UNIVERSITY OF CALIFORNIA,
Santa Barbara

Topics in Modeling and Control of Spatially Distributed Systems

A dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy in Electrical and Computer Engineering
by

Jonathan Peter Epperlein

Committee:
Professor Bassam Bamieh, Chair
Professor João Hespanha
Professor Brad Paden
Professor Andrew Teel

December 2014

The dissertation of Jonathan Peter Epperlein is approved:

_____

Professor Andrew Teel

_____

Professor Brad Paden

_____

Professor João Hespanha

_____

Professor Bassam Bamieh, Chair

December 2014

Topics in Modeling and Control of Spatially Distributed Systems

Meiner Familie; vergangen, gegenwärtig, und zukünftig.

```
Z  G  V  O  L  J  E  N  Y  O  J  E  N  V  N  V  O
Y  J  D  M  H  A  A  T  W  W  Q  J  K  T  I  T  M
B  N  L  A  H  S  M  A  R  I  O  O  N  I  C  K  A
P  U  Y  H  U  C  X  D  L  J  E  S  S  T  K  X  B
A  N  N  E  K  H  M  A  M  A  M  E  L  I  E  H  A
J  D  N  L  U  A  C  A  R  L  A  P  B  S  L  P  E
M  E  D  G  N  T  L  I  L  I  T  H  Z  J  R  A  R
O  N  M  A  T  A  V  F  P  Z  W  I  W  W  A  U  B
G  W  O  E  Z  F  B  F  B  L  A  N  E  H  J  L  E
L  A  M  P  M  K  J  A  N  I  N  E  N  I  E  A  L
I  R  O  H  E  A  Z  N  O  K  R  I  S  T  I  N  A
N  I  L  I  I  T  G  E  R  T  R  U  D  N  T  A  E
O  E  I  L  S  A  Y  W  A  N  N  I  K  E  I  L  M
T  L  A  I  T  F  A  L  Y  S  S  A  Q  Y  L  I  I
U  R  B  P  E  T  O  P  A  L  E  O  K  H  M  Z  L
W  E  G  P  R  U  B  E  N  J  U  L  I  A  A  Z  Y
R  L  G  P  A  P  A  N  D  R  E  A  S  D  N  A  N
```

---

[1] Random roommates:
Stochasticity works out
perfectly, sometimes.

Curriculum Vitae

Education

06/2011–12/2014 — **University Of California, Santa Barbara**
Ph.D. in Electrical and Computer Engineering, Emphasis in Controls

09/2009–06/2011 — **University Of California, Santa Barbara**
MSc. in Electrical and Computer Engineering

10/2002–06/2008 — **Otto-von-Guericke University | Magdeburg, Germany**
Diplom Ingenieur of Engineering Cybernetics

Research Experience

07/2008–06/2009 — **Harvard Medical School | Boston, MA**
Visiting Scholar, Sorger Lab

09/2007–05/2008 — **Yale University | New Haven, CT**
Visiting Student, Center for Systems Science

10/2006–05/2007 — **Hamilton Institute | Maynooth, Ireland**
Visiting Student, Communication Networks Group

12/2003–03/2005 — **Max-Planck-Institute | Magdeburg, Germany**
Student Assistant, Systems Biology and Systems & Control Theory Groups

Vocational Experience

04/2006–08/2006 — **Robert-Bosch-GmbH | Schwieberdingen, Germany**
Intern, Corporate Research

Awards etc.

Spring 2013 ————— Center for Control, Dynamical Systems, and Computation (CCDC) Fellowship

04/2009 ————— CCDC Outstanding Scholar Fellowship

2005–2007 ————— German National Academic Foundation Scholarship

Publications

Jonathan P. Epperlein, Bassam Bamieh, and Karl J. Åström, "*ThermoAcoustics and the Rijke tube: Experiments, identification and modeling*," Control Systems Magazine, to appear

Jonathan P. Epperlein, Bassam Bamieh, and Karl J. Åström, "*Control laboratory experiments in ThermoAcoustics using the Rijke tube*, 2014 American Control Conference

Jonathan P. Epperlein and Bassam Bamieh, "*Distributed control of spatially invariant systems over Sobolev spaces*," 2014 European Control Conference

Jonathan P. Epperlein and Bassam Bamieh, "*A frequency domain method for optimal periodic control*," 2012 American Control Conference

J. Saez Rodriguez, L. Alexopoulos, **J. Epperlein**, R. Samaga, D. A. Lauffenburger, S. Klamt and P. K. Sorger, "*Discrete logic modeling as a means to link protein signaling networks with functional analysis of mammalian signal transduction*," Molecular Systems Biology. Mol Syst Biol 5: 331, 2009

D. J. Leith, C. M. Kellett, P. Clifford, **J. Epperlein** and R. N. Shorten, "*A Non-Invasive Method for Link Upgrade Planning Using Coarse-Grained Measurements*," IEEE Communications Letters 11(12) Dec. 2007

ABSTRACT

Topics in Modeling and Control of Spatially Distributed Systems

by

Jonathan Peter Epperlein

This dissertation consists of three parts centered around the topic of spatially distributed systems.

The first part treats a specific spatially distributed system, the so-called Rijke tube, an experiment illustrating the unstable interplay of heat exchange and gas dynamics. The experiment is described in detail and it is demonstrated how closed-loop system identification tools can be applied to obtain a transfer function model, before a spatially distributed model is developed and analyzed. The model in its most idealized form can be described in the frequency domain by a matrix of non-rational transfer functions, which facilitates analysis with classical methods such as the root locus.

The second part considers the following problem: for a given plant and cost function, could there be a finite-length periodic trajectory that achieves better performance than the optimal steady state? Termed optimal periodic control (OPC), this problem has received attention over several decades, however most available methods employ state-space based methods and hence scale very badly with plant dimension. Here, the problem is approached from a frequency-domain perspective, and methods whose complexity is independent of system dimension are developed by recasting the OPC problem for linear plants with certain memoryless polynomial nonlinearities as the problem of minimizing a polynomial.

Finally, the third part extends results for a special class within spatially distributed systems, that

of spatially invariant systems, from systems defined on $\mathfrak{L}_2$ (square-integrable) spaces to systems whose state space is an inner-product Sobolev space as they arise when considering systems of higher temporal order. It is shown how standard results on exponential stability, stabilizability and LQ control can be generalized by carefully keeping track of spatial frequency weighting functions related to the Sobolev inner products, and simple recipes for doing so are given.

Contents

Introduction

This dissertation consists of three seemingly disjointed parts, but a common thread is the modeling and treatment of distributed-parameter systems; further connections between the parts are pointed out as they come up. Still, each part is presented as a chapter with its own introduction, summary and appendices, and each can be read without making any reference to the others.

The first and longest chapter concerns *thermoacoustic instability* and its simplest incarnation, the Rijke tube. Thermoacoustic instabilities can be encountered if heat is released, typically in the form of combustion, into an acoustic resonator such as a pipe; practical examples include ramjets, where the entire engine serves as a resonator, certain kinds of rockets, where the combustion chamber is the resonator, industrial furnaces, and so on. They manifest themselves as resonating sound waves, typically very powerful and audible.

In these examples, the instability is unwanted, and efforts are directed at stabilization, i.e. suppression of the resonant pressure wave. Another interpretation of the process – and the one that inspired the investigations in this dissertation – is that of *energy conversion*: The thermoacoustic instability receives energy in the form of heat and converts it to a different form of energy, namely acoustic energy.

The Rijke tube as the canonical example of a thermoacoustic instability is deceptively simple to build: a slender tube is set upright, with both ends open, and a heating element, such as a hot wire, is place in its lower third. If enough heat is supplied, a hum will start to swell

until it reaches saturation. Despite this simplicity, derivation of a model and explanation of the observed effect are far from simple; Rijke's own explanation was not sufficient.

The chapter opens with a description of the entire experiment and then proceeds to present results of a closed-loop system identification, before getting to the main contribution, the derivation of a distributed-parameter model of the Rijke tube.

Much effort went into making the exposition self-contained, and it starts from the conservation equations of fluid dynamics, which are then adapted to the Rijke tube, linearized and rewritten in dimensionless parameters. The normalization allows statements about the relative importance of modeled effects to be made, namely it can be seen that a buoyancy-driven upwards flow and wall friction play minor roles only, while diffusive effects have even less impact.

Neglecting all three leads to a very simple transcendental transfer function model which, to the best of my knowledge, has not appeared in the literature before. The model is then shown to predict the hum and its frequency by using a root-locus argument.

In the last few sections, several possible refinements of the model by including previously neglected effects are investigated numerically.

In Chapter 2, the subject of optimal periodic control (OPC) is treated, with an eye on the application of optimizing the periodic energy conversion process that is the Rijke tube; however the developed methods are not yet general enough. OPC, or the calculation of optimal periodic trajectories for a given control system, has been the subject of much research in the past, but available methods typically suffer from the "curse of dimensionality" in that the computational effort grows with the dimension of the system under investigation.

Frequency-domain methods circumnavigate this dependence on system dimension, but they also limit application to problems for which frequency-domain descriptions can be found. The methods developed in this dissertation allow for problem structures with linear, but possibly infinite-dimensional dynamics, certain memoryless polynomial nonlinearities and polynomial

2

cost.[2]

Two different approaches are taken. The first is to employ variational calculus to derive first-order optimality conditions in the form of a Hamiltonian system. It turns out that all periodic solutions of the Hamiltonian system satisfy the first-order optimality conditions. Harmonic balance, a generalization of the describing function method, is a frequency-domain-based method of identifying periodic trajectories, and due to the structure of plant and cost, the harmonic balance can be rewritten as a system of (multivariate) polynomial equations, the roots of which correspond to the Fourier coefficients of trajectories satisfying the first-order optimality conditions.

The second approach does away with variational calculus but instead uses the Fourier series representation of all involved signals and Parseval's identity to rewrite the cost function as a polynomial in the Fourier coefficients. While it is neat that the system of polynomial equations obtained using the first approach can be derived from this second approach by setting the gradient of the cost to zero, the more important observations are that: (a) minimizing a polynomial can be done in more efficient ways than by finding its extrema by setting the gradient to zero, and it is demonstrated how Sum-of-Squares programming can be applied; (b) avoiding variational calculus allows us to treat more general structures than the first approach does, in particular the structure shown in Figure 2.13 is quite similar to the Rijke tube model shown in Figure 1.15, but the method is not yet general enough to apply. Extension to more general problems is the subject of future research.

The last chapter presents an extension of results on frequency-domain treatment of spatially-invariant systems shown in [BPD02] from systems whose state space consists of square-integrable, or $\mathcal{L}_2$, functions to systems on slightly more general state spaces, namely Sobolev spaces which require functions and some of their spatial derivatives to be square-integrable. This requires to consider weighted spaces in the frequency domain, and these additional fre-

---

[2]Technically, the running cost is meant: if the cost is $J = \int L(x, u)\mathrm{d}t$, it is required that L be polynomial.

quency weights need to be carefully accounted for. Doing so enables the solution of stability, stabilizability, and LQR problems for spatially-invariant systems on Sobolev spaces by solving a family of standard Lyapunov or Riccati equations.

Commonly, such Sobolev state spaces are encountered when temporal derivatives of order 2 or higher are involved; typical examples are the wave and beam equation. A particular realization of the wave equation, but interestingly one that does not call for Sobolev spaces, is encountered in Rijke tube modeling process as (1.27), which sparked the interest in this subject.

Chapter 1

Thermoacoustics and the Rijke Tube

The term *thermoacoustics* refers to the interplay of heat exchange and the dynamics of the thermodynamic state (in particular the pressure) and velocity of the fluid(s) that the heat is exchanged with. These interconnections can be unstable, leading to large pressure oscillations, and the Rijke tube experiment is the probably simplest illustration of this phenomenon.

In this chapter, contributions to the modeling and interpretation from a control-systems perspective of thermoacoustic systems, in particular the Rijke tube, are made. A very brief introduction to the phenomenon of thermoacoustic instabilities and a rudimentary overview over the vast literature on them is given in Section 1.1; the Rijke tube experiment is described next in some detail to gain familiarity with the system (Section 1.2). Section 1.3 describes how a model for the system can be obtained by using simple closed-loop system identification tools, and, finally, in Section 1.4 a spatially distributed model of the Rijke tube is developed from first principles.

Some of the material of this chapter is the subject of the

Related Publications:

[EBÅar] Jonathan P. Epperlein, Bassam Bamieh, and Karl J. Åström. ThermoAcoustics and

the Rijke tube: Experiments, identification and modeling. *Control Systems Magazine*, to appear.

[EBÅ14] Jonathan P. Epperlein, Bassam Bamieh, and Karl J. Åström. Control laboratory experiments in ThermoAcoustics using the Rijke tube. In *2014 American Control Conference*, 2014, pp. 2550–2556.

1.1   Thermoacoustic Instabilities

When heat is released into gas in underdamped acoustic cavities, sometimes a powerful sound is generated. This is a manifestation of the heat transferred to the gas generating and sustaining a standing pressure wave, in other words an unstable coupling between heat release rate and the thermodynamic state of the gas; hence, such phenomena are known as *thermoacoustic*, or – when the heat transfer is through a flame – *combustion instabilities*.

Thermoacoustic instabilities occur e.g. in ramjets, furnaces and solid-fuel rockets [RBFB93]. In such applications, they are undesirable due to the additional mechanical strain that the pressure waves cause on the components, and much of the research in the controls community has been focussed on avoiding the instabilities by stabilizing the interaction between heat exchange and acoustic gas dynamics; model-based approaches are described in e.g. [HAFG98, BAKJ04, AFR⁺00, TD92, ZN97, Can02, DM05, Hec88].

A different interpretation of the thermoacoustic instability is that it constitutes an energy conversion process: energy supplied as heat is converted into acoustic energy in the form of pressure waves. Hence, the process can be considered a *thermoacoustic engine*. This application has received less attention in the literature, important exceptions being e.g. [BS00] and the review article [Swi88]. There is also the very accessible *American Scientist* article [GB00].

In practical applications, thermoacoustic instabilities are commonly encountered as combus-

tion instabilities, and are notoriously difficult to model due to the additional complexity of combustion dynamics [McI86, McI90]. The *Rijke tube* experiment [Rij59] on the other hand generates thermoacoustic instabilities without needing a combustion process: a slender tube is set upright and a heating element, which may be a resistive heater or a flame heating a metal mesh as in Rijke's original experiments, is placed in the lower half. After a short transient, a loud and steady hum can be heard.

The Rijke tube is thus the simplest experiment conceivable that exhibits thermoacoustic instability, yet it still illustrates a very complex phenomenon, the explanation of which is not quite that simple. Rijke's own explanation [Rij59] for instance was insufficient. Today it is known [Ray78] that the phasing between the fluctuating heat release and the pressure oscillations is such that Rayleigh's criterion [Ray96] is satisfied: more heat is added while the gas is compressed, and less while it expands. The physical model developed in Section 1.4 will show that this behavior corresponds to a linearly unstable feedback interconnection of gas velocity and heat transfer.

