Inverse problems in thermoacoustics

Hans Yu

Supervisor: Matthew P. Juniper

Advisor: Luca Magri

Department of Engineering
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Trinity College August 2020

Schläft ein Lied in allen Dingen,
Die da träumen fort und fort,
Und die Welt hebt an zu singen,
Triffst du nur das Zauberwort.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Hans Yu
August 2020

Acknowledgements

The completion of this document marks the close to a very fulfilling and perhaps even the most stimulating chapter in my life so far. There are simply too many things and especially people to be grateful for. I am having difficulty finding the right words but I have to make an attempt nonetheless. So here I go.

First of all, I would like to thank my supervisor, Prof. Matthew Juniper. Obviously, his grasp of the mathematics and physics of thermofluid mechanics is superb. Less obviously, he also serves as an inspiration to me in all other aspects of being a good academic such as lucid writing, engaging presentations and leadership. What makes him stand out as a supervisor in my opinion, it is his ability to find the right balance between hands-on and hands-off guidance, which allows all of his students including myself to thrive according to our individual personalities. Englishman that he is, I am secretly pleased to imagine him cringing when reading my unabashedly genuine appraisal.

Next, I would like to thank my advisor, Dr. Luca Magri. I think of him as a second supervisor because it is so hard to overestimate his positive involvement in my work. There was a time after my first-year assessment when I felt slightly lost. He was there for me, and pointed me towards exciting ideas, which we would flesh out together and eventually present at conferences and workshops in Europe, North America as well as hopefully Australia soon. The generosity that he has shown me is not going to be forgotten.

Every day, I am immensely grateful for being part of this research group. They make me look forward to coming to the office despite the unreliable heating in the winter and the lack of ventilation in the summer. I have always enjoyed our lively discussions about all topics related and unrelated to science and technology, especially over lunch at 12:30 pm or teatime at 4 pm. I am not further withholding their names from the curious reader: Nick, José, Jack, Petr, Denis, Tejal, Filip, Ciccio, Stefano, Max, Ushnish and Alex. Although not a student anymore, I must not omit Ubello's name from this list as the current eighth inhabitant of ISO-40. Speaking for all inhabitants, we are always grateful for his wisdom, and that he puts up with our shenanigans.

Despite my fondness for the research group that I am in, when asked for my affiliation at conferences, I always answer the following: the Hopkinson Laboratory. With many of

us from far away to pursue our ambitions in Cambridge, the ups and downs of research naturally forge a bond between us. Be it late-night shifts or weekend sessions, it is always comforting to know that there is a friendly soul around. Or more joyous occasions such as birthdays, submissions and vivas, which give us reasons to celebrate over cake, with rounds at the Snug or at fancy formal dinners. In particular, I am thinking of amazing people like Pau, Anh Khoa, Francesca, Jenna, Maxim, Karla and Isabel. And not to forget the fantastic staff, for me in particular Kate Graham, Katia Babayan and Peter Benie, without whose continued support we would not be able to do the things we do. I am grateful for the friendships with those who have come before me and those that are leaving after me. It is my sincere hope that this atmosphere of comradery and hospitality transcending research groups and topics continues to grow in the foreseeable future.

And then there is Trinity College, a microcosm that I will never fully comprehend. The time spent at formal dinners, college events, the Mayball, BA committee meetings, basketball practices, and so on – it feels like one long, lucid dream. Nevertheless, the memories that stand out to me are the great (for me at least) conversations. I remember reading Der Teil und das Ganze when I was a teenager. I envied Werner Heisenberg for his interactions with all those exceptional people that were passionate to a fault about their fields but never failed to appreciate the philosophy and the aesthetics behind their own endeavors and others'. I guess that I do not have to feel envious anymore. In particular, I like to think back to the conversations that I have had with my college dad Jacob, my college wife Lydia, my favorite Germans Marcel and Steffi as well as my friends Annelise, Amar, Paul, Matilda, Parwana and Mikwi. With everyone on different paths in life, I do not wish for these conversations to ever cease.

I would also like to mention some people from previous chapters and side sections in my life because physical distance does not make my relationships with them less significant. The most formative years of my life were probably spent in Aachen, where I found wonderful friends like Anselm, Wolf and Marvin. What brought us together in the beginning were the initials of our last names in the range of W to Z. By now, I have become confident that our friendships are going to last for a lifetime. One thing that I had always taken for granted was the freedom to travel. I visited Ann Arbor in 2016, and met Malik and Amy again in Aachen and Cambridge three years later, respectively. Or visiting Stanford in 2018, where I got to meet young researchers from Munich, Bergen, Athens, Stockholm, Birmingham, Austin and Boston. I am not sure whether these experiences would have been possible today, and when that level of freedom is going to return.

Last but not least, I would like to thank my parents, Zhangzheng and Lei, from the bottom of my heart. Without them, I would not be where I am standing today.

Somewhat anticlimactically, I would also like to acknowledge not only the people but the institutions that have been vital to my research. I gratefully acknowledge the financial support by the CAMBRIDGE TRUST and SCHLUMBERGER GOULD RESEARCH through their SCHLUMBERGER CAMBRIDGE INTERNATIONAL SCHOLARSHIP as well as by TRINITY COLLEGE through their HONORARY EXTERNAL RESEARCH STUDENTSHIP. Furthermore, I am very grateful for the travel grants received from the DEPARTMENT OF ENGINEERING and the CAMBRIDGE PHILOSOPHICAL SOCIETY.

The research leading to this document would not have been possible without the open-source and open-knowledge communities. This document was typeset in LaTeX, with version control in GIT, on a computer powered by GNU/LINUX and UBUNTU. The drawings were generated in ASYMPTOTE, GRAPHVIZ DOT and PLANTUML. PYTHON served as my Swiss army knife: NUMPY, SCIPY and PANDAS for scientific computing; MATPLOTLIB and JUPYTER for (interactive) visualization; as well as SCIKIT-LEARN and SCIKIT-IMAGE for machine learning. If I had to name a single book that had left the deepest impression upon the research in this document, it would be INFORMATION THEORY, INFERENCE, AND LEARNING ALGORITHMS by the late PROF. DAVID MACKAY. It is a tour de force in Bayesian inference, and distributed for free. Similarly to my supervisors who regularly remind me to ‘be kind to the reader’, this and his other books have taught me two more things: (i) that elegance, rigor and fun do not have to be mutually exclusive; and (ii) that ideas and knowledge are meant to be shared.

Abstract

Thermoacoustics is a branch of fluid mechanics, and is as such governed by the conservation laws of mass, momentum, energy and species. While computational fluid dynamics (CFD) has entered the design process of many applications in fluid mechanics, its success in thermoacoustics is limited by the multi-scale, multi-physics nature of the subject. In his influential monograph from 2006, PROF. FRED CULICK writes about the role of CFD in thermoacoustic modeling:[1]

> *The main reason that CFD has otherwise been relatively helpless in this subject is that problems of combustion instabilities involve physical and chemical matters that are still not well understood. Moreover, they exist in practical circumstances which are not readily approximated by models suitable to formulation within CFD. Hence, the methods discussed and developed in this book will likely be useful for a long time to come, in both research and practice.*
>
> *[...] It seems to me that eventually the most effective ways of formulating predictions and theoretical interpretations of combustion instabilities in practice will rest on combining methods of the sort discussed in this book with computational fluid dynamics, the whole confirmed by experimental results.*

Despite advances in CFD and large-eddy simulation (LES) in particular, unsteady simulations for more than a few selected operating points are computationally infeasible. The 'methods discussed in this book' refer to reduced-order models of thermoacoustic oscillations. Whether intentional or not, the last sentence anticipates the advent of data-driven methods, and encapsulates the philosophy behind this work.

This work brings together two workhorses of the design process: physics-informed reduced-order models and data from higher-fidelity sources such as simulations and experiments. The three building blocks to all our statistical inference frameworks are: (i) a hierarchical view of reduced-order models consisting of states, parameters and governing equations; (ii) probabilistic formulations with random variables and stochastic processes; and

---

[1] Culick, Fred: *Unsteady Motions in Combustion Chambers for Propulsion Systems*. NATO Research and Technology Organisation, 2006.

(iii) efficient algorithms from statistical learning theory and machine learning. While leveraging advances in statistical and machine learning, we demonstrate the feasibility of Bayes' rule as a first principle in physics-informed statistical inference. In particular, we discuss two types of inverse problems in thermoacoustics: (i) implicit reduced-order models representative of nonlinear eigenproblems from linear stability analysis; and (ii) time-dependent reduced-order models used to investigate nonlinear dynamics. The outcomes of statistical inference are improved predictions of the state, estimates of the parameters with uncertainty quantification and an assessment of the reduced-order model itself.

This work highlights the role that data can play in the future of combustion modeling for thermoacoustics. It is increasingly impractical to store data, particularly as experiments become automated and numerical simulations become more detailed. Rather than store the data itself, the techniques in this work optimally assimilate the data into the parameters of a physics-informed reduced-order model. With data-driven reduced-order models, rapid prototyping of combustion systems can feed into rapid calibration of their reduced-order models and then into gradient-based design optimization. While it has been shown, e.g. in the context of ignition and extinction, that large-eddy simulations become quantitatively predictive when augmented with data, the reduced-order modeling of flame dynamics in turbulent flows remains challenging. For these challenging situations, this work opens up new possibilities for the development of reduced-order models that adaptively change any time that data from experiments or simulations becomes available.

The reader in mind is a scientist or engineer with an interest in data-driven methods. For readers mostly interested in the results, we provide references to our ideally more self-contained publications where available. For the more methodological chapters, we provide JUPYTER notebooks so that inclined readers are able to familiarize themselves with the statistical and numerical concepts of this work. They are either available on GITLAB[2] for download or as a BINDER[3] executed within the browser. More information on JUPYTER notebooks are found online.[4]

---

[2] https://gitlab.com/hyu/phd-thesis-tutorials

[3] https://mybinder.org/

[4] https://jupyter-notebook.readthedocs.io/en/stable/

Table of contents

List of figures

List of tables

List of Algorithms

Nomenclature

Roman Symbols

adj	adjugate
$\mathbb{C}$	complex numbers
CFL	Courant-Friedrich-Lewy number
det	determinant
$\mathbb{E}$	expected value
E	identity matrix
Gr	Grashof number
k	kernel (covariance function) of a stochastic process
m	mean of a stochastic process
$\mathscr{M}$	Markstein length
$\mathscr{N}$	normal distribution
Nu	Nusselt number
P	probability; probability density
Pe	Péclet number
Pr	Prandtl number
$\mathbb{R}$	real numbers
$\mathbb{R}^+$	non-negative real numbers

Ra	Rayleigh number
Re	Reynolds number
Re	real part
sgn	sign function
tr	trace

Greek Symbols

δ	Kronecker delta
Φ	cumulative distribution function of normal distribution
φ	probability density function of normal distribution

Superscripts

$\dagger$	adjoint
a	analysis
$*$	complex conjugate
f	forecast
$+$	pseudoinverse

Subscripts

bf	base flow
ev	evidence
fr	flame response
L	laminar
obs	observations

Other Symbols

$\hat{\cdot}$	optimal
$\check{\cdot}$	pseudo

$\tilde{\cdot}$	augmented

Acronyms / Abbreviations

ARD	automatic relevance determination
CFD	computational fluid dynamics
DNS	direct numerical simulation
EM	expectation-maximization
FDF	flame describing function
FIML	field inversion and machine learning
FTF	flame transfer function
LES	large-eddy simulation
MAP	maximum-a-posteriori
MCMC	Markov-Chain Monte Carlo
MFC	mass flow controller
ML	maximum-likelihood
MP	message-passing
RANS	Reynolds-averaged Navier-Stokes
RMS	root mean square
RTS	Rauch-Tung-Striebel
SE	squared-exponential
TVD	total-variation diminishing
WENO	weighted essentially non-oscillatory

Part I

Introduction

Chapter 1

Thermoacoustics

This chapter is a literature review of thermoacoustics with an emphasis on reduced-order models. In Chapter 1.1, we discuss the progress and interplay between theoretical models, computational methods and experimental techniques. We give an outline of the thesis based on this idea. In Chapter 1.2, we provide an overview of reduced-order models beyond computational fluid dynamics. In Chapter 1.3, we discuss existing tools that elucidate the role of parameters in reduced-order models. This sets the stage for the so-called inverse problem, a statistical perspective on reduced-order modeling, which is introduced in the next chapter (Chapter 2).

1.1 Thermoacoustic modeling

Thermoacoustic oscillations (also known as combustion instabilities) are a persistent challenge in the design of jet and rocket engines [1, 2]. They arise inside combustion chambers from various interactions between flames, sound and the flow, leading to structural damage in the worst case. They must not occur during operation under any critical circumstances, and thus have to be mitigated at the design stage. An example of the costs associated with the inability to mitigate thermoacoustic oscillations is given by the development of the F-1 engine for the SATURN V rocket, which required more than 3,200 full-scale tests, 2,000 of which were to eliminate thermoacoustic oscillations [2]. Even today, the onset of thermoacoustic oscillations is difficult to predict, and their existence holds back the exploration of favorable operating regimes in gas turbine engines. Robust prediction would pave the way for the development of engines with higher performance, lower emissions and less noise.

While the notoriety of thermoacoustic oscillations began in the middle of the twentieth century with the invention of jet and rocket engines due to their high energy densities, the phenomenon had already been observed by artisans and experimentalists in the beginning of the nineteenth century [3, 4]. LORD RAYLEIGH gave an explanation in as early as 1878 [5]:

> *If heat be periodically communicated to, and abstracted from, a mass of air vibrating (for example) in a cylinder bounded by a piston, the effect produced will depend upon the phase of the vibration at which the transfer of heat takes place. If heat be given to the air at the moment of greatest condensation, or taken from it at the moment of greatest rarefaction, the vibration is encouraged. On the other hand, if heat be given at the moment of greatest rarefaction, or abstracted at the moment of greatest condensation, the vibration is discouraged.*

While the so-called Rayleigh criterion has since then been formalized and extended, the physical explanation still holds true today [6–8].

With a physical explanation of the interaction between flames and sound as well as the governing equations of reacting flows at our disposal, this raises the question as to why the reliable prediction of thermoacoustic oscillations remains so elusive. When viewed through the lens of scientific inquiry, the scientific method encounters practical challenges at each step (Fig. 1.1):

Theory. A fully resolved simulation of the governing equations of fluid mechanics, also known as direct numerical simulation (DNS), including turbulence, combustion and acoustics is computationally infeasible for industrial applications [9]. Therefore, approaches based on computational fluid dynamics (CFD) filter the governing equations according to physical length and time scales, whose a-priori identification is challenging even if only a subset of possibly interacting physical mechanisms is considered, e.g. aeroacoustics, gas dynamics, multi-phase flow, turbulent combustion or combustion noise [10–16]. In simulations based on the Reynolds-averaged Navier-Stokes (RANS) equations, the unsteadiness due to turbulence is removed by applying time or ensemble averaging and modeling the resulting Reynolds stress tensor. Despite their shortcomings for transient flows, RANS simulations are still considered state-of-the-art for industrial applications as a compromise between accuracy, performance and ease of use [17–19]. In large-eddy simulations (LES), low-pass filters are applied so that only the sub-grid scales require modeling. While the higher resolution compared to RANS simulations makes LES more suitable for targeting complex geometries, detailed chemistry and transient behavior, computational time and cost compared to non-CFD approaches hinder their widespread adoption in industry [20–22]. Non-CFD approaches are reviewed in the next section (Chapter 1.2). In summary, both CFD and non-CFD approaches to thermoacoustics are not only governed by first principles but represent reduced-order models with closures.

Prediction. The vision of computational fluid dynamics, and DNS in particular, is expressed in a two-hundred year old thought experiment popularly known as Laplace's demon [23, Chapter 2]:

> *We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes.*

In reality, predictions are compromised by a practical lack of access to governing equations as well as initial and boundary conditions. Instead, a reduced-order model is used, whose parameters are necessarily not universal constants but are based on scaling arguments and fitted from laboratory experiments. For example, turbulence closures based on eddy viscosity rely on parameters such as Prandtl mixing length, static and dynamic Smagorinsky constants or others, whose values are chosen according to the flow configuration [24, Chapters 8 to 13]. In LES of flame dynamics and thermoacoustics, it has been observed that predictions strongly depend on the choice of boundary conditions [25–27]. For acoustic boundary conditions, the widely used choice of numerically non-reflecting boundary conditions requires relaxation in order to predict realistic levels of resonance [25]. For thermal boundary conditions, widely used choices including (i) constant-temperature (Dirichlet), (ii) adiabatic (Neumann) and (iii) heat-transfer (Robin) formulations are outperformed in predicting frequency and amplitude of the unstable mode by LES coupled with conjugate heat transfer [27]. In general, systematic investigations into the impact of model parameters on the predictions provide useful insight into the reduced-order model: Predictions which depend too strongly on the choice of model parameters in order to achieve acceptable accuracy call into question the validity of the reduced-order model in the first place. Both quantitative and qualitative methods for understanding the role of model parameters are reviewed in the last section of this chapter (Chapter 1.3).

Validation. The degree to which a reduced-order model is validated is limited by the accuracy and precision of the corresponding experiment. In thermoacoustics, the following two reasons complicate the acquisition of high-quality data [2]:

1. It is very difficult to take detailed measurements of the internal flow inside combustion chambers due to the high pressures and temperatures.
2. The physical mechanisms involved in thermoacoustic oscillations, including turbulent combustion, acoustics and spray dynamics, are potentially numerous and highly unsteady.

In particular, there are difficulties associated with the (simultaneous) measurements of unsteady quantities such as fluctuating heat release rates, growth rates in the unsteady case as well as spatial distributions of gas and wall temperatures. The results are incomplete, inaccurate or imprecise data. Low-quality data poses a significant challenge, both for the scientific method with humans in the loop and data-driven methods automatically executed by machines.

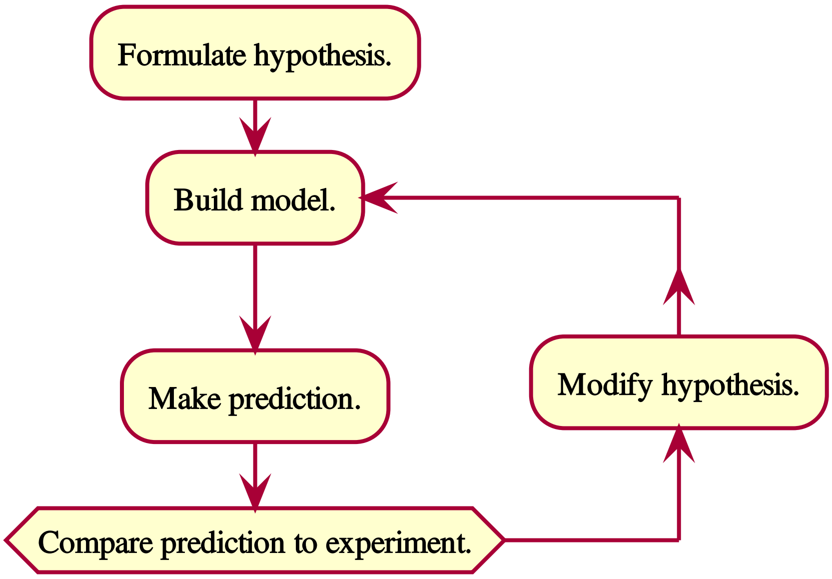


Fig. 1.1 The scientific method ideally follows three steps: (i) In a theory step, a mathematical model is derived based on a hypothesis. (ii) In a prediction step, the model is solved, either analytically or numerically. (iii) In a validation step, the outcome of an experiment informed by the hypothesis is compared to the model prediction. The three steps of the scientific method form a feedback loop: If the model and the experiment do not sufficiently agree, the hypothesis is modified and the procedure is repeated.

In summary, (i) reduced-order models, (ii) predictions based on the reduced-order models as well as (iii) experiments to validate the reduced-order models all fail to offer the highest level of resolution due to various constraints. This hinders scientists and engineers in fully understanding the behavior of the system under investigation. The aim of this work is to develop frameworks that rigorously combine the imperfect information from theory, computation and experiment in order to enable objectively more accurate and trustworthy predictions. Therefore, we adopt the abstract viewpoint of solving so-called inverse problems, which are introduced in the next chapter and formulated in the language of probability theory and statistics (Chapter 2). In Parts II and III, we explore two classes of thermoacoustic models and formulate the inverse problems, respectively. In Part IV, we summarize the results and provide an outlook for the statistical perspective on thermoacoustic modeling.

1.2 Reduced-order models

For this literature review, we distinguish between reduced-order models of thermoacoustics in the linear and nonlinear regimes, respectively. The linear regime describes the onset of the thermoacoustic oscillations, and is characterized by exponential growth in amplitude. The nonlinear regime describes thermoacoustic oscillations with saturated amplitude, and exhibits varying degrees of nonlinear dynamics. This distinction between linear and nonlinear dynamics is a common approach to flow instabilities [28, Chapter 1].

In the linear regime, we begin by linearizing the governing equations of fluid mechanics. In the simplest form, the linearized governing equations for momentum and energy in one dimension become [29]

$$\rho_0 \frac{\partial u_1}{\partial t} + \frac{\partial p_1}{\partial x} = 0 \quad , \tag{1.1}$$

$$\frac{\partial p_1}{\partial t} + \gamma p_0 \frac{\partial u_1}{\partial x} = (\gamma - 1) q_1 \quad , \tag{1.2}$$

where the density is denoted by ρ, the velocity in the x-direction by u, the pressure by p, the ratio of specific heats by γ and the heat release rate by q. The subscript indices denote the order of perturbation. Combining Eq. (1.1) and (1.2), we obtain

$$\frac{\partial^2 p_1}{\partial t^2} - c_0^2 \frac{\partial^2 p_1}{\partial x^2} = (\gamma - 1) \frac{\partial q_1}{\partial t} \quad , \tag{1.3}$$

where $c = \sqrt{\gamma p / \rho}$ denotes the speed of sound. Eq. (1.3) is a wave equation, which elucidates the acoustic nature of Eq. (1.1) and (1.2). Depending on the assumptions, more general wave equations are derived, for example including indirect noise due to entropy, vorticity or composition fluctuations [30–32]. For illustration purposes, we continue to refer to Eq. (1.1) and (1.2) throughout this section as their salient features are representative of analysis in the linear regime [29].

The heat release rate q_1 in Eq. (1.2) encompasses the complex interaction between flames, sound and the flow, and therefore requires closure. In the simplest case, the closure assumes the form of an n-τ model [33, Chapter 1]:

$$q_1(x,t) = nu_1(x,t-\tau) \quad , \tag{1.4}$$

where n denotes the interaction index, and τ denotes the time delay. Physically speaking, the flame is treated as a black box with u_1 as input and q_1 as output. Alternatively, the relationship between u_1 and q_1 is represented by a flame transfer function, where a gain and a phase are recorded at each forcing frequency [34–36]. This corresponds to a Fourier transform of the n-τ model where n and τ become functions of forcing frequency. Various extensions are possible: In a flame transfer matrix, the flame as a black box has multiple ports, e.g. for u_1 and p_1, compared to the n-τ model which has a single input port for u_1 [37]. In a flame describing function, gain and phase are not only functions of forcing frequency but also of amplitude [38–42]. This extension to the flame transfer function describes nonlinear phenomena such as saturation, bistability and hysteresis. In general, thermoacoustic oscillations in the linear regime are sufficiently well described by linear acoustics, i.e. linear in the acoustic pressure p_1, but require a nonlinear model for the heat release rate, e.g. nonlinear in the time delay τ [43].

Linear stability analysis of Eq. (1.1) and (1.2) requires the treatment of a nonlinear eigenproblem[1] [28, Chapter 2]. The three principal tools in thermoacoustics for the derivation of these eigenproblems are (i) traveling waves, (ii) Galerkin modes and (iii) Helmholtz modes [29, Tutorial 1]. Traveling waves are characteristics of the linearized governing equations subject to jump and boundary conditions [44, Chapter 11]. For acoustic elements of constant cross section with homogeneous mean flow, the computation of the characteristics is straightforward. Industrial applications with complex geometries are modeled as thermoacoustic networks, i.e. series of acoustic elements [45]. The jump conditions between the acoustic elements are derived from the linearized Euler equations. The forward- and backward-traveling waves are reflected according to the boundary conditions, and form standing waves that grow or decay in time. The eigenvalues of the thermoacoustic network, i.e. the growth rates and frequencies of the standing waves such that all jump and boundary conditions are satisfied, are found iteratively using shooting methods [46, Chapter 4]. Galerkin modes project the solution to the linearized governing equations onto a vector space formed of acoustic modes [47]. The partial differential equations become an initial value problem, i.e. a set of ordinary

---

[1] Nonlinear eigenproblems are nonlinear in the eigenvalue but linear in the eigenvector. A nonlinear eigenproblem in thermoacoustics arises when there is a time delay between the heat release rate q_1 and the acoustic pressure p_1 or the acoustic velocity u_1.

differential equations in time. In practice, Galerkin modes are difficult to compute for at least two reasons [48]: Firstly, it is not trivial to construct a basis for the vector space unless the acoustic boundary conditions are trivial, e.g. open or closed ends. Secondly, the mode shape of the thermoacoustic oscillation is not necessarily well approximated by acoustic modes, depending on the activity of the flame. Helmholtz modes are the Laplace transform of the solution to the linearized governing equations [48]. The partial differential equations become a boundary value problem, i.e. a set of partial differential equations in space but not in time. Numerical methods to solve the resulting large nonlinear eigenproblems are based on fixed-point iteration and Krylov solvers [48]. In annular combustion chambers which exhibit a discrete rotational symmetry, it is sufficient to simulate a single burner section if Bloch-periodic boundary conditions are used in the azimuthal direction [49]. Treatment of nonlinear eigenproblems as inverse problems is the subject of Part II.

In the nonlinear regime, linear stability analysis is extended by weakly nonlinear stability analysis [28, Chapter 5]. For example, an autonomous dynamical system in one dimension is given by the ordinary differential equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x) \quad . \tag{1.5}$$

In a linear stability analysis, the right-hand side is linearized around a fixed point x_0 with $f(x_0) = 0$:

$$\frac{\mathrm{d}x}{\mathrm{d}t} \approx \left.\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x_0} (x - x_0) \quad . \tag{1.6}$$

If (the real part of) $\mathrm{d}f/\mathrm{d}x$ is negative, the fixed point x_0 is linearly stable. If (the real part of) $\mathrm{d}f/\mathrm{d}x$ is positive, the fixed point x_0 is linearly unstable. In a weakly nonlinear stability analysis, the right-side is expanded into a higher-order polynomial. For example, the third-order expansion

$$\frac{\mathrm{d}x}{\mathrm{d}t} \approx x - x^3 \tag{1.7}$$

has two linearly stable fixed points at $x = 1$ and $x = -1$ as well as one linearly unstable fixed point at $x = 0$. However, the basin of attraction switches from $x = -1$ to $x = 1$ or vice versa whenever a perturbation causes x to switch signs. This serves as a prototype for thermoacoustic bistability, a phenomenon exacerbated by the presence of non-normal modes with transient growth and triggering [50–52].

Alternatively, the nonlinear regime is observed in time series obtained by numerical integration of reduced-order models, e.g. using the G-equation. The G-equation is a kinematic description of premixed flames with prescribed velocity fields [53, 39]. It generates realistic flame transfer functions and flame describing functions [35, 36, 54–56]. When coupled with a simple representation of the acoustics, e.g. one-dimensional Galerkin modes, the reduced-order model reproduces period doubling, quasiperiodicity and chaos [57]. The time-accurate calculation of the G-equation is discussed in detail in Part III. While thermoacoustic chaos has been observed in both laboratory experiments and industrial applications, thermoacoustic limit cycles characterized by an amplitude and a frequency are most relevant in practice [58–60]. Limit cycles, even the unstable ones, are efficiently identified with continuation methods and visualized in phase portraits [61, 62]. Their stability is determined by Floquet analysis, which is the equivalent to linear stability analysis for fixed points [63, Chapter 8]. More generally, the stability of a dynamical system is determined by covariant Lyapunov vector analysis [64].

1.3 Model parameters

Model parameters, as opposed to universal constants or design parameters, introduce considerable uncertainty into reduced-order models. In thermoacoustics, a model parameter particularly prone to uncertainty is the time delay τ (Eq. (1.4)). It is an amalgamation of different intricate physical mechanisms, e.g. injection, atomization, vaporization, mixing and reaction in the case of a liquid propellant [65]. As such, it is extremely sensitive to the operating conditions of the combustion chamber. In turn, the onset of thermoacoustic oscillations shifts according to the Rayleigh criterion as the time delay affects the phase between fluctuations in heat release rate and pressure [33, Chapter 1]. While extreme sensitivity is potentially exploited in the design and control of combustion chambers, all reduced-order models short of DNS suffer from the corresponding uncertainty [29].

We attribute the uncertainty in reduced-order models to two sources [66]:

Parametric uncertainty. The quantification of parametric uncertainty starts from the assumption that the reduced-order model and its parametrization are perfectly valid over the entire design space. Then, the values of the model parameters are estimated from laboratory experiments. The confidence in their values depends on the precision of the noisy data. Parametric uncertainty falls under the category of aleatoric uncertainty, also known as statistical error [67]. Although the exact values of the model parameters are unknown, there is no doubt that they are well defined and obtainable to arbitrary precision in the limit of infinite, accurate data.

Structural uncertainty. Structural uncertainty goes further than parametric uncertainty, and questions the validity of the reduced-order model itself. Even if the model parameters are estimated to arbitrary precision, structural uncertainty exists if there is significant bias between predictions made by the reduced-order model and experimental measurements. In a first instance, the model parameters become functions of the state variables, e.g. of pressure or temperature, and now vary in space and time. In a second instance, additional degrees of freedom are introduced in order to relax the equations of the reduced-order model. Structural uncertainty falls under the category of epistemic uncertainty, also known as systematic error [67]. This means that structural uncertainty is rooted in the ignorance of the modeler, and thus arguably more difficult to identify and mitigate.

One tool for uncertainty propagation, i.e. studying the effect of the uncertainty in the reduced-order model on its predictions, is sensitivity analysis. Sensitivity analysis is well-established in the study of hydrodynamic and thermoacoustic instability [68–70]. Sensitivity is traditionally understood as the response to perturbations that are either deliberate, e.g. changes in the context of design or control, or stochastic, e.g. fluctuations in turbulent flow. Nevertheless, sensitivity analysis is also meaningful in the context of reduced-order models and uncertainty quantification.

From a system-theoretic point of view, sensitivity analysis is embedded into an optimization problem with constraints [71, Chapter 2]:

$$\text{minimize} \quad J(x,\theta) \quad , \tag{1.8}$$

$$\text{subject to} \quad R(x,\theta) = 0 \quad , \tag{1.9}$$

where x denotes the variables characterizing the state of the system, and θ denotes the model parameters. The governing equations R of the reduced-order model, e.g. conservation laws or the linearized governing equations (1.1) and (1.2), constrain the minimization of the quantity of interest J, e.g. cost, error or a growth rate. In an optimization problem, the role of sensitivity analysis is to find the direction of gradient ascent (Eq. (1.8)):

$$\frac{\mathrm{d}J}{\mathrm{d}\theta} = \frac{\partial J}{\partial \theta} + \frac{\partial J}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}\theta} \quad , \tag{1.10}$$

where the Jacobian $\mathrm{d}x/\mathrm{d}\theta$ is implicitly given by (Eq. (1.9))

$$\frac{\partial R}{\partial \theta} + \frac{\partial R}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}\theta} = 0 \quad . \tag{1.11}$$

In direct sensitivity analysis, the sensitivity equation (1.11) is solved either directly or approximately via finite differences. In adjoint sensitivity analysis, the adjoint equation (1.12) is alternatively solved for the Lagrange multiplier ϕ:

$$\frac{\partial J}{\partial x} + \phi^T \frac{\partial R}{\partial x} = 0 \quad , \tag{1.12}$$

where the Lagrange multiplier ϕ is the adjoint to the state vector x. Hence, the sensitivity $\mathrm{d}J/\mathrm{d}\theta$ is given by (Eq. (1.10), (1.12), (1.11))

$$\begin{aligned} \frac{\mathrm{d}J}{\mathrm{d}\theta} &= \frac{\partial J}{\partial \theta} + \frac{\partial J}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}\theta} && (1.13) \\ &= \frac{\partial J}{\partial \theta} - \phi^T \frac{\partial R}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}\theta} && (1.14) \\ &= \frac{\partial J}{\partial \theta} + \phi^T \frac{\partial R}{\partial \theta} \quad . && (1.15) \end{aligned}$$

The details regarding the implementation, e.g. the computational costs associated with direct and adjoint sensitivity analysis, the pitfalls of continuous and discrete adjoints and the possibly automated development of adjoint solvers, are noteworthy but beyond the scope of this discussion [72]. Instead, we focus on the physical insight from sensitivity analysis with respect to reduced-order models.

The relationship between sensitivity analysis and parameteric uncertainty is obvious. If θ represents design parameters, sensitivity analysis reveals the optimal modifications that result in stable operation for example [73–75]. At the same time, it quantifies the uncertainty due to manufacturing tolerances or coarsely resolved geometric features in the reduced-order model. If θ represents model parameters, sensitivity analysis reduces the computational cost of uncertainty quantification via identification of active subspaces and construction of surrogate models [76, 77]. The insight from sensitivity analysis into the structural uncertainty of a reduced-order model is less obvious, and subject of the following discussion.

In a first instance, we assume we have neglected a physical mechanism in the reduced-order model. A first-order correction to the reduced-order model is given by (Eq. (1.9))

$$R(x,\theta) + Sx = 0 \quad , \tag{1.16}$$

where S is a zero-valued matrix in the reduced-order model without correction. One example in thermoacoustics for such a first-order correction to a reduced-order model is the concept of sensitive time-lag introduced for liquid propellants, where the time delay τ is not assumed constant but to be a function of the acoustic pressure p_1 [33, Chapter 1]. The sensitivity of J

with respect to the elements of S is given by the matrix (Eq. (1.15))

$$\frac{\mathrm{d}J}{\mathrm{d}S} = \phi x^T \quad , \tag{1.17}$$

which is also known as the structural sensitivity [78]. In hydrodynamics, structural sensitivity is able to quantify the uncertainty arising from the base flow [79]. The uncertainty in the eigenvalues is represented by a pseudospectrum, and calculated from the corresponding resolvent operator [80, Chapter 2]. In thermoacoustics, structural sensitivity has been used to devise closed-loop control based on the optimal placement of dampers [81–83]. At the same time, it is able to account for uncertainties in reduced-order models due to neglected physical mechanisms, e.g. acoustic attenuation in Eq. (1.3).

In a second instance, we call the validity of the governing equations of the reduced-order model into question. We relax the constraints by admitting a residual σ (Eq. (1.9)):

$$R(x,\theta) + \sigma = 0 \quad , \tag{1.18}$$

where σ vanishes in a perfectly valid reduced-order model. The residual σ would represent a source or a sink term in a conservation law, or forcing in a harmonic oscillator, just to name a few actuation mechanisms. The sensitivity of J with respect to the residual σ is given by (Eq. (1.15))

$$\frac{\mathrm{d}J}{\mathrm{d}\sigma} = \phi^T \quad , \tag{1.19}$$

which is also known as the receptivity [78]. In CFD, receptivity analysis reveals the regions in the computational domain where the quantity of interest reacts most sensitively to discretization error, and thus serves as an indicator for the adaptive refinement of the computational mesh in order to mitigate the structural uncertainty in the discretized governing equations [84]. In thermoacoustics, receptivity analysis has been used to devise open-loop control, e.g. based on secondary injection of fuel into a diffusion flame [82]. In general, it is able to account for uncertainties in reduced-order models due to neglected inputs and outputs, e.g. bleed flows, air films and heat loss in turbomachinery.

In summary, sensitivity analysis is a versatile tool for uncertainty propagation in the linear regime. In the nonlinear regime, stochastic approaches include Monte-Carlo simulations and polynomial chaos [85]. In a polynomial chaos expansion, the solution to the reduced-order model is considered a realization of a random variable, which is projected onto a basis of uncorrelated random variables with the corresponding coordinates. In an intrusive approach, a projection of the governing equations is solved in each coordinate. In a non-intrusive approach, the governing equations of the reduced-order model are solved for each sample of the polynomial chaos. Polynomial chaos for uncertainty quantification has been successfully adopted by the CFD, combustion and aeroelasticity communities among others [86–88]. In thermoacoustics, stochastic approaches have been applied to reduced-order models of turbulent swirl combustors [89–91]. While capable of uncertainty quantification in the nonlinear regime, polynomial chaos and, to a lesser degree, Monte-Carlo simulations suffer from the curse of dimensionality unlike adjoint sensitivity analysis. In general, they rely on dimensionality reduction techniques such as active subspaces, which are beyond the scope of this discussion [92].

We conclude that the state of the art in thermoacoustic modeling incorporates a variety of tools, combining reduced-order models based on first principles with uncertainty propagation based on variational or stochastic approaches. At the same time, the fluid mechanics community is appropriating non-traditional, data-driven approaches such as machine learning [93, 94]. In areas such as computer vision and natural language processing, machine learning has unprecedentedly accomplished human and superhuman prediction accuracy despite the lack of input in the form of governing equations or first principles. This thesis is meant to be a contribution towards the efforts of the fluid mechanics community to reconcile data-driven approaches, which are widely perceived as black boxes, with the tradition of physics-informed reduced-order modeling. As a precursor to machine learning, we introduce statistical learning to thermoacoustic models, i.e. statistical inference using Bayes' rule (Chapter 2). More or less surprisingly, statistical learning is a versatile tool whose foundation in probability theory and statistics enables it to formulate and solve a number of seemingly unrelated tasks in a natural manner, e.g. parameter estimation, uncertainty quantification and model comparison. After Parts II and III, where we perform statistical inference for two types of thermoacoustic models, we return to this discussion in the concluding chapter (Chapter 10).

Chapter 2

Inverse problems

Towards the end of Chapter 1, we commented on the uncertainty present in reduced-order models without providing a definition. This chapter reviews the language of probability theory in order to continue with a more educated discussion of statistical inference and the so-called inverse problem in terms of probability distributions and their associated statistics. This discussion is necessary because it enables the development of data-driven techniques throughout the rest of this thesis. For the interpretation of data and the incorporation into our reduced-order models, we rely on probability theory to provide us with first principles for statistical inference in contrast to ad-hoc statistics such as bias or variance. In particular, we substantiate the notions of belief, randomness and knowledge beyond thermoacoustics.

Supplementary material:

- `Chapter2/Statistical inference.ipynb` (Git, Binder).
- `Chapter2/Stochastic processes.ipynb` (Git, Binder).

2.1 Statistical inference

In science and engineering, we introduce reduced-order models due to various constraints (Chapter 1.1). The governing equations of a reduced-order model deviate from the governing equations of the natural phenomenon. Hence, the predictions by the reduced-order model are never exact. Thus, it is not meaningful to call a reduced-order model or any of its predictions true or false; they are by definition always false. Alternatively, we quantify our degree of belief within a non-Boolean framework[1] with the following requirements [95, Chapter 1]:

---

[1] A Boolean framework is characterized by binary variables holding one of two possible values, i.e. 0 or 1, true or false, etc.

- A degree of belief is represented by a real number.
- A degree of belief qualitatively varies according to 'common sense'. For example: If a prediction A ('The value lies between 0 and 1.') is *more* plausible than a prediction B ('The value lies between 1 and 2.'), then the negated prediction $\overline{A}$ ('The value *does not* lie between 0 and 1.') is *less* plausible than the negated prediction $\overline{B}$ ('The value *does not* lie between 1 and 2.').
- A degree of belief relies only on the knowledge available, not on the mode of reasoning. Although belief has a subjective component by definition, the degree of belief is well-defined once our subjective and objective knowledge is quantified.

Cox' theorem states that such a belief framework must follow the rules of probability theory [96]. The task of statistical inference is to assess a quantity of interest related to our reduced-order model based on our corresponding degree of belief.

Probability theory as the language of statistical inference is somewhat controversial. This controversy is rooted in two complementary interpretations of probability and inference [97, Chapter 1]: Frequentist probability assumes that the degree of belief is approximated by the relative frequency in a repeated experiment. The task of frequentist inference is to construct estimators for the quantity of interest from the experimental data, and to devise metrics for these estimators such as confidence intervals or p-values. Bayesian probability goes further in assuming that the degree of belief is also conditional on knowledge independent of the experiment. Bayesian inference computes the inverse probability, i.e. the degree of belief in our knowledge conditional on the experimental data. A settlement of this controversy is beyond the scope of this discussion. In practice, they are not mutually exclusive, and subject to personal view [97, Appendix B]. Nevertheless, we examine several arguments in favor of the Bayesian interpretation over the frequentist interpretation within the context of reduced-order models: Firstly, a prediction from a reduced-order model is strictly speaking not frequentist as a repetition of the derivation or computation does not necessarily increase confidence. Secondly, it is not always straightforward to construct estimators, especially if the relationship between the quantity of interest and the measured quantity is nonlinear. Thirdly, the belief in a reduced-order model does not stem from a single experiment but a variety of sources and measurements, e.g. numerical simulations, laboratory experiments as well as on-line calibration. Therefore, we adopt the Bayesian interpretation, and henceforth use the expressions 'degree of belief' and 'probability' interchangeably.

A brief summary of the relevant definitions and theorems (without proofs) from probability theory leading to Bayes' rule is provided below [95, Chapter 2]. They form the calculus of statistical inference, and are referenced extensively throughout the rest of the thesis. Note that we assume no previous knowledge of measure theory [98, Chapter 2]. For example, we do not distinguish between sample spaces and event spaces, i.e. probabilities and probability densities. For countable sample spaces, this is not problematic, and the power set of the sample space serves as an event space. For uncountable sample spaces such as $\mathbb{R}^n$, measures and thus probabilities are only defined on certain event spaces that are all strict subspaces of the power set of $\mathbb{R}^n$. Nevertheless, we exchange rigor in favor of clarity for a wider audience, and instead provide the more mathematically inclined reader with relevant footnotes.

Definition 2.1 (Random variable)**.** A random variable $X : \Omega \to \Omega'$ is a function between the two sample spaces Ω and Ω'.[2] The probability P of an event $x \in \Omega'$ is given by

$$P(X = x) = \mu\left(\{\omega \mid X(\omega) = x\}\right) \quad , \tag{2.1}$$

where $\mu : \Omega \to \mathbb{R}^+$ denotes a real measure on the sample space Ω with $\mu(\Omega) = 1$.[3]

Definition 2.2 (Joint probability)**.** The joint probability of two random variables $X : \Omega \to \Omega_X$ and $Y : \Omega \to \Omega_Y$ is defined as

$$P(X = x, Y = y) = \mu\left(\{\omega \mid X(\omega) = x\} \cap \{\omega \mid Y(\omega) = y\}\right) \quad . \tag{2.2}$$

It quantifies the degree of belief in simultaneous $X = x$ and $Y = y$.

Theorem 2.1 (Sum rule)**.** *The relationship between the joint probability distribution $P(X,Y)$ and the marginal probability distributions $P(X)$ and $P(Y)$ is given by*[4]

$$P(X = x) = \sum_{y \in \Omega_Y} P(X = x, Y = y) \quad , \tag{2.3}$$

$$P(Y = y) = \sum_{x \in \Omega_X} P(X = x, Y = y) \quad . \tag{2.4}$$

---

[2] A random variable $X : (\Omega, \mathfrak{A}) \to (\Omega', \mathfrak{B})$ is a measurable function, where Ω and Ω' are sample spaces, $\mathfrak{A}$ and $\mathfrak{B}$ are event spaces and $(\Omega, \mathfrak{A}, \mu)$ forms a probability space with the real measure $\mu : \mathfrak{A} \to \mathbb{R}^+$, i.e. $\mu(\Omega) = 1$.

[3] The real measure $P : \mathfrak{B} \to \mathbb{R}^+$ on the event space $\mathfrak{B}$ is induced by the random variable X.

[4] For uncountable sample spaces such as $\mathbb{R}^n$, replace summation with integration [98, Chapter 4]:

$$P(X = x) = \int_{\Omega_Y} P(X = x, Y = y)\,\mathrm{d}y \quad , \quad P(Y = y) = \int_{\Omega_X} P(X = x, Y = y)\,\mathrm{d}x \quad .$$

Definition 2.3 (Conditional probability)**.** The conditional probability of two random variables $X : \Omega \to \Omega_X$ and $Y : \Omega \to \Omega_Y$ is defined as

$$P(Y = y \mid X = x) = \frac{\mu\left(\{\omega \mid X(\omega) = x\} \cap \{\omega \mid Y(\omega) = y\}\right)}{P(X = x)} \quad . \tag{2.5}$$

It quantifies the degree of belief in $Y = y$ given $X = x$.

Theorem 2.2 (Product rule)**.** *The relationship between the joint probability distribution $P(X,Y)$ and the conditional probability distributions $P(Y \mid X)$ and $P(X \mid Y)$ is given by*

$$P(X = x, Y = y) = P(X = x)P(Y = y \mid X = x) \tag{2.6}$$

$$= P(Y = y)P(X = x \mid Y = y) \quad . \tag{2.7}$$

Definition 2.4 (Statistical independence)**.** Two events $X = x$ and $Y = y$ are statistically independent if

$$P(X = x, Y = y) = P(X = x)P(Y = y) \quad . \tag{2.8}$$

Two random variables X and Y are statistically independent if they are independent for all events $x \in \Omega_X$ and $y \in \Omega_Y$. More specifically, it follows from the product rule that for $P(X = x) \neq 0$ (Theorem 2.2):

$$P(Y = y \mid X = x) = P(Y = y) \quad . \tag{2.9}$$

Theorem 2.3 (Bayes' rule)**.** *The relationship between the conditional probability distribution $P(Y \mid X)$ and its so-called inverse probability distribution $P(X \mid Y)$ is given by*

$$P(X = x \mid Y = y) = \frac{P(X = x)P(Y = y \mid X = x)}{P(Y = y)} \quad , \tag{2.10}$$

where the prior (probability) is denoted by $P(X = x)$, the likelihood (function) by $P(Y = y \mid X = x)$ and the posterior (probability) by $P(X = x \mid Y = y)$.[5] *The probability $P(Y = y)$ is also known as the evidence.*

Following the discussion in Chapter 1.1, our generative picture of reduced-order modeling in the language of probability theory is presented:

---

[5] Note that, although the likelihood $P(Y = y \mid X = x)$ is a probability in $y \in \Omega_Y$, it is thought of as a function of $x \in \Omega_X$ ('the likelihood of x'):

$$\sum_{y \in \Omega_Y} P(Y = y \mid X = x) = 1 \quad , \quad \sum_{x \in \Omega_X} P(Y = y \mid X = x) \neq 1 \quad . \tag{2.11}$$

The terminology is a common cause of confusion [99, Chapter 2].

- We derive a reduced-order model M. The belief in M is given by

$$P(M) \quad . \tag{2.12}$$

- We select a set of parameters θ for our reduced-order model M. Assuming that there are multiple reduced-order models with the same parametrization under consideration[6], the belief in θ is given by

$$\begin{aligned} P(\theta) &= \sum_M P(\theta, M) && (2.13) \\ &= \sum_M P(M) P(\theta \mid M) \quad . && (2.14) \end{aligned}$$

- We find the state x corresponding to our reduced-order model M and our set of parameters θ. Assuming that there are multiple reduced-order models with the same parametrization under consideration as well as exploring a range of parameters, the belief in x is given by

$$\begin{aligned} P(x) &= \sum_M \sum_\theta P(x, \theta, M) && (2.15) \\ &= \sum_M \sum_\theta P(\theta, M) P(x \mid \theta, M) && (2.16) \\ &= \sum_M \sum_\theta P(M) P(\theta \mid M) P(x \mid \theta, M) \quad . && (2.17) \end{aligned}$$

In summary, the marginal probability distributions $P(\theta)$ and $P(x)$ representing our belief in the parameters and the state, respectively, are successively computed using the sum and product rules (Theorems 2.1, 2.2). In practice, the prior probability distribution $P(M)$ is either uniform or based on the complexity of the reduced-order model under consideration [99, Chapter 28]. The conditional probability distributions $P(\theta \mid M)$ and $P(x \mid \theta, M)$ constitute a graphical model of reduced-order modeling (Fig. 2.1).[7]

If experimental data is available in addition to a reduced-order model, we formulate the so-called forward and inverse problems of statistical inference.

---

[6] For simplicity, we assume a shared parametrization between models. In general, a parametrized model forms a subspace over the function space of all functions. If models have different parametrizations, it is generally possible to construct a subspace over the function space of all functions that contains all parametrized models. Therefore, the assumption of shared parametrization does not restrict generality.

[7] Note that a graphical model is not purely a visual illustration, but represents the causal dependence of a random variable on all nodes upstream in the graph [99, Chapter 23].

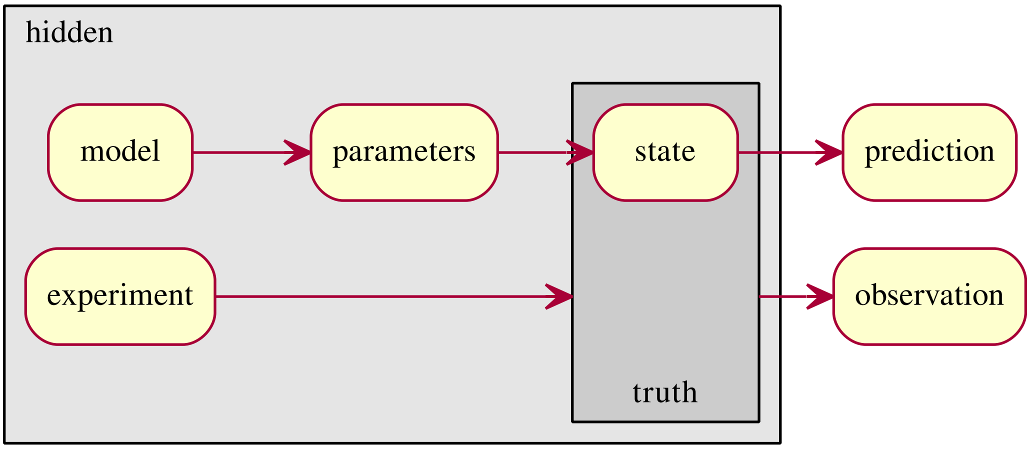


Fig. 2.1 A graphical model of reduced-order modeling. On the one hand, we have an experiment which is supposed to reproduce the truth. On the other hand, we have a reduced-order model of the system under investigation. We assign values to the parameters of the model, and solve for the state of the system. In our graphical model, we are not able to compare the state with the truth because it is hidden from us. Instead, we obtain noisy measurements from the experiment, and compare a prediction based on the reduced-order model to the imperfect observation of the truth.

Definition 2.5 (Forward problem)**.** Given a reduced-order model M, a set of parameters θ and the state x, as well as observations y of the truth from an experiment, our degree of belief is given by their likelihood

$$P(y \mid x, \theta, M) \quad . \tag{2.18}$$

Maximum-likelihood (ML) estimation finds the state x, the parameters θ and the reduced-order model M that maximize the likelihood $P(y \mid x, \theta, M)$.

Most efforts in thermoacoustics as outlined in Chapter 1, including reduced-order modeling, sensitivity analysis and uncertainty propagation, have focused on the forward problem, i.e. either (i) on how to derive reduced-order models, how to optimally calibrate the parameters in the reduced-order models, and how to solve the reduced-order models (Chapter 1.2), or (ii) on how variations in the parameters affect predictions of a quantity of interest, and how they alter the nonlinear dynamics (Chapter 1.3). In practice, ML estimation is not necessarily performed probabilistically but variationally by minimizing the negative logarithm of the likelihood, also known as the negative log-likelihood. For example, least-squares regression, in order to find the reduced-order model that best fits the experimental data, assumes a likelihood function with normally distributed observations [100, Chapter 3].

Definition 2.6 (Inverse problem)**.** Given a reduced-order model M, a set of parameters θ and the state x, as well as observations y of the truth from an experiment, our degree of belief is given by the posterior probability

$$P(x,\theta,M \mid y) \quad . \tag{2.19}$$

Maximum-a-posteriori (MAP) estimation finds the state x, the parameters θ and the reduced-order model M that maximize the inverse probability $P(x,\theta,M \mid y)$.

It follows from Bayes' rule that ML estimation is a special case of MAP estimation if the prior $P(x,\theta,M)$ is uniform and as such not a function of the state x, the parameters θ or the reduced-order model M (Theorem 2.3):

$$P(x,\theta,M \mid y) = \frac{P(x,\theta,M)P(y \mid x,\theta,M)}{P(y)} \propto P(y \mid x,\theta,M) \quad . \tag{2.20}$$

However, the distinction between forward and inverse problems is not limited to the use of a prior. Firstly, the inverse problem represents the uncertainty in the form of a posterior probability distribution over (x,θ,M), from which all their statistics are derived, whereas the forward problem focuses on a single scalar such as the log-likelihood or the root-mean-square error, for which ML estimation finds optimal values of (x,θ,M). Secondly, the inverse problem is able to address the uncertainty in a reduced-order model on multiple levels [99, Chapter 28]. For example, the degree of belief in a model M and a set of parameters θ given data y, independent of the uncertainty in the predictions based on the state x, is given by (Theorem 2.1)

$$P(\theta,M \mid y) = \sum_x P(x,\theta,M \mid y) \quad . \tag{2.21}$$

Thirdly, ML estimation based on the forward problem is prone to ill condition so that the optimal (x,θ,M) is either non-unique or very sensitive to the data y. The ad-hoc approach is regularization in order to promote weight decay or sparsity [100, Chapter 3]. The analogous approach in an inverse problem is the inclusion of a prior in agreement with Bayes' rule (Theorem 2.3).

In practice, the posterior probability distribution over (x,θ,M) is not computed directly due to its high dimensionality. Instead, the inverse problem for reduced-order models is layered into three levels of statistical inference, i.e. state estimation, parameter estimation and model estimation [99, Chapter 28].

Definition 2.7 (State estimation)**.** Given a reduced-order model M, a set of parameters θ, as well as observations y of the truth from an experiment, our belief in the state x is given by the posterior probability (Theorem 2.3)

$$P(x \mid y, \theta, M) = \frac{P(x \mid \theta, M) P(y \mid x, \theta, M)}{P(y \mid \theta, M)} \quad . \tag{2.22}$$

In general, the experiment exists independently of the modeler so that the observations y do not depend on the parameters θ or the reduced-order model M as long as the statistical inference is conditioned on the state x (Fig. 2.1):

$$P(x \mid y, \theta, M) = \frac{P(x \mid \theta, M) P(y \mid x)}{P(y \mid \theta, M)} \quad . \tag{2.23}$$

Definition 2.8 (Parameter estimation)**.** Given a reduced-order model M, as well as observations y of the truth from an experiment, our belief in a set of parameters θ is given by the posterior probability (Theorem 2.3)

$$P(\theta \mid y, M) = \frac{P(\theta \mid M) P(y \mid \theta, M)}{P(y \mid M)} \quad . \tag{2.24}$$

Note that parameter estimation uses the evidence $P(y \mid \theta, M)$ from state estimation as its likelihood (Definition 2.7).

Definition 2.9 (Model estimation)**.** Given observations y of the truth from an experiment, our belief in the reduced-order model M is given by the posterior probability (Theorem 2.3)

$$P(M \mid y) = \frac{P(M) P(y \mid M)}{P(y)} \quad . \tag{2.25}$$

Note that model estimation uses the evidence $P(y \mid M)$ from parameter estimation as its likelihood (Definition 2.8). Furthermore, the normalization of the prior $P(M)$ is problematic if the sample space of reduced-order models is infinite-dimensional. Alternatively, model comparison is performed instead of model estimation:

$$\frac{P(M' \mid y)}{P(M \mid y)} = \frac{P(M')}{P(M)} \frac{P(y \mid M')}{P(y \mid M)} \quad . \tag{2.26}$$

Thus, the calculation of the evidence $P(y)$ is avoided, and only the relative degrees of belief for the priors $P(M)$ and $P(M')$ are required.

In conclusion, probability theory is a powerful tool for addressing the uncertainty in reduced-order models. In the presence of experimental data, we treat reduced-order models as inverse problems, and perform multiple levels of Bayesian inference. In Parts II and III, we demonstrate this approach for two classes of reduced-order models in thermoacoustics, i.e. eigenproblems and time series. In the remaining sections of this chapter, we continue our discussion of reduced-order modeling from a statistical perspective in order to better understand the relationship between models and data.

2.2 Stochastic processes

As discussed in the previous section, probability theory as the language of statistical inference is somewhat controversial. In essence, we describe the unknown with the language of the random, thus equating two distinct epistemological concepts. In accepting this premise, we continue to consider families of random events in this section. Stochastic processes extend the notion of random variables from the previous section, which are defined for a single random event [101, Chapter 3].

Definition 2.10 (Stochastic process)**.** A stochastic process is a family of random variables $(X_w)_{w \in W} : \Omega \to \Omega'$ over the index set W. The probability P of an event $x \in \Omega'$ is given by (Definition 2.1)

$$P(X = x; w) = \mu\left(\{\omega \mid X_w(\omega) = x\}\right) \quad . \tag{2.27}$$

Examples of stochastic processes include time series, i.e. random variables as a function of time, and model closures, e.g. reaction rates as a function of pressure and temperature. Index sets are either discrete or continuous, with a countable or uncountable number of elements, respectively. In the context of reduced-order models, the elements of an index set represent design parameters of the experiment such as time or operating conditions which are under the modeler's control. The index set W is also referred to as the design space or the operating regime of the experiment.

Stochastic processes are defined by the joint and conditional probability distributions shared between their random variables. In particular, we discuss two types of stochastic processes: Gaussian processes and Wiener processes.

Definition 2.11 (Gaussian process)**.** A stochastic process is Gaussian if the joint probability distribution of any finite subset of random variables is Gaussian.

Thus, a Gaussian process is fully defined by its mean $m : W \to \Omega'$ and its covariance function $k : W \times W \to \Omega'$ [102, Chapter 2]:

$$m(w) = \mathbb{E}\left[X_w\right] \quad , \tag{2.28}$$

$$k(w, w') = \mathbb{E}\left[\left(X_w - m(w)\right)\left(X_{w'} - m(w')\right)\right] \quad , \tag{2.29}$$

where $\mathbb{E}[X_w]$ denotes the expected value of the random variable X_w for the index w. In practice, problems are set up in such a way that the Gaussian processes a priori have zero mean and unit variance [103]. In the following discussion, we focus on specifying the covariance function k, which is also known as the kernel of the Gaussian process [102, Chapter 4].

Definition 2.12 (Kernel)**.** A kernel $k : W \times W \to \Omega'$ is a function that correlates two points from the index set W by mapping them into the sample space $\Omega' \subset \mathbb{R}$. On top of that, we require kernels to be symmetric and positive semi-definite.

Theorem 2.4. *The sum of two kernels is a kernel:*

$$(k_1 + k_2)(w, w') = k_1(w, w') + k_2(w, w') \quad . \tag{2.30}$$

Theorem 2.5. *The (pointwise) product of two kernels is a kernel:*

$$(k_1 \times k_2)(w, w') = k_1(w, w')k_2(w, w') \quad . \tag{2.31}$$

Theorems 2.4 and 2.5 provide the calculus to construct kernels by recombination. We use three classes of kernels: constant kernels, white kernels and squared-exponential (SE) kernels. Similarly to physical models, the choice of kernel introduces structure and scaling. Therefore, we discuss the structure of the kernels in detail, although not necessarily rigorously, in order to equip the reader with an intuition regarding statistical modeling [104, Chapter 2].

Definition 2.13 (Constant kernel)**.** For $\Omega' \subset \mathbb{R}$, the constant kernel is given by

$$k(w, w') = \sigma_f^2 \quad , \tag{2.32}$$

where $\sigma_f \in \mathbb{R}$ represents the noise level.

Definition 2.14 (White kernel)**.** For $\Omega' \subset \mathbb{R}$, the white kernel is given by

$$k(w, w') = \begin{cases} 1 & w = w' \\ 0 & w \neq w' \end{cases} \quad . \tag{2.33}$$

Definition 2.15 (Squared-exponential kernel)**.** For $\Omega' \subset \mathbb{R}$, the squared-exponential (SE) kernel is given by

$$k(w, w') = \exp\left(-\frac{|w - w'|^2}{2\sigma_l^2}\right) \quad , \tag{2.34}$$

where $\sigma_l \in \mathbb{R}$ represents a characteristic unit of the index set W.

