagnostics: A comparative review,” *J. Am. Stat. Assoc.* **91**(434), 883–904 (1996).
- ⁷¹J. Gabry, D. Simpson, A. Vehtari, M. Betancourt, and A. Gelman, “Visualization in Bayesian workflow,” *J. R. Stat. Soc. Ser. A* **182**(2), 389–402 (2019).
- ⁷²D. Calvetti and E. Somersalo, “Subjective knowledge or objective belief? An oblique look to Bayesian methods,” in *Large-Scale Inverse Problems and Quantification of Uncertainty* (Wiley, 2010), pp. 33–70.
- ⁷³A. D. Gordon, T. A. Henzinger, V. AdityaNori, and K. SriramRajamani, “Probabilistic programming,” in *Future of Software Engineering Proceedings* (Association for Computing Machinery (ACM), 2014), pp. 167–181.
- ⁷⁴J.-W. van de Meent, B. Paige, H. Yang, and F. Wood, “An introduction to probabilistic programming,” [arXiv:1809.10756](https://arxiv.org/abs/1809.10756) (2018).
- ⁷⁵J. Salvatier, T. V. Wiecki, and C. Fonnesbeck, “Probabilistic programming in Python using PyMC3,” *PeerJ Comput. Sci.* **2**, e55 (2016).
- ⁷⁶E. Bingham, J. P. Chen, M. Jankowiak, F. Obermeyer, N. Pradhan, T. Karaletos, R. Singh, P. Szerlip, P. Horsfall, and N. D. Goodman, “Pyro: Deep universal probabilistic programming,” *J. Mach. Learn. Res.* **20**(1), 973–978 (2019); [arXiv:1810.09538](https://arxiv.org/abs/1810.09538) (2018).
- ⁷⁷V. JoshuaDillon, I. Langmore, D. Tran, E. Brevdo, S. Vasudevan, D. Moore, B. Patton, A. Alemi, M. Hoffman, and R. A. Saurous, “TensorFlow distributions,” [arXiv:1711.10604](https://arxiv.org/abs/1711.10604) (2017).
- ⁷⁸B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. A. Brubaker, J. Guo, P. Li, and A. Riddell, “Stan: A probabilistic programming language,” *J. Stat. Software* **76**(1), 1–32 (2017).
- ⁷⁹J. P. Kaipio and C. Fox, “The Bayesian framework for inverse problems in heat transfer,” *Heat Transfer Eng.* **32**(9), 718–753 (2011).
- ⁸⁰J. Wang and N. Zabaras, “A Bayesian inference approach to the inverse heat conduction problem,” *Int. J. Heat Mass Transfer* **47**(17–18), 3927–3941 (2004).
- ⁸¹H. R. B. Orlande, G. S. Dulikravich, M. Neumayer, D. Watzgenig, and M. J. Colaço, “Accelerated Bayesian inference for the estimation of spatially varying heat flux in a heat conduction problem,” *Numer. Heat Transfer, Part A* **65**(1), 1–25 (2014).
- ⁸²V. Isakov, *Inverse Problems for Partial Differential Equations* (Springer, 2006), Vol. 127.
- ⁸³A. H. Hasanoğlu and V. G. Romanov, *Introduction to Inverse Problems for Differential Equations* (Springer, 2021).
- ⁸⁴M. Richter, *Inverse Problems: Basics, Theory and Applications in Geophysics* (Springer Nature, 2021).
- ⁸⁵G. M. L. Gladwell, *Inverse Problems in Vibration* (Springer, 1986).
- ⁸⁶A. Ribes and F. Schmitt, “Linear inverse problems in imaging,” *IEEE Signal Process. Mag.* **25**(4), 84–99 (2008).
- ⁸⁷S. L. Cotter, M. Dashti, J. C. Robinson, and A. M. Stuart, “Bayesian inverse problems for functions and applications to fluid mechanics,” *Inverse Probl.* **25**(11), 115008 (2009).