1.2   The Rijke Tube Experiment

To observe the thermoacoustic instability Rijke observed in 1859 [Rij59], nothing more is needed than a slender tube and a heating element. The exact dimensions of the tube are not important, as long as it is much longer than wide (Rijke himself used tubes with diameter of about 34 mm and lengths between 0.2 and 0.8 m). This tube is set upright, so that both ends are open ends, and the heating element (nowadays typically a resistive heater) is placed in the lower half of the tube; the effect is strongest if the heater is placed at one quarter of the length. A photograph and a schematic diagram of the setup used for all experiments in this dissertation are shown in Figure 1.1, and details on the materials and components used are given in Section 1.A.
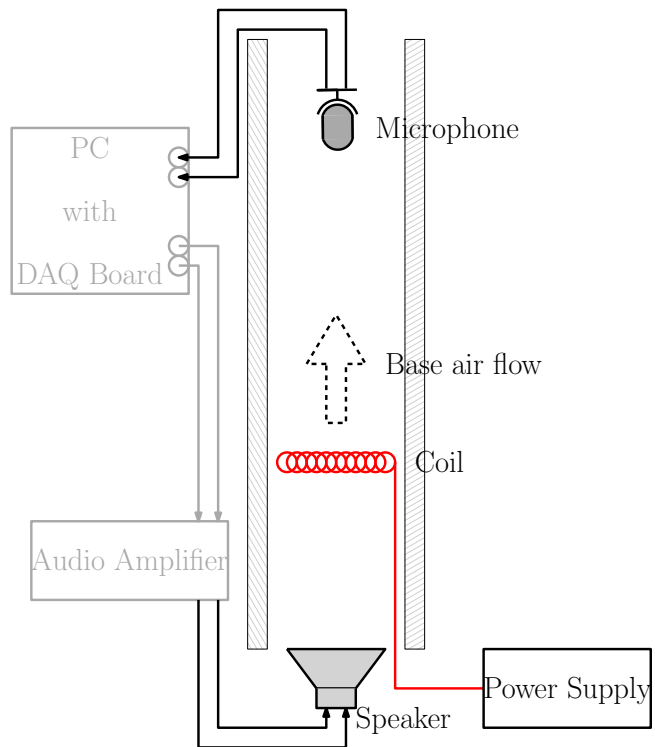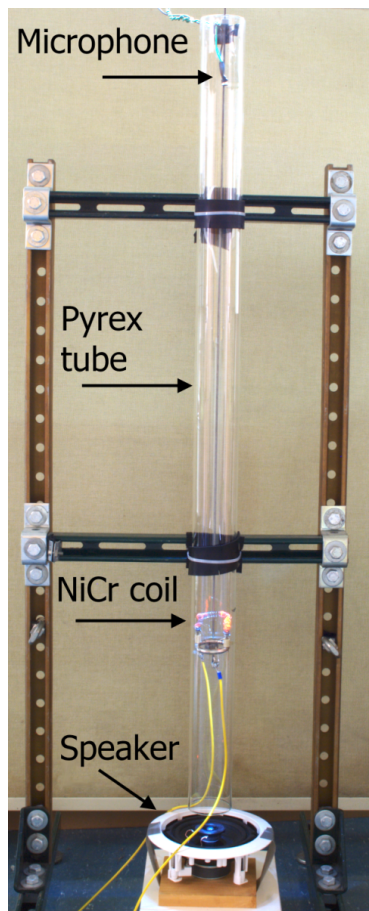
Figure 1.1 – *Photograph and diagram of the Rijke tube experimental apparatus. Only the coil and the power supply are necessary to induce the thermoacoustic instability. Microphone, speaker and PC interface are used to realize control and system identification.*
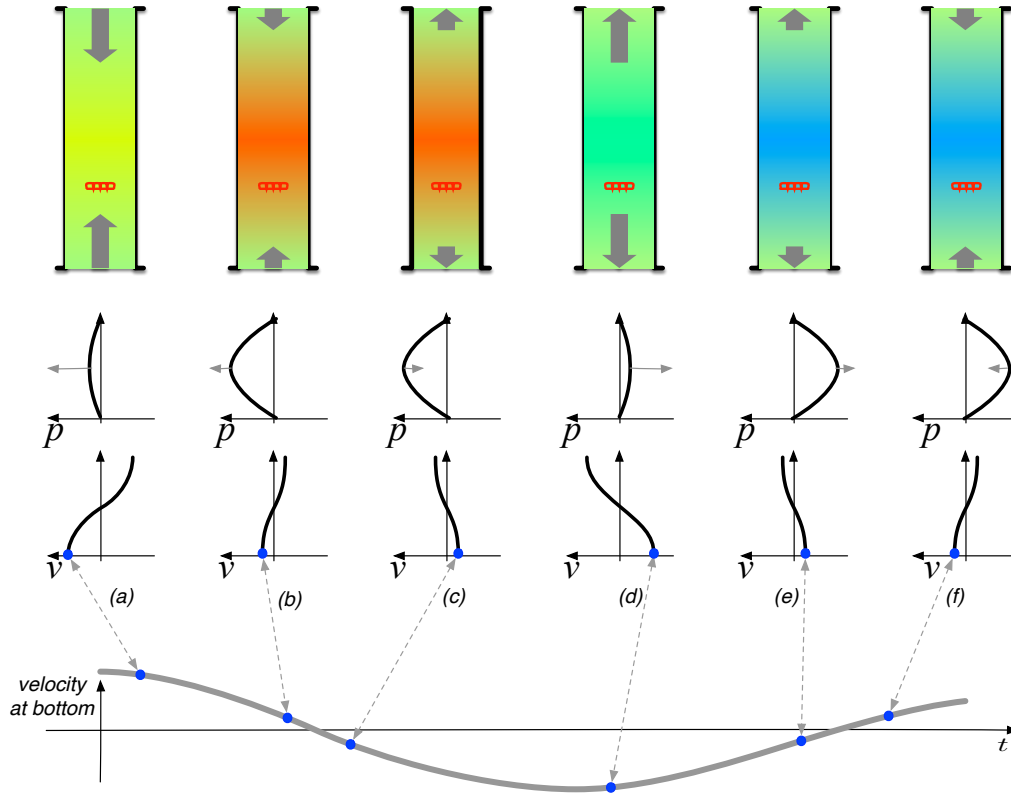
Figure 1.2 – *A diagram of the fundamental acoustic mode of the Rijke tube. In the top three rows, the spatial waveforms are shown. Acoustic pressure is shown in color and the acoustic velocity at the ends is illustrated as arrows. The bottom row depicts one period of the temporal waveform of the velocity at the bottom of the tube. In phase a, the pressure just started increasing in the center of the tube due to the air rushing in, which in b has led to the pressure achieving a maximum at the center, while simultaneously the velocity has been decreased by the resulting pressure gradient. In c, the pressure gradient has inverted the velocity, so that air now starts rushing out of the tube with d increasing velocity, until e the pressure reaches a minimum in the center and the gradient leads to f air being sucked in again, until a pressure moves towards its maximum again and the cycle repeats.*

The power to the heater is then slowly increased until the critical value is reached after which a loud and steady hum appears. The wire will be glowing red at this point. The frequency f_0 of the hum is

$$f_0 \approx \frac{c}{2L}, \tag{1.1}$$

where c is the velocity of sound and L is the tube length. In the specific setup used here, measurement yields $f = 143$ Hz. The wavelength of the sound is $\lambda = 2L$, so it corresponds to a half-wavelength standing wave in the tube, which is the fundamental mode of a tube open at both ends.[1] This mode is illustrated in Figure 1.2, and time series of the pressure are shown in Figure 1.3(a)-(c).

Covering either end of the tube will make the hum disappear instantly; the phenomenon will also not occur if the tube is horizontal. This indicates that the buoyancy-driven upwards flow that sets in as soon as the heater is turned on is vital for the effect to occur. We will return to this observation during the modeling process in Section 1.4.2. Decreasing the power to the heater below the critical value will also cause the hum to die out, but a slight hysteresis might be observed. Increasing the power further once the hum has appeared will lead to a slight increase in sound intensity, but the frequency is unaffected.

As a next step, actuation and measurement are added and a controller is applied. A speaker placed underneath the tube (but not flush with it!) serves as an actuator, while a microphone dangling into the top half of the tube provides a measurement, see Figure 1.1. The microphone measurement is amplified and fed to the speaker, in other words a simple proportional controller is employed.

As the control gain is increased, there is a critical gain value $K_{\min}$ above which the hum will die

---

[1] While it is common to assume that every mode has a pressure node exactly at the open end, in reality the nodes are located slightly outside the tube, and so the wavelength is slightly more than $2L$. One can account for this by using a tube length $L + 2\Delta L$ in the computations; the so-called end correction ΔL generally depends on the radius R of the tube and the wavelength λ of the considered mode, but for $\lambda \gg R$ it is approximately independent of λ: $\Delta L \approx 0.61 \cdot R$ [LS48]. For the tube described in Section 1.A, this yields a discrepancy of roughly 4.6 cm or 8 Hz.
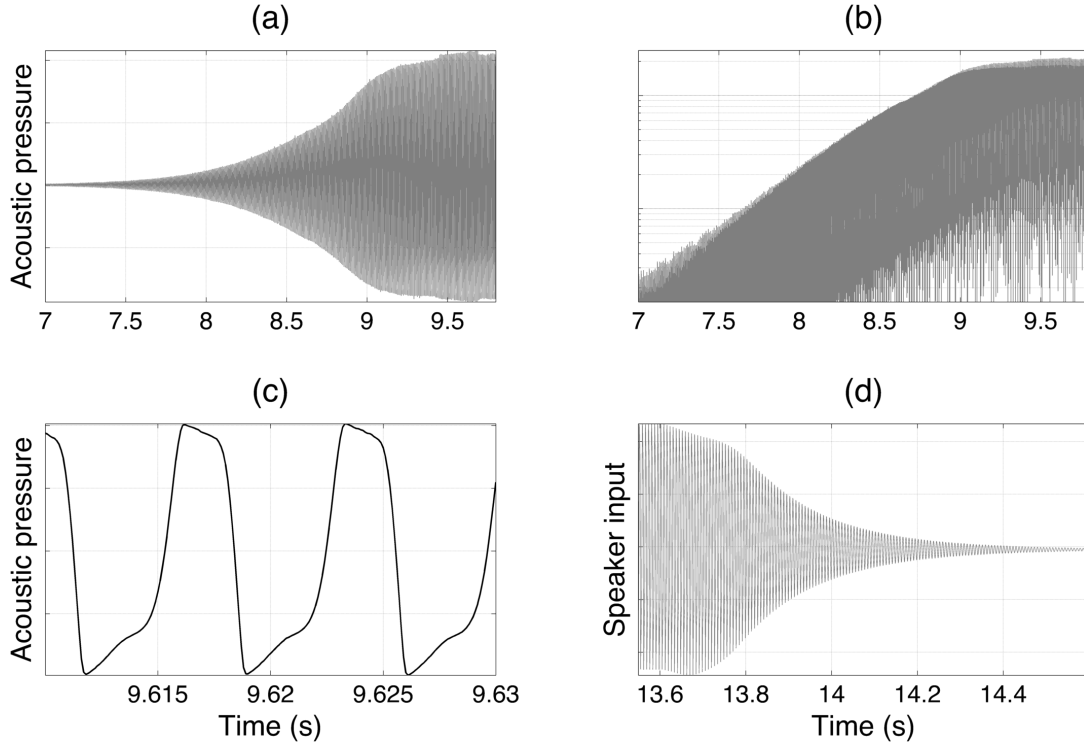
Figure 1.3 – *Time trace of the pressure (measured with a microphone) (a) at the onset of instability showing growth, and then saturation of the limit cycle. Linear growth on a semilog plot (b) of the signal's envelope confirms initial exponential growth of its amplitude. A zoomed-in picture (c) shows the periodic, but non-symmetric limit cycle behavior. With appropriate proportional feedback, the limit cycle is stabilized as this trace of the speaker's input signal (d) shows.*

out quickly. However, there is a critical higher gain value K_{max} above which a new instability is triggered. When that gain is reached, the tube will begin to "screech" loudly. The screech frequency reveals it to be a harmonic of the initial fundamental hum frequency. Exactly which harmonic it is will depend on the details of the experimental set up (in the setup described in Section 1.A it is typically the 3rd or 5th harmonic). This phenomenon is however repeatable if the experimental setup — microphone, speaker and heater locations — is unchanged. This higher gain instability and its frequency can be used to validate the model identified in the next section. The exact location of the microphone does not influence this qualitative behavior, but K_{min} and K_{max} are different for different microphone positions. Figure 1.4 sums up these observations.
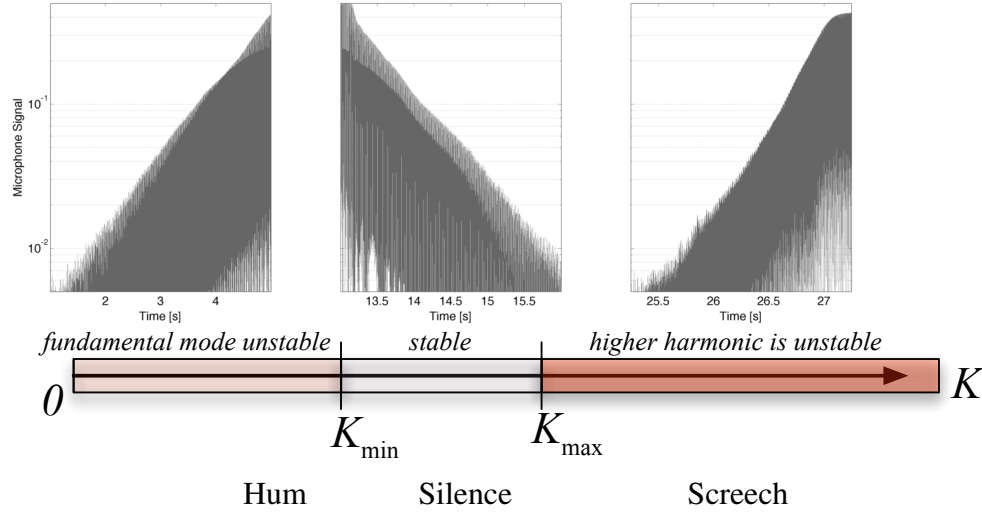
Figure 1.4 – *A depiction of the effects of proportional feedback on the Rijke tube. A minimum feedback gain K_{min} is necessary to stabilize the unstable fundamental mode. There is then a critical higher gain K_{max} beyond which a higher harmonic mode of the tube becomes unstable yielding a high-pitched screech. The upper part shows typical microphone traces of the transient phase: exponential decay for $K_{min} < K < K_{max}$ and exponential growth else.*

1.3 Rijke Tube System Identification

In this section, there is no physical modeling done; instead, a (linear) model of the Rijke tube is obtained purely by system identification based on measurements collected from the microphone inside the tube, and actuation provided through the speaker below.

1.3.1 Closed-Loop System Identification

As demonstrated in the preceding section, the Rijke tube is an unstable system, therefore it must be identified while operating in a stabilizing closed loop. Identifying systems in closed loops comes with some ramifications that have to be addressed in the identification scheme; simply recording the plant input and output and applying open-loop identification techniques, ignoring the fact that the input is the result of feedback, is definitely not a good idea.

The underlying reason is that most identification techniques start from a system model of the form $y(t) = F(q)\,u(t) + v(t)$, where y, u and v represent plant output, input and measurement noise, respectively, and F is the transfer function (in the shift operator q) of the plant to be identified. Under the assumption that noise and input are uncorrelated one can show, among other things, "consistency," i.e. that if the true plant $F_o(q)$ lies in the parametrized set of models and an infinite amount of data is collected, the identified model will converge to the true plant. However, if the system operates in closed loop the input $u(t)$ is determined from the measured output, hence it is correlated with the measurement noise, and consistency is lost.

To illustrate this fact, some quantitative detail for the case of predicition-error methods [Lju99] is provided: For the open-loop case with uncorrelated noise and input, one can show [Lju99, Ch. 8.5] that for an infinite amount of collected data, the identified frequency response $\widehat{F}(e^{j\omega})$ minimizes

$$\int_{-\pi}^{\pi} \left| F_o(e^{j\omega}) - \widehat{F}(e^{j\omega}) \right|^2 \Phi_u(\omega)\, \mathrm{d}\omega,$$

where Φ_s denotes the frequency spectrum of the signal s. If F_o is among the models that the algorithm can choose from, this minimum is 0, which happens exactly when $\widehat{F} \equiv F_o$, hence we have consistency.

On the other hand in a feedback loop, i.e. $u(t) = C(q)y(t) + w(t)$, where C is the controller, and w is an exogenous signal (sometimes referred to as "test" or "interrogation" signal), the identification process minimizes [Gev93]

$$\int_{-\pi}^{\pi} \left| \frac{F_o(e^{j\omega}) - \widehat{F}(e^{j\omega})}{1 + F_o(e^{j\omega})C(e^{j\omega})} \right|^2 \Phi_w(\omega) + \left| \frac{1 + C(e^{j\omega})\widehat{F}(e^{j\omega})}{1 + F_o(e^{j\omega})C(e^{j\omega})} \right|^2 \Phi_v(\omega)\, \mathrm{d}\omega; \tag{1.2}$$

the important feature is the new term that is multiplied by the noise spectrum, the presence of which moves the global optimum away from $\widehat{F} = F_o$. In the extreme case of no test signal $\Phi_w \equiv 0$, the estimate will instead converge to $-C^{-1}(e^{j\omega})$.

Since the idealized conditions that yield consistency typically are not satisfied anyway, one can

minimize the closed-loop effect by choosing a test signal w that has rich frequency content and hence should satisfy $\Phi_w(\omega) \gg \Phi_v(\omega)$ at all frequencies ω. By (1.2) the estimate $\widehat{F}$ then "pretty much" minimizes the error $\left|F_o(e^{j\omega}) - \widehat{F}(e^{j\omega})\right|$, weighted by the sensitivity of the closed loop. For example [IMR11] and [MD07] choose this path and successfully identify a model of their Rijke tube.

Another simple workaround, and the one chosen here, is the so-called *indirect method*. Like before, an exogenous test signal w is added into the loop, however now the *closed*-loop response is identified. The structure of the model then is $y(t) = T(q)\,w(t) + v(t)$, where T is the closed-loop transfer function, and w and v are indeed uncorrelated. The frequency response $T(e^{j\omega})$ can be identified with any open-loop technique, and the open-loop response $F(e^{j\omega})$ can be recovered by simple algebra. Of course, there are many more sophisticated methods, see e.g. the surveys [FL99, GLS77].