For a constant kernel, the two-point correlation between $w_1, w_2 \in W$ has the covariance matrix (Definition 2.13)

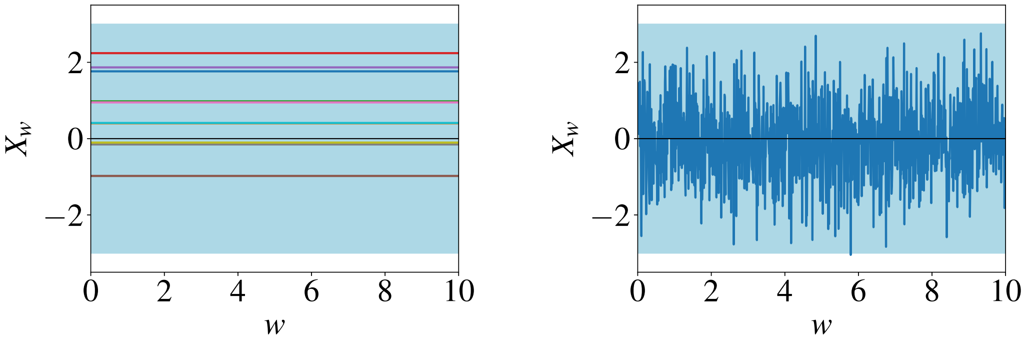
$$\Sigma_{1,2} = \begin{pmatrix} \sigma_f^2 & \sigma_f^2 \\ \sigma_f^2 & \sigma_f^2 \end{pmatrix} \quad . \tag{2.35}$$

For $X_1 = x_1$, the conditional probability distribution of the random variable X_2 is given by $\mathscr{N}(x_1, 0)$, which is a delta distribution (Corollary A.2). This means that the realization $x(w)$ of the stochastic process $(X_w)_{w \in W}$ is fully determined by the realization of a single random variable X_w. In fact, 'constant' refers to both the function values of this kernel and the shape of its realizations (Fig. 2.2, left).

For a white kernel, the two-point correlation between $w_1, w_2 \in W$ has the covariance matrix (Definition 2.14)

$$\Sigma_{1,2} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad . \tag{2.36}$$

For $X_1 = x_1$, the conditional probability distribution of the random variable X_2 is given by $\mathscr{N}(0, 1)$, which is a standard normal distribution (Corollary A.2). This means that all random variables X_w are statistically independent (Definition 2.4). In fact, 'white' refers to the realizations of this kernel sharing the characteristics of white noise (Fig. 2.2, right).



(a) Constant kernel ($\sigma_f = 1$).

(b) White kernel.

Fig. 2.2 Realizations of Gaussian processes for constant (left) and white kernels (right) with 3-σ confidence intervals. The ten realizations of the constant kernel are straight, horizontal lines. The realization of the white kernel is sampled at a resolution of $\Delta w = 0.01$.

For the SE kernel, we offer a more informal discussion of mean-square continuity and differentiability [102, Chapter 4]. The two-point correlation between $w_1, w_2 \in W$ is a function of $\delta = w_2 - w_1 > 0$, and has the covariance matrix (Definition 2.15)

$$\Sigma_{1,2} = \begin{pmatrix} 1 & \exp\left(-\delta^2/2\right) \\ \exp\left(-\delta^2/2\right) & 1 \end{pmatrix} \approx \begin{pmatrix} 1 & 1 - \delta^2/2 \\ 1 - \delta^2/2 & 1 \end{pmatrix} \quad , \tag{2.37}$$

where a second-order approximation of the exponential term is introduced for $\delta \ll 1$. For $X_1 = x_1$, the conditional probability distribution of the random variable X_2 is given by $\mathcal{N}(\mu_{2|1}, \Sigma_{2|1})$ (Corollary A.2):

$$\mu_{2|1} \approx x_1 \left[1 - \frac{\delta^2}{2}\right] \quad , \tag{2.38}$$

$$\Sigma_{2|1} \approx \delta^2 \quad . \tag{2.39}$$

For $\delta \to 0$, the random variable X_2 converges to the value x_1 in the sense that its mean $\mu_{2|1}$ converges to the value x_1 and that its standard deviation $\sqrt{\Sigma_{2|1}}$ vanishes linearly, which amounts to mean-square continuity. The three-point correlation between $w_1, w_2, w_3 \in W$ is a function of $\delta = w_3 - w_2 = w_2 - w_1 > 0$, and has the covariance matrix (Definition 2.15)

$$\Sigma_{1,2,3} = \begin{pmatrix} 1 & \exp\left(-\delta^2/2\right) & \exp\left(-2\delta^2\right) \\ \exp\left(-\delta^2/2\right) & 1 & \exp\left(-\delta^2/2\right) \\ \exp\left(-2\delta^2\right) & \exp\left(-\delta^2/2\right) & 1 \end{pmatrix} \tag{2.40}$$

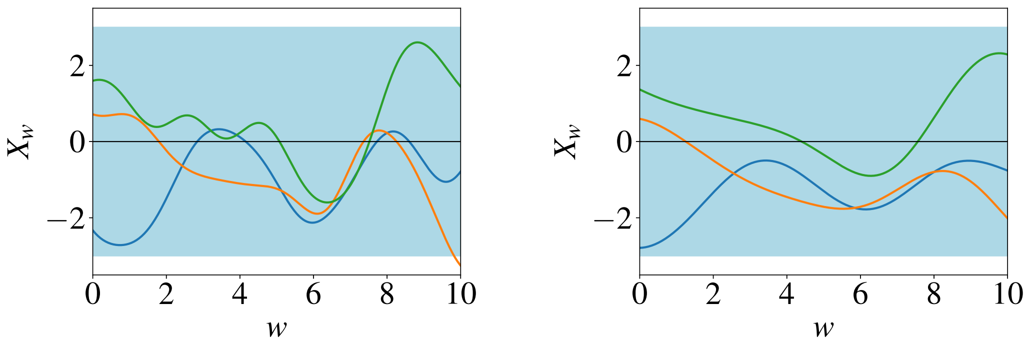
$$\approx \begin{pmatrix} 1 & 1-\delta^2/2+\delta^4/8 & 1-2\delta^2+2\delta^4 \\ 1-\delta^2/2+\delta^4/8 & 1 & 1-\delta^2/2+\delta^4/8 \\ 1-2\delta^2+2\delta^4 & 1-\delta^2/2+\delta^4/8 & 1 \end{pmatrix} \quad . \tag{2.41}$$

where a fourth-order approximation of the exponential term is introduced for $\delta \ll 1$. For $X_1 = x_1$ and $X_2 = x_2$, the conditional probability distribution of the random variable X_3 is given by $\mathcal{N}(\mu_{3|1,2}, \Sigma_{3|1,2})$ (Corollary A.2):

$$\mu_{3|1,2} \approx (2x_2 - x_1)\left[1 - \delta^2\right] \quad , \tag{2.42}$$

$$\Sigma_{3|1,2} \approx 2\delta^4 \quad , \tag{2.43}$$

For $\delta \to 0$, the random variable X_3 converges to the value $2x_2 - x_1$, which represents the linear extrapolation of the indices w_1 and w_2 to the index w_3, in the sense that its mean $\mu_{2|1}$ converges to the value $2x_2 - x_1$ and that its standard deviation $\sqrt{\Sigma_{3|1,2}}$ vanishes quadratically, which amounts to mean-square (first-order) differentiability. Without loss of generality, the n-point correlations extend to arbitrarily high orders n so that the realizations of the SE kernel are infinitely differentiable and smooth. Realizations of Gaussian processes with SE kernels are shown in Fig. 2.3. The variations are smooth, and the distance between peaks and troughs is approximately $2\sigma_l$. It follows either from visual inspection of Fig. 2.2 and 2.3 or from comparison between the two-point correlations (Eq. (2.35)–(2.37)) that the constant kernel and the white kernel are limiting cases of the SE kernel for $\sigma_l \to \infty$ and $\sigma_l \to 0$, respectively.



(a) SE kernel ($\sigma_l = 1$).

(b) SE kernel ($\sigma_l = 2$).

Fig. 2.3 Realizations of Gaussian processes for squared-exponential (SE) kernels of different characteristic units with 3-σ confidence intervals. For each characteristic unit, the variations in the three realizations are smooth. The coherence of the variations scales with σ_l.

We conclude our treatment of stochastic processes with a second example, the Wiener process [105].

Definition 2.16 (Wiener process)**.** A stochastic process $(X_w)_{w\in\mathbb{R}^+} : \Omega \to \Omega'$ is a Wiener process if it has the following properties:

Independent increments. For $s < t < u < v$, the random variables $X_t - X_s$ and $X_v - X_u$ are statistically independent (Definition 2.4).

Stationary increments. For $s < t$, the random variable $X_t - X_s$ has the same probability distribution as the random variable X_{t-s}.

Normal increments. All random variables X_w are Gaussian. In particular, the random variable X_1 at the index $w = 1$ has unit variance:

$$X_1 \sim \mathcal{N}(0,1) \quad . \tag{2.44}$$

In physics, realizations of the Wiener process are also known as Brownian motion: The sample space Ω' represents the position of a particle in space, and the index w represents a point in time. At time $w = 0$, the position of the particle is known. Afterwards, the trajectory of the particle is a so-called random walk. This means that the motion of the particle between two successive snapshots is not deterministic but subject to a probability distribution. If the snapshots are taken Δw units of time apart from each other, it follows from Definition 2.16 that the steps of the random walk follow a normal distribution with a variance of Δw:

$$X_{\Delta w} \sim \mathcal{N}(0,\Delta w) \quad . \tag{2.45}$$

In the limit of $\Delta w \to 0$, Brownian motion thus becomes the integral of white noise. Realizations of the Wiener process in one dimension ($\Omega' = \mathbb{R}$) are shown in Fig. 2.4.

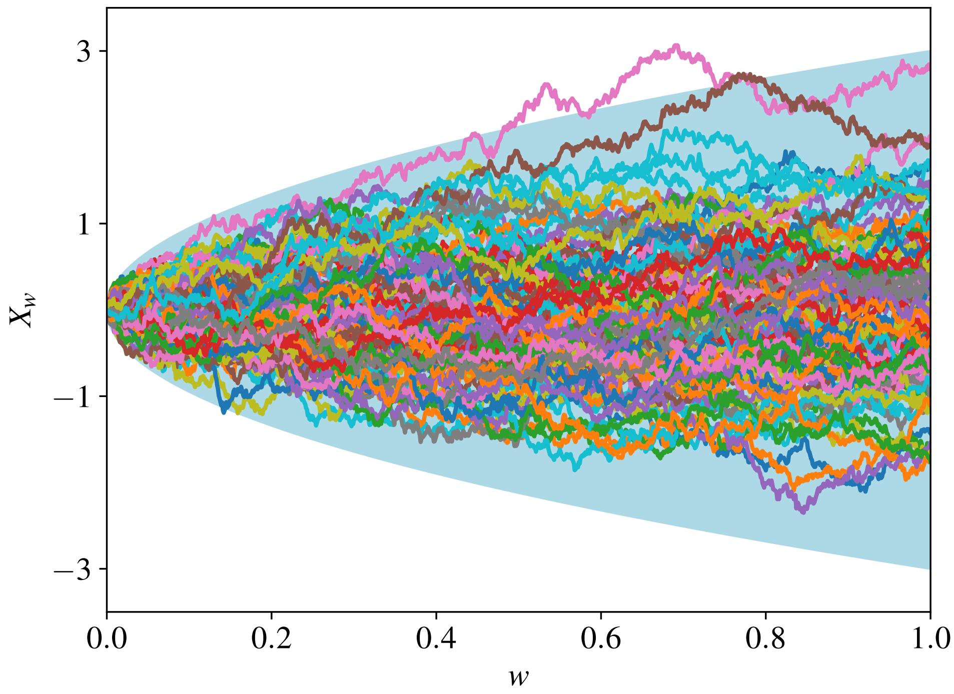


Fig. 2.4 One hundred realizations of the Wiener process, sampled at a resolution of $\Delta w = 0.001$, with 3-σ confidence intervals ($\sigma = \sqrt{w}$). Note that the realizations are continuous but nowhere differentiable.

In conclusion, we consider two types of stochastic processes for statistical inference. Firstly, Gaussian processes are characterized by their marginalization property: For a stochastic process as a possibly uncountable family of random variables, it is sufficient to specify the joint probability distributions of all finite subsets of random variables (Definition 2.11). In Part II, we use the theory of Gaussian processes to analyze experiments performed over a range of operating conditions. Secondly, the Wiener process is characterized by the Markov property: For $s < t$, the conditional probability distribution of the random variable X_t given the random variable X_s represents a step in a random walk that is independent from all the earlier steps in the random walk (Definition 2.16). In Part III, we use the theory of the Wiener process to analyze experimental data in the form of time series.

2.3 Model selection

In Chapters 2.1 and 2.2, we equipped reduced-order models, which are parametric in nature, with statistical models, i.e. random variables and stochastic processes, the latter of which are nonparametric in nature. Nevertheless, stochastic processes rely on so-called hyperparameters, in contrast to model parameters. For example, the kernel of a Gaussian process has hyperparameters such as σ_f and σ_l (Definitions 2.13, 2.15). Although hyperparameters are not physical parameters of the reduced-order model in the first place, the choice of hyperparameters significantly affects the model parameters and thus the model predictions inside and outside the design space.

In general, there are two strategies to learning parameters: In optimization, the parameters are reduced to a single set of values that maximizes or minimizes a quantity of interest, i.e. to a point estimate in the maximum-likelihood (ML) sense (Definition 2.5). In statistical inference, the parameters are represented by a probability distribution which incorporates prior knowledge. The optimal set of parameters is the most probable one, i.e. from a maximum-a-posteriori (MAP) estimate (Definition 2.6). There are multiple combinations and variations of these two strategies when it comes to parameter and hyperparameter estimation. Our strategy follows the so-called evidence framework [106]:

1. Perform statistical inference for the model parameters.
2. Optimize the hyperparameters.

As the statistical model is determined by the choice of hyperparameters, the optimization of hyperparameters is also known as model selection.

On the one hand, statistical inference of the model parameters in a Bayesian framework carefully calibrates the reduced-order model by introducing every observation exactly once, either sequentially or batchwise. This matters because model predictions are biased towards prior knowledge or observations otherwise. On the other hand, optimization of the hyperparameters is less restrained in its usage of the observations. This is because we are only interested in an optimal set of values but not their uncertainties as long as the hyperparameters maximize or minimize the quantity of interest. In practice, when observations are introduced sequentially, it takes at least one pass to learn the model parameters, and generally multiple passes to also learn the hyperparameters. This is a common thread for the algorithms developed in Parts II and III.

In general, the optimization of the hyperparameters is an iterative procedure. For a Gaussian process, we select a kernel and a set of hyperparameters θ^{hyp} such that the Gaussian process prior has zero mean and unit variance (Chapter 2.2). We denote this model by $M =$

$M(\theta^{\text{hyp}})$, and the Gaussian process prior for the model parameters θ is given by $P(\theta \mid M)$. For given observations y, the likelihood of the model parameters θ is given by $P(y \mid \theta, M)$. Bayes' rule tells us that the posterior probability distribution is given by (Theorem 2.3)

$$P(\theta \mid y, M) = \frac{P(\theta \mid M) P(y \mid \theta, M)}{P(y \mid M)} \quad . \tag{2.46}$$

The most obvious quantity of interest to maximize over the hyperparameters θ^{hyp} would be the posterior probability $P(\hat{\theta} \mid y, M)$ of the MAP point estimate $\hat{\theta}$. This would eliminate the probability distribution over the model parameters θ from the optimization problem, and thus simplify it. But it would not do near-uniform probability distributions justice where the MAP point estimate $\hat{\theta}$ barely stands out against other nearby point estimates of the model parameters θ. Therefore, we instead eliminate the probability distribution over the model parameters θ by considering the evidence (Theorem 2.3)

$$P(y \mid M) = \int P(\theta \mid M) P(y \mid \theta, M) \, \mathrm{d}\theta \quad . \tag{2.47}$$

Due to the integration of the likelihood (weighted by the prior), the evidence $P(y \mid M)$ is also known as the marginal likelihood. Another way to understand the significance of the marginal likelihood $P(y \mid M)$ is again based on Bayes' rule (Theorem 2.3). If we have no preference for any reduced-order model and use a uniform prior, then the posterior probability is given by (Eq. (2.20))

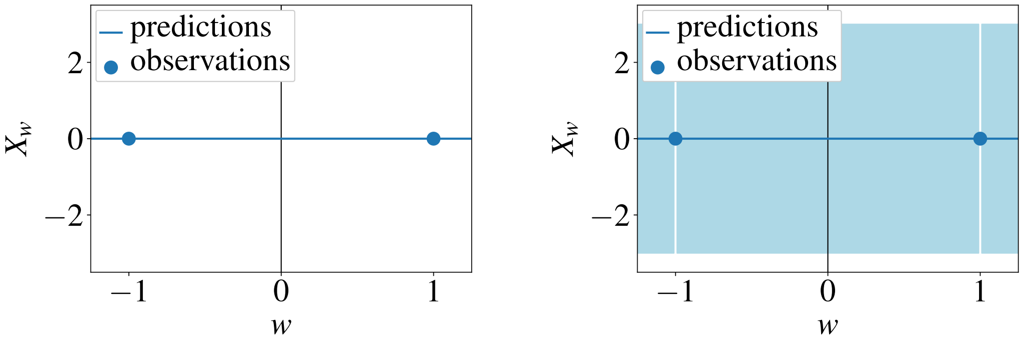
$$P(M \mid y) = \frac{P(M) P(y \mid M)}{P(y)} \propto P(y \mid M) \quad , \tag{2.48}$$

Thus, the marginal likelihood $P(y \mid M)$ allows us to compare our degrees of belief in various reduced-order models.

In practice, the logarithm of the evidence $\log(P(y \mid M))$, also known as the log-marginal likelihood, is maximized over the hyperparameters θ^{hyp}, and thus the models $M = M(\theta^{\text{hyp}})$, using gradient descent [102, Chapter 5]. The role of the evidence in learning the hyperparameters establishes the naming of our strategy as the evidence framework [106]. We conclude the discussion about model selection by showcasing two academic, noise-free examples in order to provide the reader with more intuition regarding statistical modeling.

In the first example, we consider two observations of the same value $y = 0$ at the two indices $w = \pm 1$ (Fig. 2.5). We consider two kernels, i.e. the constant kernel and the white kernel, corresponding to pure signal and noise, respectively (Definitions 2.13, 2.14). On the one hand, the constant kernel is only able to properly handle observations of the same value. This is the case here, and the marginal likelihood of the model parameters is

$1/\sqrt{2\pi} \approx \exp(-0.919)$ (Table 2.1, first row). On the other hand, the white kernel is able to handle any data. The marginal likelihood of the model parameters is $1/(2\pi) \approx \exp(-1.838)$, which is less than the value for the constant kernel (Table 2.1, second row). Therefore, the statistical model with the constant kernel is more plausible here although the likelihood of each observation is $1/\sqrt{2\pi}$ for both statistical models. This example reflects the preference of the evidence framework for simpler models with fewer degrees of freedom, all other things being equal.



(a) Constant kernel ($\sigma_f = 1$).

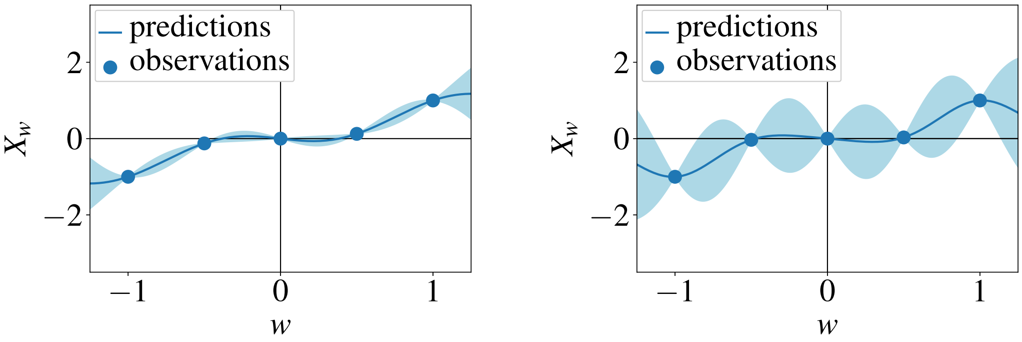
(b) White kernel.

Fig. 2.5 Model predictions of Gaussian processes for constant (left) and white kernels (right) with 3-σ confidence intervals. There is no uncertainty in the model prediction of the constant kernel. The white kernel has the same uncertainty in the model prediction as the prior except at the indices $w = \pm 1$ of the observations, where the uncertainty vanishes. The log-marginal likelihoods are given in Table 2.1.

In the second example, we consider multiple observations over the equidistant design interval $-1 \leq w \leq 1$ (Fig. 2.6). In the first case, the observations are sampled from $y = w^3$, in the second case from $y = w^5$. Both cases have similar trends in that the observation y monotonically increases with the index w and holds the same value for the indices $w = 0$ and $w = \pm 1$. But they differ in their degrees of smoothness, i.e. an observation following $y = w^n$ even becomes discontinuous at the index $w = 1$ in the limiting case of $n \to \infty$. We consider the product of a constant kernel and an SE kernel (Theorem 2.5):

$$k(w, w') = \sigma_f^2 \exp\left(\frac{|w - w'|^2}{2\sigma_l^2} \right) \quad . \tag{2.49}$$

The optimized hyperparameters and log-marginal likelihoods are shown in Table 2.1 (third and fourth row). In terms of hyperparameters, the σ_f-values are similar, and the main difference lies in the σ_l-values. Since an observation following $y = w^3$ is smoother than one following $y = w^5$, it allows for a larger characteristic unit σ_l. When Gaussian processes and distance-based kernels are involved, this example reflects the tendency of the evidence framework to consider smoother models more plausible. In the context of reduced-order modeling, this is a highly desirable property for any model selection strategy. For example, if we consider the observations of the model parameters to be derived from the same experiment over a design parameter w but using different reduced-order models and parametrizations, the evidence framework favors reduced-order models which explain the observations with parametrizations that are ideally constant, linear, etc., in this particular order. This coincides with our intuition that a model parameter, unlike the state of the system under investigation, should by definition be universal[8] over a wide range of design parameters. This is useful insight on which we will fall back in Parts II and III, e.g. in our treatment of nonparametric regression (Chapter 3.3).



(a) $y = w^3$. (b) $y = w^5$.

Fig. 2.6 Model predictions of Gaussian process and SE kernel for observations of varying smoothness with 3-σ confidence intervals. An observation following $y = w^3$ (left) is significantly smoother than one following $y = w^5$ (right), which is also evident in the confidence intervals of the interpolated and extrapolated model predictions. The log-marginal likelihoods are given in Table 2.1.

Table 2.1 Hyperparameters and log-marginal likelihoods (Fig. 2.5, 2.6).

	σ_f	σ_l	$\log(P(y \mid M))$	
constant kernel	1	—	−0.919	Fig. 2.5, left
white kernel	—	—	−1.838	Fig. 2.5, right
SE kernel	0.846	0.575	−4.487	Fig. 2.6, left
SE kernel	0.641	0.275	−4.793	Fig. 2.6, right

---

[8] Arguably at this point, universality is a somewhat vague concept. Ideally, a model parameter is constant over the design parameters. More realistically, the variation of a model parameter with respect to the design parameters is at least smooth. We return to this discussion in the context of strong and weak parametrizations (Chapter 3.3).

Part II

Bayesian regression

Chapter 3

Nonlinear regression

We continue the discussion from Chapter 2, and present a hierarchical framework for state estimation, parameter estimation and model selection based on nonlinear regression. There are two main ingredients to our hierarchical framework: (i) the shift of the observations from the level of the states to the level of the model parameters using parametric regression (Chapter 3.1); and (ii) the relaxation as well as the regularization of reduced-order models inside nonparametric regression (Chapter 3.3). We provide definitions for our notions of models, model parameters and states, which are also featured in the index at the end of this document. Regarding the reduced-order models, we put an emphasis on nonlinear eigenproblems in the context of thermoacoustics (Chapter 3.2). Nevertheless, our framework applies to all reduced-order models following the graphical model of Fig. 2.1.

Supplementary material:

- `Chapter3/Least-squares regression.ipynb` (Git, Binder).
- `Chapter3/Expectation-maximization algorithm.ipynb` (Git, Binder).

3.1 Least-squares regression

In Part II, we consider a reduced-order model of the implicit form[1]

$$F : \mathbb{C} \times \mathbb{R}^M \times \mathbb{R}^N \to \mathbb{C} \quad , \tag{3.1}$$

---

[1] Viz., the function F maps the state $x \in \mathbb{C}$, the model parameters $\theta \in \mathbb{R}^M$ and the design parameters $w \in \mathbb{R}^N$ onto a complex number $F(x, \theta; w) \in \mathbb{C}$. The reduced-order model is implicit in that the state is not given as a function $x = f(\theta; w)$ of the model and design parameters but is found by solving $F(x, \theta; w) \equiv 0$ for given θ and w.

$$(x,\theta,w) \mapsto F(x,\theta;w) \equiv 0 \quad . \tag{3.2}$$

where the (complex) state of the system[2] is denoted by x and the model parameters by the M-dimensional vector θ. The N-dimensional vector w denotes the design parameters of each experiment, e.g. the operating conditions set by the experimentalist. For a series of K experiments indexed by $1 \leq k \leq K$ with design parameters $w \in \{w_k \in \mathbb{R}^N \mid 1 \leq k \leq K\}$ and observations $y \in \{y_k \in \mathbb{C} \mid 1 \leq k \leq K\}$ of the true state $x \in \{x_k \in \mathbb{C} \mid 1 \leq k \leq K\}$, the governing equations of the reduced-order model are alternatively written as

$$F_k : \mathbb{C} \times \mathbb{R}^M \to \mathbb{C} \quad , \tag{3.3}$$

$$(x_k,\theta) \mapsto F_k(x_k,\theta) = F(x_k,\theta;w_k) \equiv 0 \quad . \tag{3.4}$$

The least-squares residual of the reduced-order model with respect to the observations is given by[3]

$$J(x,\theta) = \sum_k \frac{|y_k - x_k|^2}{\sigma_k^2} = (y_k - x_k)^* \left(\Sigma_{\text{obs}}^{-1}\right)_{kl} (y_l - x_l) \quad , \tag{3.5}$$

where a superscript asterisk denotes complex conjugate. The positive semi-definite matrix $(\Sigma_{\text{obs}})_{kl} = \sigma_k^2 \delta_{kl}$, where the symbol δ denotes the Kronecker delta, is the covariance matrix of the observations. For a positive semi-definite quadratic form J and a reduced-order model F, least-squares regression poses an optimization problem with constraints:

$$\underset{x,\theta}{\text{minimize}} \quad J(x,\theta) \quad , \tag{3.6}$$

$$\text{subject to} \quad F_k(x_k,\theta) \equiv 0 \quad , \quad 1 \leq k \leq K \quad . \tag{3.7}$$

A necessary condition for the optimal set of parameters $\hat{\theta}$ is given by [107, Chapter 7]

$$\left.\frac{\mathrm{d}J}{\mathrm{d}\theta_m}\right|_{\hat{\theta}} = \left.\frac{\partial J}{\partial \theta_m}\right|_{\hat{x},\hat{\theta}} + \left.\frac{\partial J}{\partial x_k}\right|_{\hat{x},\hat{\theta}} \left.\frac{\mathrm{d}x_k}{\mathrm{d}\theta_m}\right|_{\hat{\theta}} \equiv 0 \quad , \quad 1 \leq m \leq M \quad . \tag{3.8}$$

In general, the roots of F in x are not analytical so that the derivatives of x in θ are found via implicit differentiation [107, Chapter 8]. The first (total) derivative of x in θ is given by (Eq. (3.2))

$$\frac{\mathrm{d}F}{\mathrm{d}\theta_m} = \frac{\partial F}{\partial \theta_m} + \frac{\partial F}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}\theta_m} \equiv 0 \quad , \tag{3.9}$$

---

[2] Note that the state of the system is implicitly defined as the solution to the reduced-order model without regard towards the physical interpretation by the modeler. For example, if the reduced-order model is an eigenproblem that sets the determinant of a matrix to zero, the state of the system is represented solely by the eigenvalue but not the eigenvectors.

[3] Products over duplicate indices are subject to Einstein summation notation.

$$\frac{\mathrm{d}x}{\mathrm{d}\theta_m} = -\left(\frac{\partial F}{\partial x}\right)^{-1} \frac{\partial F}{\partial \theta_m} \quad . \tag{3.10}$$

The second (total) derivative of x in θ is given by (Eq. (3.9))

$$\frac{\mathrm{d}^2 F}{\mathrm{d}\theta_m \mathrm{d}\theta_n} = \frac{\partial^2 F}{\partial \theta_m \partial \theta_n} + \frac{\partial^2 F}{\partial x \partial \theta_n}\frac{\mathrm{d}x}{\mathrm{d}\theta_m} + \frac{\partial^2 F}{\partial x \partial \theta_m}\frac{\mathrm{d}x}{\mathrm{d}\theta_n} + \frac{\partial^2 F}{\partial x^2}\frac{\mathrm{d}x}{\mathrm{d}\theta_m}\frac{\mathrm{d}x}{\mathrm{d}\theta_n} + \frac{\partial F}{\partial x}\frac{\mathrm{d}^2 x}{\mathrm{d}\theta_m \mathrm{d}\theta_n} \equiv 0 \quad , \tag{3.11}$$

$$\frac{\mathrm{d}^2 x}{\mathrm{d}\theta_m \mathrm{d}\theta_n} = -\left(\frac{\partial F}{\partial x}\right)^{-1}\left[\frac{\partial^2 F}{\partial \theta_m \partial \theta_n} + \frac{\partial^2 F}{\partial x \partial \theta_n}\frac{\mathrm{d}x}{\mathrm{d}\theta_m} + \frac{\partial^2 F}{\partial x \partial \theta_m}\frac{\mathrm{d}x}{\mathrm{d}\theta_n} + \frac{\partial^2 F}{\partial x^2}\frac{\mathrm{d}x}{\mathrm{d}\theta_m}\frac{\mathrm{d}x}{\mathrm{d}\theta_n}\right] \quad . \tag{3.12}$$

In a linear regression, where x is a linear function of θ, the necessary condition states (Eq. (3.8)):

$$-(y_k - x_k)^* \left(\Sigma_{\mathrm{obs}}^{-1}\right)_{kl} \frac{\mathrm{d}x_l}{\mathrm{d}\theta_m} - \frac{\mathrm{d}x_k^*}{\mathrm{d}\theta_m}\left(\Sigma_{\mathrm{obs}}^{-1}\right)_{kl}(y_l - x_l) \equiv 0 \quad . \tag{3.13}$$

Substituting the Jacobian $A_{ij} = \mathrm{d}x_i/\mathrm{d}\theta_j$ into Eq. (3.13) gives

$$\mathrm{Re}\left(A_{km}^* \left(\Sigma_{\mathrm{obs}}^{-1}\right)_{kl} A_{ln}\right)\theta_n = \mathrm{Re}\left(A_{km}^* \left(\Sigma_{\mathrm{obs}}^{-1}\right)_{kl} y_l\right) \quad , \tag{3.14}$$

where Re denotes the real part of a complex number. Eq. (3.14) is a linear system of equations ($M \times M$), also known as the normal equations, whose solution is the optimal set of parameters $\hat{\theta}$ [108, Chapter 4].

In a nonlinear regression, where there is no analytical solution of x in terms of θ, Eq. (3.13) is solved iteratively, e.g. via the Newton-Raphson method [108, Chapter 5]:

$$\left.\frac{\mathrm{d}J}{\mathrm{d}\theta_m}\right|^{i} + \left.\frac{\mathrm{d}^2 J}{\mathrm{d}\theta_m \mathrm{d}\theta_n}\right|^{i}\left(\theta_n^{i+1} - \theta_n^{i}\right) \equiv 0 \quad , \tag{3.15}$$

where the superscript index denotes the iteration step. The Hessian $\mathrm{d}^2 J/(\mathrm{d}\theta_m \mathrm{d}\theta_n)$ is given by (Eq. (3.13))

$$\frac{\mathrm{d}^2 J}{\mathrm{d}\theta_m \mathrm{d}\theta_n} = 2\mathrm{Re}\left(\frac{\mathrm{d}x_k^*}{\mathrm{d}\theta_m}\left(\Sigma_{\mathrm{obs}}^{-1}\right)_{kl}\frac{\mathrm{d}x_l}{\mathrm{d}\theta_n} - \frac{\mathrm{d}^2 x_k^*}{\mathrm{d}\theta_m \mathrm{d}\theta_n}\left(\Sigma_{\mathrm{obs}}^{-1}\right)_{kl}(y_l - x_l)\right) \quad . \tag{3.16}$$

The Gauss-Newton method is a special case of the Newton-Raphson method when x is quasilinear in terms of θ ($\mathrm{d}x_k^*/(\mathrm{d}\theta_m \mathrm{d}\theta_n) \approx 0$) or when the residual J is small ($||y_l - x_l||_2 \ll 1$) [108, Chapter 6]:

$$\frac{\mathrm{d}^2 J}{\mathrm{d}\theta_m \mathrm{d}\theta_n} \approx 2\mathrm{Re}\left(\frac{\mathrm{d}x_k^*}{\mathrm{d}\theta_m}\left(\Sigma_{\mathrm{obs}}^{-1}\right)_{kl}\frac{\mathrm{d}x_l}{\mathrm{d}\theta_n}\right) \quad . \tag{3.17}$$

Note that this bilinear form is formally identical to the linear operator on the left-hand side of the normal equations (Eq. (3.14)). This shows that linear regression is a special case of nonlinear regression.

In a Bayesian regression, the likelihood function of the forward problem is given by (Definition 2.5)

$$P(y \mid x,\theta,F) = \frac{1}{\sqrt{\det(2\pi\Sigma_{\text{obs}})}} \exp\left(-\frac{J(x,\theta)}{2}\right) \quad . \tag{3.18}$$

For maximum-likelihood (ML) estimation, the posterior probability distribution in the model parameters θ is given by (Definition 2.8)

$$P(\theta \mid y,F) = \frac{\exp\left(-\frac{J(x,\theta)}{2}\right)}{\int \exp\left(-\frac{J(x,\theta)}{2}\right) \mathrm{d}\theta} \quad . \tag{3.19}$$

The likelihood function $P(y \mid x,\theta,F)$ is normal in the observations y whereas the posterior probability distribution $P(\theta \mid y,F)$ is not normal in the model parameters θ, and its exact evidence $P(y \mid F) = \int \exp(-J(x,\theta)/2)\,\mathrm{d}\theta$ is unknown. The Laplace approximation expands the unnormalized log-likelihood function into a second-order Taylor series in the model parameters θ around the optimal set of parameters $\hat{\theta}$ and the ML estimate $\hat{x}$ [99, Chapter 27]:

$$\begin{aligned} -\log\left(\sqrt{(2\pi)^K \det(\Sigma_{\text{obs}})}P(y \mid x,\theta,F)\right) \approx{}& \frac{J(\hat{x},\hat{\theta})}{2} + \frac{1}{2}\left.\frac{\mathrm{d}J}{\mathrm{d}\theta_m}\right|_{\hat{\theta}} \left(\theta_m - \hat{\theta}_m\right) \\ &+ \frac{1}{4}\left(\theta_m - \hat{\theta}_m\right) \left.\frac{\mathrm{d}^2 J}{\mathrm{d}\theta_m \mathrm{d}\theta_n}\right|_{\hat{\theta}} \left(\theta_n - \hat{\theta}_n\right) \quad . \end{aligned} \tag{3.20}$$

The zeroth-order term in Eq. (3.20) is constant in θ. As a necessary condition, the first-order term in Eq. (3.20) vanishes (Eq. (3.8)). Furthermore, the Hessian in the second-order term is symmetric and positive semi-definite [107, Chapter 7]. For ML estimation, the posterior probability distribution $P(\theta \mid y,F)$ is thus approximated by a normal distribution in the model parameters θ (Eq. (3.16), (3.19)):

$$P(\theta \mid y,F) = \frac{1}{\sqrt{\det(2\pi\Sigma_{\text{lik}})}} \exp\left(-\frac{1}{2}\left(\theta_m - \hat{\theta}_m\right)\left(\Sigma_{\text{lik}}^{-1}\right)_{mn}\left(\theta_n - \hat{\theta}_n\right)\right) \quad , \tag{3.21}$$

$$\left(\Sigma_{\text{lik}}^{-1}\right)_{mn} = \left.\frac{\mathrm{d}x_k^*}{\mathrm{d}\theta_m}\right|_{\hat{\theta}} \left(\Sigma_{\text{obs}}^{-1}\right)_{kl} \left.\frac{\mathrm{d}x_l}{\mathrm{d}\theta_n}\right|_{\hat{\theta}} - \operatorname{Re}\left(\left.\frac{\mathrm{d}^2 x_k^*}{\mathrm{d}\theta_m \mathrm{d}\theta_n}\right|_{\hat{\theta}} \left(\Sigma_{\text{obs}}^{-1}\right)_{kl} (y_l - x_l)\right) \quad , \tag{3.22}$$

where the matrix Σ_θ^{-1} is also known as the Fisher information matrix [95, Chapter 8]. The ML estimate of a Bayesian regression with Laplace approximation is by definition the same as the optimal set of parameters $\hat{\theta}$ from a nonlinear regression. On top of that, Bayesian regression quantifies the uncertainty in the model parameters θ in the form of the covariance matrix Σ_{lik}. For maximum-a-posteriori (MAP) estimation, we assume that the prior is given by a normal distribution in the model parameters θ:

$$P(\theta \mid F) = \frac{1}{\sqrt{\det\left(2\pi\Sigma_{\text{pre}}\right)}} \exp\left(-\frac{1}{2}\left(\theta_m - \theta_m^{\text{pre}}\right)\left(\Sigma_{\text{pre}}^{-1}\right)_{mn}\left(\theta_n - \theta_n^{\text{pre}}\right)\right) \quad . \tag{3.23}$$

The posterior is then also a normal distribution in the model parameters θ (Appendix A):

$$P(\theta \mid y, F) = \frac{1}{\sqrt{(2\pi)^M \det\left(\Sigma_{\text{post}}\right)}} \exp\left(-\frac{1}{2}\left(\theta_m - \theta_m^{\text{post}}\right)\left(\Sigma_{\text{post}}^{-1}\right)_{mn}\left(\theta_n - \theta_n^{\text{post}}\right)\right) \quad , \tag{3.24}$$

where the mean and the covariance matrix are given by

$$\left(\Sigma_{\text{post}}^{-1}\right)_{mn} = \left(\Sigma_{\text{pre}}^{-1}\right)_{mn} + \left(\Sigma_{\text{lik}}^{-1}\right)_{mn} \quad , \tag{3.25}$$

$$\left(\Sigma_{\text{post}}^{-1}\right)_{mn} \theta_n^{\text{post}} = \left[\left(\Sigma_{\text{pre}}^{-1}\right)_{mn} \theta_n^{\text{pre}} + \left(\Sigma_{\text{lik}}^{-1}\right)_{mn} \hat{\theta}_n\right] \quad . \tag{3.26}$$

Algorithm 3.1 Bayesian regression with Laplace approximation (ML).

```
procedure SolveF($\theta$, $w$)                                  ▷ Vectorized nonlinear solve.
    for $1 \le k \le K$ do
        Solve Eq. (3.2) for $x_k$.                                ▷ Nonlinear solve.
    end for
    return $x$
end procedure

$\hat{\theta} \leftarrow \theta^{\text{pre}}$                                              ▷ Initialize $\hat{\theta}$.
$\hat{x} \leftarrow$ SolveF($\hat{\theta}, w$)
while $\left| \frac{dJ}{d\theta}\Big|_{\hat{\theta}} \right| >$ EPS do                       ▷ Nonlinear regression.
    $\hat{\theta} \leftarrow \hat{\theta} - \left( \frac{d^2J}{d\theta^2}\Big|_{\hat{\theta}} \right)^{-1} \frac{dJ}{d\theta}\Big|_{\hat{\theta}}$          ▷ Linearized regression (Eq. (3.15)).
    $\hat{x} \leftarrow$ SolveF($\hat{\theta}, w$)
end while
$\Sigma_\theta^{-1} \leftarrow \frac{1}{2} \frac{d^2J}{d\theta^2}\Big|_{\hat{\theta}}$                  ▷ Laplace approximation (Eq. (3.22)).
$\Sigma_{\text{post}}^{-1} \leftarrow \Sigma_{\text{pre}}^{-1} + \Sigma_\theta^{-1}$                     ▷ Posterior covariance (Eq. (3.25)).
$\theta^{\text{post}} \leftarrow \Sigma_{\text{post}} \left[\Sigma_{\text{pre}}^{-1}\theta^{\text{pre}} + \Sigma_\theta^{-1}\hat{\theta}\right]$.       ▷ Posterior mean (Eq. (3.26)).
```

In conclusion, Algorithm 3.1 summarizes the steps in Bayesian regression with Laplace approximation. The procedures for evaluating F, x and J as well as their first and second derivatives are not shown. The linear and nonlinear regression procedures are essential components within the Bayesian regression algorithm. The Laplace approximation enters the algorithm for the ML estimation. Alternatively, the Laplace approximation expands the unnormalized log-posterior (instead of the log-likelihood function) into a second-order Taylor series in the model parameters θ around the optimal set of parameters θ^{post}:

$$\begin{aligned} &-\log\left(\sqrt{(2\pi)^{M+K}\det\left(\Sigma_{\text{pre}}\right)\det\left(\Sigma_{\text{obs}}\right)}P(\theta\mid F)P(y|x,\theta,F)\right) \\ &\approx \frac{J'(x^{\text{post}},\theta^{\text{post}})}{2}+\frac{1}{2}\left.\frac{\mathrm{d}J'}{\mathrm{d}\theta_m}\right|_{\theta^{\text{post}}}\left(\theta_m-\theta_m^{\text{post}}\right) \\ &\qquad\qquad +\frac{1}{4}\left(\theta_m-\theta_m^{\text{post}}\right)\left.\frac{\mathrm{d}^2J'}{\mathrm{d}\theta_m\mathrm{d}\theta_n}\right|_{\theta^{\text{post}}}\left(\theta_n-\theta_n^{\text{post}}\right) \quad , \end{aligned} \tag{3.27}$$

where the modified residual is given by

$$J'(x,\theta)=J(x,\theta)+\left(\theta_m-\theta_m^{\text{pre}}\right)\left(\Sigma_{\text{pre}}^{-1}\right)_{mn}\left(\theta_n-\theta_n^{\text{pre}}\right) \quad , \tag{3.28}$$

and the MAP estimate θ^{post} satisfies the necessary condition

$$\frac{\mathrm{d}J'}{\mathrm{d}\theta_m}=\left.\frac{\mathrm{d}J}{\mathrm{d}\theta_m}\right|_{\theta^{\text{post}}}+2\left(\Sigma_{\text{pre}}^{-1}\right)_{mn}\left(\theta_n^{\text{post}}-\theta_n^{\text{pre}}\right)\equiv 0 \quad . \tag{3.29}$$

Algorithm 3.2 summarizes the steps in Bayesian regression where the Laplace approximation enters the algorithm for the MAP estimation.

Algorithm 3.2 Bayesian regression with Laplace approximation (MAP).

```
$\hat{\theta} \leftarrow \theta^{\text{pre}}$                                          $\triangleright$ Initialize $\hat{\theta}$.
$\hat{x} \leftarrow \text{SolveF}(\hat{\theta}, w)$
while $\left|\left.\frac{\mathrm{d}J'}{\mathrm{d}\theta}\right|_{\hat{\theta}}\right| > \text{EPS}$ do                       $\triangleright$ Nonlinear regression.
    $\hat{\theta} \leftarrow \hat{\theta} - \left(\left.\frac{\mathrm{d}^2J'}{\mathrm{d}\theta^2}\right|_{\hat{\theta}}\right)^{-1}\left.\frac{\mathrm{d}J'}{\mathrm{d}\theta}\right|_{\hat{\theta}}$         $\triangleright$ Linearized regression (Eq. (3.15)).
    $\hat{x} \leftarrow \text{SolveF}(\hat{\theta}, w)$
end while
$\Sigma_{\text{post}}^{-1} \leftarrow \frac{1}{2}\left.\frac{\mathrm{d}^2J'}{\mathrm{d}\theta^2}\right|_{\hat{\theta}}$                    $\triangleright$ Laplace approximation (Eq. (3.22)).
$\theta^{\text{post}} \leftarrow \hat{\theta}$.
```

3.2 Nonlinear eigenproblems

As discussed in Chapter 1.2, the reduced-order model F for a linear stability analysis in thermoacoustics involves a nonlinear eigenproblem. For a matrix-valued function $L : \mathbb{C} \times \mathbb{R}^M \to \mathbb{C}^{n \times n}$ with the model parameters $\theta \in \mathbb{R}^M$, the eigenproblem is to find a so-called eigenvalue $s \in \mathbb{C}$ such that the matrix $L(s;\theta)$ becomes singular. The corresponding right and left eigenvectors $q, q^\dagger \in \mathbb{C}^n$ are respectively given by

$$L(s;\theta)q \equiv 0 \quad , \tag{3.30}$$

$$q^\dagger L(s;\theta) \equiv 0 \quad , \tag{3.31}$$

where $q, q^\dagger \neq 0$. For each eigenvalue s, the solutions to Eq. (3.30) and (3.31) are not unique, and form their respective eigenspaces. In a linear stability analysis, the real part of an eigenvalue s corresponds to the growth rate of a perturbation, and the imaginary part corresponds to its angular frequency. Thus, the eigenvalue s characterizes the stability of a system. If we equate the state x of a system to its stability, and perform a Bayesian regression in observations of the eigenvalue s, we require the first and second derivatives of the eigenvalue s with respect to the model parameters θ (Chapter 3.1). In this section, we describe two approaches to obtaining the eigenvalue sensitivities $\mathrm{d}s/\mathrm{d}\theta$ and $\mathrm{d}^2 s/\mathrm{d}\theta^2$.

The first approach involves a necessary and sufficient condition for singular $L(s;\theta)$ based on the determinant [109, Chapter 3]:

$$F(s,\theta) = \det\left(L(s;\theta)\right) \equiv 0 \quad . \tag{3.32}$$

Otherwise, we follow the procedure as outlined by Eq. (3.9)–(3.12) via implicit differentiation. The derivative of a determinant is given by Jacobi's formula, which directly follows from Laplace's formula [110, Chapter 0].

Theorem 3.1 (Laplace's formula)**.** *For any $1 \leq i \leq n$, the determinant of a matrix $A \in \mathbb{C}^{n \times n}$ is given by*

$$\det(A) = \sum_{j=1}^{n} A_{ij} \mathrm{adj}(A)_{ji} \quad , \tag{3.33}$$

where the adjugate of the matrix A is given by

$$\mathrm{adj}(A)_{ji} = (-1)^{i+j} \det\left((A_{kl})_{k \neq i, l \neq j} \right) \quad . \tag{3.34}$$

The submatrix $(A_{kl})_{k\neq i,l\neq j}$ *is obtained by omitting the i-th row and j-th column from the matrix A. If the matrix A is regular, then the inverse of the matrix A is given by*

$$A^{-1} = \frac{1}{\det(A)}\mathrm{adj}(A) \quad . \tag{3.35}$$

Theorem 3.2 (Jacobi's formula)**.** *The derivative of the determinant of a matrix* $A \in \mathbb{C}^{n\times n}$ *with respect to one of its entries* A_{ij} *is given by*

$$\frac{\mathrm{d}}{\mathrm{d}A_{ij}}\det(A) = \mathrm{adj}(A)_{ji} \quad . \tag{3.36}$$

If the matrix A is regular, the derivative of its adjugate is given by

$$\frac{\mathrm{d}}{\mathrm{d}A_{ij}}\mathrm{adj}(A)_{kl} = \mathrm{adj}(A)_{ji}A^{-1}_{kl} - \mathrm{adj}(A)_{ki}A^{-1}_{jl} \quad . \tag{3.37}$$

For instance, the partial derivatives of F with respect to θ are given by (Eq. (3.32))

$$\frac{\partial F}{\partial \theta_m} = \det(L(s;\theta))\,L^{-1}_{ji}(s;\theta)\frac{\partial L_{ij}}{\partial \theta_m} \quad , \tag{3.38}$$

$$\begin{aligned}\frac{\partial^2 F}{\partial \theta_m \partial \theta_n} &= \det(L(s;\theta))\,L^{-1}_{lk}(s;\theta)\frac{\partial L_{kl}}{\partial \theta_n}L^{-1}_{ji}(s;\theta)\frac{\partial L_{ij}}{\partial \theta_m} \\ &\quad - \det(L(s;\theta))\,L^{-1}_{jk}(s;\theta)\frac{\partial L_{kl}}{\partial \theta_n}L^{-1}_{li}(s;\theta)\frac{\partial L_{ij}}{\partial \theta_m} \\ &\quad + \det(L(s;\theta))\,L^{-1}_{ji}(s;\theta)\frac{\partial^2 L_{ij}}{\partial \theta_m \partial \theta_n} \quad .\end{aligned} \tag{3.39}$$

The second approach is based on perturbation theory, and bypasses the procedure as outlined by Eq. (3.9)–(3.12) [111]. Implicit differentiation of Eq. (3.30) gives

$$\frac{\partial L_{ij}}{\partial s}\frac{\mathrm{d}s}{\mathrm{d}\theta_m}q_j + \frac{\partial L_{ij}}{\partial \theta_m}q_j + L_{ij}(s;\theta)\frac{\mathrm{d}q_j}{\mathrm{d}\theta_m} \equiv 0 \quad . \tag{3.40}$$

Multiplying Eq. (3.40) with the left eigenvector $q^\dagger$ from the left gives the first-order eigenvalue sensitivity (Eq. (3.31))

$$\frac{\mathrm{d}s}{\mathrm{d}\theta_m} = -\frac{q^\dagger_i \dfrac{\partial L_{ij}}{\partial \theta_m}q_j}{q^\dagger_i \dfrac{\partial L_{ij}}{\partial s}q_j} \quad , \tag{3.41}$$

Implicit differentiation of Eq. (3.40) gives

$$\begin{aligned}
&\frac{\partial^2 L_{ij}}{\partial s^2}\frac{\mathrm{d}s}{\mathrm{d}\theta_n}\frac{\mathrm{d}s}{\mathrm{d}\theta_m}q_j+\frac{\partial^2 L_{ij}}{\partial s\partial\theta_n}\frac{\mathrm{d}s}{\mathrm{d}\theta_m}q_j+\frac{\partial L_{ij}}{\partial s}\frac{\mathrm{d}^2 s}{\mathrm{d}\theta_m\mathrm{d}\theta_n}q_j+\frac{\partial L_{ij}}{\partial s}\frac{\mathrm{d}s}{\mathrm{d}\theta_m}\frac{\mathrm{d}q_j}{\mathrm{d}\theta_n}\\
&\quad+\frac{\partial^2 L_{ij}}{\partial s\partial\theta_m}\frac{\mathrm{d}s}{\mathrm{d}\theta_n}q_j+\frac{\partial^2 L_{ij}}{\partial\theta_m\partial\theta_n}q_j+\frac{\partial L_{ij}}{\partial\theta_m}\frac{\mathrm{d}q_j}{\mathrm{d}\theta_n}\\
&\quad+\frac{\partial L_{ij}}{\partial s}\frac{\mathrm{d}s}{\mathrm{d}\theta_n}\frac{\mathrm{d}q_j}{\mathrm{d}\theta_m}+\frac{\partial L_{ij}}{\partial\theta_n}\frac{\mathrm{d}q_j}{\mathrm{d}\theta_m}+L_{ij}(s;\theta)\frac{\mathrm{d}^2 q_j}{\mathrm{d}\theta_m\mathrm{d}\theta_n}\quad\equiv 0\quad, \qquad (3.42)
\end{aligned}$$

Multiplying Eq. (3.42) with the left eigenvector $q^\dagger$ from the left gives the second-order eigenvalue sensitivity (Eq. (3.31))

$$\begin{aligned}
\frac{\mathrm{d}^2 s}{\mathrm{d}\theta_m\mathrm{d}\theta_n}=&-\frac{q_i^\dagger\left[\frac{\partial^2 L_{ij}}{\partial s^2}\frac{\mathrm{d}s}{\mathrm{d}\theta_n}\frac{\mathrm{d}s}{\mathrm{d}\theta_m}+\frac{\partial^2 L_{ij}}{\partial s\partial\theta_n}\frac{\mathrm{d}s}{\mathrm{d}\theta_m}+\frac{\partial^2 L_{ij}}{\partial s\partial\theta_m}\frac{\mathrm{d}s}{\mathrm{d}\theta_n}+\frac{\partial^2 L_{ij}}{\partial\theta_m\partial\theta_n}\right]q_j}{q_i^\dagger\frac{\partial L_{ij}}{\partial s}q_j}\\
&-\frac{q_i^\dagger\left[\frac{\partial L_{ij}}{\partial s}\frac{\mathrm{d}s}{\mathrm{d}\theta_m}+\frac{\partial L_{ij}}{\partial\theta_m}\right]\frac{\mathrm{d}q_j}{\mathrm{d}\theta_n}+q_i^\dagger\left[\frac{\partial L_{ij}}{\partial s}\frac{\mathrm{d}s}{\mathrm{d}\theta_n}+\frac{\partial L_{ij}}{\partial\theta_n}\right]\frac{\mathrm{d}q_j}{\mathrm{d}\theta_m}}{q_i^\dagger\frac{\partial L_{ij}}{\partial s}q_j} \qquad (3.43)
\end{aligned}$$

Substituting the first-order eigenvalue sensitivity into Eq. (3.40), the first-order eigenvector sensitivity is given by the linear system of equations

$$L_{ij}(s;\theta)\frac{\mathrm{d}q_j}{\mathrm{d}\theta_m}=\left[\left(\frac{q_k^\dagger\frac{\partial L_{kl}}{\partial\theta_m}q_l}{q_k^\dagger\frac{\partial L_{kl}}{\partial s}q_l}\right)\frac{\partial L_{ij}}{\partial s}-\frac{\partial L_{ij}}{\partial\theta_m}\right]q_j\quad. \qquad (3.44)$$

The choice of which approach to use for eigenvalue sensitivity analysis depends on the size of the eigenproblem and the number of parameters. On the one hand, the first approach based on Jacobi's formula is more straightforward as the implementation only requires a series of linear algebra operations as well as a nonlinear solver [112, 113]. The second approach based on perturbation theory additionally requires the computation of the left and right eigenvectors as well as the first derivative of the right eigenvector. On the other hand, once the eigenvectors and the first derivative are available, the implementation of the second approach only requires a series of matrix-vector multiplications. This is computationally inexpensive, and scales favorably with the number of parameters. Note that the second approach based on perturbation theory is identical to the second-order eigenvalue sensitivity analysis as outlined by MAGRI ET AL. up to notational differences [76]. In particular, the singularity in Eq. (3.44) is resolved by using the pseudoinverse, which is equivalent to appending the orthogonality condition $q \cdot \mathrm{d}q/\mathrm{d}\theta_m = 0$ [108, Chapter 4].

3.3 Nonparametric regression

As discussed in Chapter 1.3, uncertainty in reduced-order models is attributed to two sources, i.e. parametric and structural uncertainty. On the one hand, parametric uncertainty assumes that the form of the governing equations is correct. In Chapter 3.1, we introduced Bayesian regression as a tool to estimate the values of the model parameters and to quantify their uncertainties. Nevertheless, we have already established that reduced-order models are limited in their validity. Outside a certain regime, any particular reduced-order model is biased in its predictions and underestimates the uncertainty of its predictions. On the other hand, structural uncertainty questions the validity of the governing equations and their parametrizations. We relax the structure of the governing equations by dropping the requirement for model parameters to be constant over the entire design space. Unfortunately, parameter estimation in this setting is generally ill-posed. This section addresses structural uncertainty by formalizing the concepts of strong and weak parametrizations (see below). More specifically, we present strategies to regularize this ill-posed problem by replacing the notion of model parameters as universal constants with the more flexible notion of model parameters as stochastic processes.

In a first step, we relax the reduced-order model by elevating the model parameters to the rank of states, and thus combine state and parameter estimation:

$$\widetilde{x} = \begin{pmatrix} x \\ \theta \end{pmatrix} \quad , \tag{3.45}$$

where $\widetilde{x}$ denotes the augmented state. The augmented governing equations of a strongly parametrized reduced-order model are given by (Eq. (3.2))

$$\widetilde{F}(\widetilde{x}; w) = \begin{pmatrix} F\,(x, \theta; w) \\ \theta - \hat{\theta} \end{pmatrix} \equiv 0 \quad . \tag{3.46}$$

In terms of parameter estimation, this problem is solved by least-squares regression (Eq. (3.6), (3.7)), i.e. finding the optimal set of parameters $\hat{\theta}$ (Fig. 3.1). The number of additional equations required in the augmented governing equations is equal to the number of model parameters added to the augmented state. The parametrization is strongly constrained in the sense that the model parameters θ are identical for different design parameters w [114, Chapter 5].

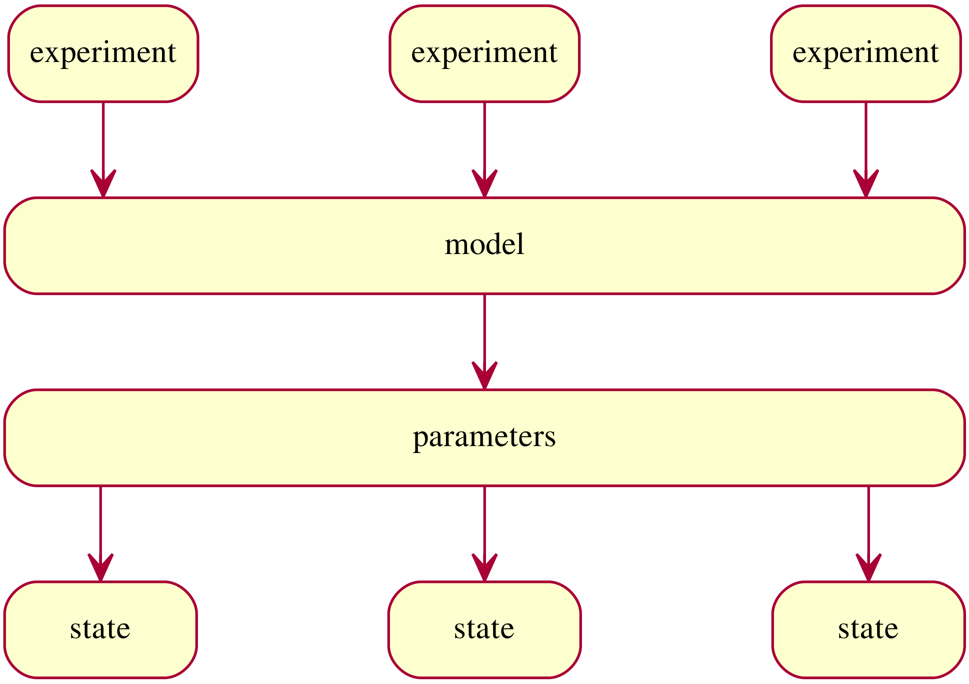


Fig. 3.1 A graphical model of a strongly parametrized reduced-order model. For experiments with different design parameters, the governing equations and model parameters stay the same while the state of the system varies accordingly. The model parameters are found using Bayesian regression (Chapter 3.1).

Following augmentation, strong parametrization is generally sufficient to close the governing equations of the reduced-order model but not necessarily the only approach. Taking a step back, the requirements towards the augmented set of governing equations is more generally described by the well-posedness of the problem [108, Chapter 12].

Definition 3.1 (Well-posed problem)**.** A problem is well-posed if the mathematical model of the physical phenomenon satisfies the following three conditions:

Existence. The problem has a solution.

Uniqueness. The solution to the problem is unique.

Stability. The solution to the problem is smooth with respect to the data.