- ⁸⁸S. Zenker, J. Rubin, and G. Clermont, “From inverse problems in mathematical physiology to quantitative differential diagnoses,” *PLoS Comput. Biol.* **15**(6), e1007155 (2019).
- ⁸⁹X. Hao, S. Cheng, D. Wu, T. Wu, X. Lin, and C. Wang, “Reconstruction of the full transmission dynamics of COVID-19 in Wuhan,” *Nature* **584**(7821), 420–424 (2020).
- ⁹⁰M. Ansari, D. Soriano-Paños, G. Ghoshal, and A. D. White, “Inferring spatial source of disease outbreaks using maximum entropy,” *Phys. Rev. E* **106**(1), 014306 (2022).
- ⁹¹M. Dowd and R. Meyer, “A Bayesian approach to the ecosystem inverse problem,” *Ecol. Modell.* **168**(1–2), 39–55 (2003).
- ⁹²M. Andrieu and A. El Badia, “On an inverse source problem for the heat equation. Application to a pollution detection problem, II,” *Inverse Probl. Sci. Eng.* **23**(3), 389–412 (2015).
- ⁹³E. Yee, F.-S. Lien, A. Keats, and R. D’Amours, “Bayesian inversion of concentration data: Source reconstruction in the adjoint representation of atmospheric diffusion,” *J. Wind Eng. Ind. Aerodyn.* **96**(10–11), 1805–1816 (2008).
- ⁹⁴J. Haslett, M. Whitley, S. Bhattacharya, M. Salter-Townshend, S. P. Wilson, J. R. M. Allen, B. Huntley, and F. J. G. Mitchell, “Bayesian palaeoclimate reconstruction,” *J. R. Stat. Soc. Ser. A* **169**(3), 395–438 (2006).
- ⁹⁵H. W. Engl, C. Flamm, P. Kügler, J. Lu, S. Müller, and P. Schuster, “Inverse problems in systems biology,” *Inverse Probl.* **25**(12), 123014 (2009).
- ⁹⁶R. Guzzi, T. Colombo, and P. Paci, “Inverse problems in systems biology: A critical review,” *Syst. Biol.* **1702**, 69–94 (2018).
- ⁹⁷F. Santosa and B. Weitz, “An inverse problem in reaction kinetics,” *J. Math. Chem.* **49**(8), 1507 (2011).
- ⁹⁸D. S. Mebane, K. S. Bhat, J. D. Kress, D. J. Fauth, McM. L. Gray, A. Lee, and D. C. Miller, “Bayesian calibration of thermodynamic models for the uptake of CO₂ in supported amine sorbents using *ab initio* priors,” *Phys. Chem. Chem. Phys.* **15**(12), 4355–4366 (2013).
- ⁹⁹P. Kugler, E. Gaubitzer, and S. Müller, “Parameter identification for chemical reaction systems using sparsity enforcing regularization: A case study for the chlorite–iodide reaction,” *J. Phys. Chem. A* **113**(12), 2775–2785 (2009).
- ¹⁰⁰C. Shih, J. Park, D. S. Sholl, M. J. Realff, T. Yajima, and Y. Kawajiri, “Hierarchical Bayesian estimation for adsorption isotherm parameter determination,” *Chem. Eng. Sci.* **214**, 115435 (2020).
- ¹⁰¹A. Hofinger and H. K. Pikkarainen, “Convergence rate for the Bayesian approach to linear inverse problems,” *Inverse Probl.* **23**(6), 2469 (2007).
- ¹⁰²A. Mohammad-Djafari, “A full Bayesian approach for inverse problems,” in *Maximum Entropy and Bayesian Methods: Santa Fe, New Mexico, USA, 1995, Proceedings of the Fifteenth International Workshop on Maximum Entropy and Bayesian Methods* (Springer, 1996), pp. 135–144.
- ¹⁰³C. W. Groetsch, “Integral equations of the first kind, inverse problems and regularization: A crash course,” *J. Phys.: Conf. Ser.* **73**(1), 012001 (2007).