1.3.2   Identification Procedure and Results
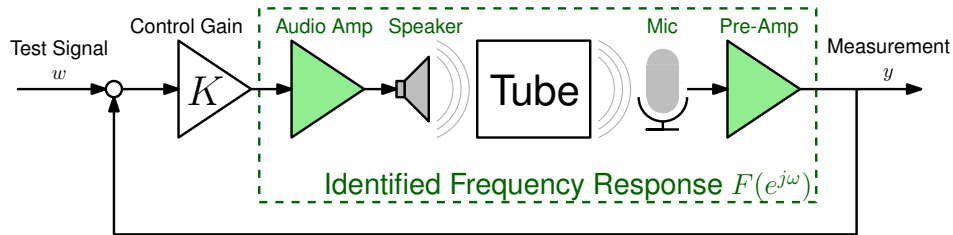
Figure 1.5 – *Equivalent block diagram of the Rijke tube with acoustic feedback. The speaker, microphone and pre-amp transfer functions are lumped together with that of the Rijke tube.*

As seen earlier, stabilization can be achieved by a proportional controller implemented by a speaker, a microphone, and amplifiers. These components are designed to pick up and transmit sounds audible to humans (20-20,000 Hz), hence their response should be reasonably flat in the frequency range of interest and it is justified to assume purely proportional control. By doing so, we implicitly lump these dynamics with the Rijke tube dynamics, as illustrated in

the conceptual block diagram in Figure 1.5, where the controller is written simply as a gain K. The closed-loop transfer function from the test signal w (added into the input of the power amplifier) to the microphone output y is

$$T(q) = \frac{KF(q)}{1 - KF(q)},$$

and once T is identified, F is obtained by inverting the above relation

$$F(q) = \frac{1}{K} \frac{T(q)}{1 + T(q)}. \tag{1.3}$$

It would be very difficult to obtain a value for K, which for instance includes the unknown conversion factor from pressure to voltage output by the microphone, but as the above equation shows, the chosen structure ensures that the poles and zeros of F can be identified from those of T without knowing K; F is identified up to an unknown gain factor $\frac{1}{K}$.

Open-Loop Identification Method The dynamics underlying the Rijke tube are a combination of acoustics and heat transfer, and are thus of relatively high order (in fact, they are infinite-dimensional). Nonparametric frequency-domain identification schemes are better suited to those types of systems than time-domain based ones, since one is not forced to select a model order a priori. Instead, the frequency response $T(e^{j\omega})$ is identified directly, and then a least-squares based method is used to fit a model of appropriate order over the identified frequency range. So-called spectral methods (see e.g. [Lju99, Ch. 6]) estimate the frequency response as the ratio of the cross spectrum $\Phi_{yw}(\omega)$ of output and test signal, and the spectrum $\Phi_w(\omega)$ of the test signal. The MATLAB System Identification Toolbox [Lju07] offers two implementa-

tions of these methods, spa and spafdr, of which the latter is the function of choice.[2] The least-squares fit is then performed using the function clsfits from the FREQID Toolbox for MATLAB [DCVdH96].

Test Signal   A test signal should have rich frequency content, while, due to actuator and sensor limitations in physical systems, amplitudes should be kept reasonably small. Popular choices include white noise, Schroeder-phased sinusoids [Bay93] and sine sweeps (also known as chirp signals). After experiments with all three types of signals, sine sweeps, which have been found to be beneficial in the identification of acoustic systems before [Bur92], emerged as the most effective choice; all shown data was collected using a sweep over the shown frequency range.

Data Collection   For the identification experiment, the tube is first brought to a hum. Then, the feedback with a stabilizing gain is turned on, and the test signal is added to the feedback signal, as shown in Figure 1.5. The microphone signal is recorded for the duration of the experiment (the data used here was collected over 120 s), and together with the applied test signal forms an input-output pair, which is all the data needed to obtain a spectral estimate. To minimize the effects of random noise, this is done several times, and an average of the estimated frequency responses is formed.

Figure 1.6 shows an averaged closed-loop frequency response along with the individual experiments. This response with several very lightly damped modes at integer multiples of a

---

[2] While both estimate the cross spectrum and input spectrum by applying a smoothing window to what roughly amounts to the discrete Fourier transforms of input and output data, spa performs the windowing in the time domain, whereas spafdr applies the window in the frequency domain. The expected sharp peaks in the frequency response require fine resolution in frequency; since a narrow frequency-domain window, which is what is required, corresponds to a wide time-domain window, using spafdr allows one to specify a small (frequency-domain) window, resulting in a drastic decrease in computation time compared to the large (time-domain) window that would have to be specified to achieve the same resolution using spa. Another important distinction between spa and spafdr is that the latter allows for frequency dependent resolution (hence the name), but this feature was not used here.
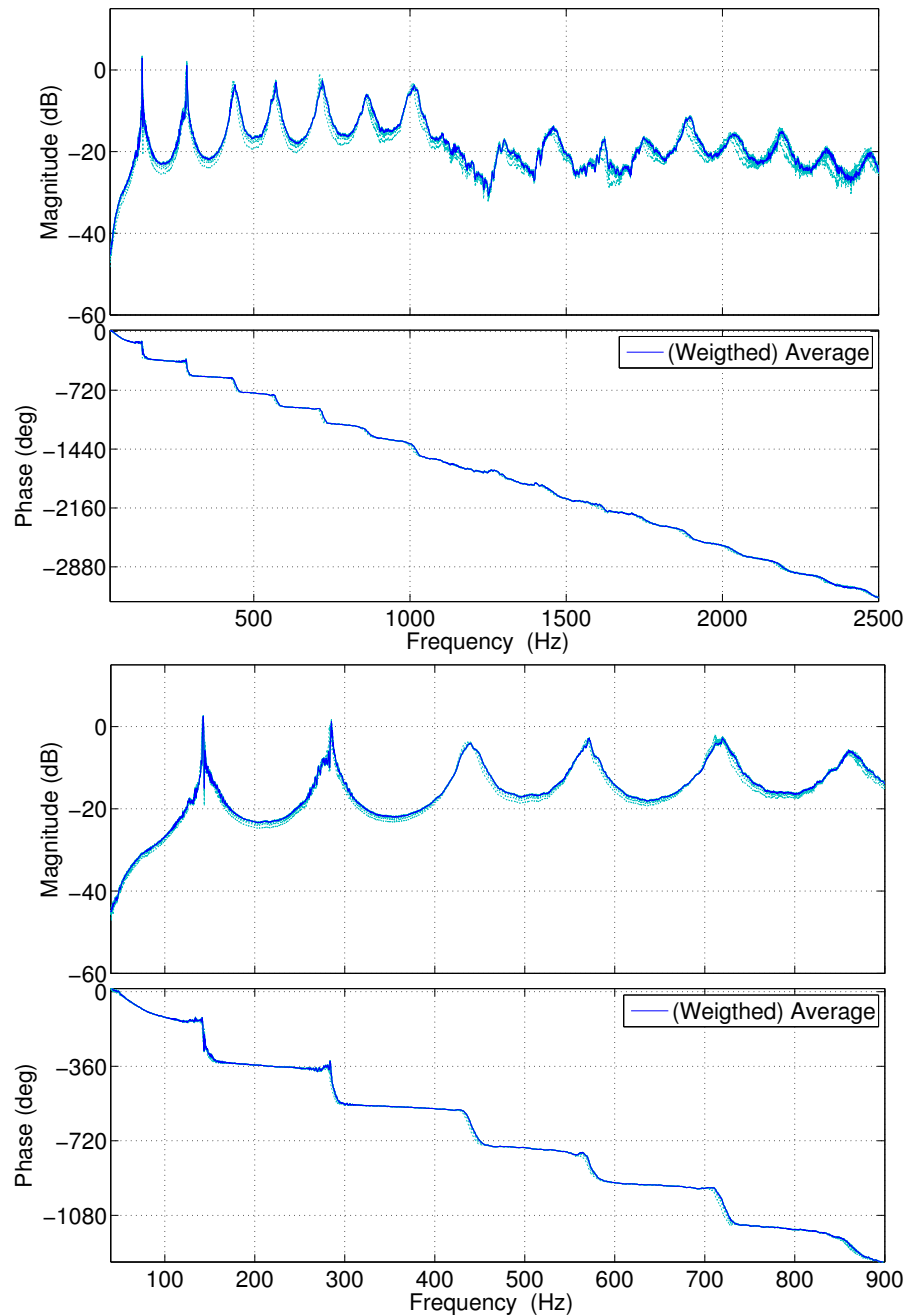
Figure 1.6 – *Closed-loop frequency response obtained with a sine sweep over the range of 0-2.5 kHz (top) and 0-900 Hz (bottom). The response below 20 Hz, which is outside the audible range, and above 1 kHz is likely dominated by microphone and speaker distortions. The range 0-1 kHz however exhibits typical wave-like dynamics with resonances occurring at multiples of the fundamental frequency. Note that here and in all Bode plots that follow, a linear frequency axis is used to emphasize the pattern of a fundamental frequency with harmonics.*
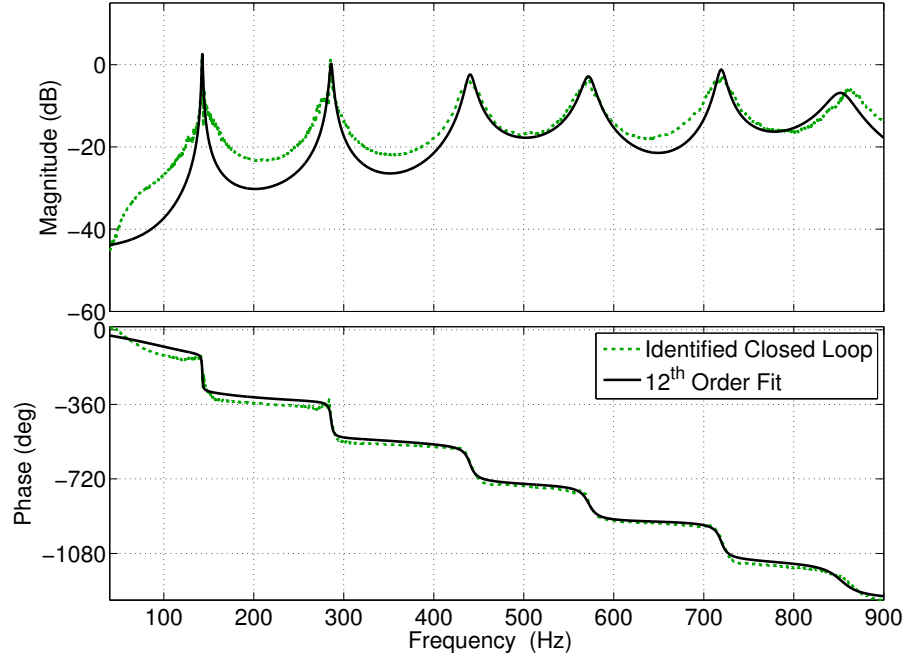
Figure 1.7 – *Closed-loop frequency response obtained by a nonparametric spectral estimation, and a 12th order least-squares fit. Note that due to the log-scale, the seemingly large deviations in the ranges between the peaks are actually very small.*

fundamental frequency has the signature of wave-like dynamics. The fundamental frequency corresponds very closely to the frequency of the hum observed in the non-stabilized Rijke tube.

In order to perform the least-squares fit of a finite dimensional transfer function model for $T(s)$, where s is the Laplace variable (we are fitting a continuous time model), to the estimated frequency response, a value for the model order needs to be selected. Figure 1.7 shows a 12th order transfer function fit, which nicely captures the first 6 harmonics in the frequency response.

To obtain the transfer function $F(s)$, a parametric model of the open loop, the fitted model T is then plugged into (1.3). Of course it is also possible to apply (1.3) to the nonparametric estimated frequency response at each frequency, thereby obtaining a nonparametric model of the open loop. Both of those possibilities are compared in Figure 1.8, they are in close agreement, which is encouraging. It is notable that, while the phase at the first peak of the closed loop T
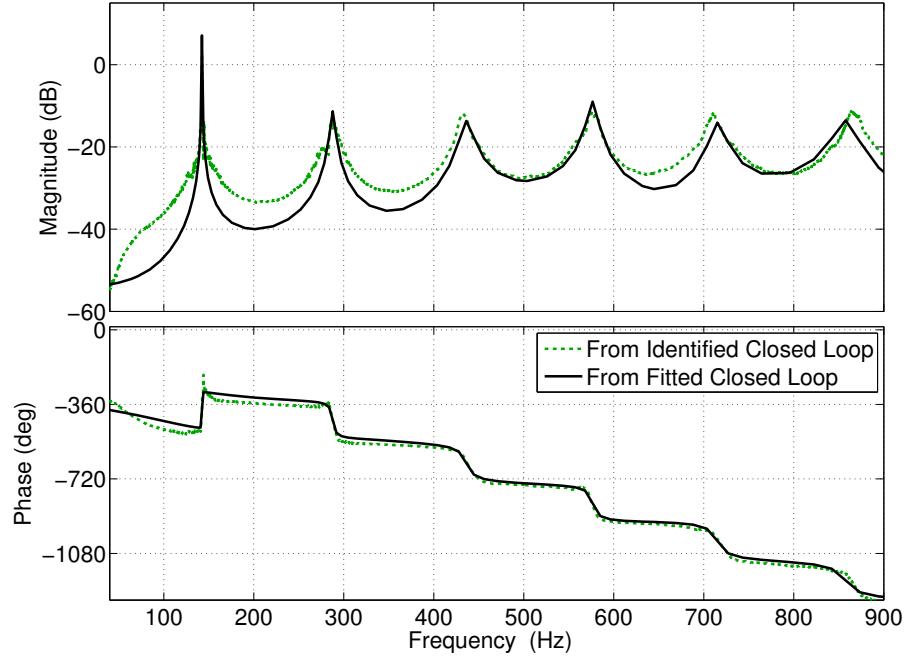
Figure 1.8 – *Open-loop frequency responses, obtained by applying* (1.3) *to the identified closed-loop response at every frequency (green) or to the fitted closed-loop response (black).*

drops by 180° indicating a stable pole slightly to the left of the imaginary axis, it increases by 180° in the open-loop response, indicating a pole slightly in the right-half plane (RHP).

1.3.3 Model Validation: Root Locus Analysis

The model $F(s)$ of the open loop obtained in the last section can be validated against the experimental observations. In particular, its root locus explains, why proportional feedback initially stabilizes the thermoacoustic instability and why a higher frequency mode becomes unstable at high gains. It also gives a quantitative prediction of that higher frequency.