In the context of partial differential equations, well-posedness refers to the existence, uniqueness and stability of solutions with respect to the specifications of initial and boundary conditions. In the context of parameter estimation, a well-posed problem consists of finding a suitably constrained parametrization for the augmented governing equations of the reduced-order model. On the one hand, it is not difficult to find any solution to the parameter estimation problem since it is always possible to fit a set of parameters to an experiment, especially if the number of model parameters is not predetermined.[4] On the other hand, uniqueness and stability of the parameter estimation problem are more difficult to achieve for at least two reasons. Firstly, the number of independent observations for a given set of design parameters is possibly less than the combined number of state variables and model parameters. For example, the number of state variables in a CFD simulation scales with the resolution of the discretization whereas the corresponding experiment is limited in its number of sensors without perturbing the flow field. Secondly, even if the number of independent observations is sufficient for any given set of design parameters, the data is generally polluted with noise. This reduces the information content of the data and affects the uniqueness and stability of the parameter estimation problem even when the number of independent observations is nominally sufficient.

In a second step, we therefore propose regularization as an alternative to strict constraints for augmenting the reduced-order model and formulating a well-posed problem (Fig. 3.2). The augmented governing equations of a weakly parametrized reduced-order model are given by (Eq. (3.46))

$$\widetilde{F}(\widetilde{x};w) = \begin{pmatrix} F\left(x,\theta;w\right) \\ \theta - \hat{\theta} \end{pmatrix} = \begin{pmatrix} 0 \\ \Theta_w \end{pmatrix} \quad , \tag{3.47}$$

where the stochastic process Θ_w is defined over the index set of the design parameters w (Definition 2.10). The parametrization is weakly constrained in the sense that the model parameters θ are allowed to differ for different design parameters w [114, Chapter 5]. In the spirit of statistical inference, we statistically model the variations in the model parameters as stochastic noise (Chapter 2.1). Most importantly, the stochastic noise is correlated such that model parameters for different design parameters are not statistically independent (Definition 2.4). In particular, we select the Gaussian process Θ_w to be a Gaussian process because of their desirable properties with respect to well-posedness (Definition 3.1):

---

[4] An instance of this is documented in the form of an anecdotal exchange between the physicists Freeman Dyson and Enrico Fermi [115]:

> *In desperation I asked Fermi whether he was not impressed by the agreement between our calculated numbers and his measured numbers. He replied, "How many arbitrary parameters did you use for your calculations?" I thought for a moment about our cut-off procedures and said, "Four." He said, "I remember my friend Johnny von Neumann used to say, with four parameters I can fit an elephant, and with five I can make him wiggle his trunk." With that the conversation was over.*

Existence. For a suitable choice of kernel functions, Gaussian processes are universal approximators, and may represent arbitrary functions θ as a function of w [116, Chapter 2].

Uniqueness. We note that a Gaussian process with a constant kernel is equivalent to strongly constrained parameter estimation, i.e. the parameters are estimated to be constant over the entire design space (Definition 2.13). Furthermore, we note that the constant kernel is a limiting case of the SE kernel for $\sigma_l \to \infty$ (Definition 2.15). This means that nonparametric regression with Gaussian processes is a generalization of parametric regression. The uniqueness in nonparametric regression is improved by increasing the value of the characteristic unit σ_l.

Stability. Similarly, we note that a Gaussian process with a white kernel is equivalent to individual estimates of the model parameters for each set of design parameters (Definition 2.14). Again, we note that the white kernel is a limiting case of the squared-exponential kernel for $\sigma_l \to 0$ (Definition 2.15). The smoothness of the Gaussian process, and hence the stability of the parameter estmation problem, is improved by increasing the value of the characteristic unit σ_l.

So far, our nonparametric regression algorithm consists of two ingredients:

1. Bayesian regression translates the observations from the state space into the parameter space. In a parametric regression, we estimate a global set of model parameters. In a nonparametric regression, we estimate a local set of model parameters for each set of design parameters and the corresponding observations. If not constrained or regularized, the weak parameter estimation problem is possibly ill-posed.
2. Our Gaussian processes live in the parameter space, and regularize the otherwise ill-posed problem. While Gaussian process regression is nonparametric, it involves a number of hyperparameters such as the hyperparameter σ_f for the constant kernel or the hyperparameter σ_l for the SE kernel (Definitions 2.13, 2.15).

The optimization of the hyperparameters inside the Gaussian processes is the third and missing ingredient to our algorithm. Unfortunately, it poses a chicken-or-egg dilemma: (i) without knowledge of the optimal hyperparameters, the Gaussian processes do not properly regularize the Bayesian regression; (ii) without a properly regularized Bayesian regression, we do not correctly infer the model parameters from the observations in the state space; and (iii) without correct estimates of the model parameters, we are not able to optimize the hyperparameters. In Algorithm 3.3, we break this cycle of infinite regress by introducing

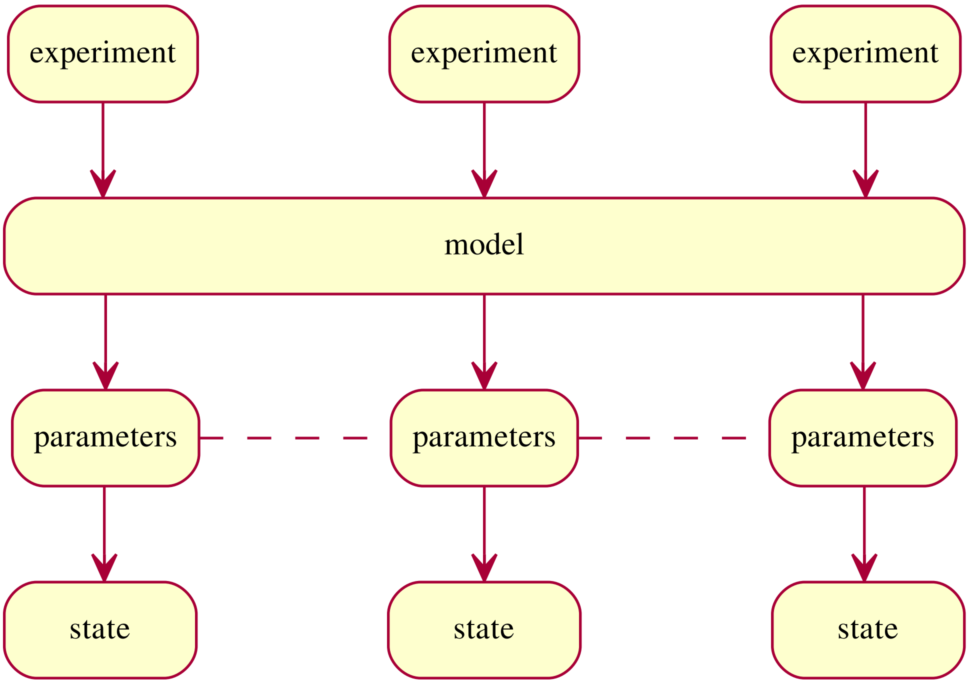


Fig. 3.2 A graphical model of a weakly parametrized reduced-order model. Unlike the strongly parametrized reduced-order model, the model parameters also vary with the design parameters of the experiment (Fig. 3.1). The model parameters are not strictly constrained but regularized by a stochastic process, e.g. a Gaussian process (dashed lines).

a prior for a single instance of model parameters. This prior forms the starting point for a recursive procedure based on the expectation-maximization (EM) algorithm [117]:

Expectation step. We take the Gaussian process from the step $k-1$ with the hyperparameters $\theta_{k-1}^{\text{hyp}}$ as well as the posterior probability distributions $P(\theta_j \mid (y_l)_{1\leq l\leq k-1}, M(\theta_{k-1}^{\text{hyp}}))$ for the steps $1 \leq j \leq k-1$ in order to predict the prior probability distribution $P(\theta_k \mid (y_l)_{1\leq l\leq k-1}, M(\theta_{k-1}^{\text{hyp}}))$ for the step k. We take the prior $P(\theta_k \mid (y_l)_{1\leq l\leq k-1}, M(\theta_{k-1}^{\text{hyp}}))$, and find the posterior probability distribution $P(\theta_k \mid (y_l)_{1\leq l\leq k}, M(\theta_{k-1}^{\text{hyp}}))$, which accounts for the observation y_k, using Bayesian regression over the observation y_k (Chapter 3.1). For the first step, when there is no Gaussian process yet, we have to directly provide a prior $P(\theta_1 \mid M)$, either from theoretical considerations or from an earlier uncertainty quantification.

Maximization step. For the steps $1 \leq j \leq k-1$, we update the posterior probability distributions $P(\theta_j \mid (y_l)_{1\leq l\leq k-1}, M(\theta_{k-1}^{\text{hyp}}))$ to $P(\theta_j \mid (y_l)_{1\leq l\leq k}, M(\theta_{k-1}^{\text{hyp}}))$ in order to account for the observation y_k:

$$\theta_j \leftarrow \theta_j + \frac{k(w_j, w_k)}{k(w_k, w_k)} \left(\theta_k^{\text{post}} - \theta_k^{\text{pre}} \right) \quad , \tag{3.48}$$

where $k(w, w')$ denotes the kernel function of the Gaussian process with the hyperparameters $\theta_{k-1}^{\text{hyp}}$ (Definition 2.12). Finally, we update the hyperparameters $\theta_{k-1}^{\text{hyp}}$ to θ_k^{hyp} to fit the posterior probability distributions $P(\theta_j \mid (y_l)_{1 \leq l \leq k}, M(\theta_{k-1}^{\text{hyp}}))$ for the steps $1 \leq j \leq k$ (Chapter 2.3).

We conclude our treatment of nonparametric regression with a discussion on the Gaussian process prior $P(\theta \mid M)$. Note that it is not the same as the prior $P(\theta_1 \mid M)$ from the first expectation step. The model parameter θ_1 is a random variable for the design parameter w_1 whereas the realization θ is a stochastic process over the entire design space W. By default, the mean and the variance of the observations are used to normalize the Gaussian process prior such that it has zero mean and unit variance (Chapter 2.2). In the case of Algorithm 3.3, the observations of the model parameters are the inferred posterior probability distributions $P(\theta_j \mid (y_l)_{1 \leq l \leq k-1}, M(\theta_{k-1}^{\text{hyp}}))$ for the steps $1 \leq j \leq k-1$. This leads to significant bias towards the earlier observations in the calculation of the Gaussian process prior. For instance, the posterior θ_1^{post}, and thus the observation y_k, enters the calculation of the Gaussian process prior at every step k whereas the posterior θ_K^{post}, and thus the observation y_K, only enters the algorithm at the very last step K. Ideally, the Gaussian process prior is consistent at every step k of the nonparametric regression algorithm.

In order to alleviate this shortcoming, we extend the expectation-maximization algorithm with a message-passing (MP) algorithm [118, Chapter 4]. Instead of a single loop over the index set, we restart our nonparametric regression from the beginning and perform multiple loops. However, we do not discard any posterior estimates of the model parameters but reuse them for the calculations of means and variances in all future loops. In the language of Bayesian networks and belief propagation, every instance of Bayesian regression would be viewed as a node in a graph, and the retained posterior estimates of the model parameters represent messages from all the past nodes. While there exists no proof of convergence for message passing over cyclical graphs, we have found the posterior estimates of the model parameters to be sufficiently converged after at most three loops. The construction of the Gaussian process prior using message passing is outlined in Algorithm 3.4.

Algorithm 3.3 Nonparametric regression with Gaussian processes and expectation maximization.

```
procedure LeastSquares($w$, $\theta^{\text{pre}}$, $y$)                    ▷ Bayesian regression (Chapter 3.1).
    // . . .                                                        ▷ Algorithm 3.2
    return $\theta^{\text{post}}$, $\Sigma_{\text{post}}$
end procedure

procedure GPFit([$w$], [$\theta^{\text{post}}$], [$\Sigma_{\text{post}}$])                     ▷ Fit Gaussian process.
    // . . .                                                        ▷ [102, Algorithm 5.1]
    return $\theta^{\text{hyp}}$
end procedure

procedure GPPredict([$w$], [$\theta^{\text{post}}$], $\theta^{\text{hyp}}$, $w$)             ▷ Predict from Gaussian process.
    // . . .                                                        ▷ [102, Algorithm 2.1]
    return $\theta$
end procedure

procedure GPCov($\theta^{\text{hyp}}$, $w_1$, $w_2$)              ▷ Compute (co-)variance between $w_1$ and $w_2$.
    $\sigma_{12} \leftarrow k(w_1, w_2)$                                  ▷ Definition 2.12.
    return $\sigma_{12}$
end procedure

$\hat{\theta} \leftarrow \theta^{\text{pre}}$
$\theta[1], \Sigma[1] \leftarrow \text{LeastSquares}(w[1], \hat{\theta}, y[1])$
$\theta^{\text{hyp}} \leftarrow \text{GPFit}([w[1]], [\theta[1]], [\Sigma[1]])$.
for $2 \leq k \leq K$ do                                            ▷ Iteration over index set.
    $\hat{\theta} \leftarrow \text{GPPredict}(w[1:k-1], \theta[1:k-1], \theta^{\text{hyp}}, w[k])$          ▷ Expectation step.
    $\theta[k], \Sigma[k] \leftarrow \text{LeastSquares}(w[k], \hat{\theta}, y[k])$
    for $1 \leq j \leq k-1$ do                                      ▷ Bayesian update.
        $\theta[j] \leftarrow \theta[j] + \frac{\text{GPCov}(\theta^{\text{hyp}}, w[j], w[k])}{\text{GPCov}(\theta^{\text{hyp}}, w[k], w[k])}(\theta[k] - \hat{\theta})$
    end for
    $\theta^{\text{hyp}} \leftarrow \text{GPFit}(w[1:k], \theta[1:k], \Sigma[1:k])$.               ▷ Maximization step.
end for
```

Algorithm 3.4 Nonparametric regression with Gaussian processes, expectation maximization and message passing.

```
$[\theta^{\mathrm{MP}}] \leftarrow []$                                                ▷ Initialize message passing.

procedure GPFit($[w]$, $[\theta^{\mathrm{post}}]$, $[\Sigma_{\mathrm{post}}]$)                         ▷ Fit Gaussian process.
    $\theta^{\mathrm{post}} \leftarrow (\theta^{\mathrm{post}} - \mathrm{mean}([\theta^{\mathrm{MP}}]))/\mathrm{std}([\theta^{\mathrm{MP}}])$                 ▷ Normalize $\theta^{\mathrm{post}}$.
    $\Sigma_{\mathrm{post}} \leftarrow \Sigma_{\mathrm{post}}/\mathrm{std}([\theta^{\mathrm{MP}}])^2$                              ▷ Normalize $\Sigma_{\mathrm{post}}$.
    // ...                                                        ▷ [102, Algorithm 5.1]
    return $\theta^{\mathrm{hyp}}$
end procedure

procedure GPPredict($[w]$, $[\theta^{\mathrm{post}}]$, $\theta^{\mathrm{hyp}}$, $w$)                ▷ Predict from Gaussian process.
    // ...                                                        ▷ [102, Algorithm 2.1]
    $\theta \leftarrow \theta * \mathrm{std}([\theta^{\mathrm{MP}}]) + \mathrm{mean}([\theta^{\mathrm{MP}}])$                       ▷ Denormalize $\theta$.
    return $\theta$
end procedure

procedure GPCov($\theta^{\mathrm{hyp}}$, $w_1$, $w_2$)          ▷ Compute (co-)variance between $w_1$ and $w_2$.
    $\sigma_{12} \leftarrow k(w_1, w_2)$                                              ▷ Definition 2.12.
    $\sigma_{12} \leftarrow \sigma_{12} * \mathrm{std}([\theta^{\mathrm{MP}}])^2$                                  ▷ Denormalize $\sigma_{12}$.
    return $\sigma_{12}$
end procedure

for $1 \le l \le \#\mathrm{MP}$ do                                           ▷ Loops of message passing.
    $\hat{\theta} \leftarrow \theta^{\mathrm{pre}}$
    $\theta[1], \Sigma[1] \leftarrow$ LeastSquares($w[1], \hat{\theta}, y[1]$)
    Append $\theta[1]$ to $\theta^{\mathrm{MP}}$.                                     ▷ Update message passing.
    $\theta^{\mathrm{hyp}} \leftarrow$ GPFit($[w[1]], [\theta[1]], [\Sigma[1]]$).
    for $2 \le k \le K$ do                                           ▷ Iteration over index set.
        $\hat{\theta} \leftarrow$ GPPredict($w[1:k-1], \theta[1:k-1], \theta^{\mathrm{hyp}}, w[k]$)          ▷ Expectation step.
        $\theta[k], \Sigma[k] \leftarrow$ LeastSquares($w[k], \hat{\theta}, y[k]$)
        for $1 \le j \le k-1$ do                                            ▷ Bayesian update.
            $\theta[j] \leftarrow \theta[j] + \frac{\mathrm{GPCov}(\theta^{\mathrm{hyp}}, w[j], w[k])}{\mathrm{GPCov}(\theta^{\mathrm{hyp}}, w[k], w[k])}(\theta[k] - \hat{\theta})$
        end for
        Append $\theta[k]$ to $\theta^{\mathrm{MP}}$.                                 ▷ Update message passing.
        $\theta^{\mathrm{hyp}} \leftarrow$ GPFit($w[1:k], \theta[1:k], \Sigma[1:k]$).                ▷ Maximization step.
    end for
end for
```

Chapter 4

Rijke tube

We apply the ideas from Chapters 1 to 3, namely reduced-order modeling and statistical inference, to the Rijke tube, a simple yet representative thermoacoustic system. We briefly describe the Rijke tube set-up by RIGAS ET AL. [119] (Chapter 4.1). The thermoacoustic models consist of two components: a base-flow model and an acoustic model. The base-flow models are compared to temperature measurements at the inlet and the outlet of the Rijke tube as well as on its outside wall (Chapter 4.2). The optimal model parameters are found using nonlinear regression (Chapter 3.1). The acoustic models are compared to growth (or decay) rates and frequencies extracted from pressure measurements (Chapter 4.3). In a first step, we estimate the model parameters as well as their uncertainties by using Bayesian regression (Chapters 3.1, 3.2). In a second step, we improve upon the results by using nonparametric regression, and compare the various reduced-order models to each other from the rigorous perspective of Bayesian inference (Chapters 2.3, 3.3). The thermoacoustic models were developed in collaboration with PROF. MATTHEW JUNIPER at the UNIVERSITY OF CAMBRIDGE. The experimental data was provided by DR. NICHOLAS JAMIESON.

4.1 Experimental set-up

The electrically heated Rijke tube is a simple and inexpensive system that contains all the essential physics of thermoacoustic instability without the complexities of industrial systems such as jet or rocket engines. Its simplicity means that it is well suited for comparing experimental observations to theoretical predictions. The Rijke tube consists of a vertical pipe with an electric heater installed inside. When hot, the electric heater causes a buoyancy-driven flow to form inside the tube. For electric powers in the range of 0 to $400\,\mathrm{W}$, velocities of the base flow are measured to be on the order of 0.1 to $1\,\mathrm{m/s}$. This corresponds to virtually zero Mach number.

Pressure fluctuations form and travel through the Rijke tube. The pressure fluctuations are reflected at the ends of the Rijke tube to form a standing wave. As a result, compression and rarefaction of the air inside the Rijke tube and the associated changes in the flow speed at the electric heater cause the convective heat transfer between the electric heater and the air to become unsteady. A small time delay exists between the velocity and heat release rate fluctuations [120]. This time delay is crucial in determining the phase difference between the pressure and heat release rate fluctuations. The boundary conditions and the heat release mechanism determine the shape of the standing wave as well as its growth or decay rate. The phase difference is used in the Rayleigh criterion to predict the onset of thermoacoustic instability [5].

On the left-hand side of Fig. 4.1, a one-dimensional schematic of the Rijke tube is presented. The Rijke tube has a length of $l_u + l_b$ and a diameter of d. The electric heater is located at a distance of l_u from the bottom. On top of the Rijke tube, an adjustable iris is installed which acts as an orifice. On the right-hand side of Fig. 4.1, a schematic of the electric heater is shown. It consists of a ceramic disc with nichrome wire wrapped around it. Measurements of the Rijke tube and the electric heater are given in Table 4.1. More details on the experimental set-up and the instrumentation are given by RIGAS ET AL. [119]. Note that the length l_b of the upper section is different from their specification because the iris is treated as an extension to the upper section.

Table 4.1 Measurements of Rijke tube.

quantity	measurement	
d	47.4 mm	inner diameter of Rijke tube
d_h	1 mm	distance between two windings
d_i	32.2 mm	inner diameter of electric heater
d_w	0.559 mm	wire diameter
l_b	784 mm	length of upper section of Rijke tube
l_h	15 mm	length of electric heater
l_u	250 mm	length of lower section of Rijke tube
n_w	12	number of wire windings
t	1.5 mm	thickness of Rijke tube

The electric power is incremented from 0 to 340 W in steps of 10 W. Twenty repetitions are performed at each step. A microphone placed 55 mm from the bottom of the Rijke tube is used to measure the pressure fluctuations over time. The complex amplitudes of the pressure fluctuations are extracted from the filtered signal via a Hilbert transform [121]. As the electric power increases, the Rijke tube exhibits a subcritical bifurcation. Therefore, two different measurement techniques are used: Below the Hopf point, the decay rates and

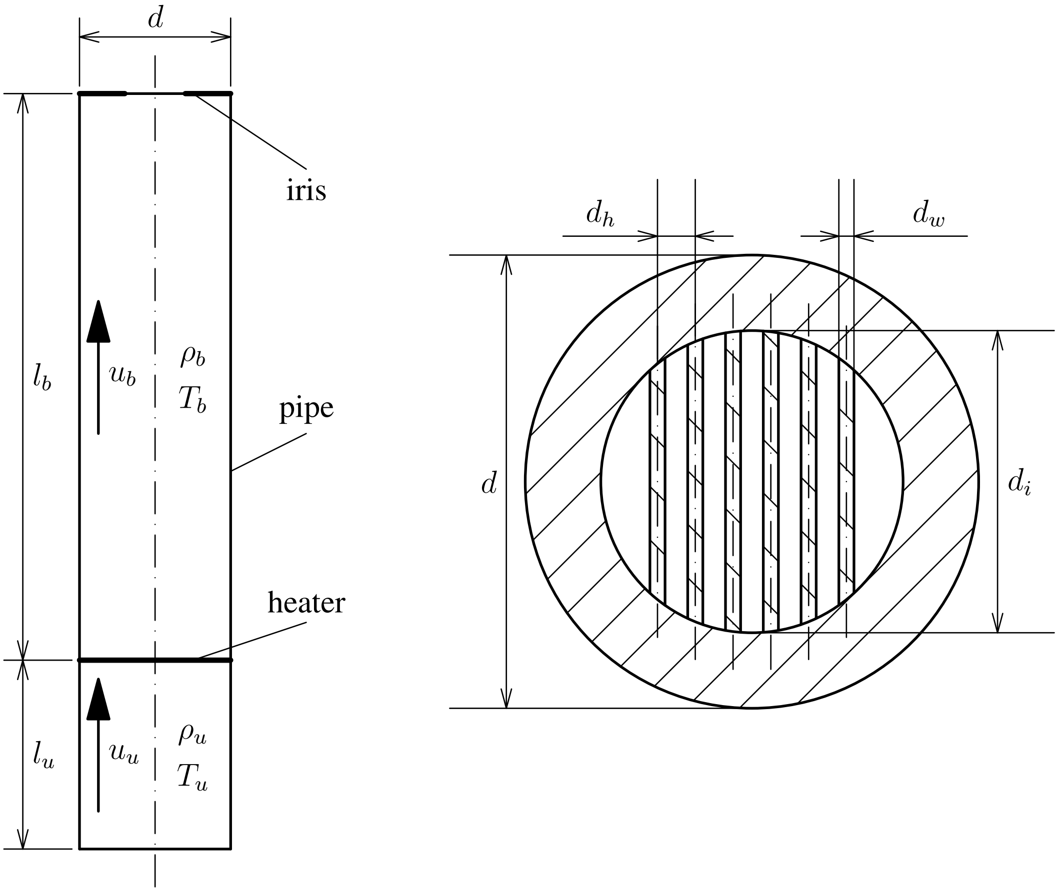


Fig. 4.1 Schematics of Rijke tube (left) and electric heater (right).

angular frequencies are measured from pulsed forcing. Above the fold point, the growth rates and angular frequencies are measured by holding the system steady with feedback control and then releasing it. In each case, a linear regime is identified where oscillations grow or decay exponentially. No measurements are taken in the bistable regime. The measured growth and decay rates as well as the frequencies are shown in Fig. 4.2. More details regarding the Hopf and fold points as well as the signal processing of the sampled pressure fluctuations are given by RIGAS ET AL. [119].

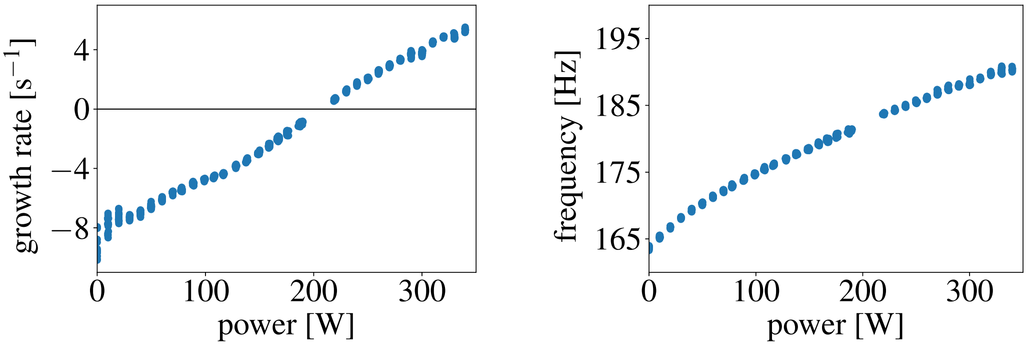


Fig. 4.2 Experimental measurements of growth rates (left) and frequencies (right) in Rijke tube. A subcritical bifurcation occurs near $\dot{Q} = 200\,\mathrm{W}$.

4.2 Base flow

The base flow inside the Rijke tube is buoyancy-driven: The electric heater increases the temperature of the air inside the pipe. The pressure difference due to the surrounding air begins to overcome the gravitational pull acting on the air inside the Rijke tube. The mass flow rate increases until the net force is in balance with the various drag mechanisms inside the Rijke tube. As the acoustics depend on the base flow, a model for the latter is required.

For a one-dimensional base flow in the vertical x-direction, mass balance is given by

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{u}) = 0 \quad . \tag{4.1}$$

In a steady flow, Eq. (4.1) reduces to

$$\rho \bar{u} A = \text{const} \quad , \tag{4.2}$$

where A denotes the cross section area of the Rijke tube. The bar indicates that the velocity u in x-direction is averaged over the cross section. Evaluated at the inlet (subscript 1), just downstream of the electric heater (subscript f) and at the outlet (subscript 2), Eq. (4.2) gives

$$\rho_1 \bar{u}_1 A_1 = \rho_f \bar{u}_f A_1 = \rho_2 \bar{u}_2 A_2 \quad , \tag{4.3}$$

where A_1 denotes the cross-section area of the Rijke tube, and A_2 denotes the cross-section area of the iris. The density and the temperature between the inlet of the Rijke tube and the electric heater are assumed to be constant. In general, the condition just downstream of the electric heater is different from the condition at the outlet of the Rijke tube due to heat loss through the pipe walls.

Energy balance is given by

$$\rho c_v \frac{\partial T}{\partial t} + \rho \vec{u} \cdot c_p \nabla T = \dot{q} \quad . \tag{4.4}$$

Effects due to viscosity and heat conduction are neglected. The substantial derivative of the pressure is also assumed to be negligible. Between the inlet of the Rijke tube and a plane just downstream of the electric heater, Eq. (4.4) gives

$$\rho_1 \bar{u}_1 A_1 c_p \left(T_f - T_1 \right) = \dot{Q} \quad , \tag{4.5}$$

where $\dot{Q}$ denotes the heat release rate due to the electric heater, which is equal to its electric power. Between a plane just downstream of the electric heater and the outlet of the Rijke tube, Eq. (4.4) gives

$$\rho_1 \bar{u}_1 A_1 c_p \frac{\mathrm{d}T}{\mathrm{d}x} = \pi d \dot{q}_x(T) \quad , \quad l_u \le x < l_u + l_b \quad , \tag{4.6}$$

$$T(l_u) = T_f \quad , \tag{4.7}$$

where $\dot{q}_x$ denotes a heat release rate times unit area. The heat release rate times unit area $\dot{q}_x$ is assumed to be only a function of the temperature T. Hence, Eq. (4.6) is a first-order ordinary differential equation. The matching boundary condition is given by Eq. (4.7).

Momentum balance is given by

$$\rho \frac{\partial \vec{u}}{\partial t} + \rho \left(\vec{u} \cdot \nabla \right) \vec{u} = -\nabla p + \rho \vec{g} \quad , \tag{4.8}$$

where $\vec{g}$ denotes gravity. Its absolute value g is $9.81\,\mathrm{m/s^2}$. Losses due to viscous and inviscid effects are treated separately. The first integral of a steady streamline between the inlet and the outlet of the Rijke tube gives

$$\frac{1}{2} \int_{t_1}^{t_2} \rho \frac{\mathrm{d}}{\mathrm{d}t} \left(u^2 \right) \mathrm{d}t = -\left(p_2 - p_1 \right) - g \int_0^{l_u + l_b} \rho \, \mathrm{d}x \quad . \tag{4.9}$$

Integrating by parts, the integral on the left-hand side of Eq. (4.9) gives

$$\frac{1}{2} \int_{t_1}^{t_2} \rho \frac{\mathrm{d}}{\mathrm{d}t} \left(u^2 \right) \mathrm{d}t = \left(\frac{1}{2} \rho_2 u_2^2 - \frac{1}{2} \rho_1 u_1^2 \right) - \frac{1}{2} \int_{t_1}^{t_2} u^2 \frac{\mathrm{d}\rho}{\mathrm{d}t} \mathrm{d}t \quad . \tag{4.10}$$

For buoyancy-driven flows, the Boussinesq approximation states that variations in density are negligible unless related to gravity [122, Chapter 14]. Strictly speaking, the Boussinesq approximation may only be applied if the temperature gradients are small. In the Rijke tube, this is clearly violated for large heat release rates $\dot{Q}$, where the temperature difference between the inlet and the outlet of the Rijke tube is of the same order of magnitude as the temperature inside the Rijke tube. Nevertheless, the Boussinesq approximation is applied because it simplifies the analysis. Eq. (4.9) becomes

$$\left(p_2+\frac{1}{2}\rho_2 u_2^2\right)-\left(p_1+\frac{1}{2}\rho_1 u_1^2\right)+g\int_0^{l_u+l_b}\rho\,\mathrm{d}x=0 \quad . \tag{4.11}$$

Since loss mechanisms are treated separately, it is assumed that the flow is isentropically accelerated from ambient conditions before entering the Rijke tube, and isentropically decelerated to ambient conditions after exiting the Rijke tube. Thus, the sum of the static pressure and the dynamic pressure at the inlet of the Rijke tube differs from the sum of the static pressure and the dynamic pressure at the outlet of the Rijke tube by

$$\left(p_1+\frac{1}{2}\rho_1 u_1^2\right)-\left(p_2+\frac{1}{2}\rho_2 u_2^2\right)=\rho_1 g\left(l_u+l_b\right) \quad . \tag{4.12}$$

By adding inviscid drag at the inlet and the outlet of the Rijke tube as well as viscous drag at the electric heater, the final form of the momentum balance is

$$g\int_{l_u}^{l_u+l_b}\left(\rho_1-\rho(T)\right)\mathrm{d}x-C_1\frac{1}{2}\rho_1\bar{u}_1^2-C_2\left(\frac{A_2}{A_1}\right)\frac{1}{2}\rho_2\bar{u}_2^2-C_3\frac{\mu l_h}{d_h^2}\bar{u}_1=0 \quad , \tag{4.13}$$

where C_1, C_2 and C_3 are dimensionless proportionality constants. Inviscid loss mechanisms do not depend on viscosity. As such, they are Reynolds-independent in theory. They scale with the dynamic pressure $\rho u^2/2$, and only depend on the geometry. The proportionality constant C_1 does not depend on the area of the cross section at the inlet A_1 because the environment is much larger than the Rijke tube. The proportionality constant C_2 is a function of the ratio between the adjustable cross-section area of the iris and the cross-section area of the Rijke tube A_2/A_1. The form of the viscous drag term is inspired by the laws for channel and pipe flows as well as the flow around a sphere which state that the drag coefficient is inversely proportional to the Reynolds number [123, Chapter 4.9]. As the density in the lower section of the Rijke tube is assumed to be equal to the ambient density, the buoyancy term only has to be evaluated in the upper section of the Rijke tube.

4.2.1 Adiabatic model

Eq. (4.3), (4.5), (4.6), (4.7) and (4.13) govern the buoyancy-driven flow inside the Rijke tube. In order to evaluate the buoyancy term in Eq. (4.13), the heat release rate times unit area $\dot{q}_x$ in Eq. (4.5) needs to be specified. In a first approximation, the base flow in the upper section of the Rijke tube is assumed to be adiabatic ($\dot{q}_x = 0$). Eq. (4.5) gives

$$T(x) = T_f \quad , \quad l_u \leq x < l_u + l_b \quad . \tag{4.14}$$

It is shown later that this assumption is not justified. Nevertheless, it yields a simple model which offers a first quantitative understanding of the buoyancy-driven flow inside the Rijke tube. This is particularly relevant for the mass flow rate and the flow speeds, for which no measurements exist. Their knowledge allows us to estimate the dimensionless numbers that characterize the base flow and to classify their respective regimes.

With uniform density in the upper section of the Rijke tube, Eq. (4.13) becomes

$$\left(\rho_1 - \rho_f\right) g l_b - C_1 \frac{1}{2} \rho_1 \bar{u}_1^2 - C_2 \left(\frac{A_2}{A_1}\right) \frac{1}{2} \rho_2 \bar{u}_2^2 - C_3 \frac{\mu l_h}{d_h^2} \bar{u}_1 = 0 \quad . \tag{4.15}$$

The flow speeds $\bar{u}_1$ and $\bar{u}_2$ are given by

$$\bar{u}_1 = \frac{\dot{Q}}{\rho_1 A_1 c_p \left(T_f - T_1\right)} \quad , \tag{4.16}$$

$$\bar{u}_2 = \frac{\dot{Q}}{\rho_2 A_2 c_p \left(T_f - T_1\right)} \quad . \tag{4.17}$$

Substituting Eq. (4.16) and (4.17) into Eq. (4.15) gives

$$(1 - \vartheta) - C_1' \left(\frac{\dot{Q}' \vartheta}{1 - \vartheta}\right)^2 - C_2' \vartheta \left(\frac{\dot{Q}'}{1 - \vartheta}\right)^2 - C_3' \frac{\dot{Q}' \vartheta}{1 - \vartheta} = 0 \quad , \tag{4.18}$$

$$C_1' = \frac{1}{2 g l_b} C_1 \quad , \quad C_2' = \frac{1}{2 g l_b} C_2 \left(\frac{A_2}{A_1}\right) \left(\frac{A_1}{A_2}\right)^2 \quad , \quad C_3' = \frac{\mu l_h}{\rho_1 g l_b d_h^2} C_3 \quad ,$$

$$\dot{Q}' = \frac{\dot{Q}}{\rho_1 A_1 c_p T_1} \quad , \quad \vartheta = \frac{\rho_f}{\rho_1} \quad .$$

Eq. (4.18) has a single unknown ϑ. The non-dimensional quantity ϑ is between zero and one. Finally, Eq. (4.18) gives a cubic equation:

$$\vartheta^3 + \left(C_1' \dot{Q}'^2 - C_3 \dot{Q}' - 3\right) \vartheta^2 + \left(C_2' \dot{Q}'^2 + C_3' \dot{Q}' + 3\right) \vartheta - 1 = 0 \quad . \tag{4.19}$$

If the drag is purely viscous ($C_1 = C_2 = 0$), Eq. (4.19) has a unique physical solution ϑ: Discarding the trivial solution $\vartheta = 1$, the cubic equation is reduced to a quadratic equation:

$$\vartheta^2 - \left(C_3'\dot{Q}' + 2\right)\vartheta + 1 = 0 \quad . \tag{4.20}$$

Due to Viète's theorem, Eq. (4.20) has two positive solutions ϑ for positive $C_3'\dot{Q}'$, one larger and one smaller than one. Because ϑ has to be between zero and one for a buoyancy-driven flow, it follows that the physical solution to Eq. (4.19) is indeed unique. The physical solution ϑ continues to exist when the inviscid drag terms are added back to Eq. (4.19).

Eq. (4.19) represents a model of the steady flow inside the Rijke tube. The input variables are the electric power $\dot{Q}$ and the cross-section area A_2 of the iris. The output variable ϑ characterizes the state of the Rijke tube. All other flow variables are calculated from ϑ. The drag coefficients C_1, C_2 and C_3 are the model parameters, which are found from a nonlinear regression (Chapter 3.1). The predicted temperature at the outlet of the Rijke tube is compared to experimental measurements for an open iris. The relationship between the observable quantity T_2 and the output variable ϑ is given by

$$T_2 = \frac{T_1}{\vartheta} \quad . \tag{4.21}$$

Within the experiment, the drag coefficients C_1, C_2 and C_3 all have a similar effect on the buoyancy-driven flow inside the Rijke tube. If the drag increases, the velocity decreases. If the heat release rate $\dot{Q}$ is held constant, a slower mass flow rate results in a larger temperature jump from T_1 to T_f. The buoyancy force increases until it is in balance with the drag. Therefore, a simultaneous regression in all three parameters is an ill-posed problem. Because the inviscid losses at the inlet and the outlet of a pipe are well-studied phenomena, they are taken from flow charts as $C_1 = 0.578$ and $C_2 = 2$ [124].

In Fig. 4.3, the results of the nonlinear regression are shown. The predicted temperature at the outlet of the Rijke tube matches the experimental measurements well. The flow speeds at the inlet and the outlet of the Rijke tube are on the order of magnitude of 0.75 and 1.3 m/s respectively, which are reasonable values. The fitting of the viscous drag coefficient gives $C_3 = 49.23$ (Table 4.2). A theoretical estimate gives $C_{3,\mathrm{th}} = 156$. This is based on the crude assumption that the flow through the electric heater behaves like a collection of $n_w + 1$ channel flows [123, Chapter 4.2]. The pipe Reynolds number is approximately 2350 in the lower section of the Rijke tube, which is just above the critical Reynolds number of 2300. But since the dynamic viscosity of air increases with temperature, the buoyancy-driven flow in the upper section of the Rijke tube is expected to be laminar. At the electric heater, the blockage ratio is approximately 50 %, which results in a wire Reynolds number of approximately 56.

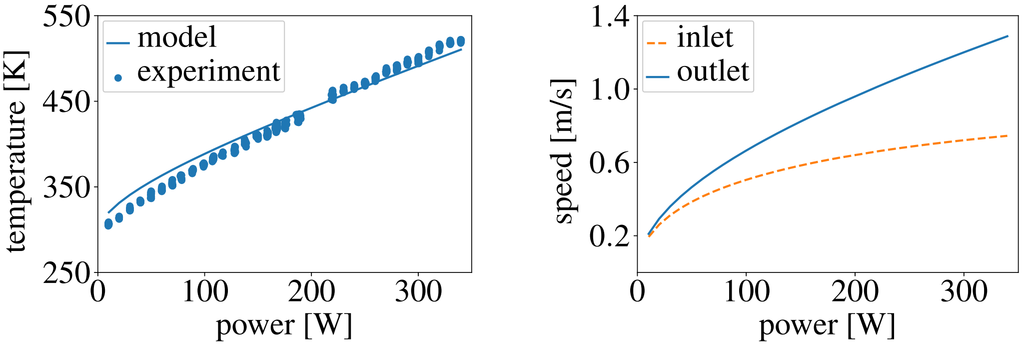


Fig. 4.3 Model predictions of outlet temperature (left) and inlet/outlet speed (right) from adiabatic base-flow model.

A wake region is expected to form, which would result in a pressure drop on the trailing side of the wire and cause inviscid drag [123, Chapter 4.12]. The wire Reynolds number is above 45, which marks the onset of vortex shedding. Nevertheless, it is assumed that the vortices quickly dissipate because of the laminar pipe flow in the upper section of the Rijke tube.

Table 4.2 Nonlinear regression of adiabatic base-flow model. The drag coefficients C_1 and C_2 are taken from flow charts [124].

model parameter	estimated value	
C_1	0.578	inviscid loss at pipe inlet
C_2	2	inviscid loss at pipe outlet ($A_2/A_1 = 1$)
C_3	49.23	viscous drag at electric heater

The magnitudes of the individual drag mechanisms in Eq. (4.13) are calculated from the parameters C_1, C_2 and C_3. The ratios between the inviscid drag at the inlet and the outlet of the Rijke tube and the viscous drag due to the electric heater are given by

$$\frac{C_1\rho_1 u_1^2 d_h^2}{2C_3\mu l_h u_1} \approx 0.020 \quad , \quad \frac{C_2\rho_2 u_2^2 d_h^2}{2C_3\mu l_h u_1} \approx 0.117 \quad . \tag{4.22}$$

This shows that the buoyancy-driven flow inside the Rijke tube is dominated by the viscous drag due to the electric heater. A similar comparison shows that the viscous drag due to the inside wall of the Rijke tube is also negligible [123, Chapter 4.2]:

$$\frac{32(l_u + l_b)u_1 d_h^2}{2C_3 l_h u_1 d^2} \approx 0.010 \quad . \tag{4.23}$$

Finally, an asymptotic analysis is performed on Eq. (4.20). Although the adiabatic base-flow model is not expected to behave pathologically in the limits of small and large heat release rates $\dot{Q}$, the same will not necessarily hold for more complicated base-flow models. Therefore, it will prove advantageous to be able to compare their asymptotic behaviors to the one of the adiabatic base flow.

Substituting Eq. (4.16) and (4.17) into Eq. (4.15) gives

$$(1-\vartheta)^3 = C_1'\dot{Q}'^2\vartheta^2 + C_2'\dot{Q}'^2\vartheta + C_3'\dot{Q}'\vartheta(1-\vartheta) \quad . \tag{4.24}$$

The left-hand side of Eq. (4.24) has a value between zero and one, which the right-hand side of Eq. (4.24) has to match. As $\dot{Q}'$ goes to infinity, the right-hand side of Eq. (4.24) has to stay finite. As ϑ goes to zero, the right-hand side of Eq. (4.24) is dominated by the term in C_2'. The results of the asymptotic analysis are

$$\vartheta \sim \frac{1}{C_2'\dot{Q}'^2} \quad , \quad \bar{u}_1 \sim \frac{1}{C_2'\dot{Q}'} \quad , \quad \bar{u}_2 \sim \dot{Q}' \quad . \tag{4.25}$$

As $\dot{Q}'$ goes to infinity, so do the temperature T_2 and the flow speed $\bar{u}_2$ at the outlet of the Rijke tube. The flow speed $\bar{u}_1$ at the outlet of the Rijke tube and the mass flow rate go to zero. This behavior is consistent within the adiabatic base-flow model: Because the density in the upper section of the Rijke tube cannot become negative, the buoyancy term in Eq. (4.15) is bounded by $\rho_1 g l_b$. As the change in specific momentum across the electric heater $\bar{u}_2 - \bar{u}_1$ goes to infinity, the mass flow rate has to go to zero.

As $\dot{Q}'$ goes to zero, ϑ must go to one. The right-hand side of Eq. (4.24) is then dominated by the term in C_3'. The results of the asymptotic analysis are

$$1-\vartheta \sim \sqrt{C_3'\dot{Q}'} \quad , \quad \bar{u}_1 \sim \bar{u}_2 \sim \sqrt{\frac{\dot{Q}'}{C_3'}} \quad . \tag{4.26}$$

As $\dot{Q}'$ goes to zero, so do the flow speeds $\bar{u}_1$ and $\bar{u}_2$ at the inlet and the outlet of the Rijke tube as well as the mass flow rate. The temperature T_2 at the outlet of the Rijke tube goes to ambient temperature. This behavior is physical since zero heat release rate should result in zero mass flow rate and no change in temperature across the electric heater.

4.2.2 Model with heat loss

In the adiabatic base-flow model, the temperature just downstream of the electric heater T_f is equal to the temperature T_2 at the outlet of the Rijke tube. While there are no measurements of the temperature or the flow speed inside the Rijke tube, thermocouples are installed on the outside wall of the Rijke tube (Table 4.3). In Fig. 4.4, the measured wall temperature profiles downstream of the electric heater are shown for three different electric powers $\dot{Q}$. At an electric power of 300 W, the wall temperatures at the electric heater ($x = 0.25\,\mathrm{m}$) and close to the outlet of the Rijke tube ($x = 0.95\,\mathrm{m}$) differ by more than 150 K. Therefore, it is not justified to assume that the flow inside the Rijke tube is adiabatic.

Table 4.3 Thermocouple positions on outside wall of Rijke tube.

	1	2	3	4	5
x	0.05 m	0.25 m	0.50 m	0.75 m	0.95 m

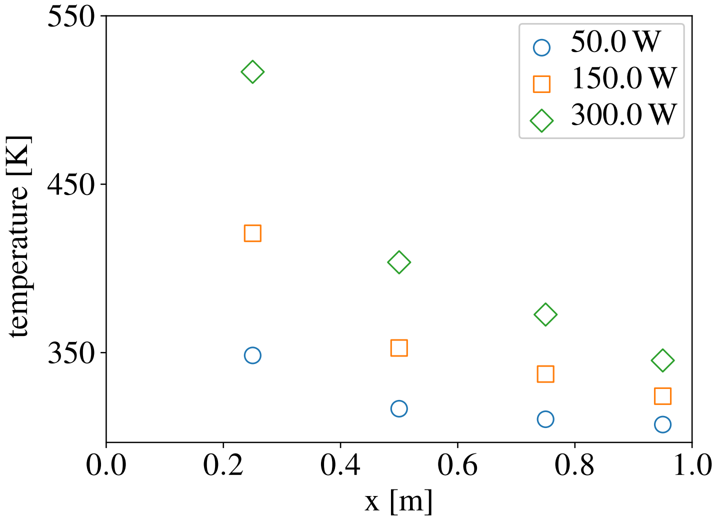


Fig. 4.4 Experimental measurements of outside wall temperature profiles downstream of electric heater. For high electric powers, the temperature just downstream of the electric heater significantly differs from the temperature at the outlet of the Rijke tube.

In a non-adiabatic base-flow model, it is assumed that heat is lost through the wall of the Rijke tube. In a first approximation, it is assumed that the heat loss is proportional to the difference between the temperature inside the Rijke tube and the ambient temperature. The heat release rate times unit area is given by

$$\dot{q}_x = -h\left[T(x) - T_1\right] \quad , \tag{4.27}$$

where h denotes the heat transfer coefficient between the buoyancy-driven flow inside the Rijke tube and the environment. The heat transfer between the buoyancy-driven flow inside the Rijke tube and the environment consists of two heat transfer mechanisms coupled in series: the heat transfer between the buoyancy-driven flow inside the Rijke tube and the pipe as well as the heat transfer between the pipe and the environment. The heat transfer coefficient is given by

$$\frac{1}{h} = \frac{1}{h_i} + \frac{1}{h_o} \quad , \tag{4.28}$$

where h_i denotes the heat transfer coefficient between the buoyancy-driven flow inside the Rijke tube and the pipe, and h_o denotes the heat transfer coefficient between the pipe and the environment. Integration of Eq. (4.6) with Eq. (4.7) as boundary condition gives

$$T(x) = T_1 + (T_f - T_1)\exp\left(-\frac{\pi d h (x - l_u)}{\rho_1 \bar{u}_1 A_1 c_p}\right) \quad , \quad l_u \le x < l_u + l_b \quad . \tag{4.29}$$

The buoyancy term in Eq. (4.13) evaluates to

$$g\int_{l_u}^{l_u+l_b} (\rho_1 - \rho(T))\,\mathrm{d}x = -\frac{\rho_1 g}{k}\log\left(\frac{T_2}{T_f}\right) \quad , \quad k = \frac{\pi d h}{\rho_1 \bar{u}_1 A_1 c_p} \quad . \tag{4.30}$$

In the adiabatic limit, the buoyancy term would evaluate to

$$-\lim_{h\to 0}\frac{\rho_1 g}{k}\log\left(\frac{T_2}{T_f}\right) = -\rho_1 g \lim_{k\to 0}\frac{1}{k}\log\left(\vartheta + (1-\vartheta)\exp(-k l_b)\right) \tag{4.31}$$

$$= \rho_1 g l_b (1-\vartheta) \quad , \tag{4.32}$$

which is in agreement with Eq. (4.15). Substituting Eq. (4.16), (4.17) and (4.30) into Eq. (4.13) gives

$$\log\left(\frac{T_2}{T_1}\vartheta\right) + C_1''\dot{Q}'\frac{\vartheta}{1-\vartheta} + C_2''\dot{Q}'\frac{\vartheta}{1-\vartheta}\frac{T_2}{T_1} + C_3'' = 0 \quad , \tag{4.33}$$

$$C_1'' = \frac{\pi d h}{2\rho_1 A_1 g c_p} C_1 \quad , \quad C_2'' = \frac{\pi d h}{2\rho_1 A_1 g c_p} C_2 \left(\frac{A_2}{A_1}\right)\left(\frac{A_1}{A_2}\right)^2 \quad , \quad C_3'' = \frac{\mu l_h \pi d h}{d_h^2 \rho_1^2 g A_1 c_p} C_3 \quad .$$

Evaluating Eq. (4.29) at $x = l_u + l_b$ gives

$$\frac{T_2}{T_1}\vartheta = \vartheta + (1-\vartheta)\exp\left(-\frac{\pi d h l_b}{\rho_1 A_1 c_p \dot{Q}'}\frac{1-\vartheta}{\vartheta}\right) \quad . \tag{4.34}$$

Eq. (4.33) and (4.34) represent a reduced-order model of the steady base flow inside the Rijke tube. The input variables are the electric power $\dot{Q}$ and the cross-section area of the iris A_2. The output variables ϑ and T_2/T_1 characterize the state of the Rijke tube. All other flow variables are calculated from ϑ and T_2/T_1. The drag coefficients C_1, C_2 and C_3 as well as the heat transfer coefficients h_i and h_o are the model parameters, which are found from a nonlinear regression (Chapter 3.1). As in the adiabatic base-flow model, the drag coefficients C_1, C_2 and C_3 all have a similar effect on the buoyancy-driven flow inside the Rijke tube. Hence, the drag coefficients C_1 and C_2 are again taken from flow charts. The heat transfer coefficients h_i and h_o account for two independent effects. Firstly, the heat transfer coefficient h determines how much heat the buoyancy-driven flow inside the Rijke tube loses to the environment. Secondly, the proportions of the inner and outer heat transfer coefficients h_i and h_o determine the temperature on the outside wall of the Rijke tube:

$$T_w(x) = T_1 + \frac{h_i}{h_i + h_o}\left(T(x) - T_1\right) \tag{4.35}$$

$$= T_1 + \frac{h_i}{h_i + h_o}(T_f - T_1)\exp\left(-\frac{\pi d h (x - l_u)}{\rho_1 A_1 c_p \dot{Q}'}\frac{1-\vartheta}{\vartheta}\right) \quad . \tag{4.36}$$

Therefore, both the temperature T_2 at the outlet of the Rijke tube and the wall temperatures T_w have to be taken into account to make the simultaneous regression in the three parameters C_3, h_i and h_o a well-posed problem. The relationship between the observable quantities T_2 and T_w and the output variables ϑ and T_2/T_1 is given by

$$T_2 = \frac{T_2}{T_1}T_1 \quad , \tag{4.37}$$

$$T_w(x) = T_1\left[1 + \frac{h_i}{h_i + h_o}\left(\frac{1}{\vartheta} - 1\right)\exp\left(-\frac{\pi d h (x - l_u)}{\rho_1 A_1 c_p \dot{Q}'}\frac{1-\vartheta}{\vartheta}\right)\right] \quad . \tag{4.38}$$

In Fig. 4.5, the results of the nonlinear regression are shown. Only the measurements from the thermocouples downstream of the electric heater are used. The four errors for T_2, $T_w(0.5\,\mathrm{m})$, $T_w(0.75\,\mathrm{m})$ and $T_w(0.95\,\mathrm{m})$ are equally weighted. The predicted temperatures at the outlet of the Rijke tube and the wall temperatures match the experimental measurements well. In Fig. 4.6, the various temperatures and flow speeds are summarized. The flow speeds at the inlet and the outlet of the Rijke tube as well as just downstream of the electric heater are on the order of magnitude of 0.25, 0.45 and $0.8\,\mathrm{m/s}$ respectively, which are reasonable values. The fitting of the viscous drag coefficient and the heat transfer coefficients gives $C_3 = 287.26$, $h_i = 5.67\,\mathrm{W/(m^2K)}$ and $h_o = 19.24\,\mathrm{W/(m^2K)}$ (Table 4.4). The viscous drag coefficient C_3 is more than five times as large as for the adiabatic base-flow model (Table 4.2). Nevertheless, it is still of the same order of magnitude as the theoretical estimate of 156 (Chapter 4.2.1). The pipe Reynolds number is approximately 800 in the lower section of the Rijke tube, which is clearly below the critical Reynolds number of 2300. Thus, the buoyancy-driven flow inside the Rijke tube is laminar everywhere.

Table 4.4 Nonlinear regression of base-flow model with heat loss. The drag coefficients C_1 and C_2 are taken from flow charts [124].

model parameter	estimated value	
C_1	0.578	inviscid loss at pipe inlet
C_2	2	inviscid loss at pipe outlet ($A_2/A_1 = 1$)
C_3	287.26	viscous drag at electric heater
h_i [W/(m^{2}K)]	5.67	heat transfer coefficient at inner wall
h_o [W/(m^{2}K)]	19.24	heat transfer coefficient at outer wall

The Nusselt number of a hydrodynamically and thermally undeveloped flow inside a circular pipe with uniform surface temperature is modeled as [125, Chapter 3.8.1.5]

$$\overline{\mathrm{Nu}}_d = \frac{0.0499 X_+^{-1} \tanh\left(X_+\right) + \dfrac{3.657}{\tanh\left(2.264 X_+^{1/3} + 1.7 X_+^{2/3}\right)}}{\tanh\left(2.432\left(\mathrm{Pr} X_+\right)^{1/6}\right)} \quad , \tag{4.39}$$

$$X_+ = \frac{L}{d \mathrm{Re} \mathrm{Pr}} \quad .$$

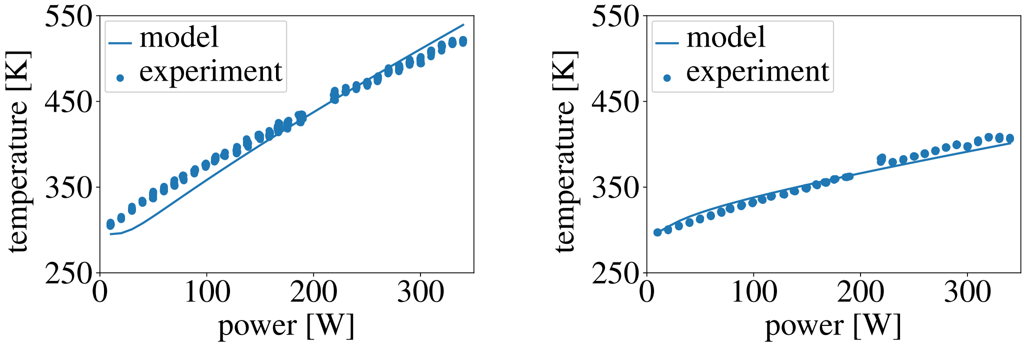
where Re denotes the Reynolds number, and Pr denotes the Prandtl number. Eq. (4.39) is valid for $0 \leq X_+ < \infty$ and $0.1 \leq \mathrm{Pr} < \infty$. In a long pipe, $\overline{\mathrm{Nu}}_d$ converges to 3.657, which is the Nusselt number for a fully developed flow inside a circular pipe with uniform surface temperature [125, Chapter 3.8.1.3]. In the upper section of the Rijke tube, the Reynolds number, the Nusselt number and the heat transfer coefficient evaluate to

$$\mathrm{Re} \approx 700 \quad , \quad \overline{\mathrm{Nu}}_d \approx 5.93 \quad , \quad \bar{h} \approx 5.05 \, \frac{\mathrm{W}}{\mathrm{m}^2\mathrm{K}} \quad . \tag{4.40}$$

The kinematic viscosity of air above 500 K is taken at $4 \times 10^{-5}\,\mathrm{m}^2/\mathrm{s}$ and the heat conductivity at $0.04\,\mathrm{W/(mK)}$ [125, Appendix B]. Eq. (4.40) is in good agreement with $h_i = 5.67\,\mathrm{W/(m^2K)}$ from Table 4.4 regarding the inner heat transfer coefficient h_i, considering that the hydrodynamic and thermal entrance lengths vary depending on the electric power, and that the temperature on the inside wall of the Rijke tube is not uniform. Nevertheless, the agreement between Eq. (4.40) and Table (4.4) confirms that the buoyancy-driven flow inside the Rijke tube is laminar.

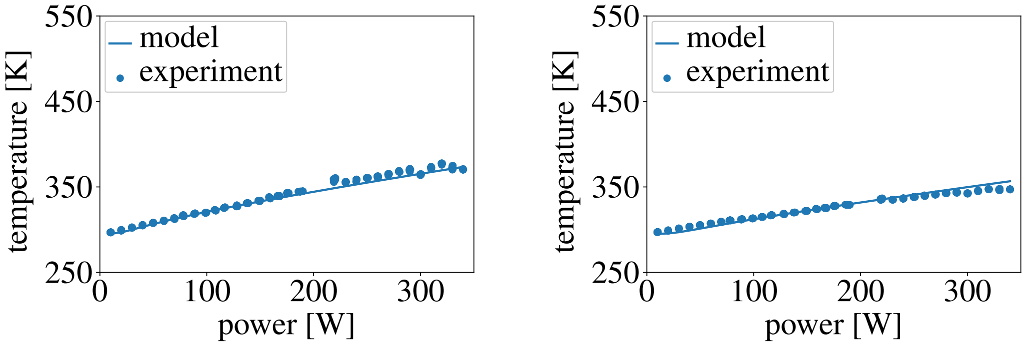
The Nusselt number for the free convection at a vertical wall is given by [126]

$$\overline{Nu}_L = 0.68 + \frac{0.67 \mathrm{Ra}^{1/4}}{\left(1 + \left(0.492/\mathrm{Pr}\right)^{9/16}\right)^{4/9}} \quad , \quad \mathrm{Ra} \leq 10^9 \quad , \tag{4.41}$$



(a) Outlet of Rijke tube.

(b) Wall at $x = 0.50\,\mathrm{m}$.



(c) Wall at $x = 0.75\,\mathrm{m}$.

(d) Wall at $x = 0.95\,\mathrm{m}$.