- ¹⁰⁴C. W. Groetsch and C. W. Groetsch, *Inverse Problems in the Mathematical Sciences* (Springer, 1993), Vol. 52.
- ¹⁰⁵M. Sunnåker, A. G. Busetto, E. Numminen, J. Corander, M. Foll, and C. Dessimoz, “Approximate Bayesian computation,” *PLoS Comput. Biol.* **9**(1), e1002803 (2013).
- ¹⁰⁶A. Gelman, “Objections to Bayesian statistics,” *Bayesian Anal.* **3**, 445–450 (2008).
- ¹⁰⁷P. Mikkola, O. A. Martin, S. Chandramouli, M. Hartmann, O. A. Pla, O. Thomas, H. Pesonen, J. Corander, A. Vehtari, S. Kaski, P.-C. Bürkner, and A. Klami, “Prior knowledge elicitation: The past, present, and future,” *Bayesian Anal.* (published online) (2023).
- ¹⁰⁸J. Brynjarsdóttir and A. O’Hagan, “Learning about physical parameters: The importance of model discrepancy,” *Inverse Probl.* **30**(11), 114007 (2014).
- ¹⁰⁹C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (MIT Press, Cambridge, MA, 2006).

- ¹¹⁰Y. Ling, J. Mullins, and S. Mahadevan, "Selection of model discrepancy priors in Bayesian calibration," *J. Comput. Phys.* **276**, 665–680 (2014).
- ¹¹¹A. A. Neath and J. E. Cavanaugh, "The Bayesian information criterion: Background, derivation, and applications," *Wiley Interdiscip. Rev.: Comput. Stat.* **4**(2), 199–203 (2012).
- ¹¹²R. E. Kass and A. E. Raftery, "Bayes factors," *J. Am. Stat. Assoc.* **90**(430), 773–795 (1995).
- ¹¹³S. Theodoridis, *Machine Learning: A Bayesian and Optimization Perspective* (Academic Press, 2015).
- ¹¹⁴A. E. Gelfand, "Gibbs sampling," *J. Am. Stat. Assoc.* **95**(452), 1300–1304 (2000).
- ¹¹⁵H. Haario, E. Saksman, and J. Tamminen, *An Adaptive Metropolis Algorithm* (Bernoulli, 2001), pp. 223–242.
- ¹¹⁶S. A. Sisson, Y. Fan, and M. Beaumont, *Handbook of Approximate Bayesian Computation* (CRC Press, 2018).
- ¹¹⁷S. Sun, "A review of deterministic approximate inference techniques for Bayesian machine learning," *Neural Comput. Appl.* **23**, 2039–2050 (2013).
- ¹¹⁸B. Efron and R. Tibshirani, "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy," *Stat. Sci.* **1**, 54–75 (1986).
- ¹¹⁹H. T. Banks, K. Holm, and D. Robbins, "Standard error computations for uncertainty quantification in inverse problems: Asymptotic theory vs. bootstrapping," *Math. Comput. Modell.* **52**(9–10), 1610–1625 (2010).
- ¹²⁰N. Zhan and J. R. Kitchin, "Uncertainty quantification in machine learning and nonlinear least squares regression models," *AIChE J.* **68**(6), e17516 (2022).
- ¹²¹M. S. Gockenbach, *Linear Inverse Problems and Tikhonov Regularization* (American Mathematical Society, 2016), Vol. 32.
- ¹²²P. C. Hansen, "Truncated singular value decomposition solutions to discrete ill-posed problems with ill-determined numerical rank," *SIAM J. Sci. Stat. Comput.* **11**(3), 503–518 (1990).
- ¹²³W. B. Muniz, H. F. de Campos Velho, and F. M. Ramos, "A comparison of some inverse methods for estimating the initial condition of the heat equation," *J. Comput. Appl. Math.* **103**(1), 145–163 (1999).
- ¹²⁴F. Waqar and C. Simon. Github repository for "A tutorial on the Bayesian statistical approach to inverse problems," <https://github.com/faaiqgwaqar/Inverse-Problems>, 2023.