Figures 1.9 and 1.10 show the root locus of the identified open-loop dynamics. The pole pattern resembles that of a damped wave equation, with imaginary parts of the poles being integer multiples of a fundamental frequency, and the real parts showing successively higher damping as the mode frequency increases. As expected, the fundamental mode is unstable,

19
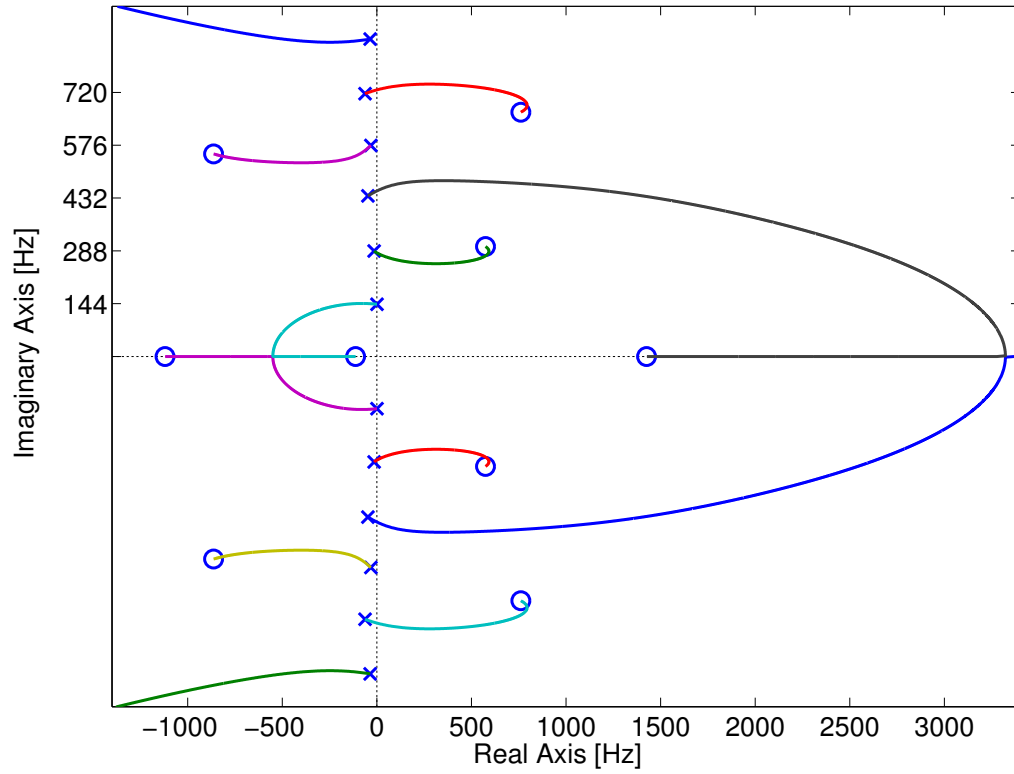
Figure 1.9 – *A full view of the root locus of the identified open-loop model with variable proportional feedback control gain. The open-loop poles show the (unstable) fundamental mode at 144 Hz and its (stable) harmonics very close to the imaginary axis. Several RHP open-loop zeros attract closed-loop poles into the RHP at high feedback control gains.*
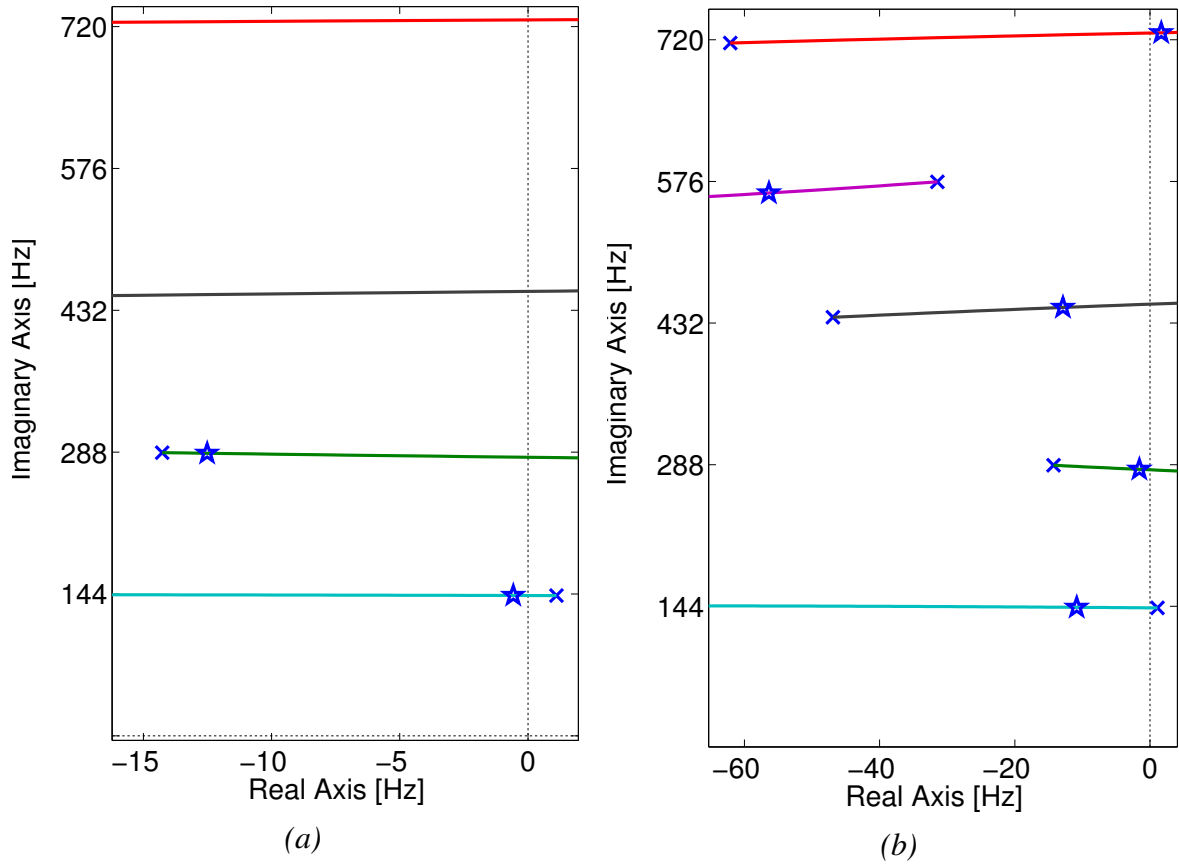
Figure 1.10 – *Close-ups of the root locus of the identified open-loop model F showing closed-loop pole locations (★) at gains that (a) just stabilize the unstable fundamental mode and that (b) make the fifth mode unstable.*

having positive real part. The imaginary part of the fundamental mode corresponds to the hum frequency heard when the tube is initially powered on.

This system also has multiple RHP zeros which ultimately attract a subset of the initially stable open-loop poles into the RHP, thus causing instability to reoccur at high gains. Figure 1.10(a) shows the locus and the pole locations at the value of the gain sufficient to initially stabilize the fundamental mode (denoted K_{min} in Figure 1.4), all poles are in the left half plane. However, due to the presence of RHP zeros, some poles will eventually cross into the RHP as the gain is increased. Figure 1.10(b) indicates that for this particular identified model, it is the fifth harmonic mode that becomes unstable at higher gain (denoted K_{max} in Figure 1.4). The frequency of this mode should correspond to the frequency of the screech heard in the experiment as the system becomes unstable again at high feedback gains, which is indeed what was observed on the setup used to collect the identification data. This serves as a useful method of model validation.

1.3.4 Common Problems in the Identification Process

Often, the phase of the open-loop frequency response will also drop, instead of increase, by 180° at the first peak, i.e. the open loop is identified as stable, while it is known from the initial experiment that the open loop must be unstable. The stability of the open loop is very sensitive to the amplitude and phase of T at the first peak. This is most easily explained with an argument based on the Nyquist criterion: From (1.3), it is seen that F has the same poles as T in negative unity feedback, so stability of F can be assessed through the Nyquist criterion. In order for T to encircle the critical point $(-1, 0j)$, one needs $|T| > 1$ and $\angle T = -180^o$ at the same frequency. Inspecting Figure 1.6 again, we see that for the presented data, the first peak reaches only about 2 dB, and the range for which it exceeds 0 dB is only about 1 Hz wide. Hence, if the peak is "cut off," the identification will result in a stable open loop. Likely culprits are insufficient

22
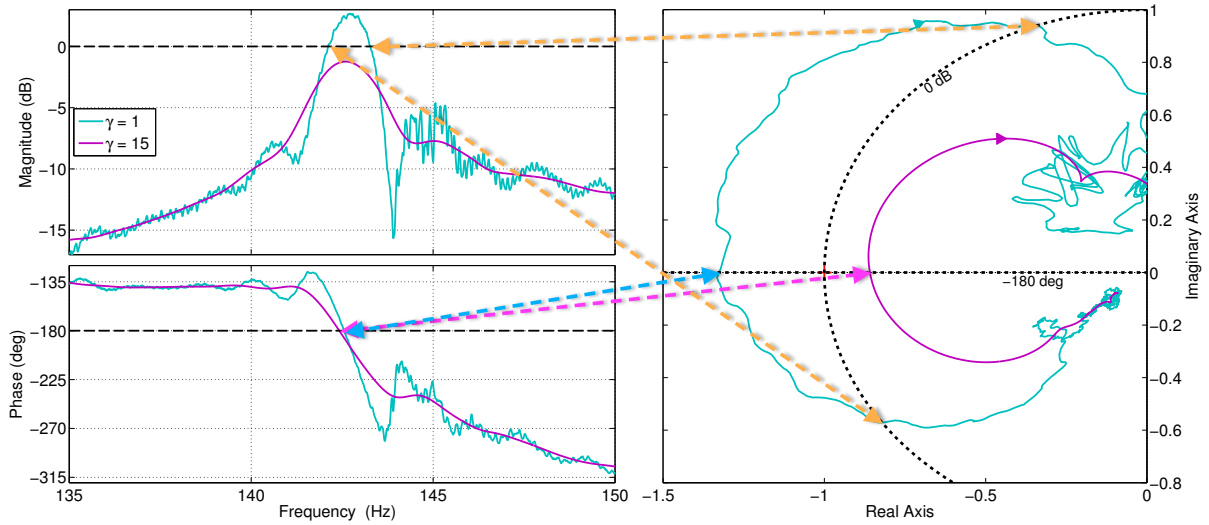
Figure 1.11 – *By comparing two frequency responses obtained from the same data with differ-*
ent amounts of smoothing (γ is passed to `spafdr` and corresponds roughly to the size
of the smallest detail [Lju07]), the difficulty in identifying an unstable open loop can
be illustrated. The cyan and magenta frequency responses shown on the left appear to
be qualitatively the same, thus one might tend to choose the magenta one as it looks
smoother while still capturing the resonant peak. However, the most important feature
– the corresponding unstable open loop – is not captured by this smoother response;
this is immediately clear when considering the Nyquist plots, shown on the right: the
magenta curve cannot loop around the critical point because it never leaves the unit
circle. Note that even γ = 15 still corresponds to relatively little smoothing.

frequency resolution and too much smoothing during the spectral estimation. This situation is detailed in Figure 1.11. If increasing the resolution and decreasing the smoothing do not help, a different speaker might be the solution; speakers were found to have quite different frequency responses, some added considerable phase lag.

More practically, if experiments are run for a long time, the tube walls, especially around the heater, absorb a lot of heat. If the identification is stopped and restarted for a new run, there might be no initial humming, due to the tube walls heating the air around the heater to the point where the heat transfer between air and heater is insufficient to support the humming. In that case, one can only wait for the tube to cool off, or, if the setup admits, increase the power to the heater to increase the coil temperature.

1.3.5 Microphone Position

The microphone position generically influences only the location of the zeros, not the poles, which is why its exact position was never stated. However, there are special locations that do yield interesting results. Figure 1.12 shows the identified open loops if the microphone is placed at three quarters, and at half the length of the tube. Doing so appears to "remove peaks," in the former case it would be every fourth, and in the latter case every even-numbered one.

This will be easily explained with the physical transfer function model (1.31) developed in the modeling section — it can in fact serve as validation for the model — but there is also a very intuitive explanation: each peak corresponds to a mode, a standing pressure wave, in the tube; the standing wave corresponding to the first peak is a half-wave, for the second peak a full wave, and so on. Placing the microphone at e.g. the center means placing it where all the even numbered modes have a pressure node, and hence their contribution is not registered by the microphone, a pressure sensor.
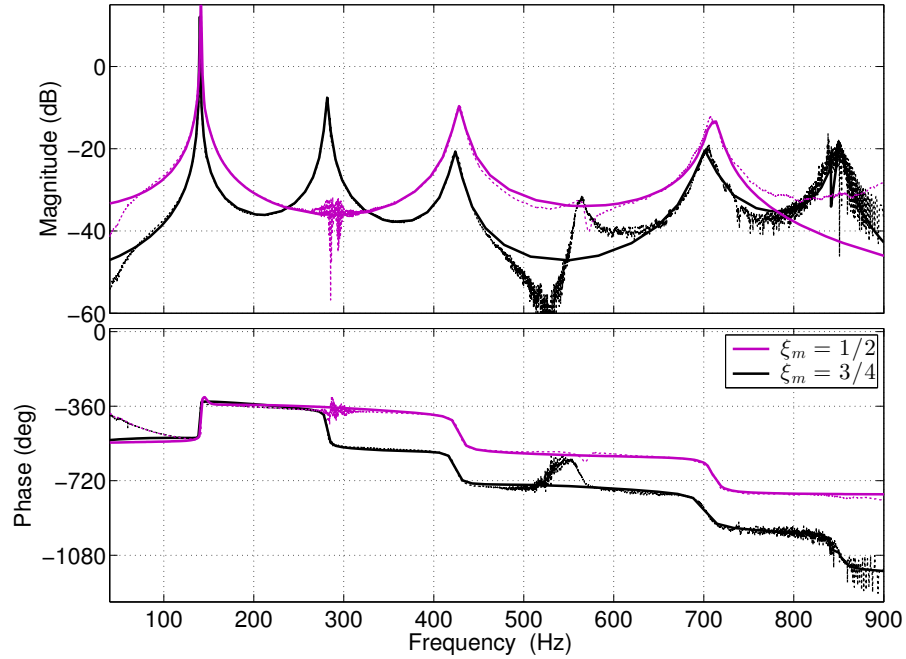
Figure 1.12 – *Bode plots for identified and fitted open-loop responses with different micro-phone positions. Placing the microphone in the middle of the tube ($\xi_m = 1/2$) seems to remove every other peak, while placing it at a quarter length from the tube attenuates the fourth peak only. The very ugly identification data at the removed peaks, and especially at the peak around 560 Hz, can be explained by the fact that perfect cancellation of a pole by a zero is virtually impossible; instead, one gets a pole and a zero very close together – notoriously difficult to identify.*
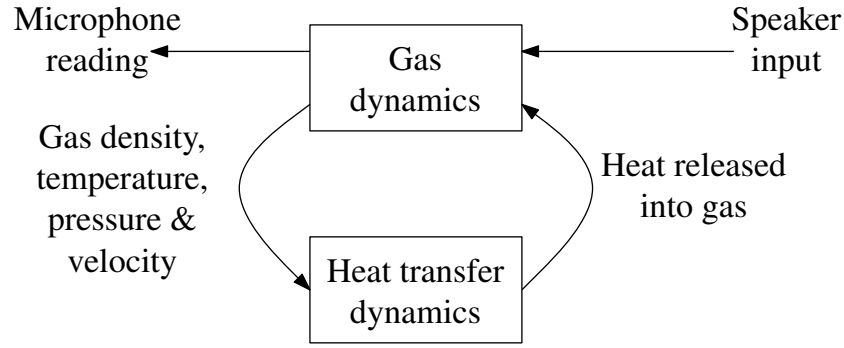
Figure 1.13 – *A sketch of the model to be developed*

1.4 A Spatially Distributed Rijke Tube Model

After the black-box approach of the previous section, we now set out to develop a model describing the interactions of the gas, the hot wire, the speaker and the microphone in the Rijke tube from physical considerations. It is natural to think of the model as consisting of two interconnected parts: the first one describing the dynamics of the gas, and the second one describing the heat exchange process between the wire and the gas.

A sketch of the model structure is shown in Figure 1.13; we will update this diagram as we go and develop models of different complexity. The structure already reveals that the heat release constitutes a feedback from the state of the gas to the released heat. It is very important to distinguish this feedback from the stabilizing proportional feedback of Section 1.3: On the one hand, the heat release process can be *interpreted* as an *internal* feedback loop, it connects signals not accessible to measurement or direct manipulation; on the other hand, the stabilizing feedback is an *external* controller connecting microphone measurement and speaker actuation. In other words: the *open* loop identified in the previous section corresponds to the transfer function from speaker signal to microphone measurement, which includes the *closed* feedback loop of the heat transfer process. This hints that the key to quantitative understanding of the thermoacoustic instability is analyzing this internal feedback loop.

It will prove convenient to introduce dimensionless parameters and variables, as is very com-

mon in fluid mechanics and thermodynamics. While that might require some "getting used to," there are several advantages if the nondimensionalization is done carefully: the number of parameters is reduced (the exact number could be quantified using the famous Buckingham Π theorem), expressions are simplified, and most importantly it allows to compare relative sizes and thus can be used to argue for or against neglecting certain terms. The transition from physical variables to dimensionless ones is done in Section 1.4.3, and the remainder of the section is written in terms of the dimensionless quantities summarized in Table 1.1.

The dynamics of the velocity and thermodynamic state[3] of the gas are derived from the conservation of mass, linear momentum and energy in Section 1.4.1. Proceeding by linearization of the nonlinear PDEs in Section 1.4.2, and nondimensionalization of the linearized equations in Section 1.4.3, we are in a position to make well-founded statements about the importance of effects such as friction and diffusion. Section 1.4.5 shows that the end result of the most idealized case is the 2×2 matrix G_s of transcendental transfer functions (1.31), describing the response of pressure and velocity fluctuation to heater and speaker inputs. The last two sections investigate how this transfer function changes as first friction, and then diffusion are considered.

Paralleling these derivations, a model for the heat release is developed. For the heat release from the wire into the gas, two aspects need to be considered. For stationary flow, i.e. in the static case, it is well known that faster fluid flow leads to improved heat transfer, and this improvement needs to be quantified. If the flow velocity is fluctuating, the question arises as to how fast this improvement is realized, which constitutes the dynamic aspect. Fortunately, both have been considered in great detail by King [Kin14] and Lighthill [Lig54], and their results simply need to be adapted to the Rijke tube. That leads to a nonlinear model at first (1.17) and after linearization (1.22), nondimensionalization (1.26) and simplification, we obtain a simple first-order lag (1.29) describing the modulation of the heat release rate by the local velocity.