Fig. 4.5 Model predictions of flow temperature at outlet and wall temperatures downstream of electric heater from base-flow model with heat loss.

where Ra denotes the Rayleigh number. Eq. (4.41) also gives the Nusselt number for the free convection at a vertical cylinder if $d/L \geq 35\mathrm{Gr}^{-1/4}$, where Gr denotes the Grashof number [125, Chapter 3.9.2]. At high electric powers, we assume $(\beta(T_w - T_\mathrm{I}))^{1/4} \approx 1$, where β denotes the thermal expansion coefficient. The kinematic viscosity of air above 400 K is taken at $3 \times 10^{-5}\,\mathrm{m}^2/\mathrm{s}$ [125, Appendix B]. With $\mathrm{Gr} \approx 5.25 \times 10^9$ and $\mathrm{Ra} \approx 3.68 \times 10^9$, the flow on the outside of the Rijke tube is in the transitional regime. Nevertheless, Eq. (4.41) is expected to hold approximately. The Nusselt number and the heat transfer coefficient evaluate to

$$\overline{\mathrm{Nu}}_L = 127.11 \quad , \quad \bar{h} = 4.86\,\frac{\mathrm{W}}{\mathrm{m^2K}} \quad . \tag{4.42}$$

The heat conductivity of air above 400 K is taken at $0.03\,\mathrm{W}/(\mathrm{mK})$ [125, Appendix B]. The value given by Eq. (4.42) is only a quarter of the outer heat transfer coefficient $h_o = 19.24\,\mathrm{W}/(\mathrm{m^2K})$ taken from Table (4.4). There are multiple possible explanations why the outer heat transfer coefficient h_o is underestimated, e.g. increased heat transfer in transitional

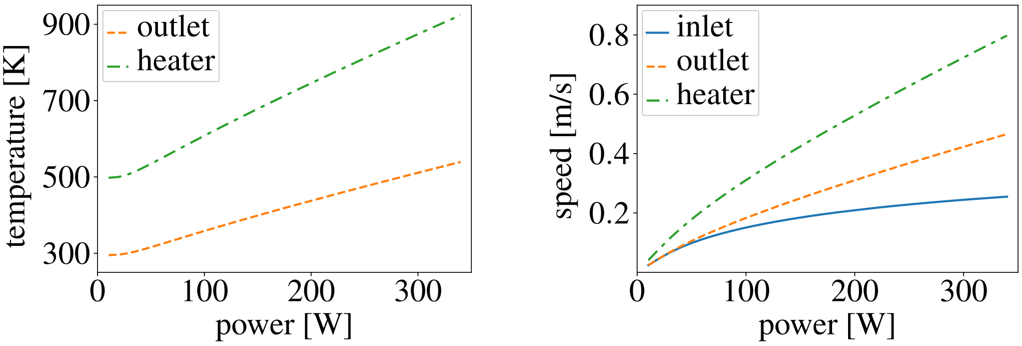


Fig. 4.6 Model predictions of temperatures (left) and flow speeds (right) from base-flow model with heat loss.

and turbulent flow, boundary layer development and heat transfer in the lower section of the Rijke tube or physical differences between a cylinder and a vertical plate.

The results in Table (4.4) are used to assess the initial assumptions about the heat transfer. Firstly, the thermal resistance of the pipe is neglected in Eq. (4.28). The thermal resistance of the pipe is given by

$$\frac{t}{\lambda_{\text{steel}}} \approx 7.5 \times 10^{-5}\,\frac{\text{m}^2\text{K}}{\text{W}} \quad . \tag{4.43}$$

The heat conductivity of stainless steel is taken at $20\,\text{W}/(\text{mK})$. This is negligible compared to the thermal resistances $1/h_i = 0.176\,\text{m}^2\text{K}/\text{W}$ and $1/h_o = 0.052\,\text{m}^2\text{K}/\text{W}$. Secondly, Eq. (4.27) implies that the heat loss of the Rijke tube depends only on the temperature at every location x, and neglects the heat conduction inside the Rijke tube in the axial direction. The heat conduction inside the pipe in the axial direction is given by

$$\pi d t \lambda_{\text{steel}} \frac{\mathrm{d}^2 T_w}{\mathrm{d}x^2} = \pi d t \lambda_{\text{steel}} \frac{h_i}{h_i + h_o} \frac{\mathrm{d}}{\mathrm{d}x^2} \left(T(x) - T_1\right) \tag{4.44}$$

$$= \pi d t k^2 \lambda_{\text{steel}} \frac{h_i}{h_i + h_o} \left(T(x) - T_1\right) \quad , \tag{4.45}$$

where k is given in Eq. (4.30). The heat transfer in the radial direction is given by

$$\pi d h (T(x) - T_1) = \pi d \frac{h_i h_o}{h_i + h_o} \left(T(x) - T_1\right) \quad . \tag{4.46}$$

The ratio between the two heat transfer mechanisms is given by

$$\frac{t \lambda_{\text{steel}}}{h(T(x) - T_1)} \frac{\mathrm{d}^2 T_w}{\mathrm{d}x^2} = \frac{\lambda_{\text{steel}} k^2 t}{h_o} \quad . \tag{4.47}$$

The ratio between the two heat transfer mechanisms does not depend on the location x. It is known from the nonlinear regression that the value of k does not exceed $10\,\mathrm{m}^{-1}$ for electric powers between 20 and 340 W. Thus, the ratio between the heat transfer mechanisms does not exceed 0.15. This justifies the assumption that the wall temperature on the outside wall of the Rijke tube only depends on the temperature of the buoyancy-driven flow inside the Rijke tube. Thirdly, the heat conduction in the buoyancy-driven flow inside the Rijke tube is considered. The ratio between the heat convection and the heat conduction in the buoyancy-driven flow inside the Rijke tube is given by

$$\mathrm{Pe} = \frac{\rho_1 \bar{u}_1 c_p}{k\lambda} \quad , \tag{4.48}$$

where Pe denotes the Péclet number. As k is inversely proportional to the mass flow rate, the Péclet number scales with the square of the mass flow rate. At low electric powers, when the mass flow rate vanishes, there is a regime where the heat conduction dominates the heat convection in the buoyancy-driven flow inside the Rijke tube.

In Fig. 4.6, note that the temperature at the electric heater does not converge to the ambient temperature in the limit of a cold electric heater. An asymptotic analysis reveals that this result is intrinsic to the model. Substituting Eq. (4.34) into Eq. (4.33) gives

$$\begin{aligned}
&-\log\left(\vartheta + (1-\vartheta)\exp\left(-\frac{\pi d h l_b}{\rho_1 A_1 c_p \dot{Q}'}\frac{1-\vartheta}{\vartheta}\right)\right) \\
&= C_3'' + C_1''\dot{Q}'\frac{\vartheta}{1-\vartheta} + C_2''\dot{Q}'\frac{\vartheta}{1-\vartheta} + C_2''\dot{Q}'\exp\left(-\frac{\pi d h l_b}{\rho_1 A_1 c_p \dot{Q}'}\frac{1-\vartheta}{\vartheta}\right) \quad .
\end{aligned} \tag{4.49}$$

Two cases are distinguished:

1. $\lim_{\dot{Q}\to 0}\dot{Q}'/(1-\vartheta) > 0$. This implies that ϑ goes to one. It follows that the left-hand side of Eq. (4.49) vanishes. This is a contradiction because at least C_3'' remains on the right-hand side of Eq. (4.49).

2. $\lim_{\dot{Q}\to 0}\dot{Q}'/(1-\vartheta) = 0$. All terms on the right-hand side of Eq. (4.49) except for C_3'' vanish. The argument of the exponential term on the left-hand side of Eq. (4.49) goes to negative infinity. This leaves

$$\log(\vartheta) = -C_3'' \quad . \tag{4.50}$$

Thus, ϑ converges to a value between zero and one. The prediction of the temperature T_f just downstream of the electric heater is 497.23 K. This is in excellent agreement with the asymptotic behavior displayed in Fig. 4.6. In Fig. 4.7, temperature profiles of the buoyancy-driven flow downstream of the electric heater are shown for four different electric powers. The temperature T_2 at the outlet of the Rijke tube converges to the ambient temperature T_1 while the temperature T_f just downstream of the electric heater converges to a value above the ambient temperature T_1. The slope in the temperature profile becomes increasingly steep near the electric heater. This is in agreement with the previous result that heat conduction in the axial direction plays a dominant role at low electric powers.

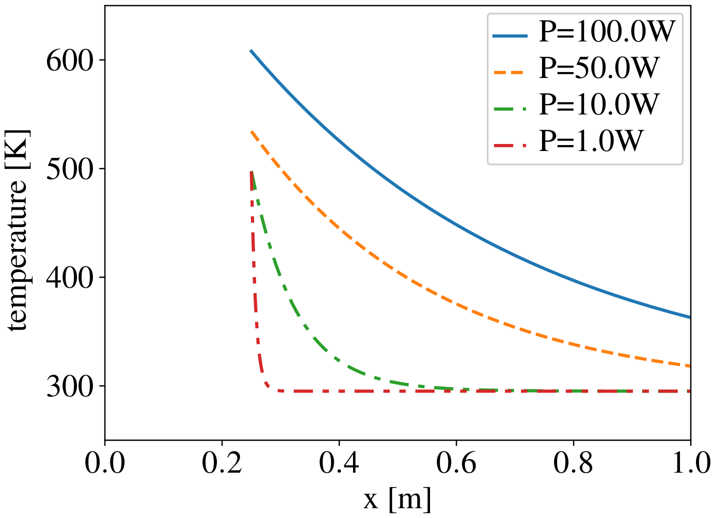


Fig. 4.7 Model predictions of flow temperature profiles downstream electric heater from base-flow model with heat loss.

4.3 Acoustics

As discussed in Chapter 1.2, there are three principal tools in thermoacoustics to execute the linear stability analysis [29, Tutorial 1]: traveling waves, Galerkin modes and Helmholtz modes. They have in common that they introduce a separation of variables into the partial differential equations. Traveling waves are characteristics of the partial differential equations, where space and time are lumped into a single coordinate. In a Galerkin framework, ansatz modes reduce the partial differential equations to a set of ordinary differential equations in time. The Helmholtz equation is the Laplace transform of a wave equation, and more specifically a partial differential equation in space and not in time. JUNIPER AND SUJITH have demonstrated that the results of the corresponding eigenproblems are identical in the limit of high resolution, with the exception of smearing or Gibbs phenomena near discontinuities [29, Tutorial 1].

In this section, we present two acoustic models: (i) a phenomenological model based on a mass-spring-damper system coupled with the adiabatic base-flow model (Chapter 4.2.1); and (ii) a network model based on traveling waves coupled with the base-flow model that includes heat loss (Chapter 4.2.2). In the mass-spring-damper model, the characteristic function for the eigenvalues is available so that the methods developed in Chapter 3.1 are directly applicable. The network model is more representative of larger industrial applications so that the Bayesian regression is based on a perturbation analysis instead of Jacobi's formula (Chapter 3.2). The emphasis of this discussion is placed on model comparison: In a first instance, we find the optimal estimates for the model parameters of the reduced-order models using least-squares regression followed by nonparametric regression. In a second instance, we assess the structures of the reduced-order models by marginalizing the model parameters within the evidence framework, drawing on many ideas from Chapters 1 to 3.

4.3.1 Mass-spring-damper model

In order to gain a first quantitative understanding of the acoustics inside the Rijke tube, we investigate the following reduced-order model:

$$s^2 + 2Bs + \Omega^2 + N\exp(-s\tau) = 0 \quad . \tag{4.51}$$

This phenomenological model resembles the description of a mass-spring-damper system. The model coefficient B represents a damping coefficient. The model coefficient Ω is the natural angular frequency of the mass-spring-damper system. The model coefficient N represents a modified interaction index that does not have the same units as the interaction index n in the n-τ model (Eq. (1.4)). The model coefficient τ represents the time delay in the n-τ model. Eq. (4.51) resembles the toy model investigated by JUNIPER AND SUJITH, which is based on a perturbation analysis of a single Galerkin mode [29, Eq. (6)]. A damping term in B has been added in order to account for acoustic radiation at the pipe ends, i.e. reflection coefficients $|R| < 1$. In the limit of a thermoacoustically passive heat source ($N = 0$), Eq. (4.51) describes a regular mass-spring-damper system:

$$s^2 + 2Bs + \Omega^2 = 0 \quad . \tag{4.52}$$

The complex eigenvalues s are given by

$$s = -B \pm \sqrt{B^2 - \Omega^2} \approx -B \pm j\Omega \quad , \tag{4.53}$$

where the final approximation is justified if $B \ll \Omega$. In summary, the purpose of this phenomenological model is to capture the essential physics of the Rijke tube: oscillatory behavior coupled with growth or decay, depending on the balance between acoustic radiation and heat release.

In order to close our phenomenological model, the model coefficients B, Ω, N and τ are empirically related to the buoyancy-driven flow inside the Rijke tube. The damping coefficient B and the natural angular frequency Ω are modeled as

$$B = k_B \quad , \tag{4.54}$$

$$\Omega = \frac{\pi c_u}{(l_u + k_\Omega) + (l_b + k_\Omega)\sqrt{\vartheta}} \quad , \tag{4.55}$$

where k_Ω denotes an end correction, assumed to be identical at both ends of the Rijke tube. This takes into account that the ends of the Rijke tube are not perfectly open in the acoustic sense. It is assumed that the pressure nodes lie slightly outside the Rijke tube. The first-order end correction is theoretically estimated to be $0.3\,d \approx 14.1\,\mathrm{mm}$ [127]. The modified interaction index N and the time delay τ are modeled as

$$N = k_N \dot{Q} \quad , \tag{4.56}$$

$$\tau = k_\tau \frac{0.2 d_w}{\bar{u}(l_u^-)} \quad . \tag{4.57}$$

The proportion of the heat converted to mechanical work is assumed to be constant. The time delay τ is assumed to scale as $0.2\,d_w/\bar{u}(l_u^-)$. The latter is the result of a perturbation analysis performed on the boundary layer of a laminar flow around a cylinder [120]. It holds for low-frequency fluctuations. For $\bar{u}(l_u^-) \approx 0.65\,\mathrm{m/s}$ (Fig. 4.3), the threshold is $20\,\bar{u}(l_u^-)/d_w \approx 23{,}000\,\mathrm{s}^{-1}$. The natural angular frequency of the Rijke tube is on the order of $\pi c_u/(l_u + l_b) \approx 1{,}000\,\mathrm{s}^{-1}$, clearly below the threshold. In the limit of high flow speed and small time delay ($|s\tau| \ll 1$), a first-order approximation of the exponential term in Eq. (4.51) representing the heat release gives

$$s^2 + (2B - N\tau)\,s + \left(\Omega^2 + N\right) \approx 0 \quad , \tag{4.58}$$

where $N\tau$ enters the damping term with the opposite sign to B. This elucidates the potentially destabilizing effect of the heat source in the limit of strong thermoacoustic activity ($N\tau/B \gg 1$), as well as the sensitivity of the growth rate to variations in the time delay τ [29].

Therefore, we begin by performing a Bayesian regression of the mass-spring-damper model coupled with the adiabatic base-flow model in the unstable operating regime ($\dot{Q} >$

200 W). The maximum-likelihood estimates of the model parameters are given in Table 4.5. The end correction k_Ω is about twice as large as the theoretical estimate of 14.1 mm, and has low uncertainty. The time delay τ is about thirty times as large as the theoretical estimate of $k_\tau = 1$. Nevertheless, the uncertainty is low, which indicates correct scaling in the unstable operating regime. The model parameters k_N and k_B, which are expected to have a direct impact on the growth rate, are marked by high uncertainty (Eq. (4.58)). The high uncertainty in k_B is justified since we expect a damping coefficient of approximately $B = 8\,\mathrm{s}^{-1}$ for a switched-off electric heater (Fig. 4.2). The sign of k_B is unphysical because it implies negative damping. However, this behavior is due to the fact that the training data so far only encompasses positive growth rates, where damping mechanisms are not evident.

Table 4.5 Bayesian regression of mass-spring-damper model coupled with adiabtic base-flow model. The uncertainty in the maximum-likelihood estimates is represented by the standard deviation in the marginal probability distributions. Note that parametric regression only quantifies the parametric uncertainty but not the structural uncertainty (Chapter 1.3).

model parameter	estimated value	standard deviation
k_B [s^{-1}]	−6.54	0.94 (14.4 %)
k_Ω [mm]	27.52	0.48 (1.7 %)
k_N [s$^{-2}W^{-1}$]	81.68	8.98 (11.0 %)
k_τ	30.80	0.97 (3.1 %)

The low uncertainties in k_Ω and k_τ as well as the high uncertainties in k_B and k_N are reflected in the confidence intervals of the associated model coefficients Ω, τ, B and N, respectively (Fig. 4.8). The corresponding confidence intervals in the model predictions of growth rates and frequencies at every operating point are found from the following optimization problem with constraints:

$$\underset{\theta}{\text{maximize}} \quad \left|s(\theta) - s(\hat{\theta})\right|^2 \quad , \tag{4.59}$$

$$\text{subject to} \quad \left(\theta_m - \hat{\theta}_m\right)\left(\Sigma_\theta^{-1}\right)_{mn}\left(\theta_n - \hat{\theta}_n\right) = r^2 \quad , \tag{4.60}$$

where r denotes the confidence level, i.e. the radius of the confidence region in units of standard deviations. A first-order approximation of Eq. (4.59) gives

$$\left|s(\theta) - s(\hat{\theta})\right|^2 \approx \left(\theta_m - \hat{\theta}_m\right)\left[\left.\frac{\mathrm{d}s^*}{\mathrm{d}\theta_m}\right|_{\hat{\theta}} \left.\frac{\mathrm{d}s}{\mathrm{d}\theta_n}\right|_{\hat{\theta}}\right]\left(\theta_n - \hat{\theta}_n\right) \quad . \tag{4.61}$$

The method of Lagrange multipliers gives the necessary condition [107, Chapter 9]:

$$\left[\left.\frac{\mathrm{d}s^*}{\mathrm{d}\theta_m}\right|_{\hat{\theta}}\left.\frac{\mathrm{d}s}{\mathrm{d}\theta_n}\right|_{\hat{\theta}}\right]\left(\theta_n-\hat{\theta}_n\right)=\lambda\left(\Sigma_\theta^{-1}\right)_{mn}\left(\theta_n-\hat{\theta}_n\right) \quad , \tag{4.62}$$

where λ denotes a so-called Lagrange multiplier. Eq. (4.62) is an eigenproblem, where the Lagrange multipliers λ are the eigenvalues and the optimal $\theta_n-\hat{\theta}_n$ is an eigenvector:

$$\max\left\{\left(\theta_m-\hat{\theta}_m\right)\left[\left.\frac{\mathrm{d}s^*}{\mathrm{d}\theta_m}\right|_{\hat{\theta}}\left.\frac{\mathrm{d}s}{\mathrm{d}\theta_n}\right|_{\hat{\theta}}\right]\left(\theta_n-\hat{\theta}_n\right)\right\} \tag{4.63}$$

$$=\max\left\{\lambda\left(\theta_m-\hat{\theta}_m\right)\left(\Sigma_\theta^{-1}\right)_{mn}\left(\theta_n-\hat{\theta}_n\right)\right\} \tag{4.64}$$

$$=r^2\max\left\{\lambda\right\} \quad . \tag{4.65}$$

The confidence intervals in the model predictions of growth rates and frequencies are shown in Fig. 4.9. The model predictions of the frequencies reflect the low uncertainty in k_Ω and k_τ. The model predictions in the growth rates are significantly more uncertain, especially towards lower and higher electric powers.

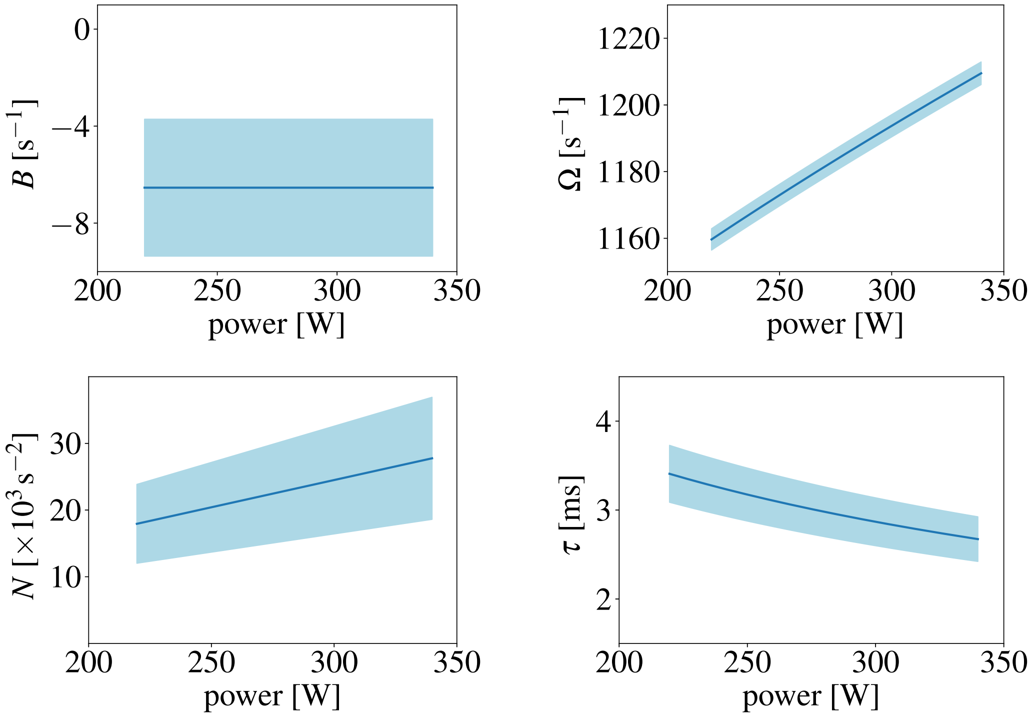


Fig. 4.8 Model predictions of mass-spring-damper model coefficients with 3-σ confidence intervals (99.7 % confidence).

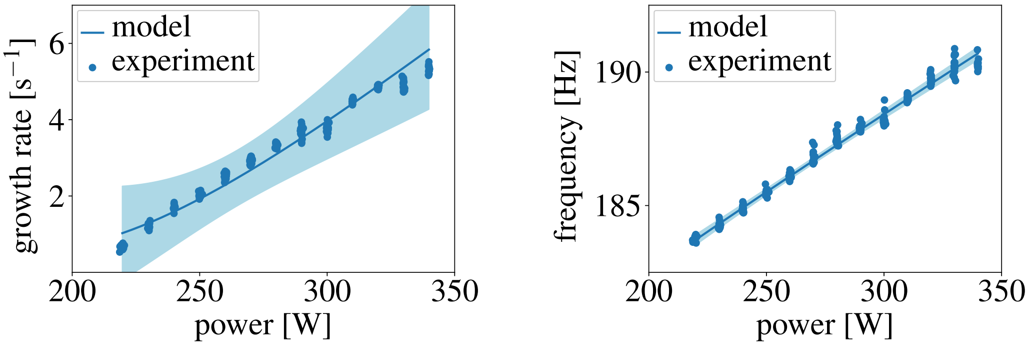


Fig. 4.9 Model predictions of growth rates (left) and frequencies (right) in Rijke tube from mass-spring-damper model with 3-σ confidence intervals (93.9 % confidence[1]).

In Fig. 4.10, the model predictions of the growth rates and frequencies over the entire operating regime ($0 \leq \dot{Q} \leq 340\,\mathrm{W}$) are shown. While the frequency is only slightly underpredicted near zero power, the growth rate makes it clear that a strongly parametrized mass-spring-damper model is not suited for the lower powers in the stable operating regime. For $\dot{Q} \to 0$, the growth rate converges to $-k_B \approx 6.54\,\mathrm{s}^{-1}$ as predicted by Eq. (4.52) and (4.53) (Table 4.5). In this particular experiment, the model predictions of the frequency are more robust since $\Omega^2/N \gg 1$, and they only deviate slightly from the natural frequencies of the Rijke tube with a thermoacoustically passive heat source. We therefore focus our attention on the model predictions of the growth rate when we turn to nonparametric regression in the next step.

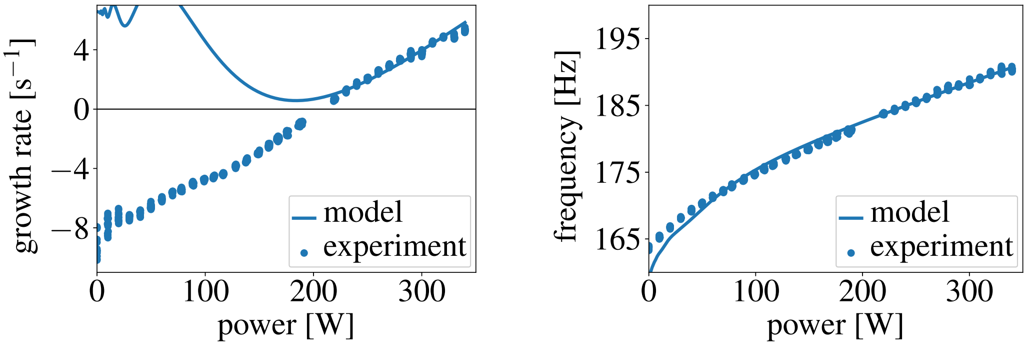


Fig. 4.10 Model predictions of growth rates (left) and frequencies (right) in Rijke tube from mass-spring-damper model over entire operating regime.

---

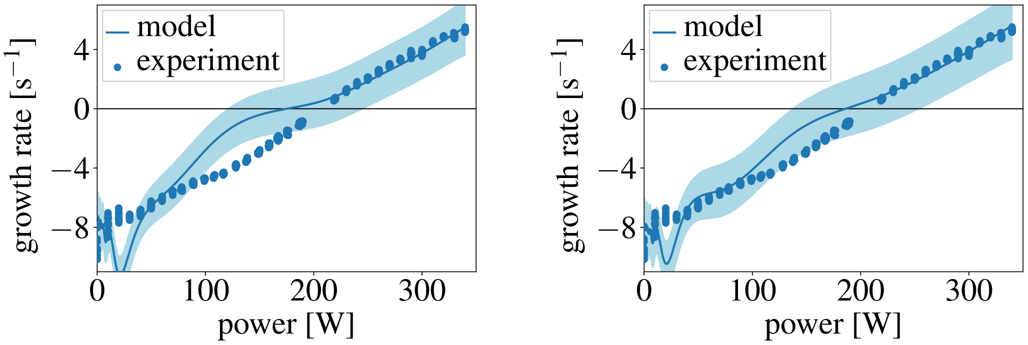
[1] The 3-σ confidence interval is based on the 3-σ confidence region in the model parameters. The well-known values 68.3 %, 95.4 % and 99.7 % are for the 1-σ, 2-σ and 3-σ confidence intervals of a one-dimensional normal distribution, respectively. For four model parameters, the corresponding confidence regions have the values 9.0 %, 59.4 % and 93.9 %, respectively.

For the nonparamateric regression, we use a linear combination of the squared-exponential kernel with a white kernel (Definitions 2.14, 2.15):

$$k(w,w') = \begin{cases} \sigma_f^2 \exp\left(\frac{|w-w'|^2}{2\sigma_l^2}\right) + \sigma_n^2 & w = w' \\ \sigma_f^2 \exp\left(\frac{|w-w'|^2}{2\sigma_l^2}\right) & w \neq w' \end{cases}, \tag{4.66}$$

which has the hyperparameters σ_f, σ_l and σ_n. This kernel has automatic relevance determination (ARD) built-in [102, Chapter 5]: Plausible models tend to have high signal-to-noise ratios over long distances, i.e. large σ_f and σ_l as well as small σ_n. The recursive expectation-maximization (EM) algorithm begins at $\dot{Q} = 340\,\mathrm{W}$, and iterates from high powers to low powers (Chapter 3.3). We use the results of the Bayesian regression over the unstable operating regime as the prior for the model parameters at $\dot{Q} = 340\,\mathrm{W}$ to initialize the EM algorithm (Table 4.5).

In Fig. 4.11, the model predictions of the growth rates using nonparametric regression based on Algorithms 3.3 and 3.4 are shown (Chapter 3.3). One loop of nonparametric regression improves the model predictions of the growth rate to the point that the growth rate at zero power is realistically predicted (Fig. 4.11, left). Multiple loops of nonparametric regression with message passing further improve the model predictions up to the point that all observations above $\dot{Q} = 50\,\mathrm{W}$ are within the 3-σ confidence interval (Fig. 4.11, right).

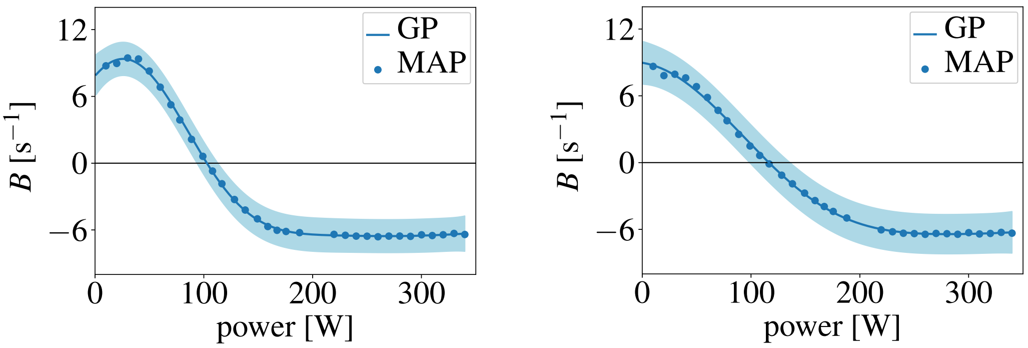


(a) One loop. (b) Three loops of message passing.

Fig. 4.11 Model predictions of growth rates in Rijke tube from mass-spring-damper model with Gaussian processes, expectation-maximization (left) as well as message passing (right).

In summary, the straightforward application of our nonlinear regression algorithms noticably improves the model predictions of the state of the system, i.e. the growth rate and the frequency. Before refining the analysis any further, we critically assess the relevance of those improvements. Legitimately speaking, if only the model predictions of the state of the system were of interest, even a linear function would do a decent job at interpolating the observations of growth rate and frequency (Fig. 4.2). The value of reduced-order and phenomenological models does not only lie in their lower computational costs compared to CFD but in the physical interpretability of their model parameters. In the design of jet and rocket engines, it is not our aim to suppress unstable modes during operation, e.g. using active control, but to identify favorable modifications to the system, and thus to system parameters such as the time delay, in order to ensure that the modes in question do not become unstable in the first place. The prerequisites are accurately estimated model parameters. For instance, the mass-spring-damper model has four model parameters while each observation of growth rate and frequency eliminates at most two degrees of freedom. For $\dot{Q} \to 0$, an improperly regularized reduced-order model might encourage $k_N \to \infty$ such that the heat source would be thermoacoustically active even at zero power. When further refining the applications of our nonlinear regression algorithms, we thus redirect part of our attention from the state level to the parameter level.

One particular strong suit of Bayesian methods is the principled addition of knowledge to reduced-order models. In general, it is not possible to exactly infer the value of the model parameters at a given operating point because there are four model parameters but only two observed quantities. At zero power however, when $N = 0$, only the model parameters B and Ω of a regular mass-spring-damper system remain so that the number of model parameters and the number of observed quantities is balanced (Eq. (4.52)). In particular, we estimate the damping coefficient B to be approximately $9.22\,\mathrm{s}^{-1}$ (Fig. 4.2, left). In the maximization steps of Algorithms 3.3 and 3.4, we hence optimize the hyperparameters of our Gaussian processes using the posterior estimates already computed as well as the information available at zero power. The difference this makes is shown in Fig. 4.12. Both physics-informed and physics-uninformed model predictions of the damping coefficient B display similar trends: (i) the values are fairly constant in the unstable operating regime; (ii) they increase in the stable regime in the direction of lower powers; and (iii) they reach the value of the regular mass-spring-damper system at zero power. From the last point of view, the physical piece of information does not significantly improve the estimate at zero power. Nevertheless, it affects the optimization of the hyperparameters from the first iteration of the expectation-maximization algorithm, and thus the regularization of the model parameters over the whole operating regime. Statistically speaking, the difference is marked by the distinction between interpolation and extrapolation. On the one hand, the only source of a-priori regularization in the physics-uninformed case is the prior used at $\dot{Q} = 340\,\mathrm{W}$ (Fig. 4.12, left). Therefore, the model predictions of the damping coefficient remain fairly constant well into the stable operating regime, and only begin to noticably increase at $\dot{Q} = 150\,\mathrm{W}$. Towards the lower powers, the physics of the Rijke tube change rapidly which is reflected by the steep slope. Consequently, the model predictions of the damping coefficient appear to slightly overshoot in the vicinity of $\dot{Q} = 30\,\text{W}$. On the other hand, the Gaussian processes in the physics-informed case are aware of the conditions at zero power throughout the nonparametric regression procedure (Fig. 4.12, right). The model predictions of the damping coeffient are less bound to the prior used at $\dot{Q} = 340\,\text{W}$, and already begin to increase before $\dot{Q} = 220\,\text{W}$ in the direction of lower powers. The slopes are flatter, and there is no overshooting near zero power.



(a) No information.

(b) $\lim_{\dot{Q}\to 0} k_B$.

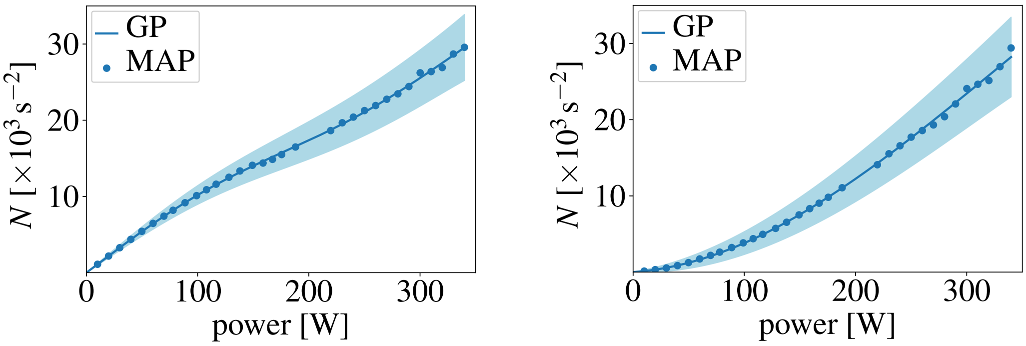
Fig. 4.12 Physics-informed model predictions of damping coefficients. On the left-hand side, no physical information is added (Fig. 4.11, right). On the right-hand side, $\lim_{\dot{Q}\to 0} k_B = 9.22\,\text{s}^{-1}$ is introduced based on the zero-power analogy to a regular mass-spring-damper system (Fig. 4.14, left).

Most notably, the damping coefficients are negative at higher powers, which is not physical. In essence, this illustrates where model estimation goes further than parameter estimation. More specifically, there are two routes here to establish that the mass-spring-damper model is imperfect: Firstly, one enforces the sign of the model parameter. Then, the model is imperfect because it does not explain the observations sufficiently well. Secondly, one does not enforce the sign of the model parameter. Now, the model is imperfect because the values of the model parameters are complicated to explain although the predictions are improved. Fig. 4.12 exposes the imperfection of the mass-spring damper model via the second route. The discussion on model assessment is continued in the concluding paragraphs of this section.

Another feature in our phenomenological model of the Rijke tube requiring special attention is the erratic behavior of the growth rate below $\dot{Q} = 50\,\text{W}$ (Fig. 4.11). The cause is a chain of modeling decisions that affect the behavior in the limit of zero power: (i) the base-flow speed $\bar{u}(l_u^-) = \bar{u}_1$ vanishes (Eq. (4.26)); (ii) our empirical closure of the time delay τ is inversely proportional to the base-flow speed $\bar{u}_1$ (Eq. (4.57)); and (iii) the time delay τ enters the governing equations of our reduced-order model as $\exp(-s\tau)$ (Eq. (4.51)). Because the growth rates, and thus the real parts of the complex eigenvalues s, are negative, our reduced-order model has a singularity at zero power. For near-zero powers, the time delay τ is mathematically sensitive to the rate at which it diverges, which is almost certainly not captured by our empirical closures. In general, there are multiple options as to how to suppress the zero-power singularity. For example, we could treat LIGHTHILL's law as an asymptotic case in the limit of high powers of a more well-behaved, empirical ansatz such as

$$\tau = k_\tau \frac{0.2 d_w}{\bar{u}(l_u^-)} \left[1 - \exp\left(k_u^2 \bar{u}(l_u^-)^2\right)\right] \quad , \tag{4.67}$$

which would introduce at least one additional model parameter k_u. Alternatively, we make the subjective observation that the model parameter k_N should go to zero so that the modified interaction index N goes to zero at a higher rate in the limit of lower powers. The difference this makes is shown in Fig. 4.13. Both physics-informed and physics-uninformed model predictions of the modified interaction index N display similar trends in that they monotonically increase from lower to higher powers. Nevertheless, the physics-uninformed model predictions have a visible inflection point near $\dot{Q} = 200\,\mathrm{W}$ whereas the physics-informed model predictions are convex in shape thoughout the entire operating regime.



(a) No information.

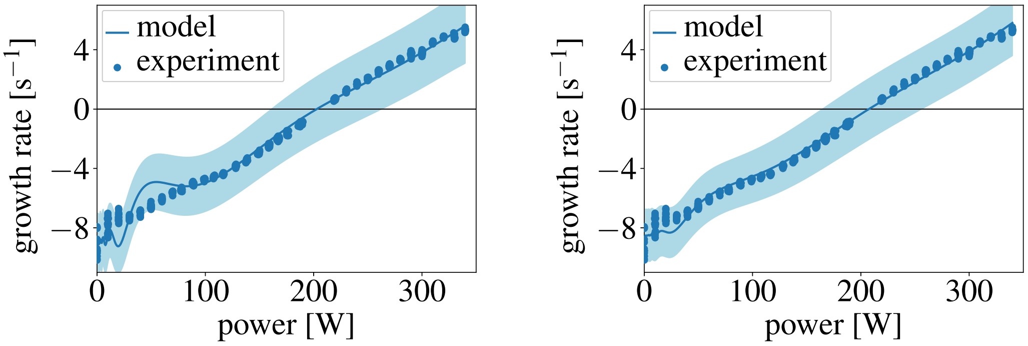
(b) $\lim_{\dot{Q}\to 0} k_B$, $\lim_{\dot{Q}\to 0} k_N$.

Fig. 4.13 Physics-informed model predictions of modified interaction indices. On the left-hand side, no physical information is added (Fig. 4.11, right). On the right-hand side, $\lim_{\dot{Q}\to 0} k_N = 0$ is introduced in addition to $\lim_{\dot{Q}\to 0} k_B = 9.22\,\mathrm{s}^{-1}$ in order to suppress the zero-power singularity $\tau \to \infty$ (Fig. 4.14, right).

In comparison, Fig. 4.14 shows how every piece of physical information also improves the model predictions of the growth rate. Physical information about the damping coefficient B leads to virtually perfect model predictions of the growth rate for the range of $\dot{Q} \geq 100\,\mathrm{W}$ (Fig. 4.14, left). Over the whole operating regime, all observations of the growth rate are within a 3-σ confidence interval. Furthermore, physical information about the modified interaction index N extends this range down to $\dot{Q} \geq 50\,\mathrm{W}$, with little inaccuracy left for high decay rates (Fig. 4.14, left). In particular, our reduced-order model picks up the waviness in the observations of the growth rate, which is difficult to discern in the raw data with the naked eye (Fig. 4.2, left).

In conclusion, the mass-spring-damper model with a time-delay term is a simple yet physical reduced-order model of the Rijke tube. For instance, the model parameter B is physically interpreted to play the role of a damping coefficient, and to control the acoustic energy lost to the environment. On a physical basis, this makes an interesting discussion regarding the quality of this reduced-order model (Fig. 4.14): On the one hand, the mass-spring-damper model with a time-delay term is a poor reduced-order model of the Rijke tube because damping coefficients are not meant to become negative. On the other hand, nonparametric regression gracefully handles the deficiency of this reduced-order model in a smooth manner. This illustrates the robustness of Bayesian inference, i.e. its ability to graciously reconcile our models with imperfect data, that is incomplete, noisy, or possibly even contradictory. In this particular case, the ability to match the observations and the expectation to keep the sign of the model parameter are at odds here: Matching the observations is more highly prioritized due to their high precision. This does not mean that the physical information is lost in the process. Rather, the sign of the model parameter is less highly prioritized. On a positive note, this reveals to the modeler that the energy-feeding mechanism is not properly captured by the time-delay term.

We provide a discussion of model selection within the evidence framework at the end of Chapter 4.3.2 when we are able to compare across different reduced-order models. Nevertheless, this serves as an entry-point demonstration for the challenges of reduced-order modeling as an inverse problem and the importance of regularization: Every piece of physical information, in addition to the model equations, serves as a beacon in the otherwise unchartered parameter landscape.



(a) $\lim_{\dot{Q}\to 0} k_B$.

(b) $\lim_{\dot{Q}\to 0} k_B$, $\lim_{\dot{Q}\to 0} k_N$.

Fig. 4.14 Physics-informed model predictions of growth rates in Rijke tube from mass-spring-damper model. On the left-hand side, $\lim_{\dot{Q}\to 0} k_B = 9.22\,\mathrm{s}^{-1}$ is introduced based on the zero-power analogy to a regular mass-spring-damper system. On the right-hand side, $\lim_{\dot{Q}\to 0} k_N = 0$ is added in order to suppress the zero-power singularity $\tau \to \infty$.

4.3.2 Network model

A more realistic as well as versatile reduced-order model is provided by network models based on traveling waves [45]. The Rijke tube is divided into two segments with uniform flow properties. The electric heater is treated as a discontinuity. This is valid because the heater Helmholtz number is approximately 0.015, which is much less than unity.[2] The segments support standing waves that are coupled by jump conditions. The forward and backward-traveling waves, that together form the standing waves, are governed by a set of hyperbolic partial differential equations. The equations are solved by finding complex eigenvalues that satisfy the jump and boundary conditions. Network models allow for greater flexibility than the phenomenological model from Chapter 4.3.1 because each section, as well as the jump and boundary conditions, are modeled independently. Most importantly, network models are more physical than our ad-hoc mass-spring-damper model.

The three-dimensional system of hyperbolic partial differential equations is in its simplest form reduced to the one-dimensional form [29]

$$\rho_0 \frac{\partial u_1}{\partial t} + \frac{\partial p_1}{\partial x} = 0 \quad , \tag{4.68}$$

$$\frac{\partial p_1}{\partial t} + \gamma p_0 \frac{\partial u_1}{\partial x} = (\gamma - 1) q_1 \quad , \tag{4.69}$$

---

[2] The Helmholtz number represents the ratio between the length of an acoustic element and the acoustic wavelength of interest.

where the subscript indices denote the order of perturbation (Eq. (1.1), (1.2)). Integration of Eq. (4.68) and (4.69) gives the jump conditions

$$[p_1]_{l_u^-}^{l_u^+} = 0 \quad , \tag{4.70}$$

$$[u_1]_{l_u^-}^{l_u^+} = \frac{\gamma - 1}{\gamma p_0} n u_1 (l_u^-, t - \tau) \quad . \tag{4.71}$$

The heat release rate perturbation q_1 is modeled with an n-τ model (Eq. (1.4)). The pressure and velocity perturbations of the traveling waves are expressed as

$$p_1(t,x) = g\left(t - \frac{x - l_u}{c_0}\right) + h\left(t + \frac{x - l_u}{c_0}\right) \quad , \tag{4.72}$$

$$u_1(t,x) = \frac{1}{\rho_0 c_0}\left[g\left(t - \frac{x - l_u}{c_0}\right) - h\left(t + \frac{x - l_u}{c_0}\right)\right] \quad . \tag{4.73}$$

The characteristics g and h represent the forward- and backward-travelling waves, respectively. The Laplace transforms of Eq. (4.72) and (4.73) are given by

$$\hat{p}(s,x) = \hat{g}(s) \exp\left(-s\frac{x - l_u}{c_0}\right) + \hat{h}(s) \exp\left(s\frac{x - l_u}{c_0}\right) \quad , \tag{4.74}$$

$$\hat{u}(s,x) = \frac{1}{\rho_0 c_0}\left[\hat{g}(s) \exp\left(-s\frac{x - l_u}{c_0}\right) - \hat{h}(s) \exp\left(s\frac{x - l_u}{c_0}\right)\right] \quad . \tag{4.75}$$

At the ends of the Rijke tube, the ingoing waves are reflected as outgoing waves. The boundary conditions are expressed as

$$\hat{g}_u(s) = R_u \hat{h}_u(s) \exp(-s\tau_u) \quad , \quad \tau_u = \frac{2l_u}{c_{0,u}} \quad , \tag{4.76}$$

$$\hat{h}_b(s) = R_b \hat{g}_b(s) \exp(-s\tau_b) \quad , \quad \tau_b = \frac{2l_b}{c_{0,b}} \quad , \tag{4.77}$$

where R_u and R_b are the complex reflection coefficients at the respective ends of the Rijke tube. The network model of the Rijke tube is illustrated in Fig. 4.15.

Substituting Eq. (4.74)–(4.77) into Eq. (4.70) and (4.71) gives the jump conditions

$$[1 + R_b \exp(-s\tau_b)]\, \hat{g}_b(s) - [1 + R_u \exp(-s\tau_u)]\, \hat{h}_u(s) = 0 \quad , \tag{4.78}$$

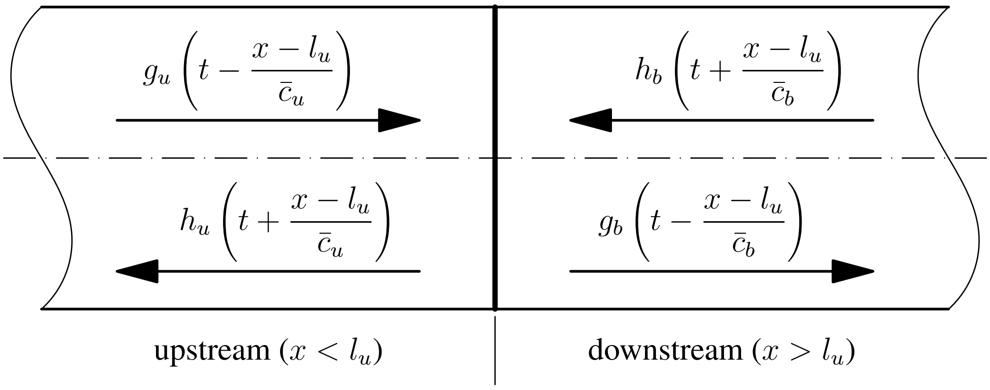


Fig. 4.15 Network model of Rijke tube.

$$\begin{aligned}\frac{1}{\rho_{0,b}c_{0,b}}\left[1-R_b\exp\left(-s\tau_b\right)\right]\hat{g}_b(s)+\frac{1}{\rho_{0,u}c_{0,u}}\left[1-R_u\exp\left(-s\tau_u\right)\right]\hat{h}_u(s)\\ =-\frac{\gamma-1}{\gamma p_0}\frac{n}{\rho_{0,u}c_{0,u}}\exp\left(-s\tau\right)\left[1-R_u\exp\left(-s\tau_u\right)\right]\hat{h}_u(s)\quad .\end{aligned} \tag{4.79}$$

The jump conditions are more conveniently expressed as a linear system of equations:

$$\begin{pmatrix}L_{11} & L_{12}\\ L_{21} & L_{22}\end{pmatrix}\begin{pmatrix}\hat{h}_u\\ \hat{g}_b\end{pmatrix}\equiv 0\quad , \tag{4.80}$$

where the matrix entries are given by

$$L_{11}=1+R_u\exp\left(-s\tau_u\right)\quad , \tag{4.81}$$

$$L_{12}=-1-R_b\exp\left(-s\tau_b\right)\quad , \tag{4.82}$$

$$L_{21}=\left(1+\frac{\gamma-1}{\gamma}\frac{n}{p_0}\exp\left(-s\tau\right)\right)\left[1-R_u\exp\left(-s\tau_u\right)\right]\quad , \tag{4.83}$$

$$L_{22}=\frac{1}{\sqrt{\vartheta}}\left[1-R_b\exp\left(-s\tau_b\right)\right]\quad . \tag{4.84}$$

The homogeneous Eq. (4.80) has exactly one solution, the trivial solution $\hat{h}_u=\hat{g}_d=0$, unless the matrix L is singular [109, Chapter 4]. In summary, our network model poses an eigenproblem in the complex eigenvalue s.

Similarly to the mass-spring-damper model, the model coefficients in Eq. (4.80) are again empirically closed with the buoyancy-driven flow inside the Rijke tube. The complex reflection coefficients at both ends of the Rijke tube are assumed to be identical. They are modeled as

$$R_u=k_{\mathrm{Re}}+\mathrm{j}\,k_{\mathrm{Im}}\quad , \tag{4.85}$$

$$R_b = k_{\mathrm{Re}} + \mathrm{j}\,k_{\mathrm{Im}} \quad . \tag{4.86}$$

The complex reflection coefficients are assumed to be frequency-independent. The interaction index n and the time delay τ are again modeled as

$$n = k_n \dot{Q} \quad , \tag{4.87}$$

$$\tau = k_\tau \frac{0.2 d_w}{u_0(l_u^-)} \quad . \tag{4.88}$$

Overall, we follow the structure of the discussion from Chapter 4.3.1.

We begin by performing a Bayesian regression of the network model coupled with the adiabatic base-flow model in the unstable operating regime ($\dot{Q} > 200\,\mathrm{W}$). The maximum-likelihood estimates of the model parameters are given in Table 4.6. The complex reflection coefficient $k_{\mathrm{Re}} + \mathrm{j}\,k_{\mathrm{Im}}$ has a value close to -1, which is the expected value for an acoustically open end. In particular, the absolute value is less than 1, which means that acoustic energy is partially radiated into the environment. The model parameters k_n and k_{Re}, which are expected to have a direct impact on the growth rate ($|k_{\mathrm{Re}} + \mathrm{j}\,k_{\mathrm{Im}}| \approx k_{\mathrm{Re}}$), as well as the model parameter k_τ are marked by high uncertainty. The model parameter k_{Im} is marked by low uncertainty because the resulting end correction is negligible compared to the travel times τ_u and τ_b ($\angle(k_{\mathrm{Re}} + \mathrm{j}\,k_{\mathrm{Im}}) \approx \pi - k_{\mathrm{Im}}$). The high uncertainties in k_{Re}, k_n and k_τ as well as the low uncertainty in k_{Im} are reflected in the model predictions of growth rate and frequency as shown in Fig. 4.16.

Table 4.6 Bayesian regression of network model coupled with adiabtic base-flow model. The uncertainty in the maximum-likelihood estimates is represented by the standard deviation in the marginal probability distributions. Note that the relative standard deviation of the model parameter k_{Re} is measured with respect to $1 - |k_{\mathrm{Re}} + jk_{\mathrm{Im}}| \approx 1 + k_{\mathrm{Re}}$.

model parameter	estimated value	standard deviation
k_{Re}	-0.979	0.004 (20.6%)
k_{Im}	0.071	0.005 (6.7%)
k_n [$\mathrm{kg\,m^{-1}s^{-2}}W^{-1}$]	73.558	11.177 (15.2%)
k_τ	15.138	1.694 (11.2%)

In Fig. 4.17, the model predictions of the growth rates and frequencies over the entire operating regime ($0 \leq \dot{Q} \leq 340\,\mathrm{W}$) are shown. The frequency is only slightly overpredicted for lower powers, whereas the growth rate is noticably underpredicted at first and eventually swings towards $\mathrm{Re}(s) \approx -6\,\mathrm{s}^{-1}$. In Fig. 4.18, the model predictions of the growth rates using nonparametric regression based on Algorithm 3.4 are shown. Multiple loops of nonparametric regression with message passing lead to virtually perfect model predictions

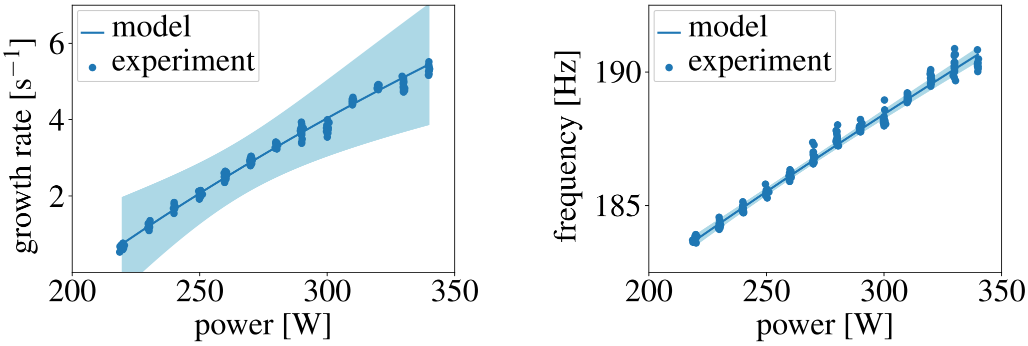


Fig. 4.16 Model predictions of growth rates (left) and frequencies (right) in Rijke tube from network model with 3-σ confidence intervals (93.9 % confidence).

of the frequency and slightly improved model predictions of the growth rate in the range of $30\,\mathrm{W} \leq \dot{Q} \leq 100\,\mathrm{W}$.

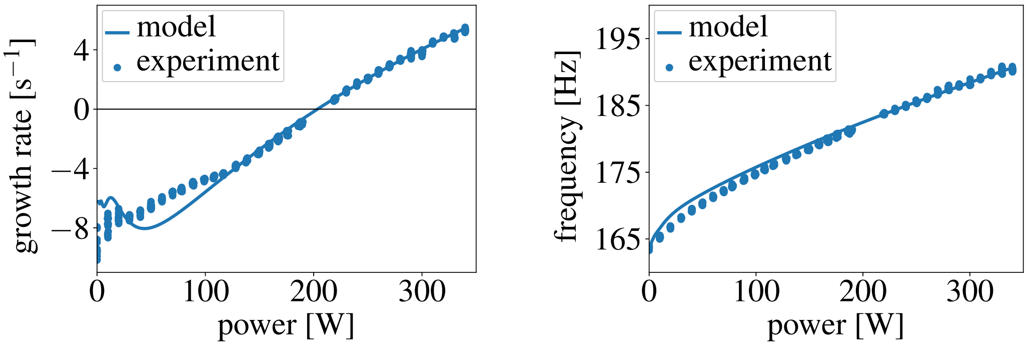


Fig. 4.17 Model predictions of growth rates (left) and frequencies (right) in Rijke tube from network model over entire operating regime.

In the limit of zero power, $\det(L) \equiv 0$ simplifies to (Eq. (3.32))

$$1 - R_u R_b \exp\left(-s\left(\tau_u + \tau_b\right)\right) \equiv 0 \quad . \tag{4.89}$$

With $n = 0$, only the complex reflections coefficients R_u and R_b remain so that the number of model parameters and the number of observed quantities is balanced. In particular, the complex reflection coefficients R_u and R_b are estimated to be $k_{\mathrm{Re}} + \mathrm{j}\,k_{\mathrm{Im}} \approx -0.971 + 0.052\mathrm{j}$ in the limit of zero power. Furthermore, we suppress the zero-power singularity in the time delay τ by making the subjective observation that the model parameter k_τ should go to zero so that $\lim_{\dot{Q}\to 0} \tau < \infty$. Fig. 4.19 shows how these two additional pieces of physical information make the model predictions of the growth rate virtually perfect. In particular, the network

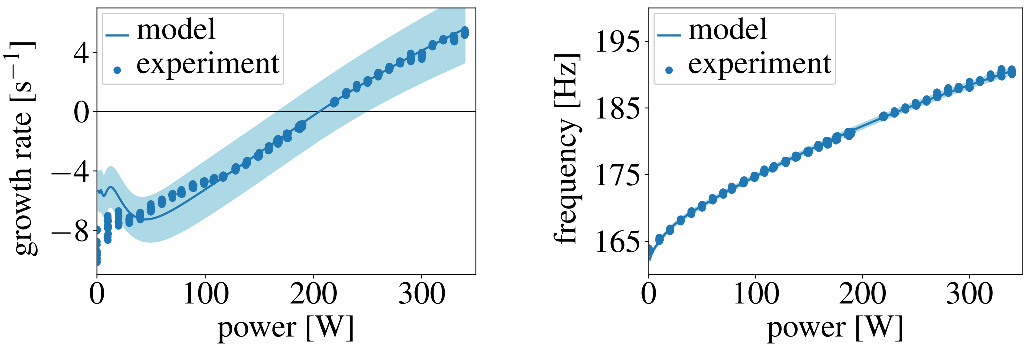


Fig. 4.18 Model predictions of growth rates (left) and frequencies (right) in Rijke tube from network model with Gaussian processes, expectation-maximization and message passing.

model, like the mass-springer-damper model, picks up the waviness in the observations of the growth rate, which is difficult to discern in the raw data with the naked eye (Fig. 4.2, left).

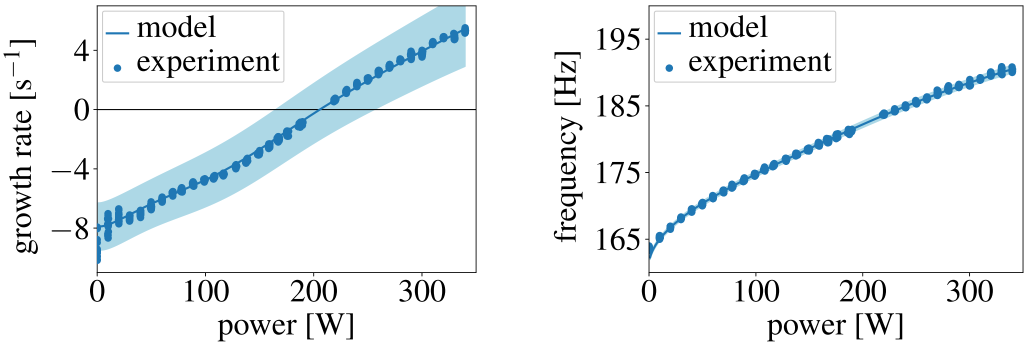


Fig. 4.19 Physics-informed model predictions of growth rates in Rijke tube from network model with 3-σ confidence intervals. $\lim_{\dot{Q}\to 0} k_{\mathrm{Re}} = -0.971$ is added in order to support a correct absolute value of the complex reflection coefficients in the limit of zero power. $\lim_{\dot{Q}\to 0} k_{\tau} = 0$ is added in order to suppress the zero-power singularity $\tau \to \infty$.

With various reduced-order models of base flow and acoustics at our disposal, we conclude our treatment with a model comparison on a statistical basis. While the discussion is based on the evidence, also known as the marginal likelihood, or the logarithm thereof, our model comparison procedure differs from our model selection procedure as outlined in Chapter 2.3: For a Gaussian process, the posterior estimates of the model parameters based on a Laplace approximation serve as observations in the parameter space. For our reduced-order models under comparison, we take the original observations from the state space. Therefore, the trained Gaussian processes of the model parameters serve as a prior, and a likelihood function reverses the Laplace approximation in order to bridge the gap between parameter space and state space:

$$P(\theta \mid M) = \frac{1}{\sqrt{\det(2\pi\Sigma_{\mathrm{GP}})}} \exp\left(-\frac{1}{2}\left(\theta_k - \theta_k^{\mathrm{GP}}\right)^T \left(\Sigma_{\mathrm{GP}}^{-1}\right)_{kl}\left(\theta_l - \theta_l^{\mathrm{GP}}\right)\right) \quad , \tag{4.90}$$

$$P(y \mid \theta, M) = \frac{1}{\sqrt{\det(2\pi\Sigma_{\mathrm{obs}})}} \exp\left(-\frac{1}{2}\left(y_k - s_k(\theta_k)\right)^* \left(\Sigma_{\mathrm{obs}}^{-1}\right)_{kl}\left(y_l - s_l(\theta_l)\right)\right) \tag{4.91}$$

$$\approx \frac{1}{\sqrt{\det(2\pi\Sigma_{\mathrm{obs}})}} \exp\Bigg(-\frac{1}{2}\left[\left(y_k - s_k\left(\theta_k^{\mathrm{GP}}\right)\right) - \frac{\mathrm{d}s_k}{\mathrm{d}\theta_k}\left(\theta_k - \theta_k^{\mathrm{GP}}\right)\right]^* \left(\Sigma_{\mathrm{obs}}^{-1}\right)_{kl}\left[\left(y_l - s_l\left(\theta_l^{\mathrm{GP}}\right)\right) - \frac{\mathrm{d}s_l}{\mathrm{d}\theta_l}\left(\theta_l - \theta_l^{\mathrm{GP}}\right)\right]\Bigg) \quad , \tag{4.92}$$

which is based on a first-order approximation of the complex eigenvalue s in the model parameters θ around the model parameters θ^{GP} as predicted by the Gaussian processes.[3] The marginal probability distribution, and thus the evidence, is retrieved from an inspection of Eq. (4.92) based on Corollary (A.2):

$$P(y \mid M) = \frac{1}{\sqrt{\det(2\pi\Sigma_{\mathrm{ev}})}} \exp\left(-\frac{1}{2}\left(y_k - s_k\left(\theta_k^{\mathrm{GP}}\right)\right)^* \left(\Sigma_{\mathrm{ev}}^{-1}\right)_{kl}\left(y_l - s_l\left(\theta_l^{\mathrm{GP}}\right)\right)\right) \quad , \tag{4.93}$$

$$\left(\Sigma_{\mathrm{ev}}\right)_{kl} = \left(\Sigma_{\mathrm{obs}}\right)_{kl} + \left(\frac{\mathrm{d}s_k}{\mathrm{d}\theta_k}\right)\left(\Sigma_{\mathrm{GP}}\right)_{mn}\left(\frac{\mathrm{d}s_l}{\mathrm{d}\theta_l}\right)^* \quad . \tag{4.94}$$

Note that the determinant and the quadratic form are real and Eq. (4.93) is well-defined because the covariance matrix Σ_{ev} as a complex matrix is Hermitian.

Before we compute the evidence for the various reduced-order models, it is worthwhile studying the structure of the covariance matrix Σ_{ev}. The likelihood of the model parameters θ^{GP} is given by

$$P(y \mid \theta^{\mathrm{GP}}, M) = \frac{1}{\sqrt{\det(2\pi\Sigma_{\mathrm{obs}})}} \exp\left(-\frac{1}{2}\left(y_k - s_k\left(\theta_k^{\mathrm{GP}}\right)\right)^* \left(\Sigma_{\mathrm{obs}}^{-1}\right)_{kl}\left(y_l - s_l\left(\theta_l^{\mathrm{GP}}\right)\right)\right) \quad . \tag{4.95}$$

The quadratic form inside the exponential term in Eq. (4.95) is the least-squares residual, and measures the statistical distance between the model predictions and the observations (Chapter 3.1). The covariance matrix Σ_{obs} represents the noise in the observations. Even if the model predictions exactly match the observations ($y_k = s_k(\theta_k^{\mathrm{GP}})$), strong noise ef-

---

[3] Note that all indices here, in particular for the model parameters θ, refer to the index of the experiment. For example, the model parameters $\theta_k = \theta(w_k)$ are associated with the design parameters w_k.

fectively diminishes our belief in this set of model parameters due to the factor in front of the exponential term in Eq. (4.95). The evidence $P(y \mid M)$ is formally identical to the likelihood $P(y \mid \theta^{\mathrm{GP}}, M)$ of the model parameters θ^{GP} except for the covariance matrix Σ_{ev} instead of Σ_{obs} (Eq. (4.93)). In addition to the covariance matrix of the observations, which measures the noise in the state space, it encompasses the covariance matrix of the model predictions Σ_{GP} as well as the eigenvalue sensitivities $\mathrm{d}s/\mathrm{d}\theta$, which are related to the model and parameter spaces (Eq. (4.94)):

- The covariance matrix Σ_{GP} represents the model complexity (Chapter 2.3). Strong correlations between the model parameters at different operating points are represented by large values in the off-diagonal entries of the covariance matrix Σ_{GP}, which in turn decreases the absolute value of the determinant $\det(\Sigma_{\mathrm{ev}})$. This leads to higher probabilities, which means that the evidence encourages reduced-order models with less complexity.

- The eigenvalue sensitivity $\mathrm{d}s/\mathrm{d}\theta$ represents the sensitivity of the state, i.e. the complex eigenvalue s, with respect to the model parameters (Chapter 3.2). The argument is similar: Large sensitivities increase the absolute value of the determinant $\det(\Sigma_{\mathrm{ev}})$. This leads to lower probabilities, which means that the evidence discourages reduced-order models with extreme sensitivity.

In summary, the evidence generalizes the likelihood of any particular set of model parameters. The likelihood function quantifies the fit of the model predictions with respect to the observations whereas the evidence additionally introduces measures to penalize model sensitivity and complexity.[4]

Table 4.7 gathers the evidence $P(y \mid M)$ computed for the various variations of the mass-spring-damper model and the network model investigated in Chapters 4.3.1 and 4.3.2. The computations for the base-flow model with heat loss follow the same procedure. The computation of the evidence formalizes the process of model comparison, which has so far been done by visual inspection, e.g. of growth rates and frequencies. For reference, the likelihood $P(y \mid \theta, M)$ computed for the various reduced-order models is also included in Table 4.7. Note that the likelihood reacts strongly to improvements in the model fit whereas the evidence is more conservative in its assessment based on model sensitivity and complexity. For example, we compare two reduced-order models in order to illustrate that likelihoods and evidence sometimes lead to different conclusions. On the one hand, our

---

[4] Note that the influences of model fit, sensitivity and complexity are not directly separable: The terms Σ_{obs}, Σ_{GP} and $\mathrm{d}s/\mathrm{d}\theta$ all have different units, and are thus not additive. While the covariance matrix Σ_{ev} is the sum of two other Hermitian matrices, there is no simple relationship between the inverse of the covariance matrix Σ_{ev} and the inverses of its summands.

arguably most simple reduced-order model is the combination of an adiabatic base-flow model and a mass-spring-damper model for the acoustics (Fig. 4.14, right). On the other hand, our arguably most complex reduced-order model is the combination of a base-flow model with heat loss and a network model for the acoustics (Fig. 4.20). The likelihoods favor the more complex reduced-order model.[5] Visual inspection confirms a good fit between model predictions and observations for all reduced-order models. However, the evidence slightly favors the more simple reduced-order model based on model sensitivity and complexity.[6] Alternatively, the better model fit of the more complex reduced-order model according to the likelihoods is invalidated by visual inspection of the large confidence intervals present in the model predictions (Fig. 4.20). In conclusion, this is an instance where the robustness of one reduced-order model outweighs the slighly better model fit of another reduced-order model, which makes the evidence, rather than the least-squares residual or the likelihood, the natural measure for model selection in Bayesian inference.

Table 4.7 Log-likelihoods $P(y \mid \theta, M)$ and log-marginal likelihoods $P(y \mid M)$ for various combinations of base-flow and acoustic models. A large[7] value of the log-likelihood $P(y \mid \theta, M)$ indicates that the model parameters θ are a good fit. A large value of the log-marginal likelihood $P(y \mid M)$ further indicates that the reduced-order model M is plausible in terms of model sensitivity and complexity (highlighted in grey color). This table summarizes the results of Chapters 4.3.1 and 4.3.2, i.e. the improvements to the mass-spring-damper (MSD) and network models due to message-passing (MP) and physics-informed (PI) Gaussian processes. The computations for the base-flow model with heat loss follow the same procedure.

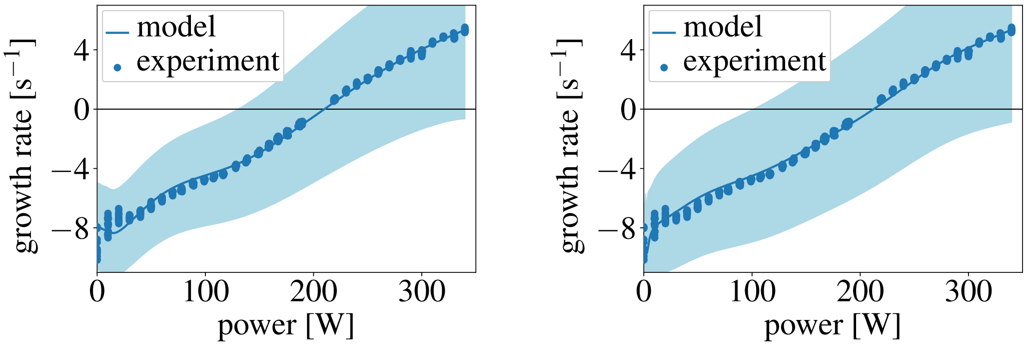
base flow	acoustics	log-evidence $\log(P(y \mid M))$	log-likelihood $\log(P(y \mid \theta, M))$	
adiabatic	MSD (w/o MP)	-104.3	-154.2	Fig. 4.11, left.
adiabatic	MSD	-68.9	-92.0	Fig. 4.11, right.
adiabatic	MSD (w/ PI)	-64.6	-82.0	Fig. 4.14, left.
adiabatic	MSD (w/ PI)	-56.3	-62.8	Fig. 4.14, right.
adiabatic	network	-50.0	-46.2	Fig. 4.18.
adiabatic	network (w/ PI)	-46.0	-34.8	Fig. 4.19.
heat loss	network	-67.0	-82.7	Not shown.
heat loss	network (w/ PI)	-57.3	-44.2	Fig. 4.20, left.
heat loss	network (w/ PI)	-59.4	-45.6	Fig. 4.20, right.

---

[5] Adiabatic & MSD (w/ PI): -62.8. Heat loss & network (w/ PI): $-44.2/45.6$.

[6] Adiabatic & MSD (w/ PI): -56.3. Heat loss & network (w/ PI): $-57.3/59.4$.

[7] Note that likelihoods and marginal likelihoods usually assume negative values because the Gaussian process priors are normalized to have standard deviations of 1. Therefore, a large value of the log-likelihood or log-marginal likelihood usually means a negative number of relatively small absolute value.



(a) $\lim_{\dot{Q}\to 0} k_{\mathrm{Re}}$, $\lim_{\dot{Q}\to 0} k_n$.

(b) $\lim_{\dot{Q}\to 0} k_{\mathrm{Re}}$, $\lim_{\dot{Q}\to 0} k_\tau$.

Fig. 4.20 Physics-informed model predictions of growth rates in Rijke tube from base-flow model with heat loss and network model with 3-σ confidence intervals.

Part III

Data assimilation

Chapter 5

Stochastic filtering and smoothing

Following the example of Chapter 3, we present another hierarchical framework for state estimation, parameter estimation and model selection. Compared to nonlinear regression, there are two main differences to stochastic filtering and smoothing[1]: Firstly, the emphasis is on high-dimensional sample spaces. The state vector of the reduced-order model in Part II is comprised of a single complex scalar, e.g. the eigenvalue of a linear stability analysis, whereas stochastic filtering and smoothing is computationally feasible for entire flow fields such as from DNS or LES. Secondly, we narrow the focus to time-dependent reduced-order models with time as the index set. The governing equations are explicit, and govern the transition from one state vector to the next in time. In Chapter 5.1, we introduce the notion of Markov chains, which formalizes the transitions between state vectors in sample space. In Chapter 5.2, we derive the Bayesian filter and smoother from first principles, i.e. the Markov property and Bayes' rule (Chapter 2.1), as well as special cases of the Kalman filter and smoother. While it is possible to derive the Kalman filter from variational principles, where the posterior estimate minimizes a least-squares residual, probability theory as the language of statistical inference is superior in that it does not preclude nonlinear transitions between non-normal probability distributions. In Chapter 5.3, we draw an analogy to nonparametric regression (Chapter 3.3), and introduce combined state and parameter estimation. Although the emphasis of Part III is on state and parameter estimation, we also briefly discuss model selection and the issue of filter divergence.

Supplementary material:

---

[1] Note that the stochastics do not refer to random fluctuations in the state as they would arise in turbulence for example. Instead, it goes back to the complementary interpretations of frequentist and Bayesian probability (Chapter 2.1). In the Bayesian interpretation, probability theory and statistics are the natural language for statistical inference regardless of random events.

- `Chapter5/Scalar advection.ipynb` (Git, Binder).
- `Chapter5/Lorenz system.ipynb` (Git, Binder).

5.1 Markov chains

In Part III, we consider a reduced-order model of the time-dependent form

$$f : \Omega \times \mathbb{R}^M \to \Omega \quad , \tag{5.1}$$

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x, \theta) \quad , \tag{5.2}$$

where the state of the system is denoted by $x \in \Omega$ and the model parameters by $\theta \in \mathbb{R}^M$. The evolution of the state vector x is governed by the reduced-order model f and the model parameters θ. In general, the sample space Ω depends on the reduced-order model, and is potentially high-dimensional or even infinite-dimensional (Definition 2.1).[2] Nevertheless, we assume without loss of generality that the sample space Ω is finite-dimensional [114, Chapter 3]. Furthermore, we discretize the reduced-order model in time, which is also known as the method of lines [108, Chapter 11]. For a sequence of timesteps t_k, with $0 \leq k \leq K$, and observations y_k of the true state $x_k = x(t_k)$, the governing equations of the reduced-order model are alternatively written as

$$G : \Omega \times \mathbb{R}^M \to \Omega \quad , \tag{5.3}$$

$$x_k = G(x_{k-1}, \theta) = x_{k-1} + \int_{t_{k-1}}^{t_k} f(x(t), \theta)\,\mathrm{d}t \quad . \tag{5.4}$$

The state of the system at a timestep t_k is unambiguously defined by the state vector x_k. Thus, our belief in the state x_k depends exclusively on our belief in (i) the previous state x_{k-1} as well as our choice of (ii) the governing equations f (or G) and (iii) the model parameters θ. At the same time, the state vector x_k is compared to noisy observations y_k through a measurement operator H. In terms of random variables, we formulate (Chapter 2.1)

$$x_k \sim P(x_k \mid x_{k-1}, \theta, f) \quad , \tag{5.5}$$

$$y_k = H(x_k) \sim P(y_k \mid x_k) \quad . \tag{5.6}$$

---

[2] In CFD for example, the obvious choice of sample space encompasses all the primitive variables, such as velocities, pressures and temperatures, for every element of the discretized computational domain. This easily adds up to at least 10^6 degrees of freedom, often more.

In brief, the transition from one state to the next is governed by the operator G. The operator G is derived from the reduced-order model f, and depends on the model parameters θ. The states x_k and the observations y_k are considered realizations of their respective random variables. The degrees of belief in each are subject to ('$\sim$') their conditional probability distributions $P(x_k \mid x_{k-1}, \theta, f)$ and $P(y_k \mid x_k)$, respectively. The degree of belief in the state x_k is conditional on the previous state x_{k-1} and the choices of the reduced-order model f as well as the model parameters θ. The degree of belief in the observation y_k is conditional on the true state x_k.

The rules of statistical inference for time-dependent reduced-order models are formalized as follows [128, Chapter 4]. Firstly, we assume that the reduced-order model is described by a Markov chain. This means that the belief in a state x_k depends only on the belief in the previous state x_{k-1}:

1. The future is independent of the past given the present:

$$P(x_k \mid x_{0:k-1}, y_{1:k-1}, \theta, f) = P(x_k \mid x_{k-1}, \theta, f) \quad , \tag{5.7}$$

   where $x_{0:k-1}$ denotes the state vectors at all timesteps from t_0 to t_{k-1}, and $y_{1:k-1}$ denotes the observations at all timesteps from t_1 to t_{k-1}.

2. The past is independent of the future given the present:

$$P(x_k \mid x_{k+1:K}, y_{k+1:K}, \theta, f) = P(x_k \mid x_{k+1}, \theta, f) \quad , \tag{5.8}$$

   where $x_{k+1:K}$ and $y_{k+1:K}$ respectively denote the state vectors and observations at all timesteps from $k+1$ to the final timestep K.

Secondly, observations are assumed to be statistically independent in time (Definition 2.4). The probability of an observation depends only on the current state:

$$P(y_k \mid x_{0:K}, y_{1:k-1}, y_{k+1:K}, \theta, f) = P(y_k \mid x_k) \quad . \tag{5.9}$$

In Fig. 5.1, the relationship between state vectors and observations is shown as well as the roles of governing equations, model parameters and measurements.

The ultimate goal of data assimilation would be to find the joint probability distribution $P(x_{0:K}, y_{1:K}, \theta, f)$. The corresponding sample space would span all state vectors x_k from timestep 0 to K and all observations y_k from timestep 1 to K, as well as the reduced-order model f and the model parameters θ. This joint probability distribution would give a complete statistical description. In principle, all probability distributions of interest would

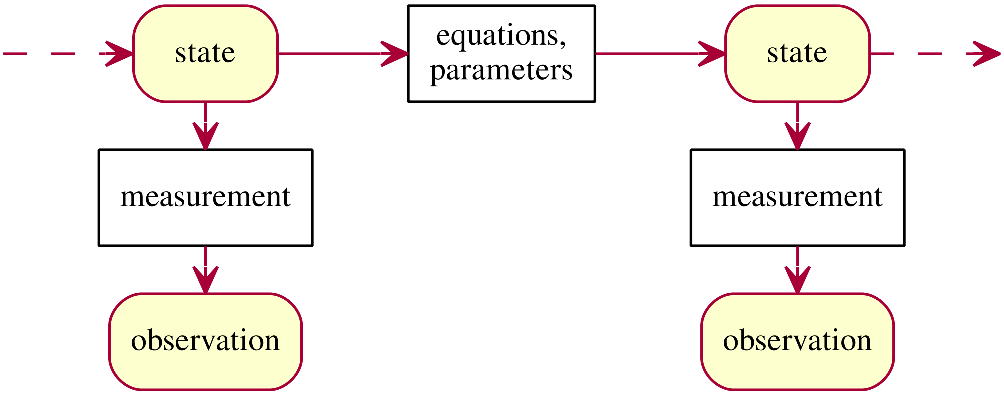


Fig. 5.1 Graphical model of time-dependent reduced-order model. The operator G and the model parameters θ govern the transition from one state to the next. The state vector x is related to the observation y through a measurement operator H.

be derived from this joint probability distribution (Table 5.1). In practice, it is difficult to compute this probability distribution because of the high dimensionality of this sample space [129, Chapter 1]. Therefore, state estimation focuses on the more direct computation of conditional probability distributions over a single timestep (Table 5.2).

Table 5.1 Conditional probability distributions in data assimilation.

task	PDF	description
state estimation	$P(x_{0:K} \mid y_{1:K}, \theta, f)$	Given a reduced-order model and its parameters, what is our belief in a series of states?
parameter estimation	$P(\theta \mid y_{1:K}, f)$	Given a reduced-order model, what is our belief in a set of parameters?
model comparison	$P(f \mid y_{1:K})$	Between two reduced-order models, in which one do we believe more?

Table 5.2 Conditional probability distributions in state estimation.

task	PDF	description
filtering	$P(x_k \mid y_{1:k}, \theta, f)$	Given all observations from the past and now, what is our belief in the current state?
smoothing	$P(x_k \mid y_{1:K}, \theta, f)$	Given all observations from the past, the future and now, what is our belief in the current state?
prediction	$P(x_{K+1} \mid y_{1:K}, \theta, f)$	Given all observations from the past and now, what is our belief in a future state?

5.2 Kalman filter and smoother

For the filtering problems in Tables 5.2 and 5.3, Bayes' rule gives (Theorem 2.3)

$$P(x_k \mid y_{1:k}, \theta, f) = \frac{P(y_k \mid x_k, y_{1:k-1}, \theta, f) P(x_k \mid y_{1:k-1}, \theta, f)}{P(y_k \mid y_{1:k-1}, \theta, f)} \tag{5.10}$$

$$= \frac{P(y_k \mid x_k) P(x_k \mid y_{1:k-1}, \theta, f)}{P(y_k \mid y_{1:k-1}, \theta, f)} \quad . \tag{5.11}$$

The prediction $P(x_k \mid y_{1:k-1}, \theta, f)$ in Eq. (5.11) is given by the Chapman-Kolmogorov equation (Theorem 2.1):

$$P(x_k \mid y_{1:k-1}, \theta, f) = \int P(x_k, x_{k-1} \mid y_{1:k-1}, \theta, f) \, \mathrm{d}x_{k-1} \tag{5.12}$$

$$= \int P(x_k \mid x_{k-1}, y_{1:k-1}, \theta, f) P(x_{k-1} \mid y_{1:k-1}, \theta, f) \, \mathrm{d}x_{k-1} \tag{5.13}$$

$$= \int P(x_k \mid x_{k-1}, \theta, f) P(x_{k-1} \mid y_{1:k-1}, \theta, f) \, \mathrm{d}x_{k-1} \quad . \tag{5.14}$$

The Chapman-Kolmogorov equation requires the statistical inverse $P(x_{k-1} \mid y_{1:k-1}, \theta, f)$ from the previous timestep t_{k-1}. In general, it is solved either numerically [105], analytically (Theorem 5.3) or by a Monte-Carlo simulation (Theorem 5.4). The steps in the Bayesian filter are summarized as follows:

Theorem 5.1 (Bayesian filter)**.**

1. *Prediction step:*

$$P(x_k \mid y_{1:k-1}, \theta, f) = \int P(x_k \mid x_{k-1}, \theta, f) P(x_{k-1} \mid y_{1:k-1}, \theta, f) \, \mathrm{d}x_{k-1} \quad . \tag{5.15}$$

2. *Update step:*

$$P(x_k \mid y_{1:k}, \theta, f) = \frac{P(y_k \mid x_k) P(x_k \mid y_{1:k-1}, \theta, f)}{P(y_k \mid y_{1:k-1}, \theta, f)} \quad . \tag{5.16}$$

At the timestep $t_k = t_K$, the filtered and smoothed distributions are identical (Table 5.2). If the smoothed distribution at a timestep t_{k+1} is known, the smoothed distribution at the previous timestep t_k is also known due to the nature of Markov chains (Eq. (5.8)). This is formalized as follows (Theorem 2.1):

$$\begin{aligned} P(x_k \mid y_{1:K}, \theta, f) &= \int P(x_k, x_{k+1} \mid y_{1:K}, \theta, f) \, \mathrm{d}x_{k+1} && (5.17) \\ &= \int P(x_k \mid x_{k+1}, y_{1:K}, \theta, f) P(x_{k+1} \mid y_{1:K}, \theta, f) \, \mathrm{d}x_{k+1} && (5.18) \\ &= \int P(x_k \mid x_{k+1}, y_{1:k}, \theta, f) P(x_{k+1} \mid y_{1:K}, \theta, f) \, \mathrm{d}x_{k+1} \quad . && (5.19) \end{aligned}$$

The probability distribution $P(x_{k+1} \mid y_{1:K}, \theta, f)$ in Eq. (5.19) is the smoothed distribution from the subsequent timestep t_{k+1}. The probability distribution $P(x_k \mid x_{k+1}, y_{1:k}, \theta, f)$ in Eq. (5.19) is computed via Bayes' rule (Theorem 2.3):

$$\begin{aligned} P(x_k \mid x_{k+1}, y_{1:k}, \theta, f) &= \frac{P(x_{k+1} \mid x_k, y_{1:k}, \theta, f) P(x_k \mid y_{1:k}, \theta, f)}{P(x_{k+1} \mid y_{1:k}, \theta, f)} && (5.20) \\ &= \frac{P(x_{k+1} \mid x_k, \theta, f) P(x_k \mid y_{1:k}, \theta, f)}{P(x_{k+1} \mid y_{1:k}, \theta, f)} \quad . && (5.21) \end{aligned}$$

Note that Eq. (5.21) involves the filtered distribution $P(x_k \mid y_{1:k}, \theta, f)$ from the current timestep t_k (Theorem 5.1). The steps in the Bayesian smoother are summarized as follows:

Theorem 5.2 (Bayesian smoother)**.**

1. *Forward loop: Bayesian filter (Theorem 5.1).*

2. *Backward loop:*

$$P(x_k \mid y_{1:K}, \theta, f) = P(x_k \mid y_{1:k}, \theta, f) \int \frac{P(x_{k+1} \mid x_k, \theta, f) P(x_{k+1} \mid y_{1:K}, \theta, f)}{P(x_{k+1} \mid y_{1:k}, \theta, f)} \, \mathrm{d}x_{k+1} \quad . \tag{5.22}$$

Note that both the Bayesian filter and smoother are sequential in nature. In the Bayesian filter, the predicted distribution $P(x_k \mid y_{1:k-1}, \theta, f)$ at a timestep t_k (Eq. (5.15)) mainly depends on the filtered distribution $P(x_{k-1} \mid y_{1:k-1}, \theta, f)$ from the previous timestep t_{k-1} (Eq. (5.16)). In the Bayesian smoother, the smoothed distribution $P(x_k \mid y_{1:K}, \theta, f)$ at a timestep t_k mainly depends on the smoothed distribution $P(x_{k+1} \mid y_{1:K}, \theta, f)$ from the subsequent timestep t_{k+1} (Eq. (5.22)). The existence of sequential algorithms and dynamic programming for the computation of filtered and smoothed distributions significantly reduces the complexity of data assimilation [129, Chapter 1].

Two additional assumptions are introduced to make the computation of filtered and smoothed distributions feasible. Firstly, the prior and the likelihood in the update step of the Bayesian filter are assumed to be normal (Eq. (5.16)):

$$x_k \sim \mathcal{N}(\psi^f, C^f_{\psi\psi}) \quad , \tag{5.23}$$

$$y_k \sim \mathcal{N}(Hx_k, C_{\varepsilon\varepsilon}) \quad , \tag{5.24}$$

where $\mathcal{N}$ denotes a normal distribution with respective mean and covariance matrix. The mean of the prior is denoted by ψ^f, its covariance matrix by $C^f_{\psi\psi}$, and the covariance matrix of the likelihood, also known as the observation error, by $C_{\varepsilon\varepsilon}$. From Eq. (5.23) and (5.24), it follows that the filtered distribution $P(x_k \mid y_{1:k}, \theta, f)$ is normal (Eq. (5.16)). Secondly, the operator G (Eq. (5.4)) in the prediction step (Eq. (5.15)) is assumed to be linear in the state vector x. The result is the well-known Kalman filter [130, 131]:

Theorem 5.3 (Kalman filter)**.** *The update step in the Kalman filter gives:*

$$x_k \sim \mathcal{N}(\psi^a, C^a_{\psi\psi}) \quad , \tag{5.25}$$

$$\psi^a = \psi^f + \left(HC^f_{\psi\psi}\right)^T \left[C_{\varepsilon\varepsilon} + HC^f_{\psi\psi}H^T\right]^{-1} \left(y_k - H\psi^f\right) \quad , \tag{5.26}$$

$$C^a_{\psi\psi} = C^f_{\psi\psi} - \left(HC^f_{\psi\psi}\right)^T \left[C_{\varepsilon\varepsilon} + HC^f_{\psi\psi}H^T\right]^{-1} \left(HC^f_{\psi\psi}\right) \quad , \tag{5.27}$$

where the superscript f denotes ‘forecast’ (everything pertaining to the prediction), and the superscript a denotes ‘analysis’ (everything pertaining to the update).

If the reduced-order model is nonlinear, the prediction of the covariance matrix $C^f_{\psi\psi}$ for the Kalman filter is approximated in different ways. In the extended Kalman filter, the predicted covariance matrix $C^f_{\psi\psi}$ is computed by linearizing the operator G [132, Chapter 6]. In strongly nonlinear dynamical systems, the predictions are found to be poor [114, Chapter 3]. Higher-order extended Kalman filters are available [133]. Nevertheless, drawbacks include their significant storage requirements, which increase exponentially with the order of approximation. An alternative is the ensemble Kalman filter [134, 135]. Instead of a mean ψ and a covariance matrix $C_{\psi\psi}$, a distribution is represented by a sample $(\psi^j)_{1\leq j\leq n}$. During the prediction step, the ensemble members ψ^j evolve in time independently. Before the update step, the statistics are recovered from the sample as follows:

$$\Psi = \left(\psi^1 - \overline{\psi} \quad , \quad \psi^2 - \overline{\psi} \quad , \quad \cdots \quad , \quad \psi^n - \overline{\psi}\right) \quad , \tag{5.28}$$

$$\overline{\psi} \approx \frac{1}{n}\sum_{j=1}^{n} \psi^j \quad , \quad C_{\psi\psi} \approx \frac{1}{n-1}\Psi\Psi^T \quad . \tag{5.29}$$

The sample covariance matrix $C_{\psi\psi}$ involves division by $n-1$ instead of n in order to avoid a sample bias.

Various implementations of the ensemble Kalman filter exist, which differ in the update step. In the straightforward implementation of the ensemble Kalman filter [134], every ensemble member is individually updated (Eq. (5.26)). Note that the observations have to be randomly perturbed in order to guarantee a statistically consistent analysis scheme [135]. In order to avoid the introduction of randomly generated numbers, the square-root filter is used here [136]. The square-root filter belongs to a larger family of ensemble Kalman filters called ensemble-transform Kalman filters [137, 138]. Unlike the straightforward implementation of the ensemble Kalman filter, the mean and the deviations of the ensemble members are updated. This requires the singular value decomposition of a symmetric, positive, semi-definite matrix ($V\Sigma V^T$, where V is orthonormal and Σ diagonal), but no spurious errors due to the random perturbation of the observations are introduced.

Theorem 5.4 (Square-root filter)**.** *The update step in the square-root filter gives:*

$$(\psi^a)^j = \overline{\psi}^a + (\Psi^a)_j \quad , \tag{5.30}$$

$$\overline{\psi}^a = \overline{\psi}^f + \Psi^f \left(H\Psi^f\right)^T \left[(n-1)C_{\varepsilon\varepsilon} + H\Psi^f \left(H\Psi^f\right)^T\right]^{-1} \left(y_k - H\overline{\psi}^f\right) \quad ,$$

$$\Psi^a = \Psi^f V \left[E - \Sigma\right]^{\frac{1}{2}} V^T \quad , \quad V\Sigma V^T = \left(H\Psi^f\right)^T \left[(n-1)C_{\varepsilon\varepsilon} + H\Psi^f \left(H\Psi^f\right)^T\right]^{-1} H\Psi^f \quad ,$$

where E is the identity matrix.

Following Eq. (5.23) and (5.24), the Bayesian smoother becomes the Rauch-Tung-Striebel (RTS) smoother, also known as the Kalman smoother [128, Chapter 8]. The smoothed distribution $P(x_k \mid y_{1:K}, \theta, f)$ becomes a normal distribution $\mathcal{N}(\psi^s, C^s_{\psi\psi})$ with mean ψ^s and covariance matrix $C^s_{\psi\psi}$. In order to again avoid the shortcomings of assuming linearity, an ensemble Kalman smoother is presented here.

Theorem 5.5 (Rauch-Tung-Striebel smoother)**.** *The backward loop in the RTS smoother has the following recursion:*

$$x_k \sim \mathcal{N}(\psi^s, C^s_{\psi\psi}) \quad , \tag{5.31}$$

$$\psi^s = \psi^a + \left((C^f_{\psi\psi})^{-1}_{k+1} G C^a_{\psi\psi} \right)^T \left[\psi^s_{k+1} - \psi^f_{k+1} \right] \quad , \tag{5.32}$$

$$C^s_{\psi\psi} = C^a_{\psi\psi} - \left((C^f_{\psi\psi})^{-1}_{k+1} G C^a_{\psi\psi} \right)^T \left[(C^f_{\psi\psi})_{k+1} - (C^s_{\psi\psi})_{k+1} \right] \left((C^f_{\psi\psi})^{-1}_{k+1} G C^a_{\psi\psi} \right) \quad . \tag{5.33}$$

Theorem 5.6 (Ensemble Kalman smoother)**.** *The backward loop in the ensemble Kalman smoother has the following recursion:*

$$(\psi^s)^j = (\psi^a)^j + \Psi^a (\Psi^f)^+_{k+1} \left[(\psi^s)^j_{k+1} - (\psi^f)^j_{k+1} \right] \quad , \tag{5.34}$$

where the superscript plus denotes the pseudoinverse of the rectangular matrix Ψ *[108, Chapter 4].*

5.3 Filter divergence

The Kalman filter and smoother address the task of state estimation, i.e. filtering and smoothing the true state from partial and noisy observations (Table 5.2). As discussed in Chapter 1.3, the uncertainty in model predictions is attributed to two sources, namely parametric and structural uncertainty. In analogy to Chapter 3.3, we relax the structure of the governing equations in two steps: (i) state augmentation and (ii) stochastic processes.

In a first step, the state vector is augmented by the model parameters so that they become subject to the same statistical inference (Fig. 5.2):

$$\widetilde{x}_k = \begin{pmatrix} x_k \\ \theta_k \end{pmatrix} \quad , \quad \widetilde{f}(\widetilde{x}(t)) = \begin{pmatrix} f(x(t), \theta) \\ 0 \end{pmatrix} \quad , \quad \widetilde{G}(\widetilde{x}_{k-1}) = \begin{pmatrix} G(x_{k-1}, \theta_{k-1}) \\ \theta_{k-1} \end{pmatrix} \quad , \tag{5.35}$$

$$\widetilde{x}_k = \widetilde{G}(\widetilde{x}_{k-1}) = \widetilde{x}_{k-1} + \int_{t_{k-1}}^{t_k} \widetilde{f}(\widetilde{x}(t)) \, \mathrm{d}t \quad . \tag{5.36}$$

The tasks in combined state and parameter estimation are given in Table 5.3. The filtered and smoothed distributions in the model parameters θ_k, $P(\theta_k \mid y_{1:k}, f)$ and $P(\theta_k \mid y_{1:K}, f)$ respectively, are retrieved by marginalizing the state vectors x_k from the augmented state vectors $\widetilde{x}_k$ (Corollary A.1). Note that the model parameters θ_k are now time-dependent as the system traverses different regimes in state space. This turns the strongly constrained parameter estimation into a weakly constrained combined state and parameter estimation [114, Chapter 5]. Thus, the results of marginalizing the probability distributions in combined state and parameter estimation (Table 5.3) are not strictly equivalent to the solutions of parameter estimation (Table 5.1).

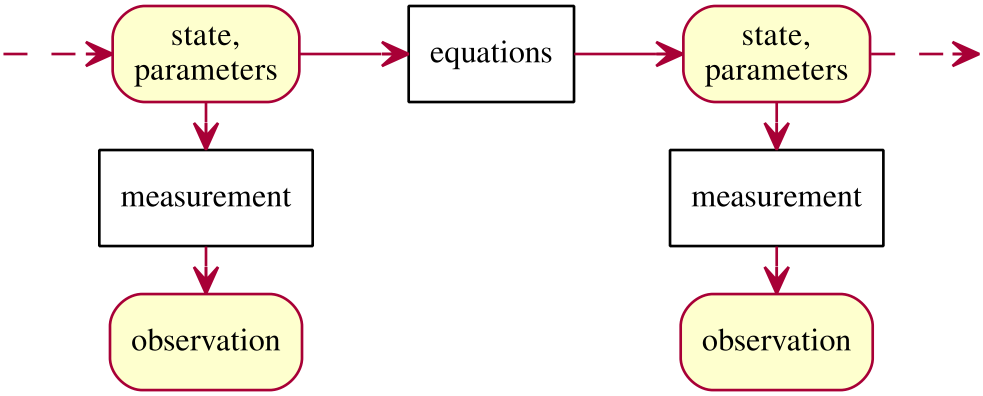


Fig. 5.2 Graphical model of time-dependent reduced-order model in combined state and parameter estimation. Unlike in state estimation, the model parameters θ are lumped with the state vector x into an augmented state vector $\widetilde{x}$ (Fig. 5.1).

Table 5.3 Conditional probability distributions in combined state and parameter estimation.

task	PDF	description
filtering	$P(x_k,\theta_k \mid y_{1:k},f)$	Given all observations from the past and now, what is our belief in the current state and set of parameters?
smoothing	$P(x_k,\theta_k \mid y_{1:K},f)$	Given all observations from the past, the future and now, what is our belief in the current state and set of parameters?

Following augmentation, combined state and parameter estimation is stable in the sense that the model parameters converge to a set of values without any further regularization (Definition 3.1). Firstly, the uncertainty after the update step as measured by the covariance matrix $C^a_{\psi\psi}$ is strictly lower than the uncertainty after the prediction step as measured by the covariance matrix $C^f_{\psi\psi}$. For example, it follows from Eq. (5.27) that the uncertainty of a fully observed one-dimensional state vector $x_k \in \mathbb{R}$ is given by

$$C^a_{\psi\psi} = C^f_{\psi\psi}\left[1-\frac{C^f_{\psi\psi}}{C_{\varepsilon\varepsilon}+C^f_{\psi\psi}}\right] < C^f_{\psi\psi} \quad . \tag{5.37}$$

Without loss of generality, the marginal uncertainty in the model parameters θ as measured by the trace of their covariance matrix $C_{\theta\theta}$ in combined state and parameter estimation is monotonically decreasing:

$$\mathrm{tr}\big(\big(C^a_{\theta\theta}\big)_{k-1}\big) = \mathrm{tr}\big(\big(C^f_{\theta\theta}\big)_k\big) > \mathrm{tr}\big(\big(C^a_{\theta\theta}\big)_k\big) \quad . \tag{5.38}$$

In particular, the trace of the covariance matrix bounds the variance of each model parameter. Secondly, the confidence in the model predictions increasingly outweighs the confidence in the observations. For example, it follows from Eq. (5.26) that the filtered estimate ψ^a of a fully observed one-dimensional state vector $x_k \in \mathbb{R}$ is given by

$$\psi^a = \frac{C_{\varepsilon\varepsilon}}{C_{\varepsilon\varepsilon} + C^f_{\psi\psi}} \psi^f + \frac{C^f_{\psi\psi}}{C_{\varepsilon\varepsilon} + C^f_{\psi\psi}} y_k \quad . \tag{5.39}$$

For $C^f_{\psi\psi} \to 0$, the observations y_k hold less and less weight.

As the series of filtered estimates converges, this raises the question whether the model parameters converge to a meaningful set of values. Note that the uncertainty $C^a_{\psi\psi}$ after the update step, unlike the filtered estimate ψ^a, improves regardless of the actual fit between the model predictions ψ^f and the observations y_k. This assumes that the reduced-order model is valid over the entire regime (Eq. (5.2)). Otherwise, the reduced-order model becomes biased in its predictions and underestimates the uncertainty in its predictions. This shortcoming of stochastic filtering and smoothing is also known as filter divergence [139].

In a second step, we further relax the augmented governing equations (Eq. (5.36)). For the lack of better knowledge, we statistically model the unknown dynamics of the model parameters as a Wiener process (Chapter 2.2):[3]

$$\widetilde{x}_k = \widetilde{G}(\widetilde{x}_{k-1}) = \begin{pmatrix} x_{k-1} + \int_{t_{k-1}}^{t_k} f(x(t), \theta)\,\mathrm{d}t \\ \theta_{k-1} + \beta(t_k - t_{k-1}) \end{pmatrix} \quad , \tag{5.41}$$

where $\beta(t)$ denotes Brownian motion as a function of time. The result is an increase in uncertainty in the model predictions. In turn, this leads to a more balanced appraisal of model predictions and observations in the update step (Eq. (5.26)). Figuratively speaking, the injection of Brownian motion leads to a memory loss: White noise accumulates and erodes over time the confidence in the model parameters θ_{k-1}, which are based on past observations $y_{l<k}$. When the observation y_k becomes available, the Kalman filter is biased

---

[3] Note that we deliberately do not formulate the natural generalization of a time-dependent reduced-order model (Eq. (3.47)):

$$\frac{\mathrm{d}\widetilde{x}}{\mathrm{d}t} = \widetilde{f}(\widetilde{x}(t)) = \begin{pmatrix} f(x(t), \theta) \\ \gamma(t) \end{pmatrix} \quad , \quad \gamma(t) \sim \mathcal{N}(0,1) \quad , \tag{5.40}$$

where $\gamma(t)$ represents white noise. While the left-hand side is an ordinary differential equation, the right-hand side is nowhere continuous. Existence and uniqueness of solutions in the classical sense are not straightforward, and require mathematical tools such as Itô or Stratonovich integrals [101, Chapter 4]. A rigorous treatment based on the theory of stochastic integration is beyond the scope of this discussion [105]. Instead, we consider Brownian motion as a series of steps $\beta(t_k - t_{k-1}) = \beta(t_k) - \beta(t_{k-1})$ in a random walk.

towards the observation y_k at the expense of the model prediction ψ^f and hence the past observations $y_{l<k}$.

In practice, structural uncertainty is statistically modeled by covariance inflation [114, Chapter 15]. In additive covariance inflation, stochastic noise is injected after the prediction step and before the update step [140]. The strength of the stochastic noise of course depends on the time interval between successive observations. In analogy to Brownian motion for example, the step size in a random walk scales with the square-root of time (Definition 2.16). In multiplicative covariance inflation, the covariance matrix $C^f_{\psi\psi}$ of the model prediction is multiplied by a factor $(1+\rho)^2$ [112]. In the square-root filter for example, this is accomplished by replacing Ψ^f with $(1+\rho)\Psi^f$ (Theorem 5.4). Optimization of the hyperparameter ρ requires a trade-off: On the one hand, large ρ mitigates filter divergence such that model predictions and the observations stay consistent. On the other hand, overly large ρ artificially distorts the confidence in the model predictions to the point that the uncertainty quantification becomes worthless. Furthermore, there are adaptive schemes for model selection, which are beyond the scope of this discussion [114, Chapter 15].

In conclusion, we consider time-dependent reduced-order models in Part III that are Markov chains. The theory of stochastic filtering and smoothing provides us with flexible algorithms for state and parameter estimation, ranging from the Bayesian filter, which is in theory the ideal stochastic filter, to the ensemble Kalman filter, which is computationally tractable for high-dimensional sample spaces at the expense of introducing statistical assumptions. Nevertheless, it is not a matter of course that any sample space is compatible with our statistical assumptions, i.e. least-squares residuals and normal distributions. In the next chapter, we encounter such a reduced-order model that allows for various sample spaces subject to different statistics, and provide an in-depth discussion on the construction of an appropriate data assimilation framework.

Chapter 6

Level-set data assimilation framework

In this chapter, the ensemble Kalman filter is combined with a narrow-band level-set method and a fast-marching method to form a computationally efficient level-set data assimilation framework. In analogy to the distinction between front-tracking and front-capturing methods in describing the motion of an interface, data can also be assimilated according to various paradigms. For the various paradigms, it is not a priori clear which will yield superior results. For this reason, the theory and the derivations behind our level-set data assimilation framework are worked out in detail.

This chapter is structured as follows: In Chapter 6.1, various formulations for the laws of motion are discussed, including the Hamilton-Jacobi equation. It is shown that the solutions to the Hamilton-Jacobi equation, the so-called generating functions, form a natural sample space for data assimilation. In Chapter 6.2, level-set methods are combined with the ensemble Kalman filter and the ensemble Kalman smoother, and thus become accessible to Bayesian inference. In Chapters 6.3 and 6.4, the level-set data assimilation framework based on the Hamilton-Jacobi equation is verified on the basis of one-dimensional and two-dimensional examples, respectively. In the following chapters, the level-set data assimilation framework is demonstrated on the nonlinear dynamics of a ducted premixed flame.

Relevant publications:

- Yu/Juniper/Magri: *Combined state and parameter estimation in level-set methods*. Journal of Computational Physics, 2019.

Supplementary material:

- `Chapter6/Level-set methods.ipynb` (Git, Binder).
- `Chapter6/Eikonal field.ipynb` (Git, Binder).

6.1 Level-set methods

A number of problems in computational physics involve the motion of interfaces, e.g. semiconductor manufacturing, multi-phase flows, crystal growth, groundwater flow, computer vision, grid generation and seismology [141]. In general, there are two approaches to calculating the motion of an interface [142]: In front-tracking methods, the interface is parametrized and discretized so that one follows the motion of the whole interface by tracking a sufficient number of points on the interface. In front-capturing methods, the interface is embedded into a function defined over the whole domain. One example of front-capturing methods are the so-called level-set methods, where the interface is embedded into a strictly monotonic function [143]. Both front-tracking and front-capturing methods have well-understood advantages, and hybrid methods exist to mitigate their respective disadvantages [144]. Nevertheless, it is worth mentioning that level-set methods provide a natural formulation for calculating the motion of an interface [141]: Level-set methods deal well with non-smooth features such as corners and cusps as well as topological merging and break-up. Furthermore, level-set methods are easily extended from two to three and higher dimensions. For more information on level-set methods, the reader is referred to many excellent expositions in the literature [145, 146, 141, 144, 142].

Despite the elegance of explaining physical phenomena by the motion of interfaces and calculating the motion of the interfaces by level-set methods, one has to remain aware that the assumption that manifolds are infinitely thin is often an asymptotic assumption. This is particularly true in fluid mechanics, which is governed by conservation laws [147]. One example is the kinematics of premixed flames [14, Chapter 2]: Depending on the combustion regime, it may be assumed that a thin reactive-diffusive layer separates the burnt and unburnt gases. While the laminar flame speed, at which the premixed flame propagates from the burnt into the unburnt region, is a well-defined thermo-chemical property of the fuel-air mixture in a one-dimensional flow, it more generally depends on the balance between heat conduction and mass diffusion, whose effects vary with flame stretch and curvature. Moreover, the turbulent flame speed scales differently depending on the interactions between turbulence and combustion length scales. Finally, level-set methods do not conserve mass, neither locally nor globally [148]. This is particularly problematic in the presence of large density gradients and in closed flows. Hybrid methods, which combine both front tracking and capturing, exist to mitigate this shortcoming [144]. Nevertheless, ad-hoc corrections only improve level-set methods at the expense of increased model complexity.

Despite the potential quantitative shortcomings in describing the kinematics of a premixed flame as the motion of an interface, it has been shown that this model successfully explains the linear and nonlinear dynamics observed in thermoacoustic instabilities of ducted premixed flames, which are relevant to the design of combustion chambers in jet and rocket engines [53, 39, 57, 62]. The discrepancies which arise when comparing to more faithful simulations or experiments are usually attributed to the unpredictable nature of turbulent flow or to uncertainties in the model and its parameters. Thus, it is relevant to assess the ability of a qualitative, physics-informed model to make quantitative, time-accurate predictions. The aim of this chapter is to develop a data-driven, statistically rigorous framework for state and parameter estimation in models using level-set methods, which has not been done before, and to apply it to oscillating flames in the following chapters (Chapters 7-9).

Inference over interfaces based on level-set methods has been performed, e.g. in the context of shape optimization, either by directly taking the functional derivative of the objective functional with respect to the level-set function [149, 150], or, equivalently, by embedding the level-set method into shape calculus [151–153]. As an alternative to these variational approaches, our framework for state and parameter estimation is based on data assimilation [114]. Data assimilation finds the statistically optimal combination of model predictions and observations. It combines concepts from control theory, probability theory and dynamic programming [101, 132, 154]. The data assimilation technique used here is the ensemble Kalman filter [134, 135]. In the ensemble Kalman filter, a Monte-Carlo approach is used to represent the necessary statistics at every timestep, which makes it a computationally efficient technique in terms of storage requirements and horizontal scalability. Compared to other data assimilation techniques based on the Kalman filter, e.g. the extended Kalman filter [133, 155], the ensemble Kalman filter is found to be particularly robust with respect to larger nonlinearities [114, Chapter 4]. This is relevant for level-set methods, due to strongly nonlinear events such as cusp formation, topological merging and break-up. A practical advantage of the ensemble Kalman filter is its non-intrusive implementation with little effort required for its parallelization. The ensemble Kalman filter has been successfully applied to a number of problems in fluid mechanics: turbulent near-wall flow [156]; transonic flows around airfoils and wings [157]; viscous flow around a cylinder [158]; model uncertainties in Reynolds-averaged Navier-Stokes (RANS) equations [159]; vortex models of separated flow [160]; and extinction and reignition dynamics in turbulent non-premixed combustion [161]. Relevant applications of data assimilation to level-set methods outside fluid mechanics include wildfire spread [162–166] and pollution on the surface of oceans and lakes [167, 168].

When data assimilation is applied to the motion of an interface, the question arises as to what constitutes its sample space. Refining the distinction between front-tracking and front-capturing methods, there are, at least, three principle ways to view the motion of an interface [141, 144]:

Geometric view. The interface is parametrized and discretized so that one follows the motion of the whole interface by solving the laws of motion for a sufficient number of points on the interface.

Set-theoretic view. A characteristic function is defined over the whole domain. The characteristic function assumes one of two values, depending on whether the point at the location in question is inside or outside the region enclosed by the interface.

Analytic view (also known as "analysis view" [144]). A level-set function is defined over the whole domain. The interface is reconstructed by identifying the position of a particular level set.

Beyond the description of motion, these three views also pertain to the assimilation of data. In fact, all three views have been employed in earlier level-set data assimilation frameworks: While using a level-set method when solving for the motion of the interface, ROUCHOUX ET AL. take each entry in the innovation vector $y_k - H\psi^f$ (Eq. (5.25)) as the distance between one point on the predicted interface and one point on the observed interface [169]. This framework corresponds to the geometric view. GAO ET AL. compute y_k and ψ^f in the innovation vector from a progress variable [170]. The progress variable assumes values between zero and one, with most intermediate values assumed near the interface. This framework is an approximation of the set-theoretic view. MORENO AND AANONSEN compute the observation y_k and the model prediction ψ^f from a level-set function [171]. This framework corresponds to the analytic view, and comes closest to the level-set data assimilation framework presented here. The motion of an interface depends on the values of the level-set function in the immediate environment of the interface, whereas the values of the level-set function are not unique away from the interface. Our level-set data assimilation framework (i) introduces an additional constraint in agreement with the analytic view, (ii) validates the choice of the constraint for two analytical test cases in one and two dimensions respectively, and (iii) addresses shortcomings in the aforementioned frameworks.

In order to construct the sample space, the derivation of the level-set method is revisited in this section. The centerpiece of this derivation is the Hamilton-Jacobi equation. In theory, the level-set method is alternatively derived as the transport of a passive scalar quantity [14, Chapter 2]. Although the value of the level-set function is well defined at the interface, the choice of values for the level-set function away from the interface is in general not unique, and would require an ad-hoc constraint. It is demonstrated how the necessary constraint to the Hamilton-Jacobi equation naturally follows from the choice of phase space. Finally, it is shown that the solutions to the Hamilton-Jacobi equation, the so-called generating functions, form the appropriate sample space for data assimilation. Two complementary level-set algorithms are combined to solve the Hamilton-Jacobi equation computationally efficiently: the narrow-band method [172] and the fast-marching method [173]. Along with the ensemble Kalman filter (Theorem 5.4), the three algorithms form the backbone of our level-set data assimilation framework.

The laws of motion of an interface shall be given by

$$\frac{\mathrm{d}r}{\mathrm{d}t} = u(r) \quad , \tag{6.1}$$

where r is the position of one point on the interface, and u is the velocity. In Hamiltonian mechanics, motion is described in phase space using generalized coordinates and momenta [174, Chapter 3]. With interfaces in mind, a natural choice for the phase space is to use the position r as generalized coordinates and the normal vector n as generalized momenta.

For the Hamilton-Jacobi equation, the generating function $G(r,t)$ shall satisfy the following properties:

- The initial interface $r(0)$ is given by a level set $G(r(0),0) = \text{const.}$
- The moving interface $r(t)$ is identified as the level set $G(r(t),t) = G(r(0),0)$.
- Huygens' principle [174, Chapter 9] states that $\nabla G = n$. Given the choice of phase space, this is a necessary relationship between the generating function and the generalized coordinates and momenta.

The first two properties are present in every front-capturing method [14, Chapter 2]. The third property, Huygens' principle, follows from Hamiltonian mechanics (Appendix B). Physically speaking, every point on the interface is interpreted as the source of a wave. As the waves propagate through the domain, they connect each point on the interface to every other point in the domain. Geometrically speaking, Huygens' principle gives an eikonal field [175].

It remains to show how the generating function translates into a level-set method. The Lagrangian $\mathscr{L}(r,\dot{r},t)$ is given by (Lemma B.1)

$$\mathscr{L}(r(t),\dot{r}(t),t) = 0 \quad , \tag{6.2}$$

$$\text{subject to} \quad G(r(t),t) - G(r(0),0) = 0 \tag{6.3}$$

$$\text{and} \quad n(t) \cdot n(t) - 1 = 0 \quad . \tag{6.4}$$

The Hamiltonian $\mathscr{H}(r,n,t)$, subject to the same constraints as the Lagrangian $\mathscr{L}(r,\dot{r},t)$ (Eq. (6.3), (6.4)), is given by (Lemma B.2)

$$\mathscr{H}(r(t),n(t),t) = u(r(t)) \cdot n(t) \quad . \tag{6.5}$$

Finally, the Hamilton-Jacobi equation is given by (Theorem B.1)

$$\frac{\partial G}{\partial t} + u(r(t)) \cdot n(t) = 0 \quad . \tag{6.6}$$

The solution to the Hamilton-Jacobi equation, the generating function $G(r,t)$, is defined over the whole domain. As such, it represents the state of the interface in the sample space. Eq. (6.6) is solved near the interface. Away from the interface, the generating function is constrained by (Eq. (6.4))

$$|\nabla G| = 1 \quad . \tag{6.7}$$

Eq. (6.7) is an eikonal equation [175]. The solutions to this eikonal equation are signed distance functions.

6.2 Data assimilation framework

Our level-set data assimilation framework combines three algorithms: (i) the narrow-band method to solve the Hamilton-Jacobi equation near the interface [172], (ii) the fast-marching method to extend the eikonal field from the narrow band to the whole domain [173], and (iii) the ensemble Kalman filter and smoother to assimilate data [114].

The narrow-band method forms the backbone of the level-set data assimilation framework [172]. It solves the Hamilton-Jacobi equation (Eq. 6.6) in a narrow band near the interface. Using a narrow band reduces the computational cost by one order of magnitude compared to finding the generating function over the whole domain [176]. In a numerical simulation, the spatial resolution shall be denoted by Δx and the temporal resolution by Δt. The temporal resolution has to satisfy the Courant-Friedrich-Lewy condition $u\Delta t < \Delta x$ at every point on the interface [172]. The narrow-band method consists of the following steps (Algorithm 6.1):

Initialize. Before the first timestep, the signed distance field has to be known in a sufficiently wide neighborhood of the initial interface. It may be obtained from either a previously computed generating function or exact knowledge of the position of the initial interface.

Localize. At the timestep k, a band

$$T^k = \{r : G(r,t_k) < \gamma\Delta x\} \tag{6.8}$$

is formed, where γ is a constant chosen according to the width of the narrow band. A second, enveloping band

$$N^k = \{r : G(r+r', t_k) < \gamma\Delta x, \exists |r'| < \Delta x\} \tag{6.9}$$

is formed. It follows from the Courant-Friedrich-Lewy condition that the band N^k contains the band T^{k+1} formed at the next timestep $k+1$.

Advance. The generating function over the band N^k is evolved by one timestep. Note that the solution $\check{G}^{k+1}(r)$ is not the generating function at the next timestep $k+1$. While the position of the interface correctly coincides with the level set of the solution $\check{G}^{k+1}(r)$, Huygens' principle is no longer satisfied [177].

Reinitialize. The following Hamilton-Jacobi equation is solved over the band N^k until steady state is reached [177]:

$$\frac{\partial \check{G}}{\partial \check{t}} + \operatorname{sgn}\left(\check{G}(r,\check{t})\right)\left(\left|\nabla \check{G}\right| - 1\right) = 0 \quad , \tag{6.10}$$

$$\check{G}(r,0) = \check{G}^{k+1}(r) \quad . \tag{6.11}$$

The steady-state solution $\check{G}(r,\check{t} \to \infty)$ gives the generating function $G(r,t_{k+1})$ over the band N^k at the timestep $k+1$. Note that this Hamilton-Jacobi equation is a numerical continuation of the eikonal equation (Eq. (6.7)), where $\check{t}$ denotes a pseudo-time [46, Chapter 8].

Algorithm 6.1 Narrow-band level-set method.

```
Initialize interface.

for $0 \leq k < K$ do
    Localize narrow band.                  ▷ Narrow band (Eq. (6.8), (6.9)).
    Advance generating function.           ▷ Hamilton-Jacobi equation (Eq. (6.6)).
    Reinitialize generating function.      ▷ Eikonal equation (Eq. (6.7), (6.10), (6.11)).
end for
```

When observations are available, the prediction step is followed by an update step (Algorithm 6.2). Like the prediction step, the update step involves localization and reinitialization so that the narrow band follows the position of the interface. Instead of the advancement step, there is an extension step and an assimilation step:

Extend. While a narrow band is sufficient to capture the motion of an interface, data assimilation has to be performed in one sample space for all interfaces. In a numerical simulation, the state vector comprises the values of the generating function at every grid point. The fast-marching method extends the generating function from the narrow band to the whole domain, thus returning a well-defined state vector in sample space. The complexity of the fast-marching method is $\mathscr{O}(m^2 \log(m))$ in two dimensions or $\mathscr{O}(m^3 \log(m))$ in three dimensions, where m denotes the number of grid points in one dimension [173]. While the generating function over the whole domain is relatively expensive to compute, the fast-marching method is only called when observations are available. If combined with the narrow-band method (complexity $\mathscr{O}(m)$ or $\mathscr{O}(m^2)$ [172]), this level-set method becomes affordable overall.

Assimilate. The ensemble Kalman filter (Theorem 5.4) and/or smoother (Theorem 5.6) are applied. The assimilation step is followed by a localization step and a reinitialization step in order to prepare the narrow band for the next prediction step.

Algorithm 6.2 Level-set data assimilation framework with filtering.

```
Initialize interface.

for $0 \le k < K$ do
    // Prediction step (Theorem 5.1).
    Localize narrow band.                 ▷ Narrow band (Eq. (6.8), (6.9)).
    Advance generating function.          ▷ Hamilton-Jacobi equation (Eq. (6.6)).
    Reinitialize generating function.     ▷ Eikonal equation (Eq. (6.7), (6.10), (6.11)).

    // Update step (Theorem 5.1).
    Localize narrow band.                 ▷ Narrow band (Eq. (6.8), (6.9)).
    Extend narrow band.                   ▷ Fast-marching method [173].
    Assimilate observations.              ▷ Ensemble Kalman filter (Theorem 5.4).
    Localize narrow band.                 ▷ Narrow band (Eq. (6.8), (6.9)).
    Reinitialize generating function.     ▷ Eikonal equation (Eq. (6.7), (6.10), (6.11)).
end for
```

Combining the building blocks of the level-set method and data assimilation gives the level-set filtering and smoothing frameworks (Algorithms 6.2, 6.3). In the following sections, the update step in the level-set filtering framework is verified using two analytical test cases: In a one-dimensional test case, the level-set filtering framework, which represents the analytic view of data assimilation, is contrasted with the geometric and set-theoretic views of data assimilation. It is shown that the analytic view of data assimilation works correctly, which sets the stage for its generalization to higher dimensions. Afterwards, the level-set filtering framework is applied to a two-dimensional test case, and the effect of the observations on the shape of the interface is discussed. It is shown how various parameters such as the number of observations and the observation error may affect the performance of the level-set data assimilation framework.

Algorithm 6.3 Level-set data assimilation framework with filtering and smoothing.

```
Initialize interface.

// Forward loop (Theorem 5.2).
for $0 \leq k < K$ do
    // Prediction step (Theorem 5.1).
    Localize narrow band.                    ▷ Narrow band (Eq. (6.8), (6.9)).
    Advance generating function.             ▷ Hamilton-Jacobi equation (Eq. (6.6)).
    Reinitialize generating function.        ▷ Eikonal equation (Eq. (6.7), (6.10), (6.11)).

    // Update step (Theorem 5.1).
    Localize narrow band.                    ▷ Narrow band (Eq. (6.8), (6.9)).
    Extend narrow band.                      ▷ Fast-marching method [173].
    Assimilate observations.                 ▷ Ensemble Kalman filter (Theorem 5.4).
    Localize narrow band.                    ▷ Narrow band (Eq. (6.8), (6.9)).
    Reinitialize generating function.        ▷ Eikonal equation (Eq. (6.7), (6.10), (6.11)).
end for

// Backward loop (Theorem 5.2).
for $K \geq k > 0$ do
    Localize narrow band.                    ▷ Narrow band (Eq. (6.8), (6.9)).
    Extend narrow band.                      ▷ Fast-marching method [173].
    Assimilate observations.                 ▷ Ensemble Kalman smoother (Theorem 5.6).
    Localize narrow band.                    ▷ Narrow band (Eq. (6.8), (6.9)).
    Reinitialize generating function.        ▷ Eikonal equation (Eq. (6.7), (6.10), (6.11)).
end for
```

6.3 One-dimensional test case

The predicted and observed positions x and y of an interface in one dimension are respectively given by (Fig. 6.1)

$$x \sim \mathscr{N}(\mu_\psi, C_{\psi\psi}) \quad , \quad \mu_\psi = 0 \quad , \quad C_{\psi\psi} = 0.64 \quad , \tag{6.12}$$

$$y = x \sim \mathcal{N}(\mu_\varepsilon, C_{\varepsilon\varepsilon}) \quad , \quad \mu_\varepsilon = 1 \quad , \quad C_{\varepsilon\varepsilon} = 0.36 \quad . \tag{6.13}$$

In one dimension, an interface reduces to a point. The normal vector on the interface is unique up to its orientation. Geometric data assimilation using the Kalman filter (Theorem 5.3) reduces to optimal interpolation of the positions μ_ψ and μ_ε (Eq. (5.26)):

$$x = \mu_\psi + \frac{C_{\psi\psi}}{C_{\varepsilon\varepsilon} + C_{\psi\psi}} \left(\mu_\varepsilon - \mu_\psi\right) = 0.64 \quad . \tag{6.14}$$

The filtered position x is closer to the observation than to the prediction in accordance with the variances $C_{\psi\psi}$ and $C_{\varepsilon\varepsilon}$ (Fig. 6.1, left).