---

[3] Two of density, pressure and temperature

Combining both parts – heat release and gas dynamics – the linearized overall model can be compactly written as a linear fractional transformation (LFT), see Figure 1.18, and root-locus analysis in Section 1.4.5 reveals that the instability results from the unstable feedback coupling between heat released into the gas, and local flow velocity around the heater. The root locus also predicts the correct frequency for the resulting limit cycle oscillations, see Figure 1.19.

While in the literature, the Rijke tube is typically modeled as consisting of two or three compartments — a cold zone below the heater, a hot zone above it, and sometimes a small zone in which the heat exchange takes place — and the heater is introduced as causing a discontinuity in the velocity field [RBFB93], in the derivations here, the tube is modeled in one piece. This approach avoids explicit introduction of time delays to account for sound waves traveling through cold and hot zones as is done e.g. in [MD07, Dow97, OCGZK13]; one can interpret the hyperbolic functions in (1.31) to contain delays, but they result naturally from the spatially distributed modeling process.

1.4.1 Nonlinear Rijke Tube Dynamics

Since the tube is much longer than wide, transverse dynamics can be neglected and the gas is modeled as a one-dimensional medium. The conservation equations yield a system of nonlinear partial differential equations (PDEs) that will be simplified by omission of terms with small influence and linearization around a steady state in the following sections.

Similarly, the initial model for the heat release contains a nonlinearity, but is decidedly simpler than the gas dynamics.

Gas Dynamics

As a first step, a nonlinear model describing the dynamics of the gas inside the tube is derived by using the conservation laws in their differential form as they can be found in virtually every textbook on fluid mechanics. Here we use the very general forms stated in [SG00, Ch. 3]. The state of the gas will be expressed in terms of its velocity v, the density ρ, pressure p and temperature T, all of which are functions of both the spatial coordinate x and the time t. Considering only a single spatial variable is justified by the assumption that the tube is much longer than wide, and hence "width-wise" phenomena have much smaller wavelengths and much higher frequencies than the lengthwise resonant modes.

Conservation of Mass The continuity equation is the simplest and probably most familiar of the three laws. It states

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho v)}{\partial x} = 0, \tag{1.4}$$

relating the density field $\rho(t,x)$ and the velocity field $v(t,x)$.

Conservation of Linear Momentum The Navier-Stokes equations are derived from the conservation of momentum, and incorporate also the pressure field $p(t,x)$:

$$\rho\frac{\partial v}{\partial t} = -\rho v\frac{\partial v}{\partial x} - \frac{\partial p}{\partial x} + \frac{4}{3}\mu\frac{\partial^2 v}{\partial x^2} - \rho g - \beta v, \tag{1.5}$$

where μ is the dynamic viscosity, which we assume to be constant, $g = 9.81$ m/s^2 is the gravitational acceleration, and β is a wall-friction factor that is included to replace the no-slip

condition at the walls.[4] Compared to [SG00, (3.42)], this equation is reduced to one spatial dimension, the body force is set to gravity, and the substantial derivative is expanded:

$$\frac{\mathrm{D}\phi}{\mathrm{D}t} := \frac{\partial \phi}{\partial t} + v\frac{\partial \phi}{\partial x}.$$

Conservation of Energy The energy equation, which is derived from the first law of thermodynamics, adds the temperature T into the mix and reads

$$\frac{\partial T}{\partial t} = -v\frac{\partial T}{\partial x} + \frac{\gamma\kappa}{\rho c_p}\frac{\partial^2 T}{\partial x^2} - \bar{\gamma}T\frac{\partial v}{\partial x} + \frac{4\gamma\mu}{3\rho c_p}\left(\frac{\partial v}{\partial x}\right)^2 + \frac{\gamma}{\rho c_p}\,q + \beta\gamma v^2/(\rho c_p), \qquad (1.7)$$

where κ is the thermal conductivity, the adiabatic ratio $\gamma = c_p/c_v$ is the ratio of the isobaric and isochoric specific heat capacities, and $\bar{\gamma} = \gamma - 1 = R/c_v$, where R is the ideal gas constant; $q(t,x)$ is exogenous added heat power per unit volume, with unit $[q] = 1$ W/m^3. Finally, the last term in the equation is *not* due to dissipation being accounted for in the energy balance, but appears when (1.5) is used to eliminate kinetic and potential energy. This suggests that all energy dissipated through friction is converted into internal energy of the gas, and none is absorbed by the walls; one could consider multiplying β in (1.7) by a factor $k_\beta \in [0, 1]$ to account for heat absorbed by the walls, however friction effects do not play a major role in what follows and this additional complication is avoided.

In deriving (1.7) from [SG00, (3.71)], we again reduced to a single spatial dimension and assumed air to be an ideal gas, from which it follows that the pressure can be eliminated via

---

[4]In one-dimensional, laminar pipe flow, also called Poiseuille flow, the developed flow profile is parabolic in shape, i.e. $v(r) = \Delta p/\Delta x\,(r^2 - D^2/4)/(4\mu)$, where r the radial coordinate and D the pipe diameter, see e.g. [PW02, Sec. 7.3]. For a Newtonian fluid (air is one), the shear stress at the wall is $\tau_0 = \mu\frac{\mathrm{d}p}{\mathrm{d}r}\big|_{r=D/2}$. Rewriting in terms of the average velocity V, that yields $\tau_0 = 8\mu V/D$, and the friction force per unit volume exerted on a cross-sectional element of length $\mathrm{d}x$ is then $\mathrm{d}F_0 = 8\pi\mu V\mathrm{d}x/(A\mathrm{d}x)$, A the cross section. That yields

$$\beta = 32\mu/D^2. \qquad (1.6)$$

This is of course only good for a very rough idea of the size of β, since no dynamic effects are included, and we most likely will not encounter purely laminar flow. The unit of β is $[\beta] = 1$ kg/(m^{3}s).

$p = \rho RT$, that the thermal expansion coefficient (β in [SG00]) equals $1/T$, and that $c_v + R = c_p$.

Choice of Thermodynamic State Variables   Now (1.4), (1.5) and (1.7) are three equations to describe the dynamics of seemingly four states – velocity v, pressure p, density ρ and temperature T – so we need to add one more equation, a so-called equation of state, relating the three thermodynamic states, e.g. as $p = p(\rho, T)$. Assuming air to be an ideal gas, this equation is $p = \rho RT$, with R the ideal gas constant, which we have used already to derive (1.7). This yields a total of three possible sets of equations of motion in velocity v and two of pressure p, density ρ and temperature T.

(ρ, v, p)-System   Using $p = \rho RT$ to eliminate T from (1.7) in favor of the pressure p we obtain

$$\frac{\partial \rho}{\partial t} = -\frac{\partial(\rho v)}{\partial x} \tag{1.8a}$$

$$\frac{\partial v}{\partial t} = -v\frac{\partial v}{\partial x} - \frac{1}{\rho}\frac{\partial p}{\partial x} + \frac{4\mu}{3\rho}\frac{\partial^2 v}{\partial x^2} - g - \beta/\rho\, v \tag{1.8b}$$

$$\frac{\partial p}{\partial t} = -v\frac{\partial p}{\partial x} - \gamma p\frac{\partial v}{\partial x}$$
$$+ \frac{\gamma \kappa p}{\rho c_p}\left[\frac{\partial^2 p}{\partial x^2}/p - \frac{\partial^2 \rho}{\partial x^2}/\rho - 2\frac{\partial \rho}{\partial x}/\rho\left(\frac{\partial p}{\partial x}/p - \frac{\partial \rho}{\partial x}/\rho\right)\right] + \frac{4\bar{\gamma}\mu}{3}\left(\frac{\partial v}{\partial x}\right)^2 + \bar{\gamma}q + \bar{\gamma}\beta v^2. \tag{1.8c}$$

Most of the mess in this pressure form of the energy equation is due to the $\frac{\partial^2 T}{\partial x^2}$ term, since

$$\frac{\partial^2 T}{\partial x^2} = \frac{p}{\rho R}\left[\frac{1}{p}\frac{\partial^2 p}{\partial x^2} - \frac{1}{\rho}\frac{\partial^2 \rho}{\partial x^2} - \frac{2}{\rho}\frac{\partial \rho}{\partial x}\left(\frac{1}{p}\frac{\partial p}{\partial x} - \frac{1}{\rho}\frac{\partial \rho}{\partial x}\right)\right]. \tag{1.9}$$

(p,v,T)-System   Eliminating ρ takes a little more effort and yields

$$\frac{\partial p}{\partial t} = -\gamma p \frac{\partial v}{\partial x} - v \frac{\partial p}{\partial x} + \bar{\gamma}\kappa \frac{\partial^2 T}{\partial x^2} + \frac{4}{3}\bar{\gamma}\mu \left(\frac{\partial v}{\partial x}\right)^2 + \bar{\gamma}q + \bar{\gamma}\beta v^2 \tag{1.10a}$$

$$\frac{\partial v}{\partial t} = -v \frac{\partial v}{\partial x} - \frac{RT}{p} \frac{\partial p}{\partial x} + \frac{4\mu}{3} \frac{RT}{p} \frac{\partial^2 v}{\partial x^2} - g - \frac{\beta RT}{p} v \tag{1.10b}$$

$$\frac{\partial T}{\partial t} = -v \frac{\partial T}{\partial x} - \bar{\gamma}T \frac{\partial v}{\partial x} + \bar{\gamma}\kappa \frac{T}{p} \frac{\partial^2 T}{\partial x^2} + \frac{4\bar{\gamma}\mu T}{3p} \left(\frac{\partial v}{\partial x}\right)^2 + \bar{\gamma}\frac{T}{p}q + \bar{\gamma}\frac{T}{p}\beta v^2 \tag{1.10c}$$

(ρ,v,T)-System   Lastly, removing p from (1.5) we get

$$\frac{\partial \rho}{\partial t} = -\frac{\partial(\rho v)}{\partial x}$$

$$\frac{\partial v}{\partial t} = -v \frac{\partial v}{\partial x} - R\frac{\partial T}{\partial x} - \frac{RT}{\rho} \frac{\partial \rho}{\partial x} + \frac{4\mu}{3\rho} \frac{\partial^2 v}{\partial x^2} - g - \beta/\rho\, v$$

$$\frac{\partial T}{\partial t} = -v \frac{\partial T}{\partial x} + \frac{\gamma\kappa}{\rho c_p} \frac{\partial^2 T}{\partial x^2} - \bar{\gamma}T \frac{\partial v}{\partial x} + \frac{4\gamma\mu}{3\rho c_p} \left(\frac{\partial v}{\partial x}\right)^2 + \frac{\gamma}{\rho c_p}\, q + \gamma\beta v^2/(\rho c_p)$$

Each of the forms has their advantages; (ρ,v,T) might be the cleanest, but it will not be used in the sequel: measurement and actuation are both expressed in terms of the pressure p thus it makes sense to keep it as one of the states. We will see later that using ρ instead of T yields, together with several simplifying assumptions, a decoupled system, whereas using T prevents the issue of (1.9) and hence is a better choice when diffusive effects are to be kept.

Boundary Conditions, Sensing and Actuation

For a complete model, an adequate number of boundary conditions is needed. What is adequate depends on the number of states and the order of the remaining spatial derivatives once simplifications are made, hence only a subset of the following boundary conditions are used in specific cases; for convenience, they are all stated here together once:

Dirichlet pressure boundary conditions The Rijke tube has both ends open, which is typically modeled by having pressure boundary conditions held at the ambient atmospheric pressure. The bottom end however is very close to the actuating speaker, and while a true model of the influence of the speaker involves accounting for inward radiating waves, a simple and reasonable approximation is to assume the speaker signal as pressure fluctuations at the bottom end. Thus the boundary conditions on the pressure are

$$p(t,0) = p_0 + u_s(t)$$
$$p(t,L) = p_0,$$
(1.11)

where p_0 is the ambient pressure, and $u_s(t)$ is the speaker signal.[5]

Neumann velocity boundary conditions At open ends, standing sound waves have pressure nodes, as (1.11) suggests, and velocity antinodes. This corresponds to vanishing spatial derivatives:

$$\frac{\partial v}{\partial x}(t,0) = \frac{\partial v}{\partial x}(t,L) = 0.$$
(1.12)

Neumann temperature boundary conditions Assuming the temperature gradient to be zero at the ends corresponds to assuming no conductive heat flux through the open ends, since by Fourier's law of heat conduction, the heat flux q_{oe} through the open ends is proportional to the temperature gradient: $q_{oe} = -\kappa \frac{\partial T}{\partial x}$. Hence

$$\frac{\partial T}{\partial x}(t,0) = \frac{\partial T}{\partial x}(t,L) = 0$$
(1.13)

is justified by noting that air is a poor conductor of heat.

For example, $\kappa_{\text{air}} = 0.024$ W/(m K) $\ll \kappa_{\text{glass}} \approx 1$ W/(m K) $\ll \kappa_{\text{iron}} \approx 80$ W/(m K).

---

[5] We note again that a correction term could be introduced into the boundary conditions to account for the fact that in reality, the nodes lie slightly outside the tube [LS48]. Since the effect of the end correction is small, additional complication is avoided by neglecting this. See also the footnote on Page 10.
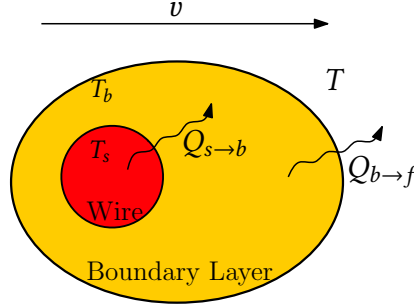
Figure 1.14 – *Illustration of the effect of the boundary layer around the hot wire.*

Note that $u_s(t)$ acts as an input to the gas dynamics through the boundary condition (1.11), but it is not distributed. The same holds for the measurement through a microphone, which is a pressure sensor whose output is (proportional to) the pressure $p(t, x_m)$ at the microphone location x_m. The controller used to stabilize the tube connects those two scalar signals via a proportional control feedback loop.

Heat Release

The model equations (1.8) and (1.10) have an external heat input $q(t, x)$ as a source term, courtesy of the energy equation (1.7). Possible sources of external heat are *(a)* the walls of the tube, *(b)* the ambient air at both ends of the tube, and of course *(c)* the hot wire. We assume that the temperature difference between walls and gas is comparatively small, so that contribution *(a)* can be neglected. Furthermore, air is a good insulator, which means that its thermal conductivity is small, and so we may neglect contribution *(b)*, as we have already done in deriving the temperature boundary condition (1.13).

It remains *(c)*, the heat transfer from the hot wire. If the coil is assumed to be located in a very narrow section, then a reasonable approximation is to say that the heat input is concentrated at a point, or that

$$q(t, x) = \frac{1}{A} \delta_m(x - x_o) Q(t),$$
(1.14)

34

where x_o denotes the position of the wire, $\delta_m(\cdot)$ is the Dirac distribution with units 1/m, $Q(t)$ is the heat power released from the coil, and A is the tube cross section.

The next step is to quantify the dependence of released heat $Q(t)$ on the thermodynamic state (ρ, p, T) of the gas and the flow velocity v near the wire. Heat is transferred from a solid to a gas due to both conduction and convection. Convection depends on the velocity of the gas, whereas conduction does not; both depend on the temperature difference. A commonly used approximation incorporating both mechanisms is King's law [Kin14], which states that the heat Q_K released from a hot circular wire of length ℓ_w in a colder fluid flow is

$$Q_K = \ell_w \left(\kappa + \kappa_v \sqrt{|v|} \, \right) (T_{\text{wire}} - T_{\text{gas}}) =: h(v)(T_{\text{wire}} - T_{\text{gas}}), \tag{1.15}$$

where κ is the fluid's thermal conductivity, and κ_v is a constant that is largely empirically determined. The first term is Fourier's law of heat conduction, while the second term expresses that convective heat transfer occurs at a rate proportional to $\sqrt{|v|}$; $h(v)$ collects these contributions as a velocity-dependent heat transfer coefficient. The formula suggests that heat transfer is enhanced by flow velocity, regardless of direction, but the "rate of enhancement" tapers off as velocity increases.