For set-theoretic data assimilation, the characteristic function χ is assumed to be bounded by ± 1 with the interface at $\chi = 0$:

$$\chi(x) = \begin{cases} -1 & x < c \\ 0 & x = c \\ 1 & x > c \end{cases} \quad , \quad c \sim \mathcal{N}(\mu_\psi, C_{\psi\psi}) \quad . \tag{6.15}$$

The characteristic function χ follows a Bernoulli distribution, which is the appropriate probability distribution for a random variable with two possible outcomes, at every location x [99, Chapter 6]:

$$P(\chi(x) = 1) = \Phi\left(\frac{x - \mu_\psi}{\sqrt{C_{\psi\psi}}}\right) \quad , \tag{6.16}$$

$$P(\chi(x) = -1) = 1 - \Phi\left(\frac{x - \mu_\psi}{\sqrt{C_{\psi\psi}}}\right) \quad , \tag{6.17}$$

where Φ denotes the cumulative density function of the normal distribution. The characteristic function χ does not follow a normal distribution whereas the Kalman filter requires the mean $\overline{\chi}$ and the covariance function $\mathscr{C}_{\chi\chi}$:

$$\overline{\chi}(x) = 2\Phi\left(\frac{x - \mu_\psi}{\sqrt{C_{\psi\psi}}}\right) - 1 \quad , \tag{6.18}$$

$$\mathscr{C}_{\chi\chi}(x_1, x_2) = 4\Phi\left(\frac{\min(x_1, x_2) - \mu_\psi}{\sqrt{C_{\psi\psi}}}\right)\left(1 - \Phi\left(\frac{\max(x_1, x_2) - \mu_\psi}{\sqrt{C_{\psi\psi}}}\right)\right) \quad . \tag{6.19}$$

Note that the mean $\overline{\chi}$ is not a characteristic function. Set-theoretic data assimilation using the Kalman filter (Theorem 5.3) gives

$$\begin{aligned}\chi(x) &= 2\Phi\left(\frac{x-\mu_\psi}{\sqrt{C_{\psi\psi}}}\right)-1\\ &\quad -\frac{4\Phi\left(\frac{\min(x,\mu_\varepsilon)-\mu_\psi}{\sqrt{C_{\psi\psi}}}\right)\left(1-\Phi\left(\frac{\max(x,\mu_\varepsilon)-\mu_\psi}{\sqrt{C_{\psi\psi}}}\right)\right)}{1+4\Phi\left(\frac{\mu_\varepsilon-\mu_\psi}{\sqrt{C_{\psi\psi}}}\right)\left(1-\Phi\left(\frac{\mu_\varepsilon-\mu_\psi}{\sqrt{C_{\psi\psi}}}\right)\right)}\left(2\Phi\left(\frac{\mu_\varepsilon-\mu_\psi}{\sqrt{C_{\psi\psi}}}\right)-1\right) \qquad (6.20)\\ &= 2\Phi(1.25x)-1-\frac{4\Phi(1.25\min(x,1))(1-\Phi(1.25\max(x,1)))}{1+4\Phi(1.25)(1-\Phi(1.25))}(2\Phi(1.25)-1) \quad . \qquad (6.21)\end{aligned}$$

The interface is localized at $x \approx 0.139$ (Fig. 6.1, middle). The filtered position x is closer to the prediction than to the observation although the prediction variance $C_{\psi\psi}$ is larger than the observation variance $C_{\varepsilon\varepsilon}$. Note that the filtered position x is independent of the variance $C_{\varepsilon\varepsilon}$.

For analytic data assimilation, the predicted and observed positions x and y are viewed as points on the level sets of the predicted and observed generating functions respectively. It follows from Eq. (6.12) that the predicted generating function is given by

$$G(x) \sim \mathcal{N}(G(x)\,|\,x-\mu_\psi, C_{\psi\psi}) \quad . \qquad (6.22)$$

The mean $\overline{G}$ and the covariance function $\mathscr{C}_{GG}$ are given by

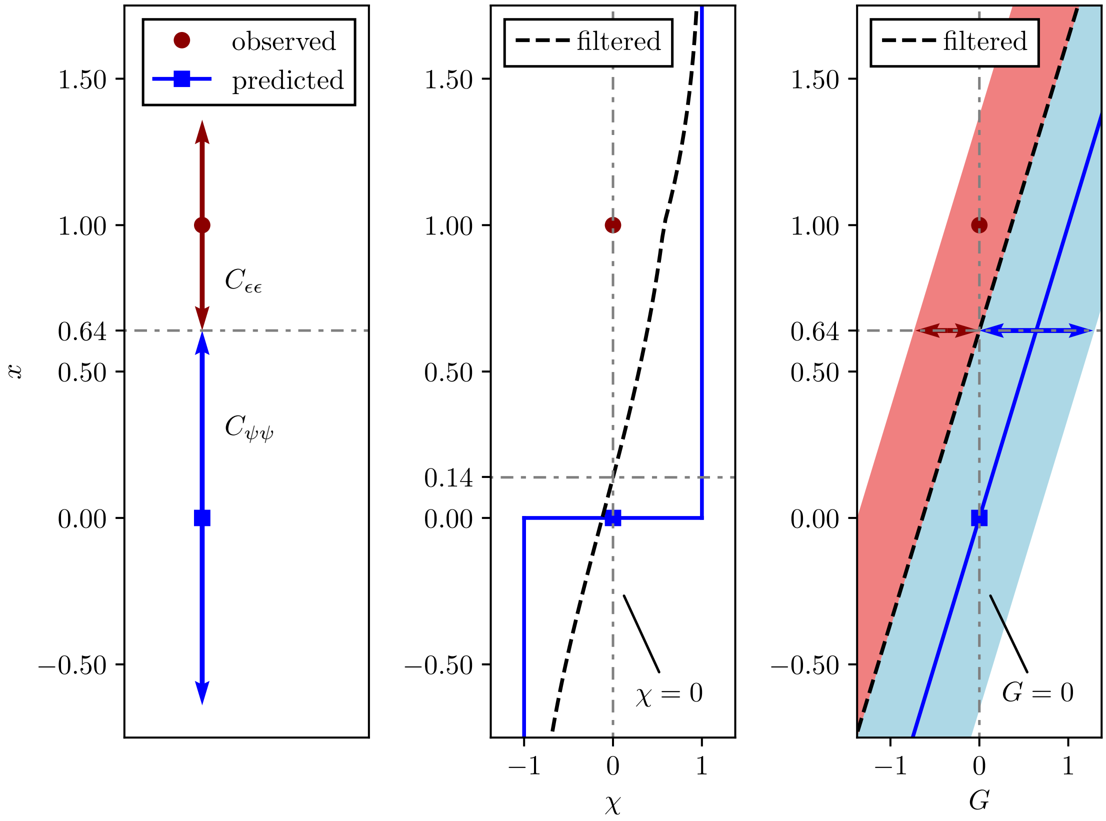
$$\overline{G}(x) = x-\mu_\psi \quad , \quad \mathscr{C}_{GG}(x_1,x_2) = C_{\psi\psi} \quad . \qquad (6.23)$$

Analytic data assimilation using the Kalman filter (Theorem 5.3) on the generating function results in

$$G(x) = x-\mu_\psi-\frac{C_{\psi\psi}}{C_{\varepsilon\varepsilon}+C_{\psi\psi}}\left(\mu_\varepsilon-\mu_\psi\right) = x-0.64 \quad . \qquad (6.24)$$

The filtered position x is given by the zero-level set of the filtered generating function G (Eq. (6.24)), which is identical to the result from geometric data assimilation (Fig. 6.1, right).

In summary, geometric, set-theoretic and analytic data assimilation are compared. Set-theoretic data assimilation gives implausible results in one dimension, and is thus discarded from any further consideration. Geometric and analytic data assimilation give consistent results in one dimension. Nevertheless, the generalization of geometric data assimilation to higher dimensions is not straightforward because it requires a one-to-one correspondence between the predicted interface and the observations. In the next test case, the generalization of analytic data assimilation, as the most feasible framework, to higher dimensions is preferentially investigated.



(a) geometric view (b) set-theoretic view (c) analytic view

Fig. 6.1 Data assimilation from geometric (left), set-theoretic (middle) and analytic view (right). The blue square at $x = 0$ marks the predicted position of the interface. The red circle at $x = 1$ marks the observed position of the interface. The grey, dash-dotted lines give the coordinates of the filtered position of the interface. In one dimension, data assimilation from the geometric view (left) reduces to optimal interpolation of the x coordinates, weighted by $C_{\psi\psi}$ and $C_{\varepsilon\varepsilon}$ respectively. The geometric solution may be considered the most intuitive one. In the middle figure, the Kalman filter is applied to the characteristic function χ (blue line). The set-theoretic solution (dashed line) localizes the interface at $x \approx 0.139$, significantly differently from the geometric solution. This shows that oft-used physical quantities, which exhibit behavior similar to a characteristic function in the vicinity of the interface of interest, are not suitable for data assimilation. In the right figure, the Kalman filter is applied to the generating function G (blue line). The analytic solution (dashed line) localizes the interface at $x = 0.64$, identically to the geometric solution.

6.4 Two-dimensional test case

The distance field S of a corner in (x_0, y_0) is given by

$$S(x,y;x_0,y_0) = \begin{cases} x - x_0 & \frac{\pi}{4} \leq \theta \leq \pi \\ y - y_0 & -\frac{\pi}{2} \leq \theta \leq \frac{\pi}{4} \\ -\sqrt{(x-x_0)^2 + (y-y_0)^2} & \pi \leq \theta \leq \frac{3\pi}{2} \end{cases} \quad . \tag{6.25}$$

For prediction, the coordinates of the corner (x_0, y_0) are taken to be statistically independent and normally distributed:

$$x_0 \sim \mathcal{N}(0,1) \quad , \quad y_0 \sim \mathcal{N}(0,1) \quad . \tag{6.26}$$

A sketch of the corner and its median distance field $S(x,y;0,0)$ are given in Fig. 6.2.

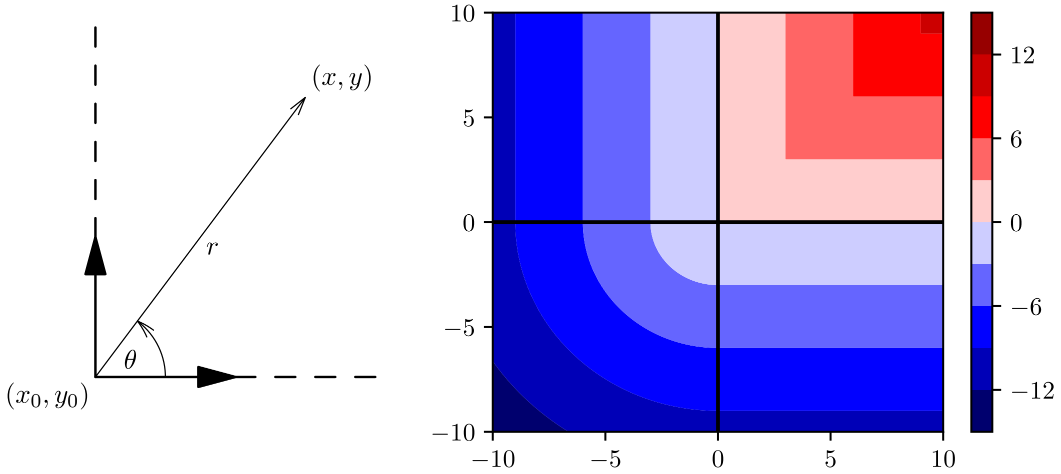


Fig. 6.2 Sketch of a corner (left) and its median distance field $S(x,y;0,0)$ (right). The walls of the corner (left, dashed line), which are represented by the zero-level set in analytic data assimilation, are parallel to the positive x and y axes and share their orientations (right). The points with $0 < \theta < \frac{\pi}{2}$ are defined to have positive distance.

For analytic data assimilation, the predicted position x is again viewed as the level set of a predicted generating function. It follows from Eq. (6.25) and (6.26) that the predicted generating function is given by

$$G(x,y) \sim \mathcal{N}\left(\overline{G}(x,y), C_{GG}(x_1,y_1,x_2,y_2)\right) \quad , \tag{6.27}$$

where the mean $\overline{G}$ and the covariance function C_{GG} are given by

$$\begin{aligned}
\overline{G}(x,y) &= \mathbb{E}_{x_0,y_0}[S](x,y) && (6.28)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S(x,y;\xi,\eta)p(x_0=\xi,y_0=\eta)\,\mathrm{d}\xi\mathrm{d}\eta \quad , && (6.29)\\
C_{GG}(x_1,y_1,x_2,y_2) &= \mathbb{E}_{x_0,y_0}\left[(S-\mathbb{E}_{x_0,y_0}[S])\times(S-\mathbb{E}_{x_0,y_0}[S])\right](x_1,y_1,x_2,y_2) && (6.30)\\
&= \mathbb{E}_{x_0,y_0}\left[S\times S\right](x_1,y_1,x_2,y_2) - \mathbb{E}_{x_0,y_0}[S](x_1,y_1)\mathbb{E}_{x_0,y_0}[S](x_2,y_2) && (6.31)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S(x_1,y_1;\xi,\eta)S(x_2,y_2;\xi,\eta)p(x_0=\xi,y_0=\eta)\,\mathrm{d}\xi\mathrm{d}\eta\\
&\quad - \mathbb{E}_{x_0,y_0}[S](x_1,y_1)\mathbb{E}_{x_0,y_0}[S](x_2,y_2) \quad , && (6.32)
\end{aligned}$$

where $\mathbb{E}_{x_0,y_0}$ denotes the expected value over the random variables x_0 and y_0, and the '×' symbol denotes the Cartesian product of two random variables. The mean $\overline{G}$ and the covariance function C_{GG} are computed from Eq. (6.29) and (6.32) using Gauss-Hermite quadrature [178, Chapter 5]. Alternatively, the mean and the variance function can be computed from the following one-dimensional integrals on the unit circle:

$$\mathbb{E}_{x_0,y_0}[S](x,y) = \int_{\pi/4}^{\pi}\cos(\theta)g(\theta)\,\mathrm{d}\theta + \int_{-\pi/2}^{\pi/4}\sin(\theta)g(\theta)\,\mathrm{d}\theta - \int_{\pi}^{3\pi/2}g(\theta)\,\mathrm{d}\theta \quad , \tag{6.33}$$

$$\mathbb{E}_{x_0,y_0}\left[S^2\right](x,y) = \int_{\pi/4}^{\pi}\cos(\theta)^2h(\theta)\,\mathrm{d}\theta + \int_{-\pi/2}^{\pi/4}\sin(\theta)^2h(\theta)\,\mathrm{d}\theta + \int_{\pi}^{3\pi/2}h(\theta)\,\mathrm{d}\theta \quad . \tag{6.34}$$

The auxiliary functions g and h are given by

$$g(\theta) = \varphi(w)\left[\left(v^2+1\right)\Phi(v) + v\varphi(v)\right] \quad , \tag{6.35}$$

$$h(\theta) = \varphi(w)\left[\left(v^2+2\right)\varphi(v) + \left(3v+v^3\right)\Phi(v)\right] \quad , \tag{6.36}$$

where φ and Φ denote the probability density function and cumulative distribution function of the normal distribution. The natural coordinates (v,w) are given by

$$v = x\cos(\theta) + y\sin(\theta) \quad , \tag{6.37}$$

$$w = x\sin(\theta) - y\cos(\theta) \quad . \tag{6.38}$$

The mean and the standard deviation of the predicted generating function are shown in Fig. 6.3. Artefacts from the interaction between the nonlinearity of distance fields and the assumption of normal distributions are visible: The mean of the predicted generating function

(Fig. 6.3, left) is rounded in the corner whereas distance fields are supposed to have right angles (Fig. 6.2). The standard deviation of the predicted generating function (Fig. 6.3, right) assumes non-unity values along $\theta = 45^\circ$ although the coordinates of the corner are statistically independent and normally distributed (Eq. (6.26)). The effects of the artefacts in the predicted generating function are now examined in various observation scenarios.

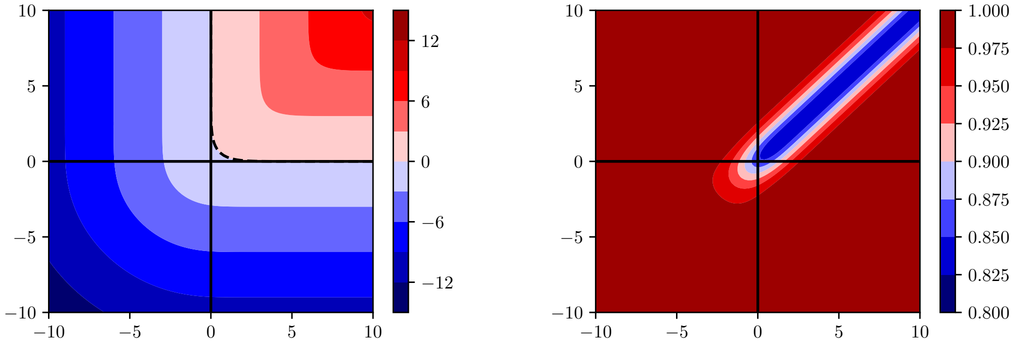


Fig. 6.3 Mean (left) and standard deviation (right) of predicted generating function. The zero-level set in the mean is marked by a dashed line. Away from $\theta = 45^\circ$, the mean of the generating function comes close to its median (Fig. 6.2, right). Note that the distance field S (Eq. (6.25)) is smooth but not differentiable on $\theta = 45^\circ$. This explains why the right angles observed in the median are not preserved but are rounded in the mean. At the origin, the right angle is thus offset by $\frac{1}{2\sqrt{\pi}} + \frac{1}{4}\sqrt{\frac{\pi}{2}} \approx 0.5954$. Away from the origin, the right angles are asymptotically offset by $\frac{1}{\sqrt{\pi}} \approx 0.5642$. The same can be observed in the standard deviation. Away from $\theta = 45^\circ$, the standard deviation is close to unity, which is identical to the standard deviation in the coordinates of the corner (Eq. (6.26)). At the origin, the standard deviation assumes a value of $\sqrt{\frac{5}{4} - \frac{3}{4\pi} - \frac{\pi}{32} - \frac{1}{4\sqrt{2}}} \approx 0.8581$. Away from the origin, the standard deviation asymptotically assumes a value of $\sqrt{1 - \frac{1}{\pi}} \approx 0.8256$.

Unlike the one-dimensional example, where one point fully characterizes an interface, observations in two (and higher) dimensions vary in resolution, ranging from the position of one point on the interface to a full sampling of the interface. In either case, the observations y_k vanish according to the definition of a level set (Chapter 6.1). Eq. (5.26) simplifies to

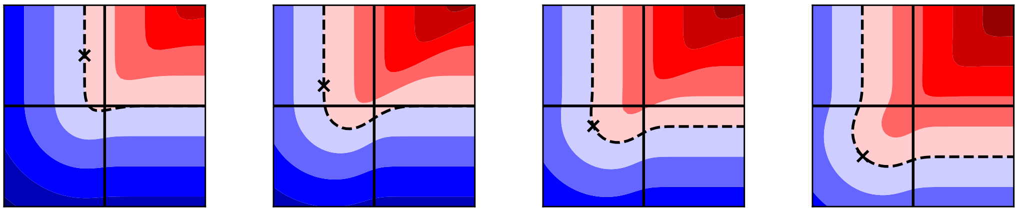
$$\psi^a = \psi^f - \left(HC^f_{\psi\psi}\right)^T \left[C_{\varepsilon\varepsilon} + HC^f_{\psi\psi}H^T\right]^{-1} H\psi^f \quad . \tag{6.39}$$

The measurement operator H is defined by

$$H\psi^f = \begin{pmatrix} \psi^f(r_1) \\ \vdots \\ \psi^f(r_m) \end{pmatrix} \quad , \tag{6.40}$$

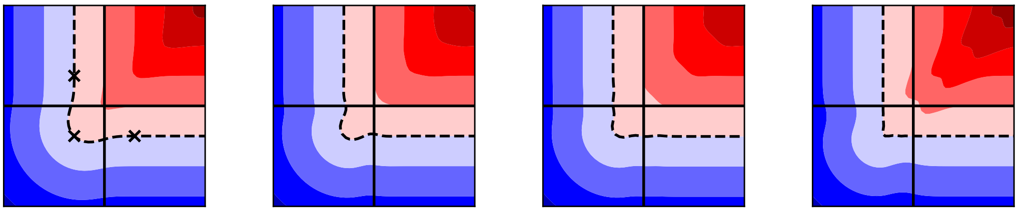
where $r_1, \ldots, r_m$ are the positions of the m observed points on the interface.

In Fig. 6.4, the influence of one observation depending on its location is investigated. It is evident that one observation is insufficient to represent a corner. Nevertheless, analytic data assimilation gives qualitatively meaningful results in that the observation shifts the vertical and horizontal edges proportionally depending on the location of the observation. This is fundamentally different from geometric data assimilation (Fig. 6.1, left), where every point on the interface is required to correspond to an observation. This makes the presented (analytic) level-set data assimilation framework fully front-capturing (rather than front-tracking), both with respect to the description of the motion of the interface and the assimilation of data. As the number of observation points increases, the right angle in the corner is retrieved (Fig. 6.5).



(a) $(-2,5)$ (b) $(-5,2)$ (c) $(-5,-2)$ (d) $(-5,-5)$

Fig. 6.4 Means of filtered generating functions for one observation of the interface at various locations. The observed location is marked by a cross. Its coordinates are specified in each caption. The zero-level set is marked by a dashed line. Without loss of generality, the observation error is set to zero. (a) The observation is located significantly closer to the vertical than the horizontal edge of the predicted generating function (Fig. 6.3, left). Hence, the observation is primarily associated with points on the vertical edge. Data assimilation thus mainly shifts the points on the vertical edge while the vertex and the horizontal edge remain largely unaffected. (b) As the location of the observation moves downwards, the observation gets more strongly associated with the vertex of the corner while the horizontal edge remains largely unaffected. (c) As the location of the observation crosses the x-axis, data assimilation also shifts the points on the horizontal edge. (d) As the location of the observation reaches the diagonal, data assimilation equally affects the vertical and horizontal axes of the corner.



(a) 3 observations. (b) 7 observations. (c) 15 observations. (d) 31 observations.

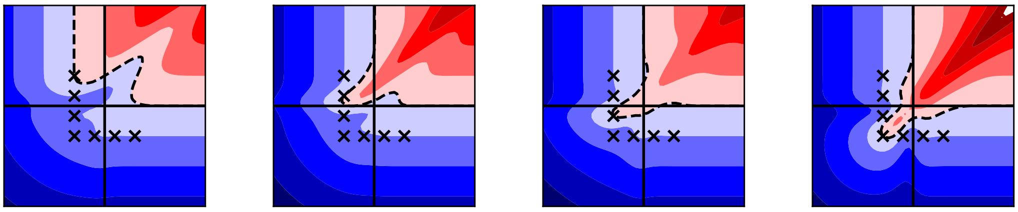
Fig. 6.5 Means of filtered generating functions for multiple observations of the interface at various resolutions. The observed locations are marked by crosses and the zero-level sets by dashed lines. The observations are equidistantly spaced between $(-3,3)$ and $(3,-3)$ via $(-3,-3)$. The number of observations is specified in each caption. For better readability, the observations are not marked in (b)-(d). Without loss of generality, the observation error is set to zero. From (a) to (d), as the number of the observation increases, the observations give a clearer image of the observed corner. Hence, the zero-level set of the filtered generating function more closely resembles a corner.

It is worth pointing out the probabilistic nature of the presented level-set data assimilation framework. Similar to soft clustering [99, Chapter 20], the points on the predicted interface are not associated with distinct observations but with all observations to varying degrees. If we consider ψ^a to be a function of ψ^f, $C^f_{\psi\psi}$, y_k, $C_{\varepsilon\varepsilon}$ and H (Eq. (5.26)), the sensitivity to data is given by

$$\frac{\partial \psi^a}{\partial y_k} = \left(HC^f_{\psi\psi}\right)^T \left[C_{\varepsilon\varepsilon} + HC^f_{\psi\psi}H^T\right]^{-1} \quad , \tag{6.41}$$

where the columns represent the sensitivities to individual observations. The shift in the zero-level set (Fig. 6.5) is the superposition of the contributions due to the individual observations. For example, while the points on the vertical edge far away from the vertex are mainly affected by the uppermost observation (Fig. 6.6a), the other observations have a vanishing effect on them as well (Fig. 6.6b-d). Nevertheless, the shape deformations are more localized compared to Fig. 6.4. While Eq. (6.41) does not formally depend on the observations y_k, it includes information about the number and the locations of the observations signified by the measurement operator H. Note that each observation of interest leaves the values of the generating function at the other observation locations intact.

In summary, analytic data assimilation is applied to a prototypical two-dimensional shape, a corner. Within the level-set data assimilation framework, the differences between distance fields, predicted generating functions and filtered generating functions are illustrated. In particular, the effects of the number of observation points and their locations are studied. In the next chapter, our level-set data assimilation framework is applied to a more realistic



(a) $(-3,3)$ (b) $(-3,1)$ (c) $(-3,-1)$ (d) $(-3,-3)$

Fig. 6.6 Means of filtered generating functions due to individual observations. The observed locations are marked by crosses and the zero-level sets by dashed lines. The observations are equidistantly spaced between $(-3,3)$ and $(3,-3)$ via $(-3,-3)$. The location of the observation of interest is specified in each caption. Without loss of generality, the observation error is set to zero.

experiment. In general, the predicted generating function is taken as a composition of straight sections, which exhibit the behavior discussed for the one-dimensional test case, and strongly curved sections insufficiently resolved by the observations, which exhibit the behavior discussed for the two-dimensional test case. The theoretical insights from this section are referenced throughout the next chapter.

Chapter 7

Twin experiment

The so-called G-equation model is a reduced-order model to study the flame dynamics which lead to heat release rate perturbations [53, 39]. The premixed flame is modeled as an interface captured by a level set. The velocity of a point in the level set is the sum of the flame speed, which is normal to the interface and points towards the unburnt gas, and the underlying flow field. The underlying flow field includes both hydrodynamic and acoustic contributions. The G-equation model is used to compute flame transfer functions (FTF) or flame describing functions (FDF), which give the heat release perturbation to a given velocity perturbation. Coupled with linear acoustics models, the G-equation model has been very successful in qualitatively characterizing the linear and nonlinear dynamics of self-excited thermoacoustic oscillations [57, 62, 179, 56].

We demonstrate our level-set data assimilation framework by performing twin experiments for the G-equation model applied to a ducted premixed Bunsen flame under acoustic forcing. In our twin experiments, both the model predictions and the observations come from G-equation simulations, but with different sets of parameters [180, Chapter 9]. This leads to uncertainties in the parameters and, as a result, in the states. In the absence of uncertainties in the reduced-order model, the twin experiment is an important benchmark in the quantitative assessment of a data assimilation framework because it becomes possible to compare filtered and smoothed solutions to a reference solution at all times [180, Chapter 8].

Relevant publications:

- Yu/Juniper/Magri: *Combined state and parameter estimation in level-set methods*. Journal of Computational Physics, 2019.

7.1 Set-up of the reference simulation

The G-equation is given by [14, Chapter 2]

$$\frac{\partial G}{\partial t} + [u(r,t) - s_{\mathrm{L}} n] \cdot \nabla G = 0 \quad , \tag{7.1}$$

where the underlying flow field is denoted by $u(r,t)$ and the laminar flame speed by s_{L}. For $|\nabla G| = 1$, the G-equation is formally equivalent to the Hamilton-Jacobi equation (Eq. 6.6). The underlying flow field $u(r,t)$ is a superposition of the steady, inviscid flow field U and the time-dependent perturbation velocity field $u'(r,t)$. In the reduced-order model of $u'(r,t)$, the perturbation has an amplitude εU, and travels downstream at the phase speed U/K, where K and ε are model parameters [181]. In the axisymmetric case, the cylindrical components of the perturbation velocity field are given by

$$u'(r,t) = \begin{pmatrix} u'_\rho \\ u'_\theta \\ u'_z \end{pmatrix} = \begin{pmatrix} \frac{1}{2} k \rho \varepsilon U \sin(kz - \omega t) \\ 0 \\ \varepsilon U \cos(kz - \omega t) \end{pmatrix} \quad , \tag{7.2}$$

where the wavenumber k and the angular frequency ω satisfy the dispersion relation $\omega(k) = Uk/K$. The perturbation velocity field $u'(r,t)$ is, mathematically speaking, divergence-free, and, physically speaking, satisfies the continuity equation.

In Fig. 7.1, a sketch of the ducted premixed flame is shown. The ducted premixed flame is forced at an angular frequency of 277.78 rad/s. Table 7.1 shows the measurements of the burner and the tube. The G-equation is numerically solved using the narrow-band level-set method with distance reinitialization [172, 177]: The computational domain is discretized using a weighted essentially non-oscillatory (WENO) scheme in space and a total-variation diminishing (TVD) version of the Runge-Kutta scheme in time, which give up to fifth-order accuracy in space and third-order accuracy in time [182–184]. At the base of the flame, a rotating boundary condition is used [61]. The G-equation solver has been verified in a number of studies [55, 57, 62, 52, 56]. These studies showed that the G-equation reproduces the dynamics of premixed flames only qualitatively, not quantitatively. With our level-set data assimilation framework, significant improvements on quantitative predictions are achieved. In axial and radial coordinates, a uniform 401×401 Cartesian grid is used, which corresponds to $30\,\mu\mathrm{m} \times 180\,\mu\mathrm{m}$ grid cells. An advection Courant-Friedrich-Lewy (CFL) number of 0.02 is chosen, which corresponds to a timestep of $3.6\,\mu\mathrm{s}$ or approximately 6,283 timesteps over one period.

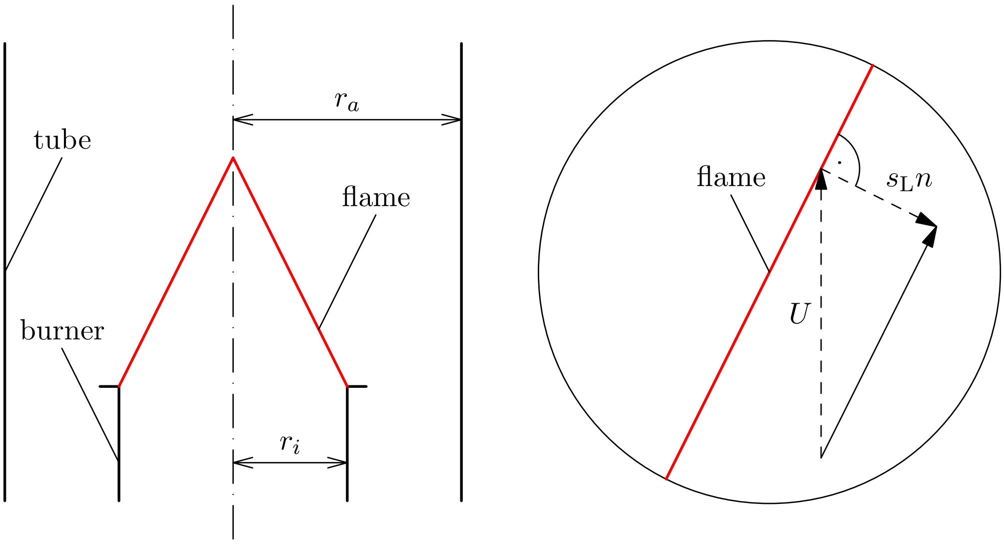


Fig. 7.1 Sketch of the ducted premixed flame without perturbations (left). The superposition of the underlying flow field and the propagation of the premixed flame into the unburnt gas gives velocity vectors, which are locally tangential to the flame surface (right).

Table 7.1 G-equation parameters for the reference simulation.

parameter	parameter value	
r_i	6 mm	burner radius
r_a	12 mm	tube radius
U	1.0 m/s	base-flow speed
s_{L}	0.164 m/s	flame speed
ω	277.78 rad/s	angular frequency of forcing
K	1.0	model parameter
ε	0.1	model parameter

Snapshots of the reference solution to the G-equation for the ducted premixed flame are shown in Fig. 7.2. The flame is attached to the burner lip, where the perturbations form and travel to the tip. When the perturbations are sufficiently large, a fuel-air pocket pinches off. In our reduced-order model, the perturbations are mainly governed by two non-dimensional parameters K and ε, which govern phase speed and amplitude respectively [181]. In practice, neither parameter is accurately known a priori, which is a major source of uncertainty.

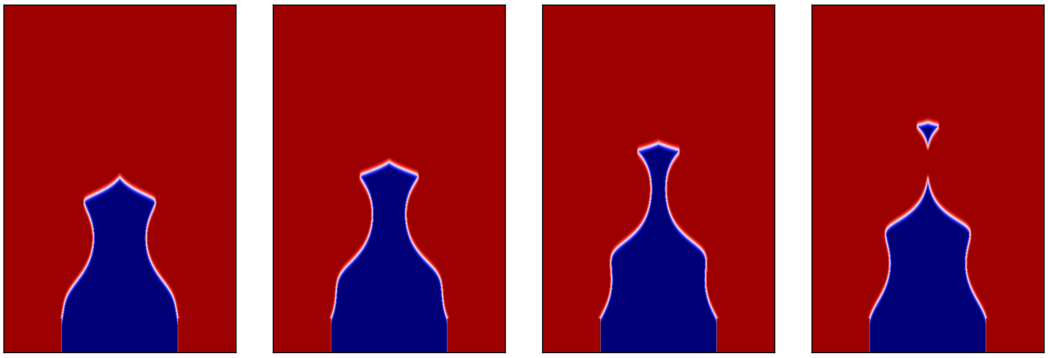


Fig. 7.2 Snapshots of G-equation simulation over one period of harmonic forcing. The fuel-air mixture leaves the burner at the bottom of each frame. The infinitely thin flame surface separates the burnt gas (red) from the unburnt gas (blue). Each snapshot from left to right is distanced by a quarter of the forcing period.

7.2 Combined state and parameter estimation

Three variations of the twin experiment are performed:

Monte-Carlo simulation. An ensemble of 32 G-equation simulations is performed. Each simulation has a different set of parameters K and ε. They are independently sampled from a normal distribution with a standard deviation of 20 %.

State estimation. In addition to the procedure described for the Monte-Carlo simulation, the ensemble Kalman filter (Theorem 5.4) is applied every 1,000 timesteps to update the discretized generating function G while leaving each set of parameters K and ε unaltered. The observations are extracted from the reference solution (Chapter 7.1): The measurement operator H is an indicator matrix which identifies all grid points in the reference solution adjacent to its zero-level set. All aforementioned grid points are treated as observations of the flame surface. As such, all entries in the observation vector y_k are set to 0. The observation error σ_ε is set to 60 μm, which corresponds to the width of two grid cells.

Combined state and parameter estimation. In addition to the procedure described for the state estimation, the discretized generating function G is augmented by appending the parameters K and ε to the state vector (Eq. (5.35), (5.36)). Thus, both the state and the parameters are updated whenever data in the form of observations from the reference solution is assimilated.

In each variation of the twin experiment, the k-th entry in the state vector ψ is marginally distributed according to

$$\psi[k] \sim \mathcal{N}\left(\overline{\psi}[k], C_{\psi\psi}[k,k]\right) \quad . \tag{7.3}$$

The mean $\overline{\psi}[k]$ and the variance $C_{\psi\psi}[k,k]$ are computed from Eq. (5.29). Explicitly, the likelihood for the flame surface to be found at the location of the k-th entry is given by

$$p[k] = \frac{1}{\sqrt{2\pi C_{\psi\psi}[k,k]}} \exp\left(-\frac{\overline{\psi}[k]^2}{2C_{\psi\psi}[k,k]}\right) \quad . \tag{7.4}$$

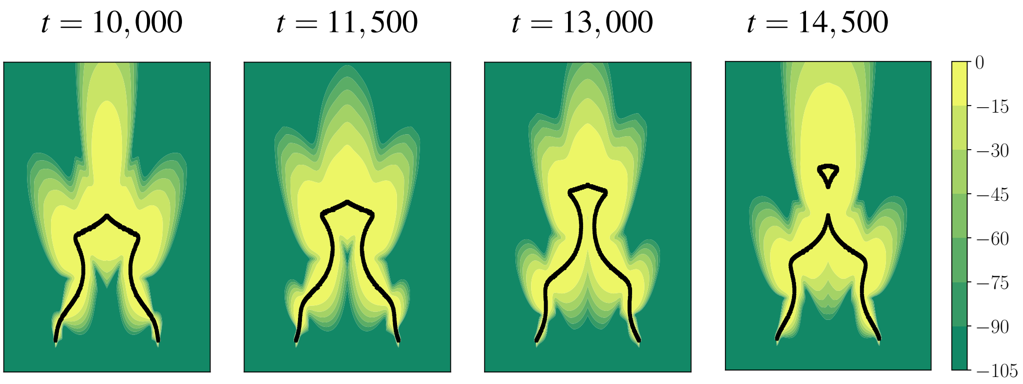
Alternatively, the logarithm of the normalized likelihood is given by [102, Chapter 2]

$$\log\left(\frac{p[k]}{p_0[k]}\right) = -\frac{\overline{\psi}[k]^2}{2C_{\psi\psi}[k,k]} \quad . \tag{7.5}$$

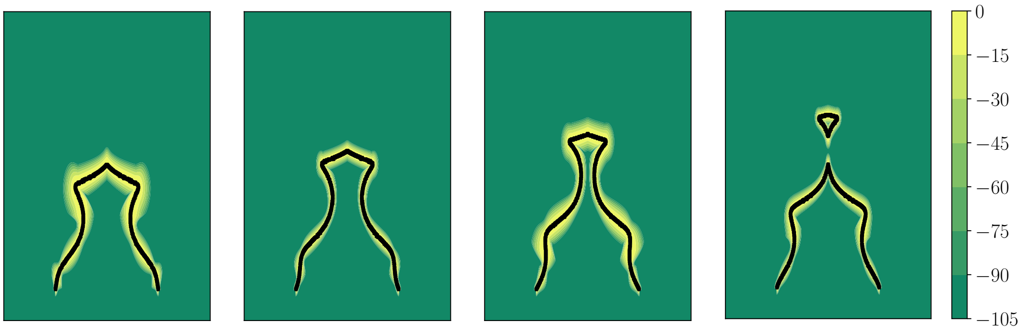
In Fig. 7.3, the logarithm of the normalized likelihood is shown for the three variations of the twin experiment. Its zero-level set gives the maximum-likelihood location of the flame surface. The more negative the value at a location is, the less likely the flame surface is to be found there. In Fig. 7.3a, the logarithm of the normalized likelihood is shown for the Monto-Carlo simulation. As the perturbation travels from the base of the flame to the tip, the high-likelihood region for the location of the flame surface spreads out. The high-likelihood region is largest when fuel-air pockets pinch off, which represents maximal uncertainty. In Fig. 7.3b, the logarithm of the normalized likelihood is shown for the twin experiment with state estimation. A qualitative comparison to Fig. 7.3a shows that the high-likelihood region for the location of the flame surface is significantly tighter. While the high-likelihood region still grows as the perturbation travels, the regular assimilation of data suppresses it to the vicinity of the observed flame surface. In Fig. 7.3c, the logarithm of the normalized likelihood is shown for the twin experiment with combined state and parameter estimation. After few data assimilation cycles, the state-augmented ensemble Kalman filter has learnt the parameters K and ε almost exactly. Knowledge of the parameters enables highly precise predictions of the location of the flame surface, even during pinch-off events. Unlike in Fig. 7.3a and b, no growth in the high-likelihood region is qualitatively discernable.

A global, more quantitative measure for the uncertainty in the location of the flame surface is the root mean square (RMS) error, which is defined as the square-root of the trace of the covariance matrix of the ensemble (Eq. (5.29)):

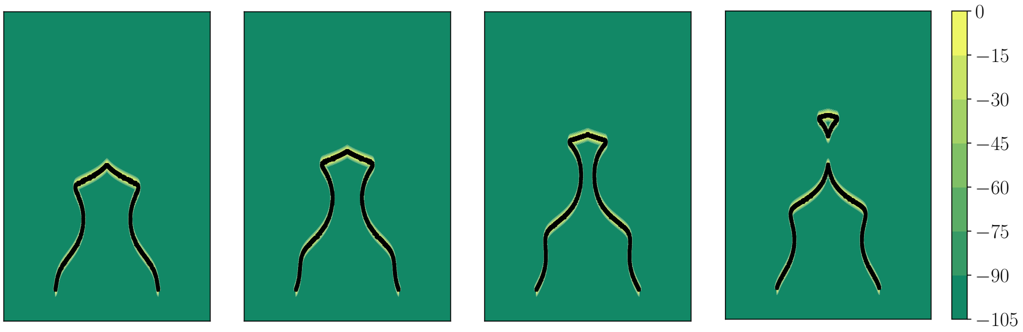
$$\text{RMS error} = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}\left(\psi_j - \overline{\psi}\right)^T\left(\psi_j - \overline{\psi}\right)} \quad . \tag{7.6}$$



(a) Monte-Carlo simulation.



(b) State estimation.



(c) Combined state and parameter estimation.

Fig. 7.3 Snapshots of the logarithm of the normalized likelihood (Eq. (7.5)) over one period for the three variations of the twin experiment respectively. The observations (black) are extracted from a reference solution to the G-equation (Fig. 7.2). High-likelihood (yellow) and low-likelihood (green) regions are shown. Each snapshot from left to right is distanced by a quarter of the forcing period.

In Fig. 7.4, the RMS error is plotted over time for the three variations of the twin experiment from $t = 0$ for eight cycles. The RMS error is initially zero in all three twin experiments because all members in each ensemble share the same initial condition. In the Monte-Carlo simulation (Fig. 7.4, blue line), the RMS error subsequently grows until it reaches a high-uncertainty plateau. Momentary spikes in the uncertainty occur approximately every 6,000 timesteps, and coincide with the pinch-off events observed in the reference solution (Fig. 7.2). With state estimation (Fig. 7.4, orange line), the assimilation of data regularly suppresses the uncertainty as qualitatively observed in Fig. 7.3b. As a result, the predictions of the location of the flame surface significantly improve. With combined state and parameter estimation (Fig. 7.4, green line), the predictions further improve. Beginning with the first instance of data assimilation at timestep 1,000, the uncertainty steadily decreases until it reaches a low-uncertainty plateau after approximately 10,000 timesteps. At this point, the state-augmented ensemble Kalman filter has optimally calibrated the parameters, and the state has reached a statistically stationary state[1]. The assessment based on the RMS error is supported by an analysis of the bias, which is plotted in Fig. 7.5 and defined as the difference between the mean prediction and the observations (Eq. (6.39)):

$$\text{bias} = \sqrt{(H\overline{\psi})^T\,(H\overline{\psi})} \quad . \tag{7.7}$$

Note that both the unknown parameters and the observation noise contribute to the bias. Hence, the bias is particularly useful as a quantitative measure in twin experiments, where the reference solution and thus the observation error are exactly known.

To gain insight into the effect of combined state and parameter estimation, it is instructive to analyze the probability distributions in the parameters. In Fig. 7.6, the marginal distributions in the normalized residuals of K and ε are shown, signified by their means and three-sigma confidence levels at every timestep. By timestep 15,000, no improvement is visible in either the estimates of the means or the uncertainties, which matches the evolution of the RMS error (Fig. 7.4). While both K and ε quickly converge to the values used in the reference simulation (Table 7.1), the low uncertainty in K compared to ε reflects the physical significance of this parameter: The nonlinear dynamics of a premixed flame are highly sensitive to the timings of pinch-off events [29], which in turn depend on the phase speed at which perturbations travel along with the flame. In our reduced-order model based on the G-equation, the phase speed is regulated by the parameter K, which makes it highly observable even in the presence of observation noise. Similar to sensitivity analysis [32, 76, 70],

---

[1] Note that the state only reaches statistical stationarity because observation noise and numerical errors are always present.

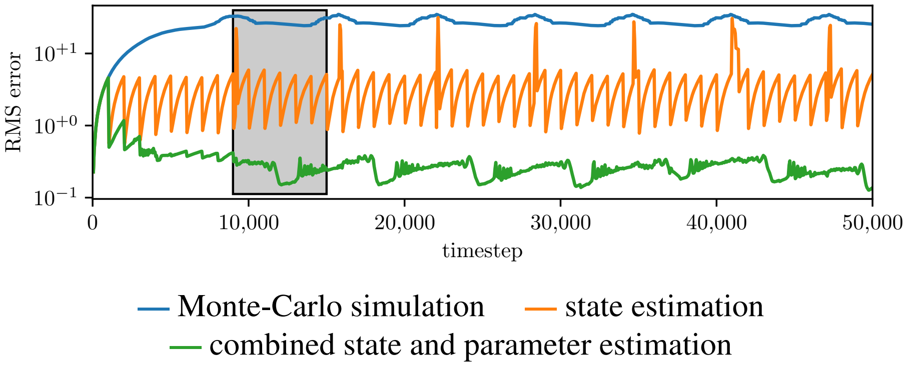


Fig. 7.4 Logarithmic plot of root mean square (RMS) error over 50,000 timesteps for the Monte-Carlo simulation (blue line) as well as the twin experiments with state estimation (orange line) and combined state and parameter estimation (green line). 6,283 timesteps correspond to one period of harmonic forcing. Data is assimilated every 1,000 timesteps. The black rectangle marks the time window depicted in Fig. 7.3.

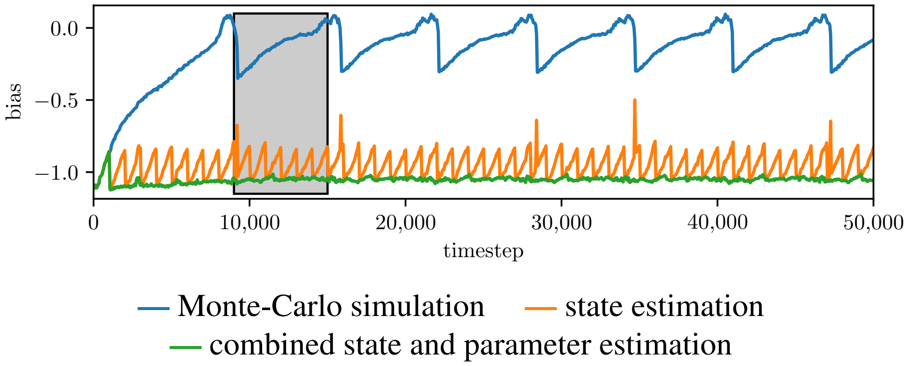


Fig. 7.5 Logarithmic plot of bias over 50,000 timesteps for the Monte-Carlo simulation (blue line) as well as the twin experiments with state estimation (orange line) and combined state and parameter estimation (green line). 6,283 timesteps correspond to one period of harmonic forcing. Data is assimilated every 1,000 timesteps. The black rectangle marks the time window depicted in Fig. 7.3.

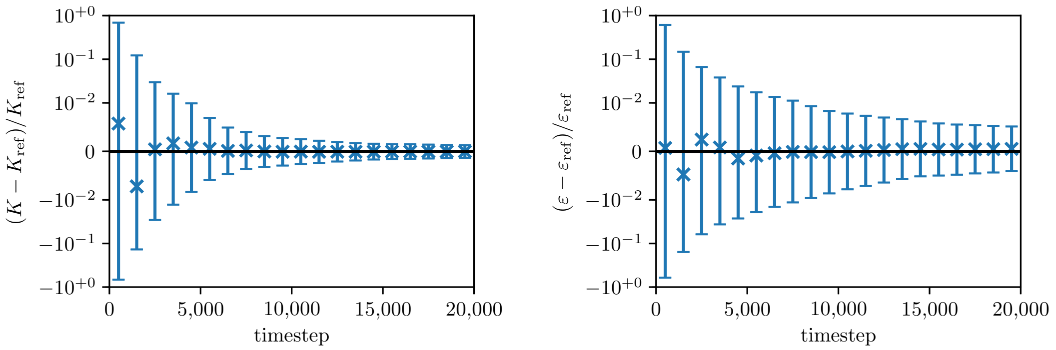


Fig. 7.6 Logarithmic plots of normalized residuals of K (left) and ε (right) over 20,000 timesteps for combined state and parameter estimation; linear scales for normalized likelihoods between $\pm 10^{-2}$. The marginal distributions in K and ε are represented by their means (crosses) and three-sigma confidence levels (error bars).

this physical insight is gained from the inspection of the uncertainties, which here exceed the residuals between the means and the set of reference values by several orders of magnitude.

Finally, combined state and parameter estimation using the ensemble Kalman smoother is performed (Theorem 5.6). In Fig. 7.6, the marginal distributions in the normalized residuals of K and ε are shown for both the ensemble Kalman filter and smoother. The effect of the ensemble Kalman filter has already been described in Fig. 7.6. In the forward-backward implementation, the ensemble Kalman smoother takes the last solution of the ensemble Kalman filter and works itself backwards in time (Theorem 5.2). In terms of information theory, the ensemble Kalman smoother takes at every timestep k all observations, from the past, present and future, into account compared to the ensemble Kalman filter, which only takes the past and the present into account. Statistically speaking, the smoothed distributions are more strictly conditioned than the filtered distributions, over $y_{1:K}$ compared to $y_{1:k}$ (Table 5.3). This surplus in information is evident in the form of lower error bars not just towards the end of the combined state and parameter estimation, but extending all the way to the very first instances of data assimilation. The evaluation of Eq. 5.6 does not rely on the solution of the governing equations, but instead relies on the predicted and filtered distributions in storage. The ensemble Kalman smoother is thus a computationally inexpensive tool to quantify the uncertainties especially at the beginning of the simulation, where little retrospective data is available. In analogy to direct-adjoint looping [50, 185], combined state and parameter estimation based on filtering and smoothing can be used to obtain otherwise inevitably ad-hoc initial conditions and parameters.

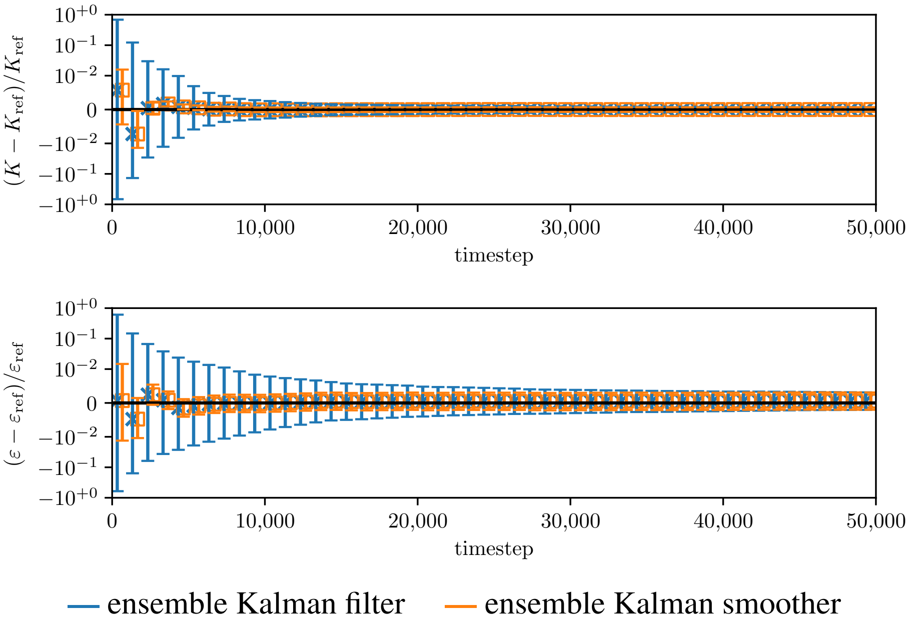


Fig. 7.7 Logarithmic plots of normalized residuals of K (top) and ε (bottom) over 50,000 timesteps for combined state and parameter estimation using the ensemble Kalman filter (blue cross) and the ensemble Kalman smoother (orange square); linear scales for normalized likelihoods between $\pm 10^{-2}$.

7.3 Hyperparametric study

In a first hyperparametric study, the observation error σ_ε is varied to investigate the effect of observation noise on the level-set data assimilation framework. In Chapter 7.2, it has been established that combined state and parameter estimation eventually leads to a statistically stationary state. In Fig. 7.8, the same RMS error is plotted over time for different observation errors σ_ε. This reveals that the sustained low uncertainties are epistemic in nature [67]: Every reduction in the observation error σ_ε by one order of magnitude reduces the RMS error by one order of magnitude. It is clear that the observation error poses a lower epistemic bound on how low the uncertainty can be reduced by combined state and parameter estimation.

In Fig. 7.9, the marginal distributions in the normalized residuals of K and ε are plotted over time for the different observation errors. The means of the residuals quickly vanish for all three observation errors σ_ε. For the lower observation errors, the error bars increasingly fail to contain the zero residual. Considering that three sigmas correspond to a confidence of 99.7 % under the assumption of normal distributions (Chapter 5.2), the frequency at which the error bars fail leads to the conclusion that the level-set data assimilation framework underpredicts the uncertainties for low observation errors. This is in agreement with the theoretical analysis of the level-set data assimilation framework for the two-dimensional

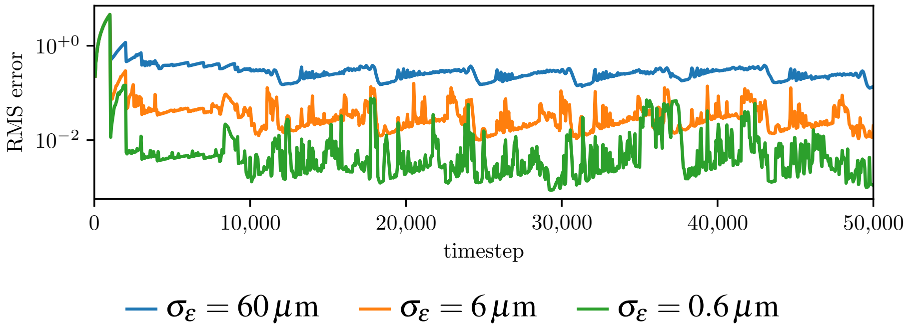


Fig. 7.8 Logarithmic plot of root mean square (RMS) error over 50,000 timesteps for combined state and parameter estimation with different observation errors ($\sigma_\varepsilon = 60\,\mu$m blue line, $\sigma_\varepsilon = 6\,\mu$m orange line, $\sigma_\varepsilon = 0.6\,\mu$m green line).

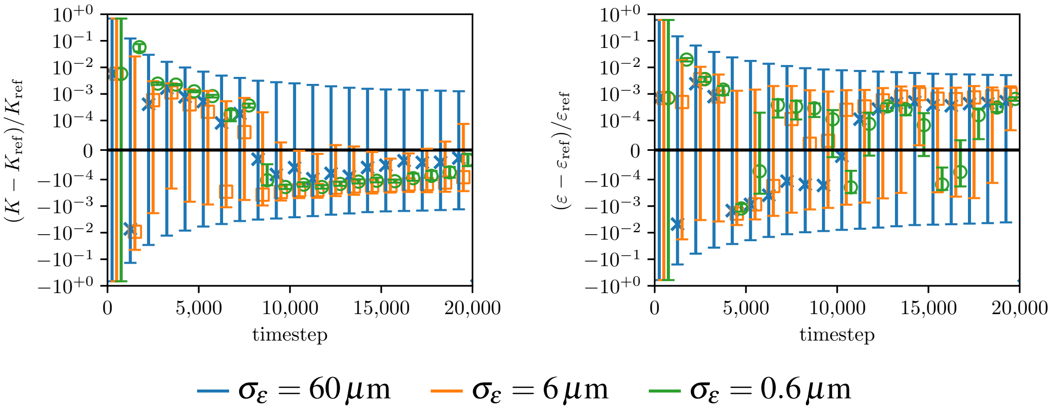


Fig. 7.9 Logarithmic plots of normalized residuals of K (left) and ε (right) over 20,000 timesteps for combined state and parameter estimation with different observation errors ($\sigma_\varepsilon = 60\,\mu$m blue cross, $\sigma_\varepsilon = 6\,\mu$m orange square, $\sigma_\varepsilon = 0.6\,\mu$m green circle); linear scales for normalized likelihoods between $\pm 10^{-4}$.

test case (Chapter 6.4): Although the individual simulations predict the formation of sharp cusps as the perturbations travel on the respective flame surfaces, the cusps are rounded in the mean of the ensemble (Fig. 6.3, left). Furthermore, the observable sharpness of a cusp depends on the local number of observations of points on the flame surface. A decrease in the observation error σ_ε without an increase in the number of observations amounts to a relative decrease in the resolution of the observed cusp (Fig. 6.5). Therefore, the local values of the eikonal fields of the individual simulations significantly deviate from normal distributions (Fig. 6.3, right), which leads to incorrect distributions and uncertainties.

In a second hyperparametric study, subsamples of the observation points are used to investigate the effect of the number of observations on the level-set data asssimilation framework. Furthermore, this serves to illustrate the superiority of our analytic view over

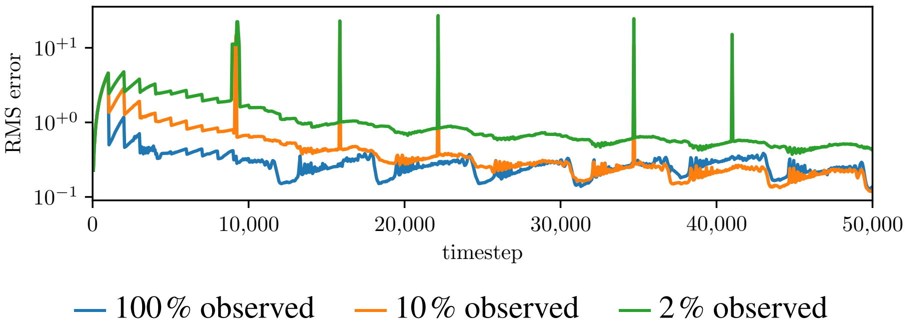


Fig. 7.10 Logarithmic plot of root mean square (RMS) error over 50,000 timesteps for combined state and parameter estimation with different subsamples of observation points (100 % blue line, 10 % orange line, 2 % green line).

the geometric view in data assimilation (Chapter 6.1): In the geometric view, the observed interface must be parametric or at least as highly resolved as the grid used in the level-set method to allow the optimal interpolation of a sufficiently large number of points on the predicted interfaces. This restriction does not exist in the analytic view. Hence, we study the effect of dispersed observation points on the level-set data assimilation framework.

In Fig. 7.10, the RMS error is plotted over time for different randomly sampled subsets of the observation points. In Chapter 7.2, it has been established that combined state and parameter estimation eventually leads to a statistically stationary state after approximately 10,000 timesteps. With 10 % of the observation points, it takes more than 30,000 timesteps to reach a comparable low-uncertainty plateau. With 2 % of the observation points, the RMS error is still decreasing after 50,000 timesteps while repeatedly failing to correctly predict the pinch-off events for the whole ensemble.

In Fig. 7.11, the marginal distributions in the normalized residuals of K and ε are plotted over time for the different subsamples. While the means of the residuals quickly vanish for all three subsamples, the confidence levels for the incomplete subsamples improve less rapidly than for the original sample of observation points. This is in agreement with the theoretical analysis of the level-set data assimilation framework for the two-dimensional test case (Chapter 6.4): In particular for the 2 %-subsample, certain features in the observed interfaces remain underresolved at the individual timesteps (Fig. 6.4). Nevertheless, the ensemble at a timestep k representing the filtered probability distribution $P(x_k, \theta \mid y_{1:k}, f)$ (Table 5.3) reflects the knowledge of the observations at all previous timesteps due to the Markov property of the reduced-order model (Chapter 5.1). In theory, the lack of resolution in the observations is compensated by more instances of data assimilation to accumulate the same amount of information. The result is a similar statistically stationary state reached at a later time.

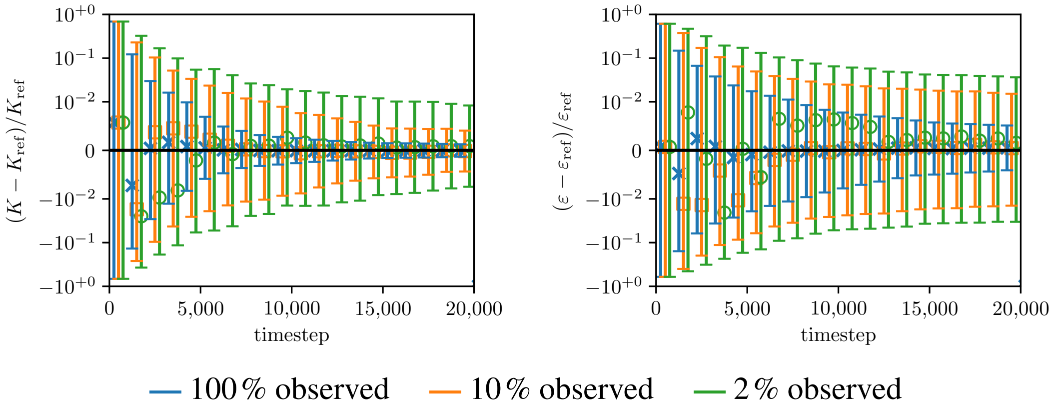


Fig. 7.11 Logarithmic plots of normalized residuals of K (left) and ε (right) over 20,000 timesteps for combined state and parameter estimation with different subsamples of observation points (100 % blue cross, 10 % orange square, 2 % green circle); linear scales for normalized likelihoods between $\pm 10^{-2}$.

Chapter 8

Simulation of a ducted premixed flame

In the previous chapter, we performed a so-called twin experiment, where the assimilated data is generated by the same reduced-order model that is used for predictions. In this chapter, we take a high-fidelity simulation, and assimilate data from this numerical experiment. The high-fidelity simulation involves a number of physical effects not covered by the reduced-order G-equation model, e.g. finite-rate kinetics and heat transfer, which affect the flame speed, and heat loss to the burner wall, which affects flame attachment, flash-back and lift-off. Although the high-fidelity simulation is not yet a real experiment, it comes significantly closer to reality than the twin experiment, and poses a more realistic challenge to the G-equation and our level-set data assimilation framework. The results were produced in collaboration with PROF. MATTHIAS IHME'S research group at Stanford university as part of the CENTER FOR TURBULENCE RESEARCH summer program in 2018. The high-fidelity simulations were performed by DR. THOMAS JARAVEL.

Relevant publications:

- Yu et al.: *Physics-informed data-driven prediction of premixed flame dynamics with data assimilation*. Center for Turbulence Research Summer Program, 2018.
- Yu et al.: *Data Assimilation and Optimal Calibration in Nonlinear Models of Flame Dynamics*. ASME Turbo Expo, 2019.
- Yu et al.: *Data Assimilation and Optimal Calibration in Nonlinear Models of Flame Dynamics*. Journal of Engineering for Gas Turbines and Power, 2019.

8.1 Set-up of the numerical experiment

In the form of a numerical experiment, the assimilated data is extracted from a direct numerical simulation (DNS) of a ducted premixed flame. The simulation is performed with the finite-volume solver CharLES^X (Cascade Technologies), solving the compressible reacting Navier-Stokes equations. For convective fluxes, a hybrid scheme is applied, consisting of a fourth-order central scheme and a second-order essentially non-oscillatory scheme [186], which minimizes the numerical dissipation [187].

The computational domain is three-dimensional (Fig. 8.1). Because of the axisymmetry of the configuration, only a quarter of the full geometry with two symmetry boundary planes is considered. The domain is 0.12 m long, and extends two diameters upstream into the burner to allow for potential flash back of the flame. It also includes a coflow of fresh air. This models the flow of fresh air that would arise in a vertical duct due to buoyancy effects. The stretched grid comprises 500,000 hexahedral elements with characteristic element sizes ranging from 0.15 mm in the flame region to 3 mm at the outlet.

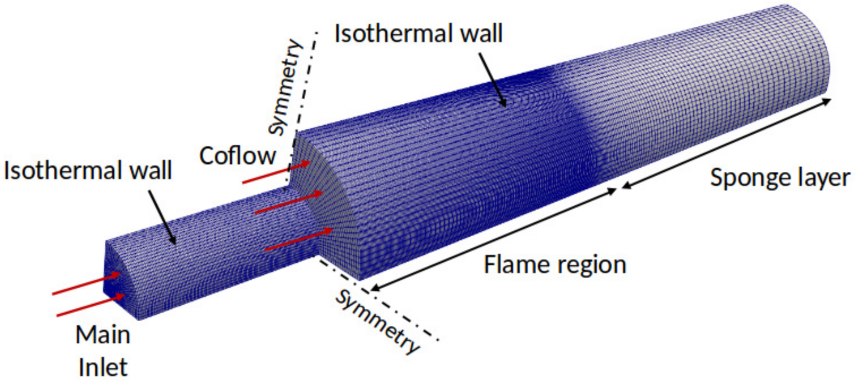


Fig. 8.1 Computational domain, computational grid and boundary conditions of direct numerical simulation (DNS). This schematic was provided by DR. THOMAS JARAVEL.

At the main inlet, a parabolic velocity profile is imposed in space along with temperature and species composition corresponding to the ethylene-methane-air mixture given in Table 8.1. For pulsed cases, a sinusoidal velocity amplitude is imposed in time. The coflow speed is precribed based on a one-dimensional analysis, inspired by the experimental configuration of a vertical Rijke tube [119]. The choice of the coflow speed does not affect the flame dynamics in a qualitative manner. Lateral and injection walls are modeled with isothermal no-slip boundary conditions with an imposed temperature $T = 300\,\mathrm{K}$ to correctly reproduce the flame stabilization. At the outlet of the domain, a constant pressure boundary is used to impose atmospheric pressure conditions, along with a sponge layer region to suppress reflected acoustic waves.

Table 8.1 Base-flow (top) and forcing (bottom) parameters for high-fidelity reacting-flow simulation at burner and coflow inlet.

Parameter	Burner inlet	Coflow inlet
Diameter [mm]	12	25
Mass-flow rate [g/s]	0.199	0.784
Temperature [K]	300	300
Mass fraction methane [-]	0.0237	-
Mass fraction ethylene [-]	0.0673	-
Mass fraction air [-]	0.909	1.0
Forcing amplitude [m/s]	0.141	-
Forcing frequency [Hz]	200	-

The reduced chemical mechanism includes 15 species and 5 quasi-steady state species, which correctly predicts laminar flame speed and auto-ignition delay for the target operating conditions [188]. In the solver, the chemical source terms are evaluated using finite-rate kinetics with a semi-implicit Rosenbrock-Krylov scheme [189, 190]. From one-dimensional laminar computations, it was verified a priori that the grid resolution in the flame region (0.15 mm) is sufficient to correctly resolve the thermal thickness of the flame (0.4 mm).

In Fig. 8.2, snapshots of the DNS are shown. The flame is attached to the burner lip on the edge of flashing back. The DNS displays the same qualitative behavior as the G-equation as perturbations are observed to travel from the base to the tip of the flame. For the given simulation set-up, no pinched-off fuel-air pockets are observed in the DNS.

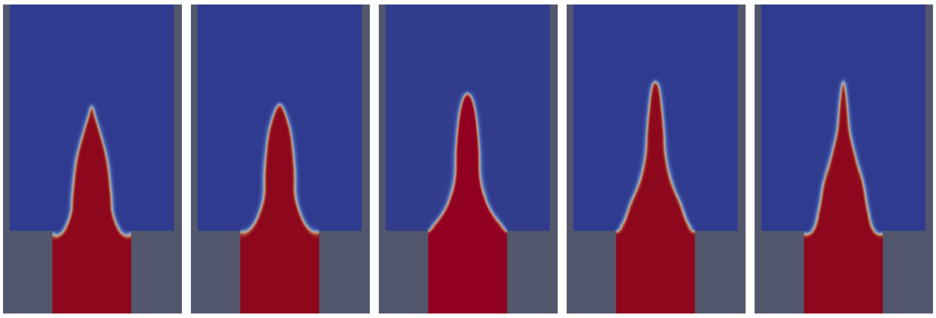


Fig. 8.2 Ethylene mass fractions in direct numerical simulation (DNS) over one period. The fuel-air mixture leaves the burner at the bottom. The flame surface separates the burnt (blue) from the unburnt (red) gas.

8.2 *G*-equation model

The kinematic description of a surface is given by its laws of motion [14, Chapter 2]:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = u - s_{\mathrm{L}} n \quad , \tag{8.1}$$

where x is the position of one point on the surface, and n is the normal vector at this point. The velocity u is the velocity field of the underlying medium, and s_{L} is the speed of the surface relative to the underlying medium. In general, s_{L} depends on geometrical properties of the surface, e.g. on the local curvature via the Markstein length $\mathscr{M}$. The laws of motion are consistent with the following Hamilton-Jacobi equation (Chapter 6.1):

$$\frac{\partial G}{\partial t} + (u \cdot n - s_{\mathrm{L}}) = 0 \quad , \tag{8.2}$$

$$\text{subject to} \quad G(x(t),t) - G(x(0),0) = 0 \quad , \tag{8.3}$$

$$\text{and} \quad \nabla G \cdot \nabla G - 1 = 0 \quad . \tag{8.4}$$

In combustion, Eq. (8.2) is also known as the *G*-equation [14, Chapter 2]. The normal vector n depends on the computation of partial derivatives in space, which makes Eq. (8.2) a partial differential equation. Eq. (8.3) and (8.4) are constraints, which make the solution to the *G*-equation unique away from the surface. For the upstream velocity forcing of the premixed flame, the radial and axial components of the underlying velocity field are prescribed as follows [181]:

$$u_r = \frac{\pi f K r e_a}{U} \cos\left(2\pi f \left(t - \frac{Kz}{U}\right)\right) \quad , \tag{8.5}$$

$$u_z = U + e_a \sin\left(2\pi f \left(t - \frac{Kz}{U}\right)\right) \quad , \tag{8.6}$$

where the base-flow speed is denoted by U, the frequency by f and the forcing amplitude by e_a. Eq. (8.5) and (8.6) are related to each other through the continuity equation. The non-dimensional parameter K is defined as the ratio between the base-flow speed and the perturbation phase speed. Hence, the perturbation phase speed is given by U/K and the wave number by $2\pi f K/U$. The solution $G(x,t)$ to the Hamilton-Jacobi equation is the so-called generating function [174, Chapter 9]. The solution $x(t)$ to Eq. (8.3) gives the location of the surface at every time t.

There are several consequences to solving the Hamilton-Jacobi equation (Eq. (8.2)-(8.4)) instead of the laws of motion (Eq. (8.1)). Firstly, the laws of motion require a parametrization of the surface [143]. The quality of the parametrization quickly deteriorates when the metric on the surface or its topology changes. The Hamilton-Jacobi equation avoids these issues by embedding the surface as a level set of the generating function, which is defined over the entire domain. Secondly, the Hamilton-Jacobi equation is in theory solved over the entire domain even though surfaces are one-codimensional (i.e. surfaces have one less dimension than the volume they inhabit) [176]. This is more computationally expensive but has the advantage that surfaces can be identified by their generating functions as vectors in state space. The choice of generating functions is highly relevant as they have to live within a sample space suitable for data assimilation.

The *G*-equation is numerically solved using a computationally inexpensive narrow-band level-set method [172]: The computational domain is discretized using a weighted essentially non-oscillatory (WENO) scheme and a total-variation diminishing (TVD) version of the Runge-Kutta scheme [182–184]. This gives third-order accuracy in time and up to fifth-order accuracy in space. Cusps and pinched-off fuel-air pockets are reliably captured. At the base of the flame, a rotating boundary condition is used [61]. The *G*-equation solver has been verified and validated in a number of studies [55, 57, 62, 52, 56]. Distance reinitialization in the narrow band is performed using a continuation method [177]. The generating function is reconstructed from the solution to the *G*-equation using a fast-marching method [173].

In the following, axisymmetric ducted premixed flames are considered. In axial and radial coordinates, a uniform 400 × 400 Cartesian grid is used. Example solutions of the *G*-equation for the ducted premixed flame are shown in Fig. 8.3. The flame is attached to the burner lip, while the perturbations are convected from the base of the flame to the tip. If the perturbations are sufficiently large, a fuel-air pocket pinches off. In our reduced-order model, the perturbations are mainly governed by two non-dimensional parameters [181]: (i) the parameter K, which governs the perturbation phase speed; and (ii) the parameter e_a, which governs the amplitude of the response of the flame surface to acoustic excitation. Both parameters enter the *G*-equation via the underlying velocity field u. Neither parameter is accurately known a priori, which is a major source of uncertainty.

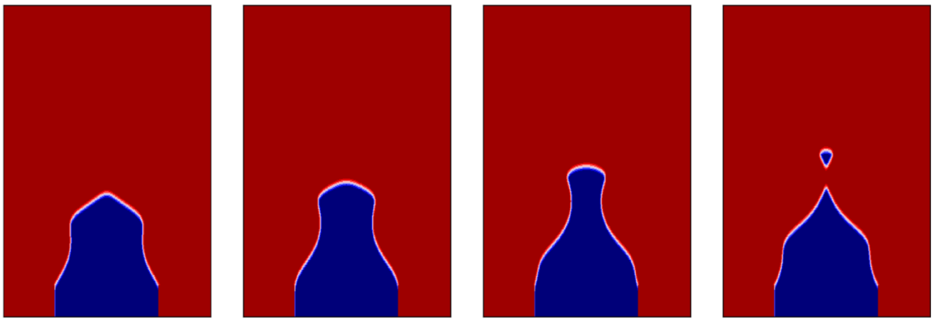


Fig. 8.3 Solutions of G-equation over one period of harmonic forcing of velocity at flame base. The fuel-air mixture leaves the burner at the bottom of each frame. The infinitely thin flame surface separates the burnt (red) from the unburnt (blue) gas.

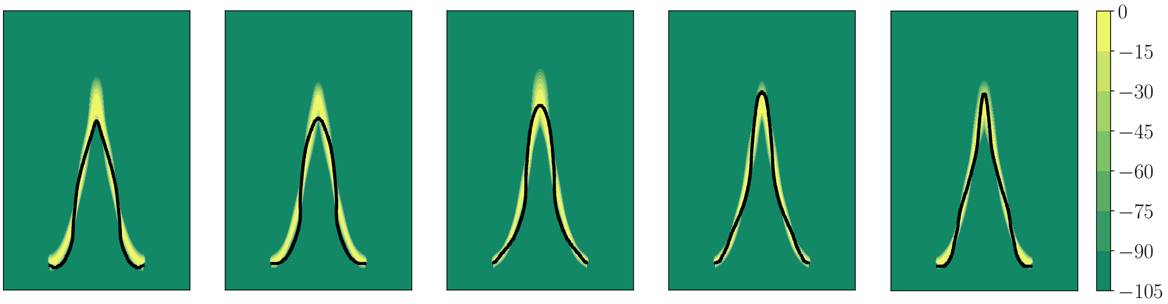
8.3 Combined state and parameter estimation

For the model predictions, an ensemble of twenty-four G-equations is solved. The same initial condition is used for the whole ensemble, but a different set of parameters K and e_a is chosen for each simulation. They are sampled from two independent normal distributions with 20 % standard deviation respectively. The covariance matrix $C_{\varepsilon\varepsilon}$ is a diagonal matrix, i.e. the experimental errors σ_ε are assumed to be statistically independent. The simulation parameters are given in Table 8.2. In Fig. 8.4, the logarithm of the normalized likelihood is shown for the forward problem and the inverse problems, the latter involving either state estimation or combined state and parameter estimation. As in Fig. 7.3, data assimilation significantly improves the ability of the reduced-order G-equation model to capture the motion of the flame surface. In a qualitative comparison, it can be seen that the high-likelihood region from state estimation (Fig. 8.4b) is a better envelope to the flame surface observed in the DNS than the unfiltered stochastic process (Fig. 8.4a). Combined state and parameter estimation (Fig. 8.4c) provides an even narrower envelope, which is testament to the reduced uncertainty. This assessment is quantitatively confirmed by the RMS error shown in Fig. 8.5. While state estimation does not appreciably reduce the RMS error, it provides a dynamic, truly a-posteriori estimate of the uncertainty that is rigorously based on Bayesian statistics. On the contrary, the RMS error of the unfiltered stochastic process only gives an a-priori estimate of the uncertainty based on an educated initial guess for the possible ranges of values for the parameters. In Fig. 8.6, the values of K and e_a with their error bars are plotted over time for combined state and parameter estimation. After a burn-in phase of approximately 4,000 timesteps, the parameters are optimally calibrated to values which are in agreement with the experimental observations. This is reflected in Fig. 8.5 by the low-uncertainty plateau, which is only interrupted by the moments when perturbations reach the tip of the flame and neutralize. The outlier around timestep 3,500, where combined state and parameter estimation performs worse in terms of RMS error than the unfiltered stochastic process and state estimation, is explained by the formation of a pinched-off fuel-air pocket that is not observed in the DNS.

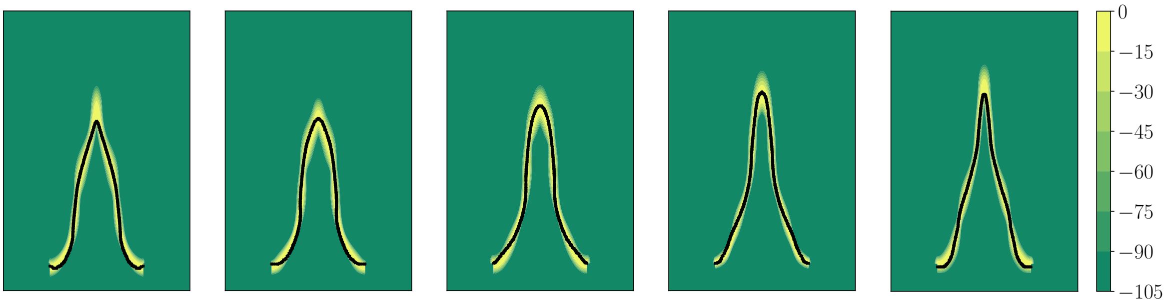
Table 8.2 Parameter values for *G*-equation simulations and assimilation of data from high-fidelity simulation.

Parameter	DNS assimilation
U	$1.41\,\mathrm{m/s}$
s_L	$0.329\,\mathrm{m/s}$
$\mathscr{M}$	$0.02\,\mathrm{m}$
e_a	$0.5\,\mathrm{m/s}$
K	0.214

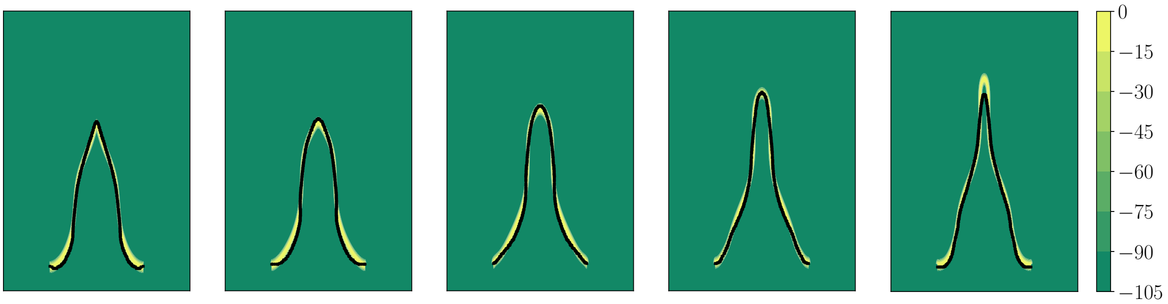
Finally, we build on our understanding of the state and parameter uncertainties to attempt model uncertainty quantification. Unlike optimization-based approaches, Bayesian statistics do not focus on the minimization of one cost functional but revolve around probability distributions. As such, they can be analyzed using histograms. In Fig. 8.7, two histograms, which are representative of the twin experiment (Chapter 7) and the assimilation of DNS data, are respectively shown. In the twin experiment (Fig. 8.7a), most observation points are located in the high-likelihood region. The likelihood of finding the flame surface in the low-likelihood region decays exponentially. When assimilating DNS data (Fig. 8.7b), the histogram has a heavy tail. While the likelihood of finding the flame surface in the high-likelihood region decays quickly, the likelihood of finding the flame surface in the low-likelihood region does not vanish, reminiscent of a power-law distribution. The deviation from an exponentially decaying distribution, a piece of information detected by our Bayesian approach but absent from an optimization-based approach, signifies the presence of model errors [191]. Such a deviation informs us of the inherent limitations of this reduced-order model, even when optimally calibrated. Model errors can only be decreased by adding complexity, i.e., more physics, to the reduced-order model assumptions. Further analysis reveals that experimental observations located in low-likelihood regions are predominantly found at the base of the flame, in particular when the flame is on the edge of flashing back. It is well known that the dynamics at the base of a premixed flame are crucial to the flame dynamics [192]. Although the motion of the flame surface is accurately captured towards the tip of the flame, level-set methods, which do not obey conservation laws, cannot model heat transfer between the flame and the burner lip or the momentum transfer in the wake region behind the burner lip. This shows how model uncertainty, which is traditionally the type of



(a) Monte-Carlo simulation.



(b) State estimation.



(c) Combined state and parameter estimation.

Fig. 8.4 Snapshots of the logarithm of the normalized likelihood over one period for (a) the forward problem and the inverse problems with either (b) state estimation or (c) combined state and parameter estimation, respectively. For the data assimilation, the experimental observations are extracted from a DNS (black line). High-likelihood (yellow) and low-likelihood (green) regions are shown. Each snapshot from left to right is distanced by one fifth of the forcing period.

uncertainty that is the most difficult to quantify, can be detected by Bayesian statistics to inspire more predictive models.

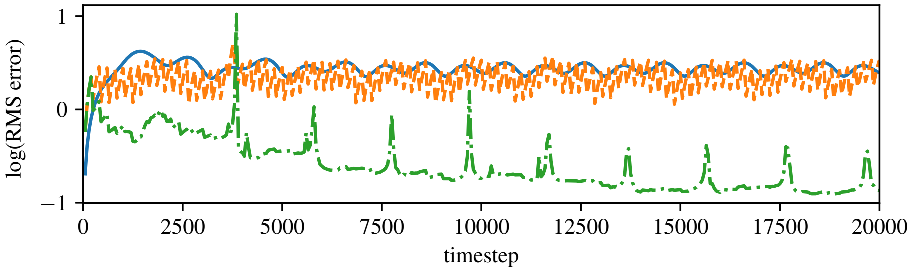


Fig. 8.5 Root mean square (RMS) error over time for the forward problem (blue line) and the inverse problems with either state estimation (orange dash) or combined state and parameter estimation (green dash-dot). One forcing period corresponds to 2,000 timesteps. Data is available every 200 timesteps.

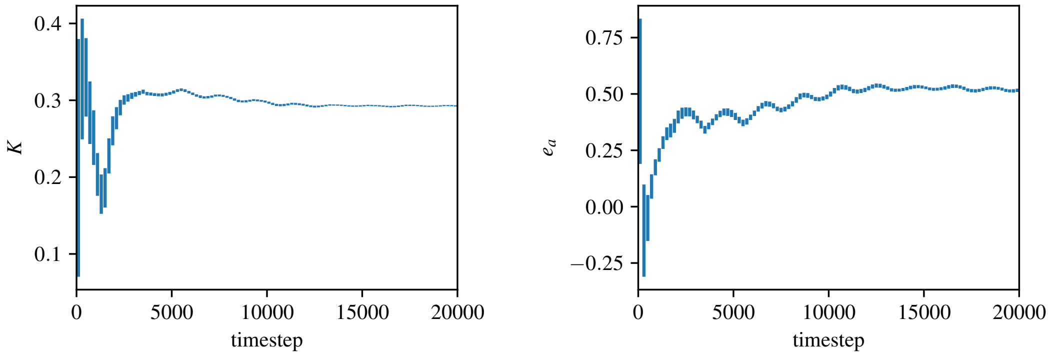
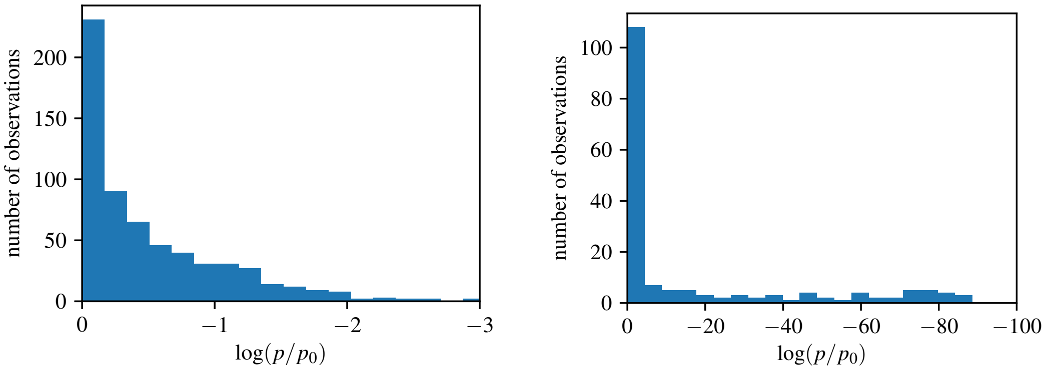


Fig. 8.6 Three-sigma error bars for K (left) and e_a (right) in combined state and parameter estimation. The convergence of the parameters shows that the reduced-order flame model is optimally calibrated.



(a) Twin experiment.

(b) DNS assimilation.

Fig. 8.7 Histograms of experimental observations over logarithm of normalized likelihood for (a) twin experiment and (b) assimilation of DNS data. High-likelihood bins are located on the left of each histogram, low-likelihood bins on the right.

Chapter 9

Experiment of a ducted premixed flame

We continue the analysis of Chapters 7 and 8, and apply our level-set data assimilation framework to experimental images of a ducted premixed flame. Compared to the twin and numerical experiments, this set-up of the ducted premixed flame is significantly more challenging. For example, while the heat loss to the burner wall is not covered by the reduced-order G-equation, we at least know in the numerical experiment that the thermal boundary condition is isothermal. In the current set-up, the heat loss to the burner wall becomes a problem of conjugate heat transfer, for which no analytical solution exists. Furthermore, certain information about the flow field such as the mean speed of the buoyancy-driven coflow air and the velocity profiles in the downstream mixing layer are inaccessible, and have to be inferred. The experiments were set up by PROF. MATTHEW JUNIPER on a rig built by DR. LARRY LI. The acquisition was performed by USHNISH SENGUPTA.

The structure of this chapter follows the procedures in Chapters 7 and 8: After a brief description of the experimental set-up (Chapter 9.1), we derive our reduced-order G-equation model (Chapter 9.2). Firstly, we provide an ansatz for the base flow, which is more complex than for the twin and numerical experiments, and compare the base flow parameters to CANTERA simulations. Secondly, we perform an asymptotic analysis of the flame response in order to anticipate the results of parameter estimation for different forcing frequencies. Finally, we discuss combined state and parameter estimation, and provide qualitative and quantitative visualizations of the outcome (Chapter 9.3).

Statistically rigorous uncertainty quantification is a relatively novel area of research. This chapter ties together all challenges as well as the solutions developed in Part III:

- High-dimensional sample spaces, i.e. high resolution of the flame surface coupled with correlations in space and time. (Chapter 5)

- Construction of an appropriate sample space for comparing level sets (viz. flame interfaces) and quantifying uncertainties. (Chapter 6)
- Imperfections of real data, which hinder the reconstruction of observed flame interfaces. (Chapters 7, 8, 9)

Relevant publications:

- Yu/Juniper/Magri: *A data-driven kinematic model of a ducted premixed flame*. International Symposium on Combustion, 2020 (accepted).

9.1 Experimental set-up

A schematic view of the experimental set-up is given in Fig. 9.1. The core of the experiment consists of a Bunsen burner inside a tube. The Bunsen burner consists of a straight metal pipe with an inner diameter of 10 mm. A tube with an integrated glass window for optical access acts as a cylindrical enclosure to shield the flame from gusts. Experiments are carried out with a premixed methane-ethene-air mixture with laminar flow rates set to 0.9, 0.25 and 8 normal liters per minute. The composition of the mixture is controlled with BRONKHORST EL-FLOW mass flow controllers (MFC). A loudspeaker, driven by an amplified sinusoidal signal, is mounted upstream of the Bunsen burner for acoustic forcing. The flame dynamics are recorded with a PHANTOM V4.2 CMOS camera with a glass lens at a resolution of $1,280 \times 800$ pixels and a frame rate of $2,800$ frames/s. The resolution is sufficiently fine to resolve the flame surface (Fig. 9.2a).

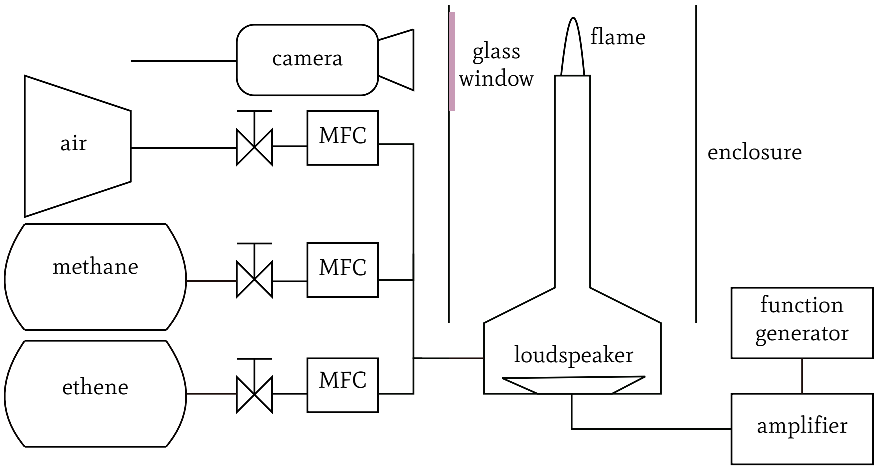


Fig. 9.1 Experimental set-up. This schematic was provided by USHNISH SENGUPTA.

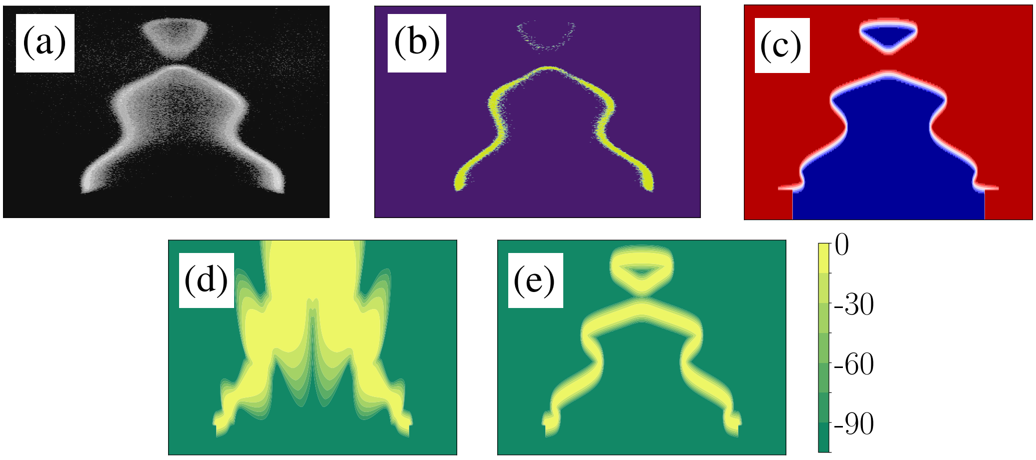


Fig. 9.2 Experimental, computational and statistical visualisations of premixed flame at reduced aspect ratios. (a): Experimental image of acoustic forcing at 200 Hz. Perturbations of the flame surface form at the base and travel to the tip. If the amplitude of the perturbation is large enough, a fuel-air pocket pinches off. (b): Experimental image after postprocessing. Pixels associated with the flame surface are colored yellow. (c): Snapshot of *G*-equation simulation for $K = 0.5$ and $\varepsilon = 0.36$. The zero-level set (white) separates the burnt (red) from the unburnt (blue) region. (d): Snapshot of log-normalized likelihood for initial guess of $K = 0.5$ and $\varepsilon = 0.2$ with 10 % standard deviation in each. Highly likely positions of the flame surface are shown in yellow; less likely positions in green. (e): Snapshot of log-normalized likelihood after combined state and parameter estimation. The model parameters K and ε are optimally calibrated such that they reproduce the propagation of perturbations along the flame surface as well as the formation of pinched-off fuel-air pockets. The spread of the high-likelihood locations (yellow) resembles the precision of the edge detection (b). The computational and statistical results of panels (c,d,e) are thoroughly explained in Chapters 9.2 and 9.3.

In order to perform data assimilation, we extract the coordinates of the pixels associated with the flame surface, which is identified by its natural luminescence in the visible range in the experimental images (Fig. 9.2a). A number of edge and ridge detection algorithms are readily available from SCIKIT-IMAGE [193], all of which give comparable results. In Fig. 9.2b, the result of applying the Sobel filter is shown.

9.2 *G*-equation model

The evolution of a premixed flame is modeled here by the kinematics of a surface. The position of the flame surface is captured by the zero-level set of a continuous scalar variable G.

It is governed by the G-equation [14, Chapter 2]:

$$\frac{\partial G}{\partial t} + (u - s_{\mathrm{L}} n) \cdot \nabla G = 0 \quad , \tag{9.1}$$

where u is the underlying flow field, s_{L} is the laminar flame speed, and n is the unit normal vector. The underlying flow field u is the superposition of a base flow U (Chapter 9.2.1) and a velocity perturbation u' (Chapter 9.2.2). The laminar flame speed is

$$s_{\mathrm{L}} = s_{\mathrm{L}}^0 (1 - \kappa \mathscr{M}) \quad , \tag{9.2}$$

where s_{L}^0 denotes the adiabatic flame speed. The Markstein length $\mathscr{M}$ makes the flame speed a function of the local curvature κ. The normal vector n and the curvature κ are given in terms of G:

$$n = \frac{\nabla G}{\|\nabla G\|} \quad , \quad \kappa = -\nabla \cdot n \quad . \tag{9.3}$$

Fig. 9.3 shows a schematic of this reduced-order model.

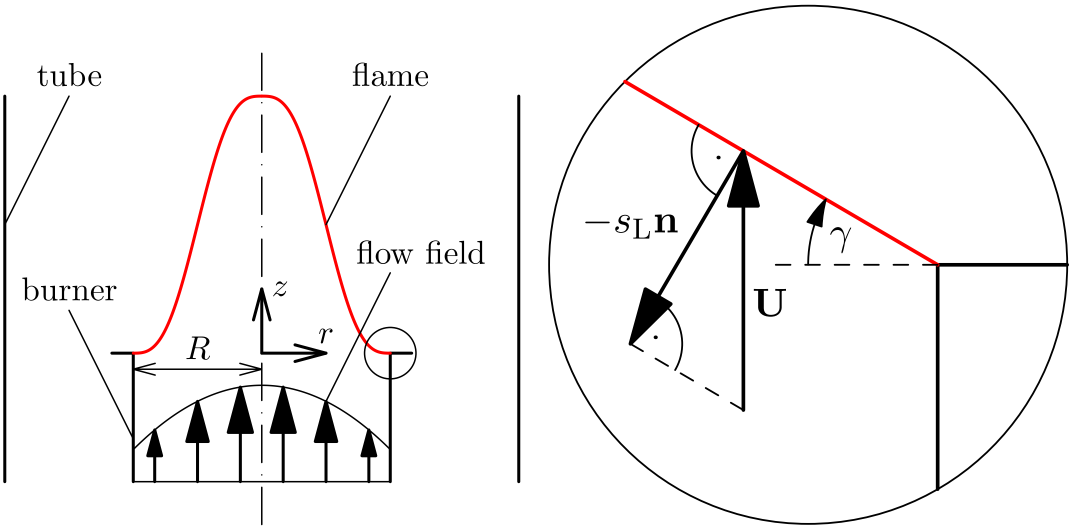


Fig. 9.3 Reduced-order model of the ducted premixed flame (left). In the absence of velocity perturbations, there is a kinematic balance between the base flow and the flame surface, e.g. at the burner lip (right).

The G-equation is numerically solved using a narrow-band level-set method with distance reinitialization. The computational domain is discretized using a fifth-order, weighted essentially non-oscillatory (WENO) scheme in space and a third-order, total variation diminishing (TVD) Runge-Kutta scheme in time. At the burner lip, a rotating boundary condition is used [61]. For details on the G-equation solver, the reader is referred to [194].

9.2.1 Base flow

For a time-independent base flow $u = U$, the G-equation (Eq. (9.1)) becomes

$$U \cdot n - s_{\mathrm{L}}^{0}(1 - \kappa \mathscr{M}) = 0 \quad . \tag{9.4}$$

For an axisymmetric burner, we parametrize the zero-level set, i.e. the flame surface, by setting its height h above the burner outlet as a function of the radius r:

$$G(r, z = h(r)) = 0 \quad , \quad 0 \leq r \leq R \quad . \tag{9.5}$$

The normal vector n and the curvature κ (Eq. (9.3)) in terms of h are

$$n = \frac{1}{[1 + h'(r)^2]^{0.5}} \begin{pmatrix} -h'(r) \\ 1 \end{pmatrix} \quad , \tag{9.6}$$

$$\kappa = \frac{h''(r)}{[1 + h'(r)^2]^{1.5}} + \frac{h'(r)}{r\,[1 + h'(r)^2]^{0.5}} \quad . \tag{9.7}$$

Substituting Eq. 9.6 and 9.7 into Eq. (9.4) gives a one-dimensional boundary value problem with $h'(0) = 0$ and $h(R) = 0$.

For simplicity, we assume that the base flow U is only a function of the radius r, but not of the height h. While the conditions at the burner outlet are theoretically known to be those of a Poiseuille-type pipe flow for a given mass flow rate, we introduce two additional parameters to account for the imperfections of this reduced-order model. Firstly, the velocity profile deviates from that of a Poiseuille flow towards a uniform flow the further we move away from the burner outlet [195, Chapter 5]. Secondly, we observe that, as the base-flow speed vanishes near the burner wall due to the no-slip condition, the flame speed also decreases due to heat loss to the burner wall [192], which is not properly modeled by the constant adiabatic flame speed s_{L}^{0} (Eq. (9.2)). Hence, we introduce two additional parameters to the reduced-order model of the base flow: The aspect ratio β gives the ratio between the flame length and the burner radius in a uniform flow without curvature effects:

$$\left(\frac{\bar{U}}{s_{\mathrm{L}}^{0}} \right)^2 = \beta^2 + 1 \quad , \tag{9.8}$$

where $\bar{U}$ denotes the mean flow speed. The shape parameter α linearly determines the velocity profile such that the mass flow rate is conserved:

$$\frac{U}{\bar{U}} = 1 + \alpha \left[1 - 2 \left(\frac{r}{R} \right)^2 \right] \quad , \tag{9.9}$$

where $\alpha = 0$ corresponds to a uniform flow, and $\alpha = 1$ corresponds to a Poiseuille flow. In summary, the base-flow model has three parameters θ_{bf} ('base flow'): the shape parameter α, the aspect ratio β and the Markstein length $\mathscr{M}$. Note that β replaces the parameter $s_{\mathrm{L}}^0/\bar{U}$ due to Eq. (9.8).

For a given set of parameters, we solve the boundary value problem (Eq. (9.4)-(9.7)) by using a finite-difference method and a Newton-Raphson solver [46, Chapter 5]. In iteration step k, the residual $R^k = R(h^k)$ is computed by evaluating the left-hand side of Eq. (9.4) at every grid point. The Jacobian $J^k = J(h^k)$ is computed by applying the chain rule to differentiate R^k with respect to h^k. Updates are performed by repeatedly solving

$$J^k \Delta h^k = R^k \quad \Longrightarrow \quad h^{k+1} = h^k - \Delta h^k \quad . \tag{9.10}$$

To infer the values of the parameters, we embed the boundary value problem into a least-squares problem with a cost functional $\mathfrak{C}$:

$$\mathfrak{C} = \sum_{m=1}^{M} \left[z_m - L_m(h(\theta_{\mathrm{bf}})) \right]^2 \quad , \tag{9.11}$$

where L_m is a suitable linear interpolation operator for the m-th measurement (r_m, z_m) of the flame surface. This optimization problem is solved by line search [196, Chapter 3]. The sensitivity of the cost functional $\mathfrak{C}$ to the parameters θ_{bf} is calculated using the adjoint variable λ [71, Chapter 2]:

$$\lambda^{\mathrm{T}} \frac{\partial R}{\partial h} = \frac{\partial \mathfrak{C}}{\partial h} \quad \Longrightarrow \quad \frac{\mathrm{d}\mathfrak{C}}{\mathrm{d}\theta_{\mathrm{bf}}} = \lambda^{\mathrm{T}} \frac{\partial R}{\partial \theta_{\mathrm{bf}}} \quad . \tag{9.12}$$

In Fig. 9.4, the results are shown for $\alpha = 0.84$, $\beta = 15.1$ and $\mathscr{M} = 3\,\mathrm{mm}$. The base-flow model agrees with the experiment. Furthermore, the base-flow model is confirmed using CANTERA simulations [197], which provide $\beta \approx 15.8$, based on the calculated adiabatic flame speed as well as the mass flow rate set in the experiment, and a flame thickness of $1.2\,\mathrm{mm}$, the latter on the same order of magnitude as the inferred Markstein length, in agreement with the theory [198, 199]. Finally, $\alpha = 0.84$ indicates a velocity profile close to Poiseuille flow as expected. While our model with three parameters covers a variety of base flows, including the one presented, more complex base-flow models are equally feasible without loss of generality regarding the data assimilation framework.

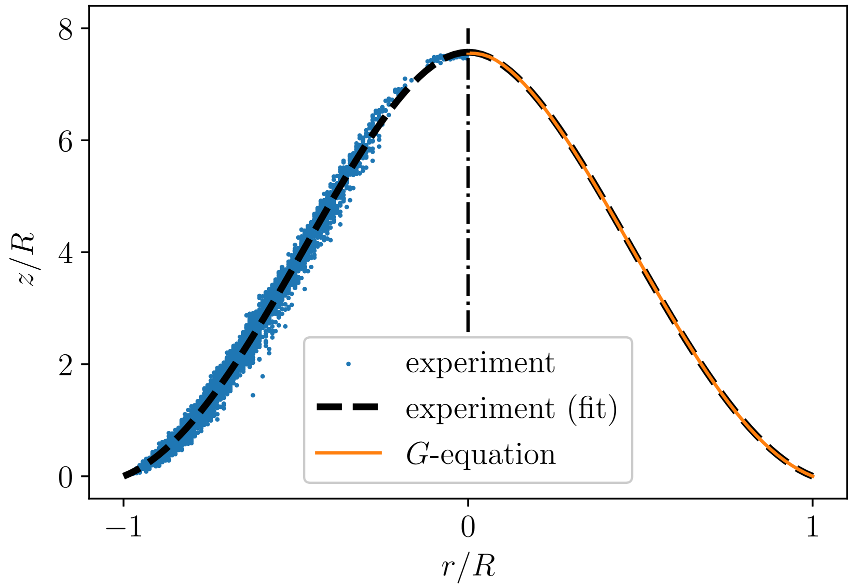


Fig. 9.4 Edge detection (blue dots, left) and least-squares fit from base-flow model (orange line, right). For comparison, the fourth-order polynomial fit $z/R = a_4(r/R)^4 + a_3(r/R)^3 + a_2(r/R)^2 - (a_4 + a_3 + a_2)$ respecting the boundary conditions is shown (black dashed line, both left and right). The base-flow model reproduces the flame surface observed in the experimental images.

9.2.2 Flame response

For the forcing of the premixed flame, the radial and axial components of the velocity perturbation u' are [181]

$$\frac{u'_r}{\bar{U}} = -\frac{\varepsilon \pi f K r}{\bar{U}} \cos\left(2\pi f \left(\frac{Kz}{\bar{U}} - t\right)\right) \quad , \tag{9.13}$$

$$\frac{u'_z}{\bar{U}} = \varepsilon \sin\left(2\pi f \left(\frac{Kz}{\bar{U}} - t\right)\right) \quad , \tag{9.14}$$

where u'_r and u'_z satisfy the continuity equation. The forcing has a frequency f and an amplitude $\varepsilon\bar{U}$. The non-dimensional parameter K is the ratio between the mean flow speed and the phase speed at which perturbations on the flame surface travel in the axial direction. The model parameters θ_{fr} (‘flame response’) are K and ε. In Fig. 9.2c, the result from a simulation based on our reduced-order model with parameters chosen to qualitatively match Fig. 9.2a,b is shown.

In preparation for data assimilation (Chapter 9.3), the G-equation has to be synchronized with the experimental images. This is not straightforward because we control the phase of the velocity perturbations in the G-equation (Eq. (9.13), (9.14)), whereas the experimental images only depict the flame surface without any velocity information. Therefore we derive an analytical relationship between the velocity perturbations and the motion of the flame surface at the base of the flame. As shown in Fig. 9.3 (right), the relationship between the base flow (Eq. (9.9)) and the flame surface (Eq. (9.8)) at the burner lip ($r = R$, $z = 0$), neglecting curvature effects, is

$$\cos(\gamma) = \left.\frac{s_\mathrm{L}}{U}\right|_R \approx \frac{s_\mathrm{L}^0}{U(R)} \approx \frac{1}{\beta(1-\alpha)} \quad , \tag{9.15}$$

where the last approximation is justified for $\beta = 15.1 \gg 1$ as observed in the experiment. The normal vector n is

$$n_r \approx 1 \quad , \quad n_z = \frac{1}{\beta(1-\alpha)} \quad . \tag{9.16}$$

We consider a small-amplitude perturbation γ' around the angle γ as a result of the corresponding velocity perturbation u':

$$u' \cdot n \approx -\varepsilon \pi f K R \sin(2\pi f t + \Delta\varphi) \quad , \tag{9.17}$$

where it is assumed that $(\pi f K R \beta(1-\alpha)/\bar{U})^2 \gg 1$ and $\tan(\Delta\varphi) = \pi f K R \beta(1-\alpha)/\bar{U}$, which is justified by the inferred values for the model parameters θ_bf and the judicious choice of frequencies f in Chapter 9.3. By observing the motion of the flame surface near the burner lip, $\Delta\varphi$ is calibrated to synchronize the G-equation with the experimental images.

Before turning to the data-driven estimation of K and ε in the next section, we summarize the a-priori insights about the model parameters. Under the assumption that the velocity perturbation felt at the base of the flame, i.e. $\dot{\gamma}'$, only depends on the volume of the loudspeaker, it follows from Eq. (9.17) that the amplitude, ε, is inversely proportional to the frequency, f. Consequently, Eq. (9.17) implies a low-pass filter for the flame response [34]. The assumption that K only depends on the base flow, not the forcing frequency, is expected to be true at small amplitudes due to linear stability analysis [200]. The behavior at larger amplitudes can be investigated with the approach in this paper. While $K \approx 1$ is reasonable for open flames in quiescent environments [35, 36], we additionally have to take into account the entrainment due to the buoyancy-driven flow surrounding the burner as well as the confinement due to the enclosing tube. Hence we anticipate a frequency-independent phase speed for the velocity perturbation with $K < 1$.

9.3 Combined state and parameter estimation

The Kalman filter provides a statistically optimal estimate ψ^a ('analysis') of the unknown state ψ of a system from a model prediction ψ^f ('forecast') and experimental observations d [114, Chapter 3]. The model prediction is mapped from its state space to the observation space through a measurement operator H. The prediction uncertainties and the experimental errors are represented by covariance matrices $C^f_{\psi\psi}$ and $C_{\varepsilon\varepsilon}$, respectively.

The application of the Kalman filter to the proposed reduced-order model is challenging for at least two reasons: Firstly, the G-equation is highly nonlinear, which is exemplified by the occurrence of cusps and pinched-off fuel-air pockets. This complicates the treatment of the time-dependent covariance matrix $C^f_{\psi\psi}$. Secondly, the proposed reduced-order model has $\mathscr{O}(10^5)$ degrees of freedom after discretization, which makes the computation and inversion of covariance matrices computationally intractable. To make the analysis statistically and computationally feasible, we instead generate an ensemble of N model predictions ψ^f_i with $i = 1,\ldots,N$. This variation of the Kalman filter, the ensemble Kalman filter, gives for ψ^a_i and its statistics (Theorem 5.3):

$$\psi^a_i = \psi^f_i + \left(HC^f_{\psi\psi}\right)^{\mathrm{T}} \left[C_{\varepsilon\varepsilon} + HC^f_{\psi\psi}H^{\mathrm{T}}\right]^{-1} \left(d - H\psi^f_i\right) \quad , \tag{9.18}$$

$$\bar{\psi} = \frac{1}{N}\sum_{i=1}^{N} \psi_i \quad , \quad C_{\psi\psi} = \frac{1}{N-1}\sum_{i=1}^{N} (\psi_i - \bar{\psi})(\psi_i - \bar{\psi})^{\mathrm{T}} \quad . \tag{9.19}$$

Finally, the question remains as to whether the (ensemble) Kalman filter preserves the kinematic nature of our reduced-order model. The choice of a suitable sample space, i.e. the construction of ψ from G, is crucial and not straightforward. Hence, we adopt our level-set data assimilation framework, which is based on the Hamilton-Jacobi formalism (Chapter 6).

In state estimation, the ensemble ψ_i represents one realization from a probability distribution in ψ with mean $\bar{\psi}$ and covariance matrix $C_{\psi\psi}$. As such, the marginal probability distribution in the k-th entry of ψ is given by the mean $\bar{\psi}[k]$ and the variance $C_{\psi\psi}[k,k]$. Consequently, the likelihood of finding the flame surface at the location corresponding to the k-th entry of ψ, regardless of the position of the flame surface elsewhere, is

$$p[k] = \frac{1}{\sqrt{2\pi C_{\psi\psi}[k,k]}} \exp\left(-\frac{\bar{\psi}[k]^2}{2C_{\psi\psi}[k,k]}\right) \quad . \tag{9.20}$$

Alternatively, the log-normalized likelihood is given by

$$\log\left(\frac{p[k]}{p_0[k]}\right) = -\frac{\bar{\psi}[k]^2}{2C_{\psi\psi}[k,k]} \quad , \tag{9.21}$$

where $\log(p/p_0) = 0$ identifies the most likely position of the flame surface. In Fig. 9.2d, the log-normalized likelihood is visualized for an educated guess of $K \approx 0.5$ and $\varepsilon \approx 0.2$ before any data assimilation. The position of the flame surface becomes highly uncertain for just a modest amount of standard deviation in the model parameters.

For combined state and parameter estimation, an augmented $\widetilde{\psi}$ is obtained by appending the model parameters θ_{fr}, i.e. K and ε, to ψ, and applying the ensemble Kalman filter to $\widetilde{\psi}$ (Chapter 5.1). In Fig. 9.2e, the results are shown for the same initial guess of K and ε as in Fig. 9.2d. In comparison, the identification of high-likelihood positions of the flame surface has significantly improved after combined state and parameter estimation. In Fig. 9.5, the joint probability distribution in K and ε is visualized. It is computed by marginalizing ψ from the probability distribution in $\widetilde{\psi}$. Although the means have the same order of magnitude, the standard deviation in K is three times smaller. The parameter K is easier to infer because the pinch-off timing strongly depends on K and is captured accurately by the proposed data assimilation method. Moreover, K and ε are only weakly correlated, which confirms their distinct roles within the proposed reduced-order model.

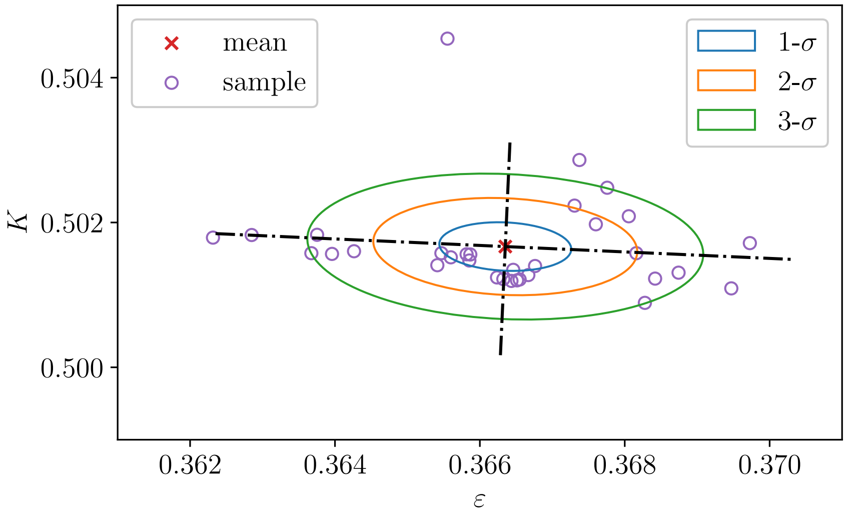


Fig. 9.5 Sample of K and ε and reconstructed probability distribution after combined state and parameter estimation for $f = 200\,\mathrm{Hz}$. The 1-, 2- and 3-σ (blue/orange/green, respectively) confidence regions correspond to 39, 86 and 99 % probabilities of sampling a given set of parameters, respectively.

To assess the proposed reduced-order model over a range of operating conditions, combined state and parameter estimation is performed for multiple frequencies ($f = 200 - 400\,\mathrm{Hz}$). Due to the low-pass nature of premixed flames, low frequency forcing leads to large fuel-air pockets, which form cusps and pinch off, while high frequency forcing leads to small perturbations which travel downstream without cusps. The range of forcing frequencies is chosen as to avoid flame blow-off at lower frequencies as well as vanishing flame response at higher frequencies. An ensemble of 32 G-equation simulations with the same initial condition is considered each time. The model parameters K and ε are sampled from independent normal distributions based around educated guesses of their values with 10 % standard deviation in each. At first, the G-equation simulations are solved without data assimilation to make sure that the dynamics are consistent with the parameters and are free of transient effects. After 10 periods of forcing, the postprocessed experimental images ($f = 2,800\,\mathrm{Hz}$, equivalent to $7 - 14$ assimilations per cycle) are assimilated for 5 periods. The covariance matrix $C_{\varepsilon\varepsilon}$, which represents the experimental errors, is a diagonal matrix with σ_ε^2 on its diagonal. The choice of $\sigma_\varepsilon = 1\,\mathrm{mm}$ is based on the thickness of the detected edges from postprocessing (Fig. 9.2b). In general, the precision of the experimental data does not affect the accuracy of the ensemble in the long run but reduces the uncertainty in the ensemble overall (Chapter 7.3). Finally, the G-equation simulations are solved for another 5 periods to observe the performance with optimally calibrated model parameters and without any data assimilation. The overall computational time is less than 40 minutes for 20 periods on a cluster node with 1 processor per G-equation simulation. This includes the very frequent output of solution files in order to study time series and statistics. This is marginally longer than a single simulation of the G-equation as the ensemble Kalman filter is parallel by design.

In Fig. 9.6, the root mean square (RMS) error,

$$\text{RMS error} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\psi_i - \bar{\psi})^{\mathrm{T}}(\psi_i - \bar{\psi})} \quad , \tag{9.22}$$

is plotted over time for combined state and parameter estimation at 200, 300 and 400 Hz. Within 5 periods of forcing, the dynamics for the sampled sets of parameters fully form. Between 10 and 15 periods, the ensemble Kalman filter gradually improves the estimates by up to two orders of magnitude. After 15 periods, the dynamics adapt to the optimally calibrated model parameters with low uncertainty. The remaining spurious oscillations are the result of noise in the data. The reduction in RMS error corresponds to the characteristic displacement due to data assimilation. 10 grid units equal 5 % of the flame length scale in either direction.

In Fig. 9.7, the optimally calibrated model parameters and their confidence intervals are shown for $f = 200 - 400\,\text{Hz}$. In agreement with theory, the model parameter K remains nearly constant while the model parameter ε decreases in inverse proportion to the frequency f (Chapter 9.2.2). This highlights a favorable feature of physics-informed reduced-order models: While the state of the system naturally varies with the operating conditions, the model itself and its parametrization are purposefully designed to be universal, so either widely constant or at least following a certain scaling. This is clearly the case for the model parameters K and ε in Fig. 9.7. Therefore, we expect the reduced-order model to interpolate well despite a limited amount of data. The confidence intervals at the different frequencies are of comparable height. In agreement with Fig. 9.5, the parameter K is significantly more certain than the parameter ε.

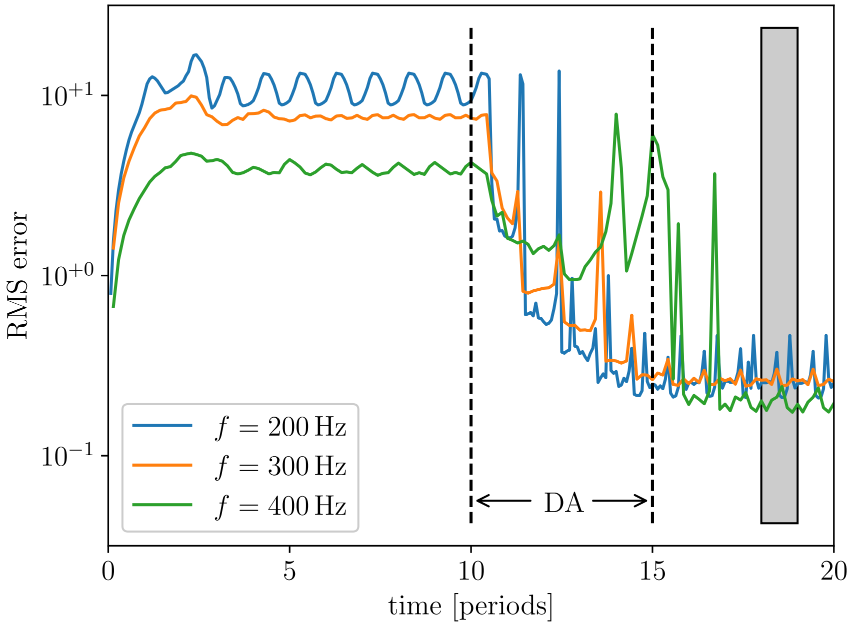


Fig. 9.6 Root mean square (RMS) error for forcing at 200, 300 and 400 Hz (blue / orange / green, respectively). Data assimilation (DA) takes place between 10 and 15 periods. The grey window is representative of the period depicted in Fig. 9.8.

In Fig. 9.8, the outcome of combined state and parameter estimation is visualized for forcing at 200, 300 and 400 Hz, respectively. The optimally calibrated reduced-order model accurately captures the perturbations traveling along the flame surface as well as the fuel-air pockets pinching off. While no individual G-equation simulation captures the motion of the flame surface completely, the G-equation simulations as an ensemble form an envelope in which the flame surface is fully contained, thus quantifying the uncertainty in the reduced-order model. While the pinched-off fuel-air pockets are clearly detectable in the experimental

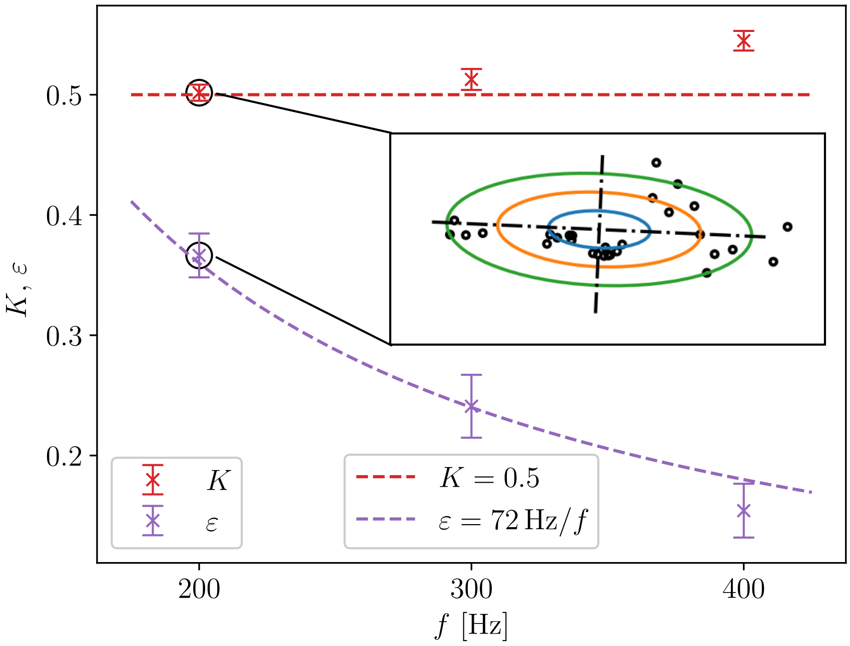


Fig. 9.7 Optimally calibrated estimates and their uncertainties for K (red) and ε (purple). The 10-σ (chosen for clearer visualisation) confidence intervals are computed by marginalizing the corresponding joint probability distributions (Fig. 9.5). Every joint probability distribution is reconstructed from an ensemble of 32 G-equation simulations, 96 simulations in total. The dashed lines show the behavior estimated from theory (Chapter 9.2.2) for K (constant) and ε (inversely proportional to f).

images for 200 Hz, the pinched-off fuel-air pockets are smaller in size for higher frequencies due to the low-pass nature of the flame response, and exist for shorter periods of time. Although the light intensity is diminished towards the tip of the flame surface, which complicates edge detection and observation in general, the optimally calibrated reduced-order model correctly infers the precise flame dynamics (300 Hz, left/middle left) that lead to the short-lived fuel-air pocket pinching off (300 Hz, middle right). As the perturbations traveling along the flame surface decrease in magnitude, so does the signal-to-noise ratio. Combined with the reduced relative frame rate at higher frequencies, the experimental images are ambiguous as to whether a fuel-air pocket pinches off, or the tip of the flame surface only retracts so rapidly because of the high local curvature (400 Hz, middle right). This ambiguity is reflected in the ensemble of G-equation simulations, especially in the elevated uncertainty towards the end of the assimilation window, where some exhibit pinched-off fuel-air pockets with lifespans below the frame rate while others do not (Fig. 9.6).