King's law does not include any temporal dynamics, yet if the gas velocity fluctuates sufficiently fast, there are dynamic effects to be considered; Lighthill [Lig54] analyzed these in great detail.[6] Whenever flowing fluid comes into contact with a solid, a boundary layer is formed. An intuitive picture of its effect on the heat transfer is to imagine this boundary layer as a blob of stagnant air, as shown in Figure 1.14, where the boundary layer is depicted as an orange ellipse. Heat transfer from the boundary layer into the free stream, denoted by $Q_{b \rightarrow f}$, reacts instantly to changes in the free stream velocity v according to King's law (1.15), but heat $Q_{s \rightarrow b}$ transferred from the wire needs to propagate through the boundary layer before reaching

---

[6]Incidentally, Lighthill was inspired to this investigation by efforts to model the Rijke tube.
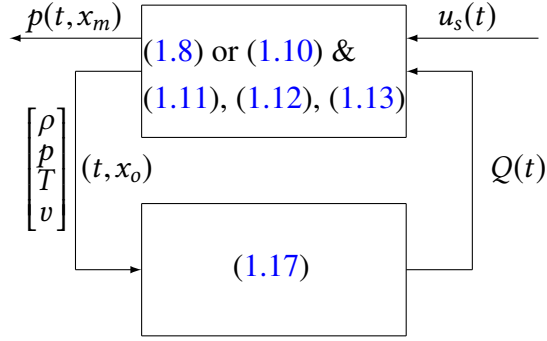
Figure 1.15 – *The conceptual diagram of Figure 1.13, now filled with some concrete equations.*

the free stream. Lighthill found that the thermal inertia of the boundary layer can be modeled by a simple first order lag, i.e. a transfer function of the form $1/(t_{hr}s + 1)$. An estimate for the corresponding time constant is also given as

$$t_{hr} = 0.2 \frac{d_{\text{wire}}}{\bar{v}} = \frac{\text{diameter of the wire}}{5 \cdot \text{steady free stream velocity}}. \tag{1.16}$$

As observed in [Lig54], this lag, even if it is very small, is crucial for the model to be valid. We confirm this observation later, see Figure 1.20.

Cascading the static and dynamic dependence of the heat transfer on the velocity, we get

$$t_{hr}\dot{Q}(t) = -Q(t) + Q_K(t) \tag{1.17a}$$

$$Q_K(t) = \ell_w \left(T_{\text{wire}} - T(t, x_o)\right) \left(\kappa + \kappa_v \sqrt{|v(t, x_0)|}\right), \tag{1.17b}$$

which constitutes the lower block in Figure 1.13.

1.4.2 Linearized Dynamics in Physical Variables

In order to linearize the nonlinear equations of the section, all states are considered as steady state plus (small) fluctuation. We will always denote steady-state, or nominal, values by $\overline{\bullet}$ and the small deviations by $\widetilde{\bullet}$, hence we have e.g. $p(t,x) = \overline{p} + \widetilde{p}(t,x)$. Since what the ear registers as "sound" are exactly those deviations, $\widetilde{p}$ and $\widetilde{v}$ are also called *acoustic pressure* and *acoustic velocity*, respectively. We remark that the steady states are only required to be steady in time, in general they could be steady spatial profiles; however, we consider only steady states that are constant in space, hence the discussion is reduced to this case.

As an illustration of the size of deviation we may expect, consider that a Vuvuzela blown in a distance of 1 m has a sound intensity of 120 dB, which corresponds to a (RMS) pressure deviation of approximately 20 Pa [SHIK10].[7] Compared to the atmospheric pressure of $p_0 = 100$ kPa, this indeed constitutes a small fluctuation, and since the Rijke tube experiment can be conducted without damage to one's hearing, we may assume small pressure deviations in the Rijke tube. Since sound waves of moderate frequencies can be assumed to be isentropic,[8] it also follows that the deviations in ρ and T are of similar order. The case is different for the velocity, which should be compared to the speed of sound; we will return to this point later in Section 1.4.3.

Formally, the linearization can be obtained by regarding the right-hand sides as functions of the states and their derivatives and truncating their Taylor series after the linear term, for instance

---

[7]Sound pressure levels are given as the root mean square of the pressure fluctuation relative to a reference level: SPL $= 20 \log_{10}(\widetilde{p}_{\mathrm{rms}}/\widetilde{p}_{\mathrm{ref}})$. This reference level is typically $\widetilde{p}_{\mathrm{ref}} = 20\ \mu$Pa, the lower threshold of human hearing.

[8]Assuming isothermal sound waves instead led Newton to a calculation that was off by about 20%. Laplace later corrected this error, see [Fin64] for some of the history.

for (1.4) we get

$$\frac{\partial \rho}{\partial t} = \overbrace{\frac{\partial \overline{\rho}}{\partial t}}^{=0} + \frac{\partial \widetilde{\rho}}{\partial t} = f\left(\rho, \frac{\partial \rho}{\partial x}, v, \frac{\partial v}{\partial x}\right) =$$

$$\underbrace{f\left(\overline{\rho}, \frac{\partial \overline{\rho}}{\partial x}, \overline{v}, \frac{\partial \overline{v}}{\partial x}\right)}_{=0,\ \text{since steady state}} + \underbrace{\left[\frac{\partial f}{\partial \rho} \quad \frac{\partial f}{\partial \left(\frac{\partial \rho}{\partial x}\right)} \quad \frac{\partial f}{\partial v} \quad \frac{\partial f}{\partial \left(\frac{\partial v}{\partial x}\right)}\right]\Big|_{(\rho, \frac{\partial \rho}{\partial x}, v, \frac{\partial v}{\partial x})=(\overline{\rho}, \frac{\partial \overline{\rho}}{\partial x}, \overline{v}, \frac{\partial \overline{v}}{\partial x})}}_{=:\overline{J}_f} \begin{bmatrix} \widetilde{\rho} \\ \frac{\partial \widetilde{\rho}}{\partial x} \\ \widetilde{v} \\ \frac{\partial \widetilde{v}}{\partial x} \end{bmatrix} + \mathcal{O}(2),$$

where $\mathcal{O}(2)$ denotes terms of order 2 or higher in the deviations; after neglecting the higher order terms we get

$$\frac{\partial \widetilde{\rho}}{\partial t} \approx \overline{J}_f \begin{bmatrix} \widetilde{\rho} \\ \frac{\partial \widetilde{\rho}}{\partial x} \\ \widetilde{v} \\ \frac{\partial \widetilde{v}}{\partial x} \end{bmatrix} = -\overline{v}\frac{\partial \widetilde{\rho}}{\partial x} - \overline{\rho}\frac{\partial \widetilde{v}}{\partial x}.$$

One would obtain the same result by inserting $p = \overline{p} + \widetilde{p}$ and so on and neglecting all terms of order 2 or higher in the deviation variables.

Of course, linearization requires knowing the steady state conditions. An accurate calculation of the steady state requires incorporating buoyancy effects (to model steady upward flow due to steady heat release from the coil) as well as the steady temperature and density variations along the tube length (e.g. gas in the upper section is hotter than that in lower section). However since the important parameter in acoustic dynamics is the speed of sound, and the above variations have relatively minor effect on the speed of sound, an alternative and much simpler equilibrium can be used for the linearization: Spatially constant density, velocity, and pressure fields trivially satisfy the nonlinear PDEs (1.8) and (1.10) if there is no heat input ($q \equiv 0$), and friction and gravity are neglected. From the boundary condition (1.11) it then follows that $\overline{p} = p_0$ (there must be no speaker actuation, i.e. $u \equiv 0$); choosing $\overline{T}$ to be the temperature of the surroundings satisfies the temperature boundary condition (1.13) and the ideal gas law then

specifies $\bar{\rho}$.

Once the steady states are determined, the linearization process is a mere exercise in calculus and can be done using a computer algebra system such as MATHEMATICA; the following expressions were found in this manner.

Linearized (ρ,v,p)-System

$$\frac{\partial}{\partial t}\begin{bmatrix}\widetilde{\rho}\\\widetilde{v}\\\widetilde{p}\end{bmatrix}=\beta\begin{bmatrix}0 & 0 & 0\\\bar{v}/\bar{\rho}^2 & -1/\bar{\rho}^2 & 0\\0 & 2\bar{\gamma}\bar{v} & 0\end{bmatrix}\begin{bmatrix}\widetilde{\rho}\\\widetilde{v}\\\widetilde{p}\end{bmatrix}-\begin{bmatrix}\bar{v} & \bar{\rho} & 0\\0 & \bar{v} & 1/\bar{\rho}\\0 & \gamma\bar{p} & \bar{v}\end{bmatrix}\frac{\partial}{\partial x}\begin{bmatrix}\widetilde{\rho}\\\widetilde{v}\\\widetilde{p}\end{bmatrix}$$

$$+\frac{\mu}{\bar{\rho}}\begin{bmatrix}0 & 0 & 0\\0 & 4/3 & 0\\-c^2/\mathrm{Pr} & 0 & \gamma/\mathrm{Pr}\end{bmatrix}\frac{\partial^2}{\partial x^2}\begin{bmatrix}\widetilde{\rho}\\\widetilde{v}\\\widetilde{p}\end{bmatrix}+\bar{\gamma}\begin{bmatrix}0\\0\\1\end{bmatrix}q,\quad(1.18)$$

where for convenience, we introduced the first dimensionless number, the Prandtl number $\mathrm{Pr}:= c_p\mu/\kappa$ and used the relation $c^2=\gamma\bar{p}/\bar{\rho}=\gamma R\overline{T}$ for the nominal speed of sound c.

Linearized (p,v,T)-System

$$\frac{\partial}{\partial t}\begin{bmatrix}\widetilde{p}\\\widetilde{v}\\\widetilde{T}\end{bmatrix}=\beta\begin{bmatrix}0 & 2\bar{\gamma}\bar{v} & 0\\R\overline{T}\bar{v}/\bar{p}^2 & -R\overline{T}/\bar{p} & -R\bar{v}/\bar{p}\\-\bar{\gamma}\overline{T}\bar{v}^2/\bar{p}^2 & 2\bar{\gamma}\overline{T}\bar{v}/\bar{p} & \bar{\gamma}\bar{v}^2/\bar{p}\end{bmatrix}\begin{bmatrix}\widetilde{p}\\\widetilde{v}\\\widetilde{T}\end{bmatrix}-\begin{bmatrix}\bar{v} & \gamma\bar{p} & 0\\R\overline{T}/\bar{p} & \bar{v} & 0\\0 & \gamma\overline{T} & \bar{v}\end{bmatrix}\frac{\partial}{\partial x}\begin{bmatrix}\widetilde{p}\\\widetilde{v}\\\widetilde{T}\end{bmatrix}$$

$$+\bar{\gamma}\kappa\begin{bmatrix}0 & 0 & 1\\0 & \frac{4\mathrm{Pr}\overline{T}}{3\gamma\bar{p}} & 0\\0 & 0 & \overline{T}/\bar{p}\end{bmatrix}\frac{\partial^2}{\partial x^2}\begin{bmatrix}\widetilde{p}\\\widetilde{v}\\\widetilde{T}\end{bmatrix}+\bar{\gamma}\begin{bmatrix}1\\0\\\overline{T}/\bar{p}\end{bmatrix}q,\quad(1.19)$$

The pressure boundary conditions (1.11) become

$$\widetilde{p}(t,0) = u_s(t)$$

$$\widetilde{p}(t,L) = 0,$$

(1.20)

while the other two stay the same, with each state replaced by its deviation.

The above equations are here for completeness' sake, and to drive home the point that diligently defined dimensionless quantities can make life a lot easier. In the next section we will do exactly that: non-dimensionalize (1.18) and (1.19). But before we can embark on that, the heat transfer model (1.17) needs to be linearized, too.

Fortunately, the dynamic part (1.17a) is linear already, hence we only need to linearize what amounts to King's law. Following the same recipe as above, we obtain

$$\widetilde{Q}_K(t) = \left[h'(\overline{v})(T_w - \overline{T}) \quad -h(\overline{v})\right] \begin{bmatrix} \widetilde{v} \\ \widetilde{T} \end{bmatrix},$$

(1.21)

where $h(v) = \ell_w \left(\kappa + \kappa_v \sqrt{|v|}\right)$ is the velocity dependent heat transfer coefficient, as in (1.15), and $h'(\cdot)$ is its first derivative.

It should be stressed here that the significance of the steady buoyancy-induced upward velocity $\overline{v}$ is very different for the heat release than for the gas dynamics. While we will see that $\overline{v}$ only contributes to the gas dynamics in form of the steady-state Mach number $\mathrm{Ma} = \overline{v}/c$ and can be neglected in deriving the wave-like equation (1.27), for the heat transfer assuming no steady upward component in the velocity would result in an invalid linearization, since the derivative of $\sqrt{|\cdot|}$ is discontinuous at 0. The importance of the offset in velocity lies in moving to the "linearizable part" of the square root function, see Figure 1.16 for an illustration.

If we would like to use ρ and p as the thermodynamic state variables – and we will – then the
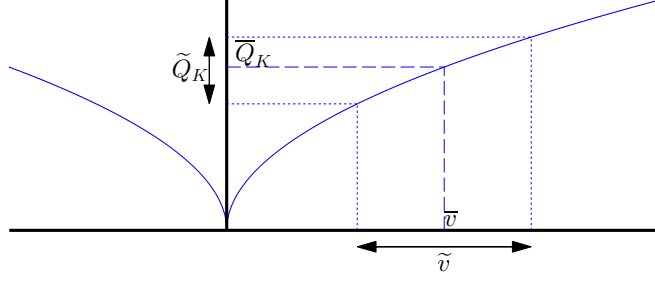
Figure 1.16 – *An illustration of the meaning of the small steady upward velocity $\bar{v}$ for the heat release process: without it, the linearization is not valid because King's law is not differentiable at 0. Also note how the square root flattens as the velocity increases.*

temperature fluctuation $\widetilde{T}$ has to be removed using the ideal gas law. That yields

$$\widetilde{Q}_K(t) = \begin{bmatrix} h(\bar{v})\frac{\overline{T}}{\bar{\rho}} & h'(\bar{v})\left(T_{\text{wire}} - \overline{T}\right) & -h(\bar{v})\frac{\overline{T}}{\bar{p}} \end{bmatrix} \begin{bmatrix} \widetilde{\rho} \\ \widetilde{v} \\ \widetilde{p} \end{bmatrix}, \tag{1.22}$$

where $\overline{T}$ can be interpreted as simply a parameter.

Thus, by pairing either (1.21) or (1.22) with the first-order lag (1.17a), the linearized heat release dynamics are either a two-input single-output, or a three-input single-output system. In anticipation of Section 1.4.5 we mention that the contributions of the thermodynamic states can be neglected compared to the influence of the velocity fluctuations $\widetilde{v}$, so that the heat transfer model becomes a single-input single-output system.

1.4.3 Dimensionless Linearized Gas Dynamics

Making variables dimensionless can also be interpreted as a normalization, and so the objective of introducing dimensionless variables and parameters is threefold: to reduce the number of parameters, to make equations look more appealing, and to yield normalized variables of similar sizes to enable comparisons and arguments like the ones we will make to neglect certain

41

Quantity		Quantity		Quantity	
Density	$r = \widetilde{\rho}/\overline{\rho}$	Pressure	$\psi = \widetilde{p}/\overline{p}$	Temperature	$\theta = \widetilde{T}/\overline{T}$
Velocity	$m = \widetilde{v}/c$	Time	$\tau = ct/L$	Space	$\xi = x/L$
Heat release	$\phi = \frac{\overline{\gamma}qL}{\overline{\rho}c^3}$	Heat release	$\Phi = \frac{QL/c}{\overline{\rho}c_v V_{\text{tube}}\overline{\overline{T}}}$		
Steady velocity	$\text{Ma} = \overline{v}/c$	Friction	$f = \beta L/(\overline{\rho}c)$	Diffusion	$\text{Pe} = \frac{\overline{\rho}c_p Lc}{\kappa}$
Diffusion	$\text{Pr} = c_p\mu/\kappa$				

Table 1.1 – *Dimensionless quantities and parameters introduced in Section 1.4.3.*

contributions to the gas dynamics in Section 1.4.5. We begin by introducing

$$\psi := p/(\gamma\overline{p}), \qquad \theta := T/(\overline{\gamma}\,\overline{T}), \quad \text{and} \quad r := \rho/\overline{\rho}$$

and arguing that, since sound waves are isentropic (see also Footnote 8 on Page 37), and we are looking at mainly sound waves, all three of them can be expected to be of similar size.

For an ideal isentropic process it holds that $p/\rho^{\gamma} = $ const., which is known as Poisson's law. Let this constant be K. Then we have:

$$\frac{\overline{p} + \widetilde{p}}{(\overline{\rho} + \widetilde{\rho})^{\gamma}} = \frac{1 + \widetilde{p}/\overline{p}}{(1 + \widetilde{\rho}/\overline{\rho})^{\gamma}} \frac{\overline{p}}{\overline{\rho}^{\gamma}} = K$$

$$\iff \quad \frac{1 + \widetilde{p}/\overline{p}}{(1 + \widetilde{\rho}/\overline{\rho})^{\gamma}} = 1 \quad \iff \quad 1 + \widetilde{p}/\overline{p} = (1 + \widetilde{\rho}/\overline{\rho})^{\gamma} = 1 + \gamma\widetilde{\rho}/\overline{\rho} + \mathcal{O}(2)$$

$$\implies \widetilde{p}/(\gamma\overline{p}) \approx \widetilde{\rho}/\overline{\rho}$$

An analogous computation can be done for θ.