In summary, the twin experiment as well as the simulated and experimental ducted premixed flames show that our level-set data assimilation framework is capable of assimilating

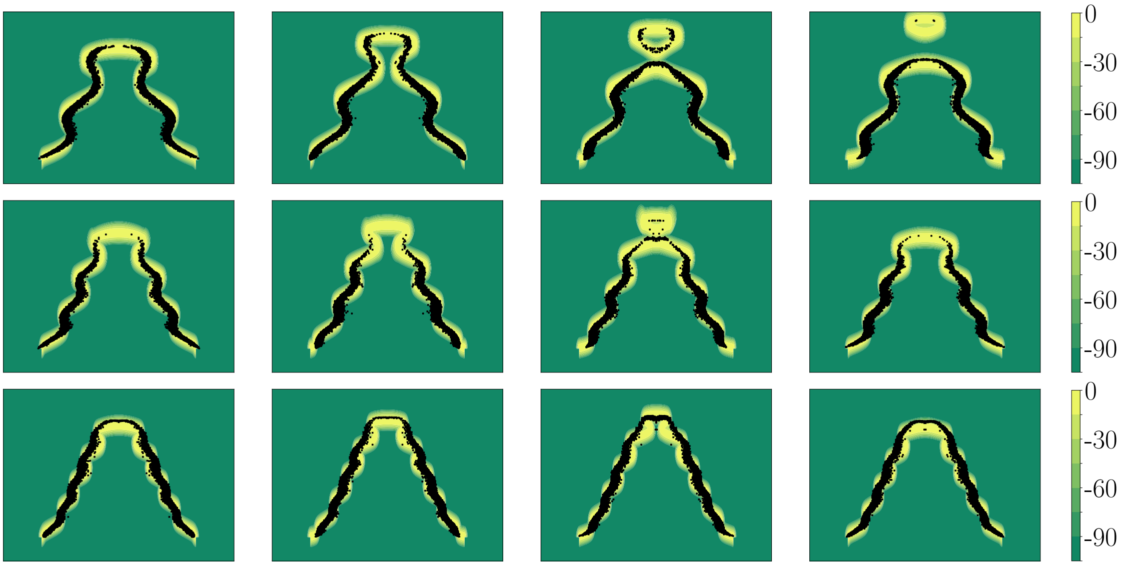


Fig. 9.8 Snapshots of log-normalized likelihood (Eq. (9.21)) over one forcing period after combined state and parameter estimation for 200, 300 and 400 Hz (top/middle/bottom row, respectively). Highly likely positions of the flame surface are shown in yellow; less likely positions in green. The flame surface as detected from experimental images is included (black dots).

data into our reduced-order model. The two key aspects of the analysis are the following: Firstly, the G-equation is a fully nonlinear model. This includes non-smooth features, e.g. cusps on the flame surface, and discontinuities such as topological merging and break-up. Unlike classical, sensitivity-based approaches designed under linear assumptions, this probabilistic, ensemble-based approach successfully models the nonlinearities and delivers an optimally calibrated reduced-order model. Secondly, the proposed level-set data assimilation framework is based on Bayesian inference. As such, all estimates are equipped with statistically rigorous uncertainty quantification. This is highly relevant to the design of combustion systems: In thermoacoustics, for example, slight errors in the model deduced from a single burner could have a large impact on predictions in a different configuration, such as inside an annular combustor [29, 70].

The data assimilation framework was developed for level-set methods in general. As such, it is readily generalizable to other flame shapes modeled by the G-equation. Therefore, we plan to conduct combined state and parameter estimation for a wider array of operating conditions. In addition to the range of frequencies, this includes forcing amplitudes, mass flow rates, fuel-air compositions as well as different burner geometries eventually.

Part IV

Conclusions

Chapter 10

Thermoacoustics and inverse problems

In this chapter, we summarize the results of treating reduced-order models of flame dynamics and thermoacoustic oscillations as inverse problems. Using probability theory as the lingua franca of statistical inference, we draw connections to two related applications of the data-driven methods developed in Parts II and III: (i) field inversion and machine learning for improved predictions; and (ii) physics-informed reduced-order models as encoders and decoders for increasingly large datasets.

10.1 Summary

The key concepts introduced in Part I are the reduced-order model and the inverse problem. Our reduced-order models are hierarchical in structure, i.e. they represent graphical models, and consist of three components: (i) the governing equations; (ii) the parameters within the governing equations; and (iii) the state vector that solves the governing equations given the parameters. In general, neither of the three is fully observable, and measurements are subject to noise. The objective of the inverse problem is to infer knowledge about the state, the parameters and the governing equations from observations. While we solve the inverse problem in a probabilistic formulation, the choice of normal distributions is convenient but not necessary. Normal distributions are convenient because they have representations in calculus and linear algebra. In calculus, the inverse problem is treated as an optimization problem, e.g. a least-squares regression, and is subject to gradient-based and adjoint methods. In linear algebra, a centered multivariate normal distribution is fully characterized by its covariance matrix, which is efficiently manipulated based on low-rank representations such as eigendecompositions and the Woodbury formula. The salient feature of our probabilisitic formulation is the introduction of random variables and stochastic processes. We formulate and solve two types of inverse problems in thermoacoustics.

In Part II, we focus on reduced-order models of implicit form with state vectors consisting of a single complex scalar. With the stability of jet and rocket engines in mind, this type of reduced-order model is not restricted to linear stability analysis or nonlinear eigenproblems. Our solution to this inverse problem involves two steps: In a first step, we estimate the model parameters using a Laplace approximation. Once the gradient and the Hessian of the state vector with respect to the model parameters are known, the statistical inversion becomes a straightforward application of Bayes' rule. In a second step, we introduce Gaussian processes for the model parameters. The simultaneous statistical inference of the model parameters and the optimization of the hyperparameters of the Gaussian processes follow the evidence framework and utilize the expectation-maximization algorithm as well as message passing. The results are statistically rigorous uncertainty quantification of the model parameters and predictions as well as comparisons regarding model sensitivity and complexity.

The highlight of our hierarchically structured statistical inference is the liberty to replace components. In place of a Laplace approximation, it is possible to translate the statistics from the observations to the model parameters using a Markov-chain Monte-Carlo (MCMC) simulation, e.g. based on Metropolis-Hastings or Gibbs sampling [201, Chapter 4]. While MCMC simulations are more computationally expensive than a Laplace approximation equipped with adjoint methods for a large number of parameters, they are quickly implemented and not restricted to normal distributions. In place of Gaussian processes, it is possible to perform the nonparametric regression using any other supervised learning algorithm, e.g. support vector machines or feed-forward neural networks [100, Chapters 5-7]. While the marginalization property and automatic relevance determination (ARD) are useful features in Gaussian processes, other supervised learning algorithms offer comparable features at more affordable computational cost with respect to the number of observations.

In Part III, we focus on time-dependent reduced-order models with high-dimensional state vectors and an emphasis on state and parameter estimation. In particular, we investigate the G-equation with level-set methods. It is highly nonlinear with non-differentiable features such as cusps and topological changes such as pinched-off fuel-air pockets. The main component of our statistical inference framework is the ensemble Kalman filter with its derivative-free formulation. The prediction step is performed in the form of a Monte-Carlo simulation, and is as such parallel by design. The update step is executed as a series of matrix-vector operations, and is as such computationally inexpensive.

The challenge lies in the choice of a sample space that respects the physics of the problem. For our level-set data assimilation framework, the physics are given by the kinematics of an interface. We invoke the Hamilton-Jacobi formalism because its generating functions form such an appropriate sample space. Even then, there is no guarantee that the state vector after the update step still lies in the same sample space, which is why we perform a distance reinitialization after the update step as well. Thus, the narrow-band method and the fast-marching method complement the ensemble Kalman filter to form our level-set data assimilation framework.

The challenge regarding the choice of sample space is not exclusive to data assimilation in level-set methods. For example, CFD simulations are required to enforce their associated set of conservation laws [202, Chapter 4]. For incompressible flow simulations, if the predictions step relies on a projection method, it is possible to use the same Poisson equation after the update step as well [203]. It shares the philosophy with our level-set data assimilation framework of using the same predictor-corrector methodology in both prediction and update steps [204]. It is worth mentioning that the sophistication regarding the choice of sample space is not always necessary. For example, Lagrangian formulations such as front-tracking methods and vortex methods allow for a straightforward application of the update step [160]. Another example are compressible flow simulations, where the straightforward application of the update step apparently improves predictions despite the possible production of spurious acoustics [161].

10.2 Outlook

The introduction of random variables and stochastic processes allows flexible yet rigorous formulations of inverse problems tailored to the reduced-order models under investigation. Most importantly, probability theory as the lingua franca of statistical inference promotes the exchange of ideas between different fields. For instance, both Parts II and III utilize state augmentation and weak constraints in order to facilitate parameter estimation. In Part II, we extend these ideas into a nonlinear regression framework using Gaussian processes to predict model parameters. In Part III, the emphasis is on combined state and parameter estimation for high-dimensional systems, where the model parameters are stationary during the prediction step (Eq. (5.35), (5.36)). This means that it is possible to adapt the machine learning ideas from Part II for the high-dimensional systems in Part III.

Indeed, the much-discussed field inversion and machine learning (FIML) framework by PARISH ET AL. is an independent attempt to couple a Bayesian regression with Gaussian processes or neural networks [205, 206, 93]. While an in-depth treatment of the FIML framework is beyond the scope of this discussion, we would like to point out a handful of modifications in line with the evidence framework and the expectation-maximization algorithm from Part II that promise principled improvements: Firstly, the a-priori estimation of the prior variance is arguably the most computationally expensive step in the FIML framework. In the evidence framework, the prior variance is treated as a hyperparameter and incrementally improved in the maximization step. Since the FIML framework already relies on the availability of an adjoint solver, it is also possible to replace the repeated sampling of the prior distribution with an optimization problem that, for example, involves typicality as the quantity of interest [99, Chapter 4]. Secondly, the FIML framework is mostly applied to fully developed flows whereas the emphasis of Part III is on dynamical systems. The straightforward generalization of the FIML framework from the spatial setting to the spatiotemporal setting would require a sample space spanning the flow fields over all the timesteps. The theory of stochastic filtering and smoothing instead yields conditional probability distributions at each timestep of a Markov chain. Furthermore, ensemble filters and smoothers provide low-rank representations of prior and posterior covariance matrices, which is potentially relevant for highly resolved simulations. Thirdly, the FIML framework is mostly applied to well-posed problems that involve a single model parameter such as eddy viscosity. The potentially ill-posed problem of a single quantity of interest with multiple model parameters requires regularization as outlined in Part II. In the following step of the FIML framework, where the functional relationship is established, the choice of input features is up to the modeler.

The role of evidence-based model comparison is expected to play a more decisive role in less academic and more practical test cases. In Part II, the evidence is used for model selection: It answers the question as to which combination of base-flow and acoustics model most plausibly explains the observed data and is expected to extrapolate well. Alternatively, it is possible to use the evidence of each reduced-order model as a weight for model averaging [207]. In applications where the physical mechanisms are not fully understood, model averaging is expected to lead to more robust and less extreme predictions. This is an instance of ensemble learning, of which an overview is beyond the scope of this discussion. Although the evidence framework and the expectation-maximization algorithm are recognized as principled approaches to statistical inference, it is worth mentioning that more ad-hoc alternatives exist [208].

This work highlights the role that data can play in the future of combustion modeling for thermoacoustics. It is increasingly impractical to store data, particularly as experiments become automated [119] and numerical simulations become more detailed. Rather than store the data itself, the techniques in this work optimally assimilate the data into the parameters of a physics-informed reduced-order model. In Part III, combined state and parameter estimation essentially compresses terabytes of flame images to a handful of model parameters with remaining epistemic uncertainties. In Part II, Bayesian regression with Gaussian processes fits reduced-order models to a wide range of operating conditions. While Gaussian processes scale unfavorably with the number of observations, their ARD feature identifies data points whose knowledge does not significantly improves predictions, and whose dismissal leads to further compression.

With data-driven reduced-order models using Bayesian regression or data assimilation, rapid prototyping of combustion systems can feed into rapid calibration of their reduced-order models and then into gradient-based design optimization. While it has been shown, e.g. in the context of ignition and extinction, that large-eddy simulations become quantitatively predictive when augmented with data [161], the reduced-order modeling of flame dynamics in turbulent flows remains challenging. For these challenging situations, this work opens up new possibilities for the development of reduced-order models that adaptively change any time that data from experiments or simulations becomes available.

References

[1] Timothy C. Lieuwen and Vigor Yang, editors. *Combustion Instabilities in Gas Turbine Engines: Operational Experience, Fundamental Mechanisms and Modeling*. Progress in Astronautics and Aeronautics. American Institute of Aeronautics and Astronautics, 2005.

[2] F. E. C. Culick. Unsteady Motions in Combustion Chambers for Propulsion Systems. Technical Report AG-AVT-039, NATO Research and Technology Organisation, December 2006.

[3] C. Sondhauss. über die Schallschwingungen der Luft in erhitzten Glasröhren und in gedeckten Pfeifen von ungleicher Weite. *Annalen der Physik und Chemie*, 155(1):1–34, 1850.

[4] P. L. Rijke. Notiz über eine neue Art, die in einer an beiden Enden offenen Röhre enthaltene Luft in Schwingungen zu versetzen. *Annalen der Physik und Chemie*, 183(6):339–343, 1859.

[5] John William Strutt. The explanation of certain acoustical phenomena. *Nature*, 18(455):319–321, July 1878.

[6] Boa-Teh Chu. On the energy transfer to small disturbances in fluid flow (Part I). *Acta Mechanica*, 1(3):215–234, September 1965.

[7] F. Nicoud and T. Poinsot. Thermoacoustic instabilities: Should the Rayleigh criterion be extended to include entropy changes? *Combustion and Flame*, 142(1-2):153–159, July 2005.

[8] Luca Magri, Matthew P. Juniper, and Jonas P. Moeck. Sensitivity of the Rayleigh criterion in thermoacoustics. *Journal of Fluid Mechanics*, 882:R1, January 2020.

[9] L.Y.M. Gicquel, G. Staffelbach, and T. Poinsot. Large Eddy Simulations of gaseous flames in gas turbine combustion chambers. *Progress in Energy and Combustion Science*, 38(6):782–817, December 2012.

[10] Peter Jordan and Tim Colonius. Wave Packets and Turbulent Jet Noise. *Annual Review of Fluid Mechanics*, 45(1):173–195, January 2013.

[11] Guillaume A. Brès, Frank E. Ham, Joseph W. Nichols, and Sanjiva K. Lele. Unstructured Large-Eddy Simulations of Supersonic Jets. *AIAA Journal*, 55(4):1164–1184, April 2017.

[12] Michael S. Dodd and Antonino Ferrante. On the interaction of Taylor length scale size droplets and isotropic turbulence. *Journal of Fluid Mechanics*, 806:356–412, November 2016.

[13] Said Elghobashi. Direct Numerical Simulation of Turbulent Flows Laden with Droplets or Bubbles. *Annual Review of Fluid Mechanics*, 51(1):217–244, January 2019.

[14] Norbert Peters. *Turbulent Combustion*. Cambridge University Press, 2000.

[15] Heinz Pitsch. Large-Eddy Simulation of Turbulent Combustion. *Annual Review of Fluid Mechanics*, 38(1):453–482, January 2006.

[16] Matthias Ihme. Combustion and Engine-Core Noise. *Annual Review of Fluid Mechanics*, 49(1):277–310, January 2017.

[17] C.A. Armitage, R. Balachandran, E. Mastorakos, and R.S. Cant. Investigation of the nonlinear response of turbulent premixed flames to imposed inlet velocity oscillations. *Combustion and Flame*, 146(3):419–436, August 2006.

[18] Xingsi Han and Aimee S. Morgans. Simulation of the flame describing function of a turbulent premixed flame using an open-source LES solver. *Combustion and Flame*, 162(5):1778–1792, May 2015.

[19] Audrey Valreau. *Computational Fluid Dynamics Investigation of Oscillatory Behaviour in Annular Combustors*. PhD thesis, University of Cambridge, October 2018.

[20] L. Selle, G. Lartigue, T. Poinsot, R. Koch, K.-U. Schildmacher, W. Krebs, B. Prade, P. Kaufmann, and D. Veynante. Compressible large eddy simulation of turbulent combustion in complex geometry on unstructured meshes. *Combustion and Flame*, 137(4):489–505, June 2004.

[21] G. Staffelbach, L.Y.M. Gicquel, G. Boudier, and T. Poinsot. Large Eddy Simulation of self excited azimuthal modes in annular combustors. *Proceedings of the Combustion Institute*, 32(2):2909–2916, 2009.

[22] A. Giusti and E. Mastorakos. Turbulent Combustion Modelling and Experiments: Recent Trends and Developments. *Flow, Turbulence and Combustion*, October 2019.

[23] Pierre Simon Laplace. *A philosophical essay on probabilities*. Dover Publications, New York, 1995.

[24] Stephen B. Pope. *Turbulent Flows*. Cambridge University Press, Cambridge, 2000.

[25] Laurent Selle, Franck Nicoud, and Thierry Poinsot. Actual Impedance of Nonreflecting Boundary Conditions: Implications for Computation of Resonators. *AIAA Journal*, 42(5):958–964, May 2004.

[26] M. Bauerheim, G. Staffelbach, N.A. Worth, J.R. Dawson, L.Y.M. Gicquel, and T. Poinsot. Sensitivity of LES-based harmonic flame response model for turbulent swirled flames and impact on the stability of azimuthal modes. *Proceedings of the Combustion Institute*, 35(3):3355–3363, 2015.

[27] Christian Kraus, Laurent Selle, and Thierry Poinsot. Coupling heat transfer and large eddy simulation for combustion instability prediction in a swirl burner. *Combustion and Flame*, 191:239–251, May 2018.

[28] P. G. Drazin. *Introduction to Hydrodynamic Stability*. Cambridge University Press, Cambridge, 2002.

[29] Matthew P. Juniper and R.I. Sujith. Sensitivity and Nonlinearity of Thermoacoustic Oscillations. *Annual Review of Fluid Mechanics*, 50(1):661–689, January 2018.

[30] F.E. Marble and S.M. Candel. Acoustic disturbance from gas non-uniformities convected through a nozzle. *Journal of Sound and Vibration*, 55(2):225–243, November 1977.

[31] N.A. Cumpsty. Jet engine combustion noise: Pressure, entropy and vorticity perturbations produced by unsteady combustion or heat addition. *Journal of Sound and Vibration*, 66(4):527–544, October 1979.

[32] Luca Magri, Jeff O'Brien, and Matthias Ihme. Compositional inhomogeneities as a source of indirect combustion noise. *Journal of Fluid Mechanics*, 799:R4, July 2016.

[33] Luigi Crocco and Xinyi Zheng. *Theory of Combustion Instability in Liquid Propellant Rocket Motors*. Butterworths Scientific, 1956.

[34] Sébastien Ducruix, Daniel Durox, and Sébastien Candel. Theoretical and experimental determinations of the transfer function of a laminar premixed flame. *Proceedings of the Combustion Institute*, 28(1):765–773, January 2000.

[35] Thierry Schuller, Sébastien Ducruix, Daniel Durox, and Sébastien Candel. Modeling tools for the prediction of premixed flame transfer functions. *Proceedings of the Combustion Institute*, 29(1):107–113, January 2002.

[36] T. Schuller, D. Durox, and S. Candel. A unified model for the prediction of laminar flame transfer functions. *Combustion and Flame*, 134(1-2):21–34, July 2003.

[37] C. O. Paschereit, B. Schuermans, W. Polifke, and O. Mattson. Measurement of transfer matrices and source terms of premixed flames. *Journal of Engineering for Gas Turbines and Power*, 124(2):239, 2002.

[38] Ann P. Dowling. Nonlinear self-excited oscillations of a ducted flame. *Journal of Fluid Mechanics*, 346:271–290, September 1997.

[39] Ann P. Dowling. A kinematic model of a ducted flame. *Journal of Fluid Mechanics*, 394:51–72, September 1999.

[40] A.S. Morgans and S.R. Stow. Model-based control of combustion instabilities in annular combustors. *Combustion and Flame*, 150(4):380–399, September 2007.

[41] N. Noiray, D. Durox, T. Schuller, and S. Candel. A unified framework for nonlinear combustion instability analysis based on the flame describing function. *Journal of Fluid Mechanics*, 615:139, November 2008.

[42] D. Durox, T. Schuller, N. Noiray, and S. Candel. Experimental analysis of nonlinear flame transfer functions for different flame geometries. *Proceedings of the Combustion Institute*, 32(1):1391–1398, 2009.

[43] G. J. Bloxsidge, A. P. Dowling, and P. J. Langhorne. Reheat buzz: an acoustically coupled combustion instability. Part 2. Theory. *Journal of Fluid Mechanics*, 193:445, August 1988.

[44] John William Strutt. *The Theory of Sound*, volume 2. Cambridge University Press, 2011.

[45] Simon R. Stow and Ann P. Dowling. Thermoacoustic Oscillations in an Annular Combustor. In *Volume 2: Coal, Biomass and Alternative Fuels; Combustion and Fuels; Oil and Gas Applications; Cycle Innovations*, page V002T02A004, New Orleans, Louisiana, USA, June 2001. American Society of Mechanical Engineers.

[46] Uri M. Ascher, Robert M. M. Mattheij, and Robert D. Russell. *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations*. Society for Industrial and Applied Mathematics, January 1995.

[47] Ben T. Zinn and Manuel E. Lores. Application of the Galerkin Method in the Solution of Non-linear Axial Combustion Instability Problems in Liquid Rockets. *Combustion Science and Technology*, 4(1):269–278, September 1971.

[48] Franck Nicoud, Laurent Benoit, Claude Sensiau, and Thierry Poinsot. Acoustic Modes in Combustors with Complex Impedances and Multidimensional Active Flames. *AIAA Journal*, 45(2):426–441, February 2007.

[49] Georg A. Mensah, Giovanni Campa, and Jonas P. Moeck. Efficient Computation of Thermoacoustic Modes in Industrial Annular Combustion Chambers Based on Bloch-Wave Theory. *Journal of Engineering for Gas Turbines and Power*, 138(8):081502, August 2016.

[50] Matthew P. Juniper. Triggering in the horizontal Rijke tube: non-normality, transient growth and bypass transition. *Journal of Fluid Mechanics*, 667:272–308, January 2011.

[51] Priya Subramanian, R. I. Sujith, and P. Wahi. Subcritical bifurcation and bistability in thermoacoustic systems. *Journal of Fluid Mechanics*, 715:210–238, January 2013.

[52] Alessandro Orchini, Georgios Rigas, and Matthew P. Juniper. Weakly nonlinear analysis of thermoacoustic bifurcations in the Rijke tube. *Journal of Fluid Mechanics*, 805:523–550, October 2016.

[53] M. Fleifil, A. M. Annaswamy, Z. A. Ghoneim, and A. F. Ghoniem. Response of a laminar premixed flame to flow oscillations: A kinematic model and thermoacoustic instability results. *Combustion and Flame*, 106(4):487–510, September 1996.

[54] Santosh H. Preetham and Tim C. Lieuwen. Response of turbulent premixed flames to harmonic acoustic forcing. *Proceedings of the Combustion Institute*, 31(1):1427–1434, January 2007.

[55] Preetham, H. Santosh, and Tim Lieuwen. Dynamics of Laminar Premixed Flames Forced by Harmonic Velocity Disturbances. *Journal of Propulsion and Power*, 24(6):1390–1402, November 2008.

[56] Bernhard Semlitsch, Alessandro Orchini, Ann P. Dowling, and Matthew P. Juniper. G-equation modelling of thermoacoustic oscillations of partially premixed flames. *International Journal of Spray and Combustion Dynamics*, 9(4):260–276, December 2017.

[57] Karthik Kashinath, Iain C. Waugh, and Matthew P. Juniper. Nonlinear self-excited thermoacoustic oscillations of a ducted premixed flame: bifurcations and routes to chaos. *Journal of Fluid Mechanics*, 761:399–430, December 2014.

[58] Hiroshi Gotoda, Hiroyuki Nikimoto, Takaya Miyano, and Shigeru Tachibana. Dynamic properties of combustion instability in a lean premixed gas-turbine combustor. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 21(1), March 2011.

[59] Lipika Kabiraj, R. I. Sujith, and Pankaj Wahi. Bifurcations of Self-Excited Ducted Laminar Premixed Flames. *Journal of Engineering for Gas Turbines and Power*, 134(3):031502, March 2012.

[60] Lipika Kabiraj, Aditya Saurabh, Pankaj Wahi, and R. I. Sujith. Route to chaos for combustion instability in ducted laminar premixed flames. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(2):023129, June 2012.

[61] Iain Waugh, Simon Illingworth, and Matthew Juniper. Matrix-free continuation of limit cycles for bifurcation analysis of large thermoacoustic systems. *Journal of Computational Physics*, 240:225–247, May 2013.

[62] Iain C. Waugh, K. Kashinath, and Matthew P. Juniper. Matrix-free continuation of limit cycles and their bifurcations for a ducted premixed flame. *Journal of Fluid Mechanics*, 759:1–27, November 2014.

[63] Steven H. Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. Westview Press, Boulder, CO, second edition, 2015. OCLC: ocn842877119.

[64] Francisco Huhn and Luca Magri. Stability, sensitivity and optimisation of chaotic acoustic oscillations. *Journal of Fluid Mechanics*, 882:A24, January 2020.

[65] William A. Sirignano. Driving Mechanisms for Combustion Instability. *Combustion Science and Technology*, 187(1-2):162–205, January 2015.

[66] Marc C. Kennedy and Anthony O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, August 2001.

[67] Armen Der Kiureghian and Ove Ditlevsen. Aleatory or epistemic? Does it matter? *Structural Safety*, 31(2):105–112, March 2009.

[68] Paolo Luchini and Alessandro Bottaro. Adjoint Equations in Stability Analysis. *Annual Review of Fluid Mechanics*, 46(1):493–517, January 2014.

[69] Peter J. Schmid and Luca Brandt. Analysis of Fluid Systems: Stability, Receptivity, Sensitivity. *Applied Mechanics Reviews*, 66(2):021003, March 2014.

[70] Luca Magri. Adjoint Methods as Design Tools in Thermoacoustics. *Applied Mechanics Reviews*, 71(2):020801, March 2019.

[71] Max D. Gunzburger. *Perspectives in Flow Control and Optimization*. Society for Industrial and Applied Mathematics, January 2002.

[72] Matthew P. Juniper. Sensitivity analysis of thermoacoustic instability with adjoint Helmholtz solvers. *Physical Review Fluids*, 3(11), November 2018.

[73] José G. Aguilar, Luca Magri, and Matthew P. Juniper. Adjoint-based sensitivity analysis of low-order thermoacoustic networks using a wave-based approach. *Journal of Computational Physics*, 341:163–181, July 2017.

[74] Petr V. Kungurtsev and Matthew P. Juniper. Adjoint-based shape optimization of the microchannels in an inkjet printhead. *Journal of Fluid Mechanics*, 871:113–138, July 2019.

[75] Jack Brewster and Matthew P. Juniper. Shape sensitivity of eigenvalues in hydrodynamic stability, with physical interpretation for the flow around a cylinder. *European Journal of Mechanics - B/Fluids*, 80:80–91, March 2020.

[76] Luca Magri, Michael Bauerheim, and Matthew P. Juniper. Stability analysis of thermo-acoustic nonlinear eigenproblems in annular combustors. Part I. Sensitivity. *Journal of Computational Physics*, 325:395–410, November 2016.

[77] Luca Magri, Michael Bauerheim, Franck Nicoud, and Matthew P. Juniper. Stability analysis of thermo-acoustic nonlinear eigenproblems in annular combustors. Part II. Uncertainty quantification. *Journal of Computational Physics*, 325:411–421, November 2016.

[78] Flavio Giannetti and Paolo Luchini. Structural sensitivity of the first instability of the cylinder wake. *Journal of Fluid Mechanics*, 581:167, June 2007.

[79] Alessandro Bottaro, Peter Corbett, and Paolo Luchini. The effect of base flow variation on flow stability. *Journal of Fluid Mechanics*, 476:293–302, February 2003.

[80] Lloyd N. Trefethen and Mark Embree. *Spectra and pseudospectra: the behavior of nonnormal matrices and operators*. Princeton University Press, Princeton, N.J, 2005. OCLC: ocm58789286.

[81] Luca Magri and Matthew P. Juniper. Sensitivity analysis of a time-delayed thermo-acoustic system via an adjoint-based approach. *Journal of Fluid Mechanics*, 719:183–202, March 2013.

[82] Luca Magri and Matthew P. Juniper. Global modes, receptivity, and sensitivity analysis of diffusion flames coupled with duct acoustics. *Journal of Fluid Mechanics*, 752:237–265, August 2014.

[83] Dong Yang, Francesca M. Sogaro, Aimee S. Morgans, and Peter J. Schmid. Optimising the acoustic damping of multiple Helmholtz resonators attached to a thin annular duct. *Journal of Sound and Vibration*, 444:69–84, March 2019.

[84] Krzysztof J. Fidkowski and Philip L. Roe. An Entropy Adjoint Approach to Mesh Refinement. *SIAM Journal on Scientific Computing*, 32(3):1261–1287, January 2010.

[85] Mass Per Pettersson, Gianluca Iaccarino, and Jan Nordström. *Polynomial Chaos Methods for Hyperbolic Partial Differential Equations*. Springer International Publishing, 2015.

[86] Habib N. Najm. Uncertainty Quantification and Polynomial Chaos Techniques in Computational Fluid Dynamics. *Annual Review of Fluid Mechanics*, 41(1):35–52, January 2009.

[87] Hai Wang and David A. Sheen. Combustion kinetic model uncertainty quantification, propagation and minimization. *Progress in Energy and Combustion Science*, 47:1–31, April 2015.

[88] Philip Beran, Bret Stanford, and Christopher Schrock. Uncertainty Quantification in Aeroelasticity. *Annual Review of Fluid Mechanics*, 49(1):361–386, January 2017.

[89] M. Bauerheim, A. Ndiaye, P. Constantine, S. Moreau, and F. Nicoud. Symmetry breaking of azimuthal thermoacoustic modes: the UQ perspective. *Journal of Fluid Mechanics*, 789:534–566, February 2016.

[90] Alexander Avdonin, Stefan Jaensch, Camilo F. Silva, Matic Češnovar, and Wolfgang Polifke. Uncertainty quantification and sensitivity analysis of thermoacoustic stability with non-intrusive polynomial chaos expansion. *Combustion and Flame*, 189:300–310, March 2018.

[91] Camilo F. Silva, Per Pettersson, Gianluca Iaccarino, and Matthias Ihme. Generalized chaos expansion of state space models for uncertainty quantification in thermoacoustics. Technical report, Center for Turbulence Research, 2018.

[92] Paul G. Constantine. *Active Subspaces: Emerging Ideas for Dimension Reduction in Parameter Studies*. Society for Industrial and Applied Mathematics, Philadelphia, PA, March 2015.

[93] Karthik Duraisamy, Gianluca Iaccarino, and Heng Xiao. Turbulence Modeling in the Age of Data. *Annual Review of Fluid Mechanics*, 51(1):357–377, January 2019.

[94] Steven L. Brunton, Bernd R. Noack, and Petros Koumoutsakos. Machine Learning for Fluid Mechanics. *Annual Review of Fluid Mechanics*, 52(1):477–508, January 2020.

[95] E. T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge, 2003.

[96] R. T. Cox. Probability, Frequency and Reasonable Expectation. *American Journal of Physics*, 14(1):1–13, January 1946.

[97] D. R. Cox. *Principles of Statistical Inference*. Cambridge University Press, Cambridge, 2006.

[98] Otto Forster. *Analysis 3*. Vieweg+Teubner Verlag, 2012.

[99] David J. C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[100] Christopher M. Bishop. *Pattern recognition and machine learning*. Information science and statistics. Springer, New York, 2006.

[101] Andrew H. Jazwinski. *Stochastic Processes and Filtering Theory*. Dover Publications, 2007.

[102] Carl E. Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, 2006.

[103] Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Edouard Duchesnay. Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, November 2011.

[104] David Kristjanson Duvenaud. *Automatic Model Construction with Gaussian Processes*. PhD thesis, University of Cambridge, June 2014.

[105] Desmond J. Higham. An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations. *SIAM Review*, 43(3):525–546, January 2001.

[106] David J. C. MacKay. Hyperparameters: Optimize, or Integrate Out? In Glenn R. Heidbreder, editor, *Maximum Entropy and Bayesian Methods*, pages 43–59. Springer Netherlands, Dordrecht, 1996.

[107] Otto Forster. *Analysis 2*. Springer Fachmedien Wiesbaden, 2013.

[108] Wolfgang Dahmen and Arnold Reusken. *Numerik für Ingenieure und Naturwissenschaftler*. Springer-Lehrbuch. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[109] Gerd Fischer. *Lineare Algebra*. Springer Fachmedien Wiesbaden, 2014.

[110] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge University Press, New York, NY, second edition, 2017.

[111] Anne Greenbaum, Ren-Cang Li, and Michael L. Overton. First-Order Perturbation Theory for Eigenvalues and Eigenvectors. *SIAM Review*, 62(2):463–482, January 2020.

[112] E. Anderson, Z. Bai, C. Bischof, L. S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen. *LAPACK Users' Guide*. Society for Industrial and Applied Mathematics, third edition, January 1999.

[113] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, March 2020.

[114] Geir Evensen. *Data Assimilation*. Springer Berlin Heidelberg, 2009.

[115] Freeman Dyson. A meeting with Enrico Fermi. *Nature*, 427(6972):297–297, January 2004.

[116] Radford M. Neal. *Bayesian Learning for Neural Networks*. PhD thesis, University of Toronto, March 1995.

[117] A. P. Dempster, N. M. Laird, and D. M. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.

[118] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems*. Elsevier, 1988.

[119] Georgios Rigas, Nicholas P. Jamieson, Larry K. B. Li, and Matthew P. Juniper. Experimental sensitivity analysis and control of thermoacoustic systems. *Journal of Fluid Mechanics*, 787, January 2016.

[120] M. J. Lighthill. The response of laminar skin friction and heat transfer to fluctuations in the stream velocity. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 224(1156):1–23, June 1954.

[121] Michael Schumm, Eberhard Berger, and Peter A. Monkewitz. Self-excited oscillations in the wake of two-dimensional bluff bodies and their control. *Journal of Fluid Mechanics*, 271:17, July 1994.

[122] D. J. Tritton. *Physical Fluid Dynamics*. Oxford Science Publications. Clarendon Press, Second edition, 1988.

[123] G. K. Batchelor. *An Introduction to Fluid Dynamics*. Cambridge University Press, 2000.

[124] R. P. Benedict, N. A. Carlucci, and S. D. Swetz. Flow losses in abrupt enlargements and contractions. *Journal of Engineering for Power*, 88(1):73, 1966.

[125] Hans Dieter Baehr and Karl Stephan. *Wärme- und Stoffübertragung*. Springer Berlin Heidelberg, 2013.

[126] Stuart W. Churchill and Humbert H.S. Chu. Correlating equations for laminar and turbulent free convection from a vertical plate. *International Journal of Heat and Mass Transfer*, 18(11):1323–1329, November 1975.

[127] Harold Levine and Julian Schwinger. On the radiation of sound from an unflanged circular pipe. *Physical Review*, 73(4):383–406, February 1948.

[128] Simo Sarkka. *Bayesian Filtering and Smoothing*. Cambridge University Press, 2013.

[129] Richard Bellman. *Dynamic programming*. Dover Publications, 2003.

[130] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35, 1960.

[131] R. E. Kalman and R. S. Bucy. New Results in Linear Filtering and Prediction Theory. *Journal of Basic Engineering*, 83(1):95, 1961.

[132] Arthur Gelb, editor. *Applied optimal estimation*. MIT Press, 1974.

[133] Robert N. Miller, Michael Ghil, and François Gauthiez. Advanced data assimilation in strongly nonlinear dynamical systems. *Journal of the Atmospheric Sciences*, 51(8):1037–1056, April 1994.

[134] Geir Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5):10143–10162, May 1994.

[135] Gerrit Burgers, Peter Jan van Leeuwen, and Geir Evensen. Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, 126(6):1719–1724, June 1998.

[136] Jeffrey S. Whitaker and Thomas M. Hamill. Ensemble Data Assimilation without Perturbed Observations. *Monthly Weather Review*, 130(7):1913–1924, July 2002.

[137] Michael K. Tippett, Jeffrey L. Anderson, Craig H. Bishop, Thomas M. Hamill, and Jeffrey S. Whitaker. Ensemble Square Root Filters. *Monthly Weather Review*, 131(7):1485–1490, July 2003.

[138] David M. Livings, Sarah L. Dance, and Nancy K. Nichols. Unbiased ensemble square root filters. *Physica D: Nonlinear Phenomena*, 237(8):1021–1028, June 2008.

[139] F. H. Schlee, C. J. Standish, and N. F. Toda. Divergence in the Kalman filter. *AIAA Journal*, 5(6):1114–1120, June 1967.

[140] Thomas M. Hamill and Jeffrey S. Whitaker. Accounting for the Error due to Unresolved Scales in Ensemble Data Assimilation: A Comparison of Different Approaches. *Monthly Weather Review*, 133(11):3132–3147, November 2005.

[141] J. A. Sethian. Evolution, Implementation, and Application of Level Set and Fast Marching Methods for Advancing Fronts. *Journal of Computational Physics*, 169(2):503–555, May 2001.

[142] Frederic Gibou, Ronald Fedkiw, and Stanley Osher. A review of level-set methods and some recent applications. *Journal of Computational Physics*, 353:82–109, January 2018.

[143] Stanley Osher and James A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79(1):12–49, November 1988.

[144] J. A. Sethian and Peter Smereka. Level Set Methods for Fluid Interfaces. *Annual Review of Fluid Mechanics*, 35(1):341–372, January 2003.

[145] James Albert Sethian. *Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*. Number 3 in Cambridge monographs on applied and computational mathematics. Cambridge University Press, second edition, 1999.

[146] Stanley Osher and Ronald P. Fedkiw. Level Set Methods: An Overview and Some Recent Results. *Journal of Computational Physics*, 169(2):463–502, May 2001.

[147] Wolfgang Kollmann. *Fluid Mechanics in Spatial and Material Description*. University Readers, July 2010.

[148] William J. Rider and Douglas B. Kothe. Reconstructing volume tracking. *Journal of Computational Physics*, 141(2):112–152, 1998.

[149] Hong-Kai Zhao, T. Chan, B. Merriman, and S. Osher. A Variational Level Set Approach to Multiphase Motion. *Journal of Computational Physics*, 127(1):179–195, August 1996.

[150] Stanley J. Osher and Fadil Santosa. Level Set Methods for Optimization Problems Involving Geometry and Constraints. *Journal of Computational Physics*, 171(1):272–288, July 2001.

[151] J. A. Sethian and Andreas Wiegmann. Structural Boundary Design via Level Set and Immersed Interface Methods. *Journal of Computational Physics*, 163(2):489–528, September 2000.

[152] Grégoire Allaire, François Jouve, and Anca-Maria Toader. Structural optimization using sensitivity analysis and a level-set method. *Journal of Computational Physics*, 194(1):363–393, February 2004.

[153] Frédéric Chantalat, Charles-Henri Bruneau, Cédric Galusinski, and Angelo Iollo. Level-set, penalization and cartesian meshes: A paradigm for inverse problems and optimal design. *Journal of Computational Physics*, 228(17):6291–6315, September 2009.

[154] Robert F. Stengel. *Optimal control and estimation*. Dover Publications, 1994.

[155] D. Rozier, F. Birol, E. Cosme, P. Brasseur, J. M. Brankart, and J. Verron. A Reduced-Order Kalman Filter for Data Assimilation in Physical Oceanography. *SIAM Review*, 49(3):449–465, January 2007.

[156] C. H. Colburn, J. B. Cessna, and T. R. Bewley. State estimation in wall-bounded flow systems. Part 3. The ensemble Kalman filter. *Journal of Fluid Mechanics*, 682:289–303, September 2011.

[157] Hiroshi Kato, Akira Yoshizawa, Genta Ueno, and Shigeru Obayashi. A data assimilation methodology for reconstructing turbulent flows around aircraft. *Journal of Computational Physics*, 283:559–581, February 2015.

[158] V. Mons, J.-C. Chassaing, T. Gomez, and P. Sagaut. Reconstruction of unsteady viscous flows using data assimilation schemes. *Journal of Computational Physics*, 316:255–280, July 2016.

[159] H. Xiao, J.-L. Wu, J.-X. Wang, R. Sun, and C.J. Roy. Quantifying and reducing model-form uncertainties in Reynolds-averaged Navier-Stokes simulations: A data-driven, physics-informed Bayesian approach. *Journal of Computational Physics*, 324:115–136, November 2016.

[160] Darwin Darakananda, André Fernando de Castro da Silva, Tim Colonius, and Jeff D. Eldredge. Data-assimilated low-order vortex modeling of separated flows. *Physical Review Fluids*, 3(12), December 2018.

[161] Jeffrey W. Labahn, Hao Wu, Bruno Coriton, Jonathan H. Frank, and Matthias Ihme. Data assimilation using high-speed measurements and LES to examine local extinction events in turbulent flames. *Proceedings of the Combustion Institute*, June 2018.

[162] Jonathan D. Beezley and Jan Mandel. Morphing ensemble Kalman filters. *Tellus A: Dynamic Meteorology and Oceanography*, 60(1):131–140, January 2008.

[163] Jan Mandel, Lynn S. Bennethum, Jonathan D. Beezley, Janice L. Coen, Craig C. Douglas, Minjeong Kim, and Anthony Vodacek. A wildland fire model with data assimilation. *Mathematics and Computers in Simulation*, 79(3):584–606, December 2008.

[164] Jan Mandel, Jonathan D. Beezley, and Adam K. Kochanski. Coupled atmosphere-wildland fire modeling with WRF-Fire. *Geoscientific Model Development*, 4(3):591–610, July 2011. arXiv: 1102.1343.

[165] M.C. Rochoux, A. Collin, C. Zhang, A. Trouvé, D. Lucor, and P. Moireau. Front Shape Similarity Measure for Shape-Oriented Sensitivity Analysis and Data Assimilation for Eikonal Equation. *ESAIM: Proceedings and Surveys*, 63:258–279, 2018.

[166] C. Zhang, A. Collin, P. Moireau, A. Trouvé, and M.C. Rochoux. Front shape similarity measure for data-driven simulations of wildland fire spread based on state estimation: Application to the RxCADRE field-scale experiment. *Proceedings of the Combustion Institute*, 37(3):4201–4209, 2019.

[167] Long Li, Francois-Xavier Le Dimet, Jianwei Ma, and Arthur Vidard. A Level-Set-Based Image Assimilation Method: Potential Applications for Predicting the Movement of Oil Spills. *IEEE Transactions on Geoscience and Remote Sensing*, 55(11):6330–6343, November 2017.

[168] Long Li, Arthur Vidard, François-Xavier Le Dimet, and Jianwei Ma. Topological data assimilation using Wasserstein distance. *Inverse Problems*, 35(1):015006, January 2019.

[169] Mélanie C. Rochoux, Bénédicte Cuenot, Sophie Ricci, Arnaud Trouvé, Blaise Delmotte, Sébastien Massart, Roberto Paoli, and Ronan Paugam. Data assimilation applied to combustion. *Comptes Rendus Mécanique*, 341(1-2):266–276, January 2013.

[170] X. Gao, Y. Wang, N. Overton, M. Zupanski, and X. Tu. Data-assimilated computational fluid dynamics modeling of convection-diffusion-reaction problems. *Journal of Computational Science*, 21:38–59, July 2017.

[171] David L. Moreno and Sigurd I. Aanonsen. Stochastic Facies Modelling Using the Level Set Method. In *Petroleum Geostatistics*, Cascais, Portugal, September 2007.

[172] Danping Peng, Barry Merriman, Stanley Osher, Hongkai Zhao, and Myungjoo Kang. A PDE-Based Fast Local Level Set Method. *Journal of Computational Physics*, 155(2):410–438, November 1999.

[173] J. A. Sethian. A fast marching level set method for monotonically advancing fronts. *Proceedings of the National Academy of Sciences of the United States of America*, 93(4):1591–1595, February 1996.

[174] V. I. Arnold. *Mathematical Methods of Classical Mechanics*, volume 60 of *Graduate Texts in Mathematics*. Springer New York, second edition, 1989.

[175] J. A. Sethian. Fast Marching Methods. *SIAM Review*, 41(2):199–235, January 1999.

[176] David Adalsteinsson and James A. Sethian. A Fast Level Set Method for Propagating Interfaces. *Journal of Computational Physics*, 118(2):269–277, May 1995.

[177] Mark Sussman, Peter Smereka, and Stanley Osher. A Level Set Approach for Computing Solutions to Incompressible Two-Phase Flow. *Journal of Computational Physics*, 114(1):146–159, September 1994.

[178] G. Szego. *Orthogonal Polynomials*, volume 23 of *Colloquium Publications*. American Mathematical Society, Providence, Rhode Island, December 1939.

[179] Alessandro Orchini and Matthew P. Juniper. Flame Double Input Describing Function analysis. *Combustion and Flame*, 171:87–102, September 2016.

[180] Sebastian Reich and Colin Cotter. *Probabilistic Forecasting and Bayesian Data Assimilation*. Cambridge University Press, Cambridge, 2015.

[181] Karthik Kashinath, Santosh Hemchandra, and Matthew P. Juniper. Nonlinear thermoacoustics of ducted premixed flames: The influence of perturbation convection speed. *Combustion and Flame*, 160(12):2856–2865, December 2013.

[182] Xu-Dong Liu, Stanley Osher, and Tony Chan. Weighted Essentially Non-oscillatory Schemes. *Journal of Computational Physics*, 115(1):200–212, November 1994.

[183] Guang-Shan Jiang and Chi-Wang Shu. Efficient Implementation of Weighted ENO Schemes. *Journal of Computational Physics*, 126(1):202–228, June 1996.

[184] Chi-Wang Shu and Stanley Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes. *Journal of Computational Physics*, 77(2):439–471, August 1988.

[185] R.R. Kerswell. Nonlinear Nonmodal Stability Theory. *Annual Review of Fluid Mechanics*, 50(1):319–345, January 2018.

[186] Peter C. Ma, Yu Lv, and Matthias Ihme. An entropy-stable hybrid scheme for simulations of transcritical real-fluid flows. *Journal of Computational Physics*, 340:330–357, July 2017.

[187] Yaser Khalighi, Frank Ham, Joseph Nichols, Sanjiva Lele, and Parviz Moin. Unstructured Large Eddy Simulation for Prediction of Noise Issued from Turbulent Jets in Various Configurations. In *17th AIAA/CEAS Aeroacoustics Conference (32nd AIAA Aeroacoustics Conference)*, Portland, Oregon, June 2011. American Institute of Aeronautics and Astronautics.

[188] K. Narayanaswamy, G. Blanquart, and H. Pitsch. A consistent chemical mechanism for oxidation of substituted aromatic species. *Combustion and Flame*, 157(10):1879–1898, October 2010.

[189] Paul Tranquilli and Adrian Sandu. Rosenbrock–Krylov Methods for Large Systems of Differential Equations. *SIAM Journal on Scientific Computing*, 36(3):A1313–A1338, January 2014.

[190] Hao Wu, Peter C. Ma, and Matthias Ihme. Efficient time stepping for reactive turbulent simulations with stiff chemistry. *arXiv:1712.00953 [physics]*, December 2017.

[191] Thomas M. Hamill. Interpretation of Rank Histograms for Verifying Ensemble Forecasts. *Monthly Weather Review*, 129(3):550–560, March 2001.

[192] Alexis Cuquel, Daniel Durox, and Thierry Schuller. Impact of flame base dynamics on the non-linear frequency response of conical flames. *Comptes Rendus Mécanique*, 341(1-2):171–180, January 2013.

[193] Stéfan van der Walt, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, and Tony Yu. scikit-image: image processing in python. *PeerJ*, 2:e453, June 2014.

[194] Santosh Hemchandra. *Dynamics of Turbulent Premixed Flames in Acoustic Fields Dynamics of Turbulent Premixed Flames in.* PhD thesis, Georgia Institute of Technology, August 2009.

[195] Herrmann Schlichting and Klaus Gersten. *Grenzschicht-Theorie*. Springer, 2006.

[196] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, 2006.

[197] David G Goodwin, Raymond L Speth, Harry K Moffat, and Bryan W Weber. Cantera: An Object-oriented Software Toolkit for Chemical Kinetics, Thermodynamics, and Transport Processes, August 2018.

[198] P. Clavin and F. A. Williams. Effects of molecular diffusion and of thermal expansion on the structure and dynamics of premixed flames in turbulent flows of large scale and low intensity. *Journal of Fluid Mechanics*, 116:251–282, March 1982.

[199] M. Matalon and B. J. Matkowsky. Flames as gasdynamic discontinuities. *Journal of Fluid Mechanics*, 124:239, November 1982.

[200] L. Landau. On the theory of slow combustion. In Pierre Pelcé, editor, *Dynamics of Curved Fronts*, pages 403–411. Academic Press, San Diego, 1988.

[201] Radford M. Neal. Probabilistic Inference using Markov Chain Monte Carlo Methods. Technical report, University of Toronto, 1993.

[202] Charles D. Pierce. *Progress-Variable Approach for Large-Eddy Simulation of Turbulent Combustion*. PhD thesis, Stanford University, June 2001.

[203] M. Meldi and A. Poux. A reduced order model based on Kalman filtering for sequential data assimilation of turbulent flows. *Journal of Computational Physics*, 347:207–234, October 2017.

[204] Alexandre Joel Chorin. A numerical method for solving incompressible viscous flow problems. *Journal of Computational Physics*, 2(1):12–26, August 1967.

[205] Eric J. Parish and Karthik Duraisamy. A paradigm for data-driven predictive modeling using field inversion and machine learning. *Journal of Computational Physics*, 305:758–774, January 2016.

[206] Anand Pratap Singh, Shivaji Medida, and Karthik Duraisamy. Machine-Learning-Augmented Predictive Modeling of Turbulent Separated Flows over Airfoils. *AIAA Journal*, 55(7):2215–2227, July 2017.

[207] Leo Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, August 1996.

[208] Samuel H. Rudy, Steven L. Brunton, and J. Nathan Kutz. Smoothing and parameter estimation by soft-adherence to governing equations. *Journal of Computational Physics*, 398:108860, December 2019.

Appendix A

Normal distributions

Definition A.1 (Normal distribution)**.** For the mean $\mu \in \mathbb{R}^n$ and the covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$, the probability density function of a normally distributed random variable $X \sim \mathcal{N}(\mu, \Sigma)$ is given by

$$P(X = x) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu)^{\mathrm{T}}\Sigma^{-1}(x-\mu)\right) \quad . \tag{A.1}$$

Lemma A.1 (Woodbury formula)**.** *For $A_{11} \in \mathbb{R}^{p \times p}$, $A_{12} \in \mathbb{R}^{p \times q}$, $A_{21} \in \mathbb{R}^{q \times p}$, $A_{22} \in \mathbb{R}^{q \times q}$, the following identity holds:*

$$\left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right)^{-1} = A_{11}^{-1} - A_{11}^{-1}A_{12}\left(A_{21}A_{11}^{-1}A_{12} - A_{22}\right)^{-1}A_{21}A_{11}^{-1} \quad . \tag{A.2}$$

Theorem A.1 (Joint normal distribution)**.** *For a normally distributed random variable $X \sim \mathcal{N}(\mu, \Sigma)$ with*

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \in \mathbb{R}^n \quad , \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \in \mathbb{R}^{n \times n} \quad , \tag{A.3}$$

where $\mu_1 \in \mathbb{R}^m$ and $\mu_2 \in \mathbb{R}^{n-m}$, as well as $\Sigma_{11} \in \mathbb{R}^{m \times m}$, $\Sigma_{12} \in \mathbb{R}^{m \times (n-m)}$, $\Sigma_{21} \in \mathbb{R}^{(n-m) \times m}$, $\Sigma_{22} \in \mathbb{R}^{(n-m) \times (n-m)}$, the probability density function $P(X)$ of the joint probability distribution is given by

$$P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)P(X_2 = x_2 \mid X_1 = x_1) \quad , \tag{A.4}$$

where both X_1 and X_2 are normally distributed random variables.

Corollar A.1 (Marginal normal distribution)**.** *For a normally distributed random variable* $X \sim \mathcal{N}(\mu, \Sigma)$*, the probability density function* $P(X_1)$ *of the marginal probability distribution is given by*

$$P(X_1 = x_1) = \frac{1}{\sqrt{\det(2\pi\Sigma_{11})}} \exp\left(-\frac{1}{2}(x_1 - \mu_1)^{\mathrm{T}} \Sigma_{11}^{-1} (x_1 - \mu_1)\right) \quad . \tag{A.5}$$

Corollar A.2 (Conditional normal distribution)**.** *For a normally distributed random variable* $X \sim \mathcal{N}(\mu, \Sigma)$*, the probability density function* $P(X_2 \mid X_1)$ *of the conditional probability distribution is given by*

$$P(X_2 = x_2 \mid X_1 = x_1) = \frac{1}{\sqrt{\det\left(2\pi\Sigma_{2|1}\right)}} \exp\left(-\frac{1}{2}\left(x_2 - \mu_{2|1}\right)^{\mathrm{T}} \Sigma_{2|1}^{-1} \left(x_2 - \mu_{2|1}\right)\right) \quad , \tag{A.6}$$

where the conditional mean $\mu_{2|1}$ *and the conditional covariance matrix* $\Sigma_{2|1}$ *are given by*

$$\mu_{2|1} = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}\left(x_1 - \mu_1\right) \quad , \tag{A.7}$$

$$\Sigma_{2|1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \quad . \tag{A.8}$$

Appendix B

Hamiltonian mechanics

Lemma B.1 (Lagrangian)**.**

$$\mathscr{L}(r(t),\dot{r}(t),t) = 0 \quad . \tag{B.1}$$

Proof. The generating function is formally the same as the action integral [174, Chapter 9]. Thus, the Lagrangian $\mathscr{L}(r,\dot{r},t)$ is the total derivative of the generating function:

$$\mathscr{L}(r(t),\dot{r}(t),t) = \frac{\mathrm{d}}{\mathrm{d}t}G(r(t),t) = \frac{\mathrm{d}}{\mathrm{d}t}G(r(0),0) = 0 \quad . \tag{B.2}$$

□

Lemma B.2 (Hamiltonian)**.**

$$\mathscr{H}(r(t),n(t),t) = u(r(t)) \cdot n(t) \quad . \tag{B.3}$$

Proof. A Legendre transform of the Lagrangian $\mathscr{L}(r,\dot{r},t)$ gives the Hamiltonian $\mathscr{H}(r,n,t)$ (Eq. (B.1)):

$$\mathscr{H}(r(t),n(t),t) = \dot{r}(t) \cdot n(t) - \mathscr{L}(r(t),\dot{r}(t),t) \tag{B.4}$$

$$= u(r(t)) \cdot n(t) \quad . \tag{B.5}$$

It remains to be shown that $\mathscr{H}(r,n,t)$ is indeed a Hamiltonian. The constraints, Eq. (6.3) and (6.4), are only functions in r and n respectively. Hence, a suitable choice of Lagrange multipliers λ_r and λ_n gives

$$\frac{\partial \mathscr{H}}{\partial n} = u + \lambda_n n = u = \frac{\mathrm{d}r}{\mathrm{d}t} \quad , \tag{B.6}$$

$$\frac{\partial \mathscr{H}}{\partial r} = n \cdot \nabla u + n \times (\nabla \times u) + \lambda_r \frac{\partial G}{\partial r} \tag{B.7}$$

$$= n \cdot \nabla u + n \times (\nabla \times u) - [n \cdot (n \cdot \nabla u)]\, n \tag{B.8}$$

$$= -\frac{\mathrm{d}n}{\mathrm{d}t} \quad . \tag{B.9}$$

Eq. (B.6) follows from the laws of motion (Eq. (6.1)). Eq. (B.9) follows from tensor calculus:

$$\frac{\mathrm{d}n}{\mathrm{d}t} = \frac{\partial n}{\partial t} + u \cdot \nabla n + \mu_n n \quad , \tag{B.10}$$

where μ_n is the Lagrange multiplier for the constraint imposed by Eq. (6.4). The gradient of the total derivative of the generating function gives an expression for the partial derivative of the normal vector (Lemma B.1):

$$\nabla\left(\frac{\mathrm{d}G}{\mathrm{d}t}\right) = \nabla\left(\frac{\partial G}{\partial t} + u \cdot \nabla G\right) = \frac{\partial n}{\partial t} + \nabla (u \cdot n) = 0 \quad . \tag{B.11}$$

The dot product of the normal vector with the total derivative of the normal vector gives an expression for the Lagrange multiplier μ_n:

$$n \cdot \frac{\mathrm{d}n}{\mathrm{d}t} = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}(n \cdot n) = 0 \quad , \tag{B.12}$$

$$n \cdot \frac{\mathrm{d}n}{\mathrm{d}t} = n \cdot \left(\frac{\partial n}{\partial t} + u \cdot \nabla n + \mu_n n\right) \tag{B.13}$$

$$= n \cdot \frac{\partial n}{\partial t} + \frac{1}{2}(u \cdot \nabla)(n \cdot n) + \mu_n n \cdot n \tag{B.14}$$

$$= -n \cdot \nabla (u \cdot n) + \mu_n \quad . \tag{B.15}$$

This gives:

$$\frac{\mathrm{d}n}{\mathrm{d}t} = -\nabla (u \cdot n) + u \cdot \nabla n + [n \cdot \nabla (u \cdot n)]\, n \tag{B.16}$$

$$\begin{aligned} = & -u \cdot \nabla n - n \cdot \nabla u - u \times (\nabla \times n) - n \times (\nabla \times u) + u \cdot \nabla n + [(n \cdot \nabla u) \cdot n]\, n \\ & + [(n \cdot \nabla n) \cdot u]\, n \end{aligned} \tag{B.17}$$

$$= -n \cdot \nabla u - n \times (\nabla \times u) + n\,[n \cdot (n \cdot \nabla u)] \quad . \tag{B.18}$$

Note that $\nabla \times n = 0$ because $n = \nabla G$ [107, Chapter 5]. This concludes the proof that $\mathscr{H}(r, n, t)$ is a Hamiltonian. □

Theorem B.1 (Hamilton-Jacobi equation)**.**

$$\frac{\partial G}{\partial t} + u(r(t)) \cdot n(t) = 0 \quad . \tag{B.19}$$

Proof. The Hamilton-Jacobi equation is given by [174, Chapter 9]

$$\frac{\partial G}{\partial t} + \mathscr{H} = \frac{\partial G}{\partial t} + u \cdot n = 0 \quad . \tag{B.20}$$

□

Index