Next, we introduce the dimensionless velocity as

$$m := \widetilde{v}/c,$$

where c is the nominal speed of sound, and the letter m is chosen because of the similarity

of this definition with the Mach number $\mathrm{Ma} = v/c$. It is known that m and ψ as defined here are of the same order in acoustic waves, e.g. [Dow97], but it can also be shown quickly, see Appendix 1.C.

Apart from the four dependent variables ρ, T, p, and v, we have the two independent variables x and t. We choose the tube length L as the characteristic length to scale x by, and the passage time L/c of a sound wave through the tube as the time scale, i.e.

$$\xi := x/L, \quad \frac{\partial}{\partial \xi} = L\frac{\partial}{\partial x}, \qquad \tau := ct/L, \quad \frac{\partial}{\partial \tau} = \frac{L}{c}\frac{\partial}{\partial t}.$$

Lastly, the (for now) exogenous input q needs to be scaled. Defining

$$\phi := \frac{q\,L/c}{\bar{\rho}c_v\bar{T}} = \frac{\bar{\gamma}q\,L/c}{\gamma\bar{p}} = \frac{\bar{\gamma}qL}{\bar{\rho}c^3}$$

makes for the prettiest equations, and ϕ can be interpreted as the ratio of heat transferred during the characteristic time span and specific internal energy at steady state.

Without further ado, rewriting (1.18) yields

$$\frac{\partial}{\partial \tau}\begin{bmatrix} r \\ m \\ \psi \end{bmatrix} = f\begin{bmatrix} 0 & 0 & 0 \\ \mathrm{Ma} & -1 & 0 \\ 0 & 2\bar{\gamma}\mathrm{Ma} & 0 \end{bmatrix}\begin{bmatrix} r \\ m \\ \psi \end{bmatrix} - \begin{bmatrix} \mathrm{Ma} & 1 & 0 \\ 0 & \mathrm{Ma} & 1 \\ 0 & 1 & \mathrm{Ma} \end{bmatrix}\frac{\partial}{\partial \xi}\begin{bmatrix} r \\ m \\ \psi \end{bmatrix}$$

$$+ \frac{1}{\mathrm{Pe}}\begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{4\mathrm{Pr}}{3} & 0 \\ -1 & 0 & \gamma \end{bmatrix}\frac{\partial^2}{\partial \xi^2}\begin{bmatrix} r \\ m \\ \psi \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\phi, \quad (1.23)$$

and doing the same to (1.19) we obtain

$$
\frac{\partial}{\partial \tau}
\begin{bmatrix} \psi \\ m \\ \theta \end{bmatrix}
= f
\begin{bmatrix}
0 & 2\bar{\gamma}\mathrm{Ma} & 0 \\
\gamma\mathrm{Ma} & -1 & -\bar{\gamma}\mathrm{Ma} \\
-(\gamma\mathrm{Ma})^2 & -2\gamma\mathrm{Ma} & \gamma\bar{\gamma}\mathrm{Ma}^2
\end{bmatrix}
\begin{bmatrix} \psi \\ m \\ \theta \end{bmatrix}
-
\begin{bmatrix}
\mathrm{Ma} & 1 & 0 \\
1 & \mathrm{Ma} & 0 \\
0 & 1 & \mathrm{Ma}
\end{bmatrix}
\frac{\partial}{\partial \xi}
\begin{bmatrix} \psi \\ m \\ \theta \end{bmatrix}
$$

$$
+ \frac{1}{\mathrm{Pe}}
\begin{bmatrix}
0 & 0 & \bar{\gamma} \\
0 & \frac{4\mathrm{Pr}}{3} & 0 \\
0 & 0 & \gamma
\end{bmatrix}
\frac{\partial^2}{\partial \xi^2}
\begin{bmatrix} \psi \\ m \\ \theta \end{bmatrix}
+
\begin{bmatrix} 1 \\ 0 \\ \gamma/\bar{\gamma} \end{bmatrix}
\phi, \quad (1.24)
$$

where

- $$\mathrm{Ma} := \frac{\bar{v}}{c}$$

  is the famous Mach number, relating the velocity of matter to the velocity of sound.

- $$f := \frac{L\beta}{\bar{\rho}c} = \frac{L\beta c}{\gamma \bar{p}}$$

  is a dimensionless friction factor. It can be interpreted as the ratio of mechanical impedance βL to characteristic acoustic impedance $\bar{\rho}c$ of the tube.

- $$\mathrm{Pr} := \frac{\mu/\bar{\rho}}{\kappa/(\bar{\rho}c_p)} = \frac{c_p\mu}{\kappa}$$

  is the Prandtl number, the ratio of diffusion of momentum to diffusion of heat.

- $$\mathrm{Pe} := \frac{Lc}{\kappa/(\bar{\rho}c_p)}$$

  has the form of the Peclet number, which relates advective to diffusive transport rates. However, the nominator in our definition does not correspond to the actual transport of matter, so the interpretation as a Peclet number is purely formal. We also have $\mathrm{Pe} = \mathrm{Re}\,\mathrm{Pr}$, with $\mathrm{Re} := (\bar{\rho}cL)/\mu$ the famed Reynolds number which has to be interpreted with the same caveat as the Peclet number.

Now we shift attention to the heat release model. Using the definition of ϕ and (1.14), we

obtain

$$\phi(\tau,\xi) = \frac{\bar{\gamma}L}{\bar{\rho}c^3}q(L\tau/c,L\xi) = \frac{\bar{\gamma}L}{\bar{\rho}c^3}\frac{1}{A}\delta_m\big(L(\xi-x_o/L)\big)Q(L\tau/c) = L\delta_m\big(L(\xi-x_o/L)\big)\frac{\bar{\gamma}}{\bar{\rho}Ac^3}Q(L\tau/c).$$

From the scaling property of the Dirac distribution it follows that $L\delta_m\big(L(\xi-x_o/L)\big) = \delta(\xi-\xi_o)$, where $\xi_o := x_o/L$. We define the dimensionless total heat power

$$\Phi(\tau) := \frac{\bar{\gamma}Q(L\tau)}{\bar{\rho}Ac^3} = \frac{Q(L\tau/c)L/c}{\bar{\rho}c_v V_{\text{tube}}\bar{T}};$$

according to the last expression, Φ could be interpreted as the ratio of heat added during the characteristic time L/c to the internal energy in the tube at steady state. We obtain:

$$\phi(\tau,\xi) = \delta(\xi-\xi_o)\Phi(\tau). \tag{1.25}$$

The entire dimensionless linearized version of (1.17) is then

$$\tau_{\text{hr}}\dot{\Phi}(\tau) = -\Phi(\tau) + \Phi_K(\tau) \tag{1.26a}$$

$$\Phi_K(\tau) = \begin{bmatrix} h_t & h_v & -\gamma h_t \end{bmatrix}\begin{bmatrix} r \\ m \\ \psi \end{bmatrix}(\tau,\xi_o) = h_v\begin{bmatrix} \bar{h}_t\text{Ma} & 1 & -\gamma\bar{h}_t\text{Ma} \end{bmatrix}\begin{bmatrix} r \\ m \\ \psi \end{bmatrix}(\tau,\xi_o),\ \text{or} \tag{1.26b}$$

$$\Phi_K(\tau) = \begin{bmatrix} 0 & h_v & \gamma h_t \end{bmatrix}\begin{bmatrix} \psi \\ m \\ \theta \end{bmatrix}(\tau,\xi_o) = h_v\begin{bmatrix} 0 & 1 & \gamma\bar{h}_t\text{Ma} \end{bmatrix}\begin{bmatrix} \psi \\ m \\ \theta \end{bmatrix}(\tau,\xi_o), \tag{1.26c}$$

where

$$h_v := \frac{h'(\bar{v})}{\bar{\rho}c_v A}\frac{T_{\text{wire}}-\bar{T}}{\bar{T}}, \qquad h_t := \frac{h(\bar{v})/c}{\bar{\rho}c_v A}, \qquad \text{and} \quad \bar{h}_t := 2\frac{\bar{T}}{T_{\text{wire}}-\bar{T}}\left(\frac{\kappa}{\kappa_v\sqrt{v}}+1\right).$$

The second forms in (1.26b) and (1.26c) would be very informative if $\bar{h}_t$ can be expected to be

45
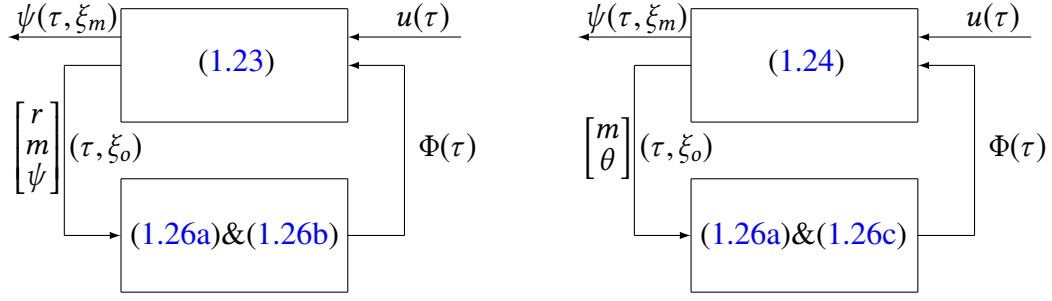
Figure 1.17 – *The diagram in Figure 1.13 populated with the linearized dimensionless models derived in Sections 1.4.2 and 1.4.3.*

of order 1 or less, because then the contributions of density, pressure and temperature fluctuations to the heat release are of order Ma compared to the contribution of velocity fluctuations. The first term is the (inverse) dimensionless temperature difference $\frac{\overline{T}}{T_{\text{wire}}-\overline{T}}$, which, using the estimate (1.38) from Appendix 1.B, can be asserted to be around 1 or even less. The other term in question is $\kappa/(\kappa_v \sqrt{\overline{v}})$, the ratio between conductive and convective heat transfer; these are typically comparable — see [Kin14, Tables V & VI] and note that velocities are measured in cm/s — and hence this term is of order unity, too, and we conclude that velocity fluctuations have by a factor of approximately Ma greater influence on the heat transfer than the thermodynamic state of the gas.

Figure 1.17 puts the work of this section together into a complete model, which now has the form of a lower linear fractional transformation (LLFT).

A great advantage of the dimensionless equations, apart from the reduction of clutter and number of parameters, is that all effects are encoded in a single parameter each: Ma quantifies buoyancy, or the steady-state upstream velocity, f quantifies the friction effect on the walls, and $1/$Pe quantifies the importance of diffusive effects. That means we can now obtain models of the gas dynamics at different levels of complexity by neglecting one effect or the other, simply by setting the corresponding parameter to 0. This is the subject of the following sections.

Quantity	Symbol	Value used	Comment
Temperature	$\overline{T}$	293 K	
Density	$\overline{\rho}$	1.2 kg/m^3	
Pressure	$\overline{p}$	101 kPa	
Adiabatic Ratio	γ	1.4	$\gamma = c_p/c_v$
Speed of Sound	c	343 m/s	$c^2 = \gamma\overline{p}/\overline{\rho}$
Specific Heat Capacities	$c_p,\ c_v$	1008, 718 J/(kg K)	
Ideal Gas Constant	R	290 J/(K kg)	$R = c_p - c_v$
Dynamic Viscosity	μ	$2 \cdot 10^{-5}$ Pa s	
Thermal Conductivity	κ	0.024 W/(m K)	
Wire Temperature	T_{wire}	933 K	see (1.38)
Tube Length	L	1.219 m	
Tube Diameter	D	0.076 m	
Mach Number	Ma	10^{-3}	
Prandtl Number	Pr	0.84	
Peclet Number	Pe	$2.1 \cdot 10^{7}$	
Friction Coefficient	f	$3.2 \cdot 10^{-4}$	see (1.6)
Heat Release Time Constant	τ_{hr}	0.94	see (1.16)

Table 1.2 – *Parameters and values*

1.4.4 Rough Parameter Estimates

In this section, we provide some rough estimates of the various parameters that have appeared in the equations so far. For mechanical and thermal properties of the gas, e.g. μ and κ, we assume air at room temperature and that they do not vary appreciably with temperature or pressure. The values are summarized in the first two parts of Table 1.2.

Certainly a lot harder to quantify is the steady-state velocity $\bar{v}$: One could try to measure it, however measuring very small air velocities is tricky and requires extra equipment; instead, we approximate $\bar{v}$ by the characteristic velocity of the free convection that is driven by the hot wire. In [SC67], we find

$$u_c = \sqrt{g \frac{\Delta T}{T} L_c},$$

where $g = 9.81 \text{ m/s}^2$ is the gravitational acceleration, we assumed an ideal gas, hence is the thermal expansion coefficient β equals $1/T$, and L_c is a characteristic length, for which we choose the coil diameter $L_c \approx 6$ mm, because the outer surface of the coil windings approximately constitutes a horizontal cylinder as considered in [SC67]. Using the estimate (1.38) for the wire temperature, and standard conditions $T = 293$ K for the gas temperature, this gives an estimate of

$$\bar{v} \approx \sqrt{9.81 \, \text{m/s}^2 \frac{933\,\text{K} - 293\,\text{K}}{293\,\text{K}} 6 \cdot 10^{-3}\text{m}} = 0.36 \, \text{m/s}.$$

That yields $\text{Ma} \approx 10^{-3}$, which is the same order of magnitude as the Mach numbers assumed in the literature, e.g. [Mat03, Hec88].

Using (1.16) to obtain an estimate of the heat release time constant, we get

$$\tau_{\text{hr}} \approx c/L \cdot 0.2 d_{\text{wire}}/\bar{v} \approx 0.94.$$

It is again unclear what wire diameter should be used here. Since [Lig54] treats the case of a straight horizontal wire, whereas we have a coil in which only portions of the wire are

horizontal, it does not make sense to use the diameter of the used NiCr wire. Since the windings of the coil are closer together than twice the boundary layer thickness,[9] it is reasonable to assume that the coil develops a boundary similar to a cylinder with the same radius. However, the exact value of τ_{hr} will not influence our analysis much, see e.g. Figure 1.19.

To get an idea of the size of the friction coefficient f, we return to the explanations in Footnote 4 on Page 30 and estimate

$$\beta \approx \frac{32\mu}{D^2} \approx 0.11 \text{ kg}/(\text{m}^3\,\text{s})$$

$$\Rightarrow f \approx 3 \cdot 10^{-4}.$$

All estimates presented here and in Table 1.2 should be taken with a tablespoon of salt; the hope is that they capture the real quantity to within an order of magnitude.

1.4.5   Analytic Solution of the Most Simplified Dynamics

In this section, we neglect as much as defendable, and by doing so obtain a compact explicit expression for the transfer functions of the gas dynamics and the heat release. Despite its simplicity, the model allows to predict the instability and the correct frequency of the limit cycle oscillations causing the hum. However, there is a trade-off between simplicity and how well the model describes the physical reality: since all damping effects are neglected, the poles of the gas dynamics move onto the imaginary axis, which presents a clinical case that makes analysis and comparison to the identification results difficult.

It should be obvious that in this experiment we have Ma $\ll$ 1. Friction and diffusive effects are typically also very small in gas dynamics, hence the first approximation we make is

---

[9] An approximate (laminar) boundary layer thickness at 90° from the stagnation point is given by $\delta_{90} = 2.33\sqrt{\mu\pi d_{\text{wire}}/\bar{v}} \approx 0.6$ mm, see e.g. [PW02, Sec. 8.6.3]. This is of a similar order as the distance between the windings in our coil.

Ma, $1/\text{Pe}$, $f \approx 0$. Application to (1.23) yields

$$\frac{\partial}{\partial \tau} \begin{bmatrix} r \\ m \\ \psi \end{bmatrix} = - \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \frac{\partial}{\partial \xi} \begin{bmatrix} r \\ m \\ \psi \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \phi.$$

We see that the density r does not couple into the dynamics of velocity and pressure, and thus can be dropped from consideration.

Incorporating also the separated form (1.25) for the heat input ϕ, we are left with

$$\frac{\partial}{\partial \tau} \begin{bmatrix} m \\ \psi \end{bmatrix} = - \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \frac{\partial}{\partial \xi} \begin{bmatrix} m \\ \psi \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \delta(\xi - \xi_o) \Phi(\tau) \tag{1.27}$$

with the boundary conditions

$$\psi(\tau, 0) = u(\tau) \qquad \psi(\tau, 1) = 0, \tag{1.28}$$

which are derived from (1.11) by setting $u(t) = u_s(t)/(\gamma \bar{p})$.

Equations (1.27) and (1.28) can be identified as a realization of an undamped wave equation, driven by the *scalar* terms for the exogenous input Φ and boundary control u. Sound waves are commonly assumed to obey a wave equation, that we find one as a special case of our model is thus reassuring.

Applying Ma ≈ 0 also to the heat release dynamics (1.26) reduces them to a single input system, namely

$$\tau_{\text{hr}} \Phi(\tau) = -\Phi(\tau) + \Phi_K(\tau) \tag{1.29a}$$

$$\Phi_K(\tau) = h_v \, m(\tau, \xi_o), \tag{1.29b}$$

and the model presents itself as the feedback interconnection of the gas dynamics and a single-

50
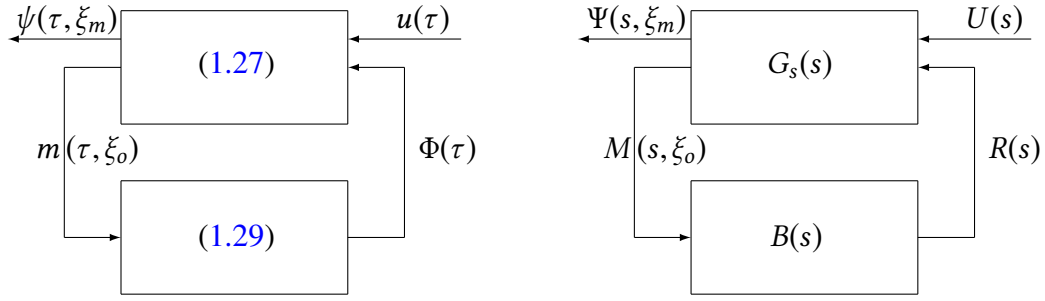
Figure 1.18 – *The diagram in Figure 1.13, updated with the simplest form of the linearized dimensionless models as derived in Section 1.4.5.*

input single-output system as shown in Figure 1.18 on the left.

It should be stressed again that the linearized heat release models (1.26c) and (1.26b) have been derived under the assumption of a small, but positive steady velocity $\bar{v}$. For $\bar{v} = \text{Ma} = 0$, the linearization would not be valid. This underlines the different importance the buoyancy-driven steady component of the gas velocity has for the gas dynamics and the heat release.

To make further progress, we will now derive transfer functions for the blocks in Figure 1.18. As can be seen there, the gas dynamics have two inputs and two outputs: the heat input $\Phi(\tau)$ and the speaker actuation $u(\tau)$ drive the pressure at ξ_m as the measurement, and the velocity at ξ_o which modulates the heat release. Applying the Laplace transform (in time) and rearranging (1.27) yields

$$\begin{bmatrix} M' \\ \Psi' \end{bmatrix}(s,\xi) = -\begin{bmatrix} 0 & s \\ s & 0 \end{bmatrix}\begin{bmatrix} M \\ \Psi \end{bmatrix}(s,\xi) + \begin{bmatrix} 1 \\ 0 \end{bmatrix}\delta(\xi - \xi_o)R(s) \tag{1.30a}$$

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}U(s) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} M \\ \Psi \end{bmatrix}(s,0) + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} M \\ \Psi \end{bmatrix}(s,1), \tag{1.30b}$$

where M, Ψ, R and U are the Laplace transforms of m, ψ, Φ and u, respectively, and the prime $\bullet'$ denotes differentiation with respect to the spatial variable ξ.

In Appendix 1.D it is shown, how transfer functions for two-point boundary value problems in

51

the form (1.30) can be computed. Following the recipe there, we obtain the following transfer function:

$$
G_s : \begin{bmatrix} u(\cdot) \\ \Phi(\cdot) \end{bmatrix} \mapsto \begin{bmatrix} \psi(\cdot, \xi_m) \\ m(\cdot, \xi_o) \end{bmatrix} \quad G_s(s) = \begin{bmatrix} \dfrac{\sinh\left((1-\xi_m)s\right)}{\sinh(s)} & \dfrac{\sinh\left((1-\xi_m)s\right)\sinh\left(\xi_o s\right)}{\sinh(s)} \\[2ex] \dfrac{\cosh\left((1-\xi_o)s\right)}{\sinh(s)} & -\dfrac{\sinh\left((1-2\xi_o)s\right)}{2\sinh(s)} \end{bmatrix}, \tag{1.31}
$$

where we choose

$$
m(\tau, \xi_o) = \frac{1}{2}\left(\lim_{\xi \searrow \xi_o} m(\tau, \xi) + \lim_{\xi \nearrow \xi_o} m(\tau, \xi) \right),
$$

i.e. the value of m at the jump discontinuity caused by the assumption of a concentrated heat release is taken to be the average of the left- and right-sided limit. This choice is made mainly because it yields more appealing expressions, the difference between the right- and left-sided limits is constant, so choosing either limit instead of their average would yield the same analysis results.

If $\xi_o = 1/4$, i.e. the coil is at a quarter length from the bottom of the tube, $G_s(s)$ takes a particularly simple form:

$$
G_s(s) = \begin{bmatrix} \dfrac{\sinh\left((1-\xi_m)s\right)}{\sinh(s)} & \dfrac{\sinh\left((1-\xi_m)s\right)\sinh(s/4)}{\sinh(s)} \\[2ex] \dfrac{\cosh(3s/4)}{\sinh(s)} & -\dfrac{1}{4\cosh(s/2)} \end{bmatrix}. \tag{1.32}
$$

The boundary layer's transfer function is, according to (1.29),

$$
B : m(\cdot, \xi_o) \mapsto \Phi(\cdot) \qquad B(s) := \frac{h_v}{\tau_{\mathrm{hr}} s + 1}, \tag{1.33}
$$

and we can represent the entire model as a linear fractional transformation of explicitly known transfer functions, as shown in Figure 1.18 on the right.

52
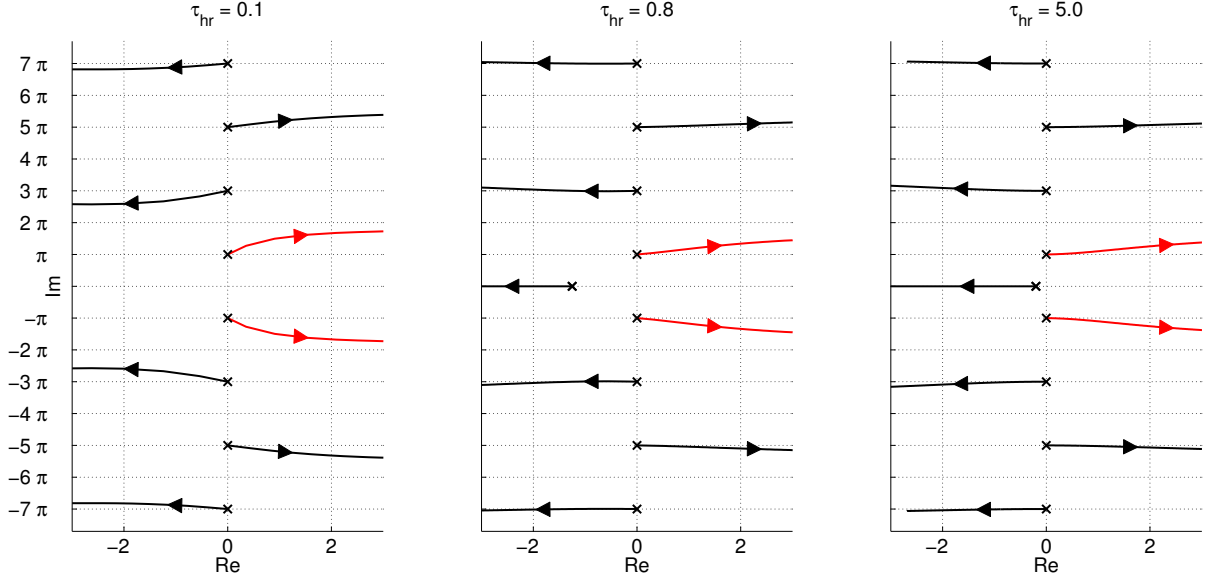
Figure 1.19 – *Root loci for the feedback interconnection of $G_{s,22}(s)$ and the linearized heat release dynamics $B(s)$ for different values of the boundary layer time constant τ_{hr}. The plot should be interpreted as if it were shifted slightly to the left, as would be the case if realistic damping effects were included. (Regarding this argument, see also Sections 1.4.6 and 1.4.7.) The branches shown in red correspond to the thermoacoustic instability. See Table 1.2 for the parameter values used.*

Root Locus Analysis

Since $B(s)$ is stable, and all transfer functions of $G_s(s)$ are at least marginally stable, the only possible mechanism of instability is the feedback loop involving $G_{s,22}$ and B. Since open-loop poles and zeros are known, but several involved gains are only known to be positive, considering the root locus is a natural choice.

The open-loop transfer function of the feedback loop is

$$G_{s,22}(s)B(s) = -\frac{1}{4\cosh(s/2)}\frac{h_v}{\tau_{hr}s+1} = \frac{-h_v}{4}\overbrace{\frac{1}{\cosh(s/2)\,(\tau_{hr}s+1)}}^{=:L(s)}, \tag{1.34}$$

and since the feedback is positive, the negative DC gain means that we have to consider the standard negative-gain ("180º") root locus with open loop $L(s)$.

The transfer function L has infinitely many poles on the imaginary axis at

$$\omega_n = \pi + n\, 2\pi, \qquad n \in \mathbb{Z},$$

and one at $s = -1/\tau_{\mathrm{hr}}$. Hence a little care is needed in generating the root locus, since not all of the classical rules for drawing a root locus apply to systems with infinitely many poles. A simple criterion to compute the departure angles from open-loop poles is derived in Appendix 1.E, as the departure angles turn out to be the most important feature.

The root loci shown in Figure 1.19 are generated numerically by continuation over the gain k: the (nonlinear) equation $1 + k_{i+1}L(s) = 0$ is solved with a Newton method, using the poles computed in iteration i as initial values. At $k_0 = 0$, the initial poles are the open-loop poles. This method has the disadvantage that effects of inserting poles and zeros into the open loop cannot be predicted easily anymore, however for our purposes of predicting the frequency of the instabilities, it is suitable.

Considering the root loci in Figure 1.19, we see that, starting from the origin, every other pole moves to the right-half plane right away, indicating that, no matter how small $h_v/4$ might be, the heat release feedback renders the Rijke tube an unstable system. The root locus is "robust" with respect to the actual value of τ_{hr} in the sense that larger values of τ_{hr} rotate the branches closer to horizontal, i.e. the departure angles get closer to 0 and π, respectively. If damping and/or radiative effects were to be included, the imaginary axis poles would be shifted to the left in a similar manner to those in Figure 1.9.[10] Therefore the open-loop dynamics are stable, but have an array of very lightly damped poles. With the positive feedback from coil heat release, half of the acoustic poles eventually move into the right-half plane (RHP). The first pole (pair) to cross into the RHP corresponds to the fundamental frequency and is depicted by the red branches in Figure 1.19; this is a reasonable assumption, since lower frequencies

---

[10]This issue is further elaborated on in Section 1.4.7, where the spectrum of the model including diffusion and friction is treated.

are typically damped less and hence those branches originate closer to the imaginary axis. Its dimensionless imaginary axis crossing frequency $\omega_0 \approx \pi$ corresponds to a physical (angular) frequency of $\omega_0^{\text{ph}} = \omega_0 c/L$, or $f_0^{\text{ph}} = c/(2L)$, exactly as observed in the experiment, see (1.1).

Once the instability is triggered, the system exhibits growing oscillations and leaves the linear regime (since the $\widetilde{\bullet}$ variables need to be small for the linear analysis to apply). Intuitively, the square root term in the heat release model (1.15) then offers diminishing returns, i.e. at higher acoustic velocities, it almost saturates, and even larger oscillations in velocity are not supported by the enhanced heat transfer anymore. The system is brought into a limit cycle, the shape of which can be seen in Figure 1.3(c).

The observations that odd-numbered (even-numbered) branches depart towards the left (towards the right) and that the angles of departure are closer to horizontal, the larger τ_{hr} is, can be confirmed applying the root locus rule derived in Appendix 1.E. Applying (1.51) to $L(s)$, we get:

$$\varphi_{\text{dep}}(s) = -\angle\left(\sinh(s/2)\frac{\tau_{\text{hr}}s + 1}{2} + \tau_{\text{hr}}\cosh(s/2)\right) + \angle(1)^{\,0} - \pi \tag{1.35}$$

$$\Rightarrow \varphi_{\text{dep},n} = \varphi_{\text{dep}}\left(j(\pi + n\,2\pi)\right) = -\angle\left((-1)^n(\tau_{\text{hr}}\pi\frac{2n + 1}{2} - j/2)\right) - \pi =$$

$$\pi - \arctan_2\left((-1)^n, (-1)^{n+1}\tau\pi(2n + 1)\right) = \begin{cases} \pi - \arctan\left(\frac{1}{\tau\pi(2n+1)}\right) & \text{if } n \text{ is odd} \\[2ex] \arctan\left(\frac{1}{\tau\pi(2n+1)}\right) & \text{if } n \text{ is even,} \end{cases}$$

where $\arctan_2$ is the two-argument (also "four-quadrant") inverse tangent. In Table 1.3, a few numerical values are given, and we see that *(a)* the branches go left for even and right for odd indices, *(b)* the larger τ_{hr}, the closer to horizontal are the angles, and *(c)* the higher order the pole, the closer to horizontal the angles. All three observations can also be made in the numerically obtained root loci in Figure 1.19.

n τ_{hr}	0	1	2	3
0.01	88.20º	-95.38º	81.07º	-102.40º
0.1	72.56º	-133.30º	32.48º	-155.55º
0.4	38.51º	-165.14º	9.04º	-173.51º
1	17.66º	-173.94º	3.64º	-177.40º
5	3.64º	-178.78º	0.73º	-179.48º

Table 1.3 – *Departure angles for the first 4 open-loop acoustic poles and several values of τ_{hr}, computed using (1.51).*
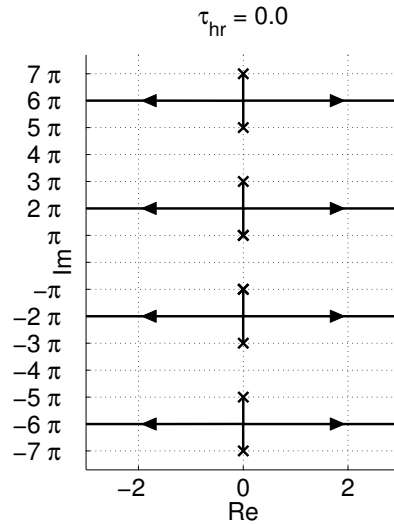
Figure 1.20 – *What the root loci in Figure 1.19 would look like if the wire's thermal boundary layer effect was ignored, which amounts to setting $\tau_{hr} = 0$. The fundamental frequency branch moving into the right-half plane is now predicted to correspond to twice the observed frequency, illustrating that inclusion of the first-order delay is crucial. The parameters used are the same as in Figure 1.19.*

Significance of the Heat Release Time Constant τ_{hr}   Figure 1.20 shows what has been observed in [Lig54] already, namely that the slight delay introduced by the boundary layer is crucial for the model to be valid. The observation can also be confirmed using formula (1.35) in the limit where $\tau_{hr} \to 0$, in which case $\varphi_{dep}(j\pi(1+2n)) = (-1)^n\pi/2$.

Comparison to Identified Model

With all the simplifications made, how closely does the model of Figure 1.18 correspond to reality – or rather the identified model as a proxy for reality – beyond explaining the thermo-acoustic oscillations? It is important to understand first the relationship between the modeled and identified transfer functions. The open-loop Bode plot of Figure 1.8 corresponds to the identified response $F(s)$ of pressure fluctuations as measured by the microphone to pressure fluctuations induced by the speaker; the model depicted in Figure 1.18 describes the transfer function from normalized boundary pressure fluctuations $u(\tau) = \psi(\tau, 0)$ to normalized pressure fluctuations $\psi(\tau, \xi_m)$ at the microphone position, including the (linearized) heat release feedback loop:

$$\check{F} : u(\cdot) \mapsto \psi(\cdot, \xi_m)$$

$$\check{F}(s) = \mathscr{F}_\ell(G_s; B) := G_{s,11}(s) + \frac{G_{s,12}(s)B(s)G_{s,21}(s)}{1 - G_{s,22}(s)B(s)} =$$

$$\frac{\sinh((1-\xi_m)s)}{\sinh(s)} + \frac{\cosh(3s/4)\sinh((1-\xi_m)s)\sinh(s/4)B(s)}{\sinh^2(s)} \frac{1}{1 + \frac{B(s)}{4\cosh(s/2)}} =$$

$$\frac{\sinh((1-\xi_m)s)}{\sinh(s)}\left(1 + \frac{(2\cosh(s/2)-1)B(s)}{B(s) + 4\cosh(s/2)}\right),$$

where $\mathscr{F}_\ell(P; K)$ denotes the lower linear fractional transformation of P with parameter K, see e.g. [SP07]. In contrast, the identified closed-loop transfer function $T(s)$ in Figure 1.7 corre-
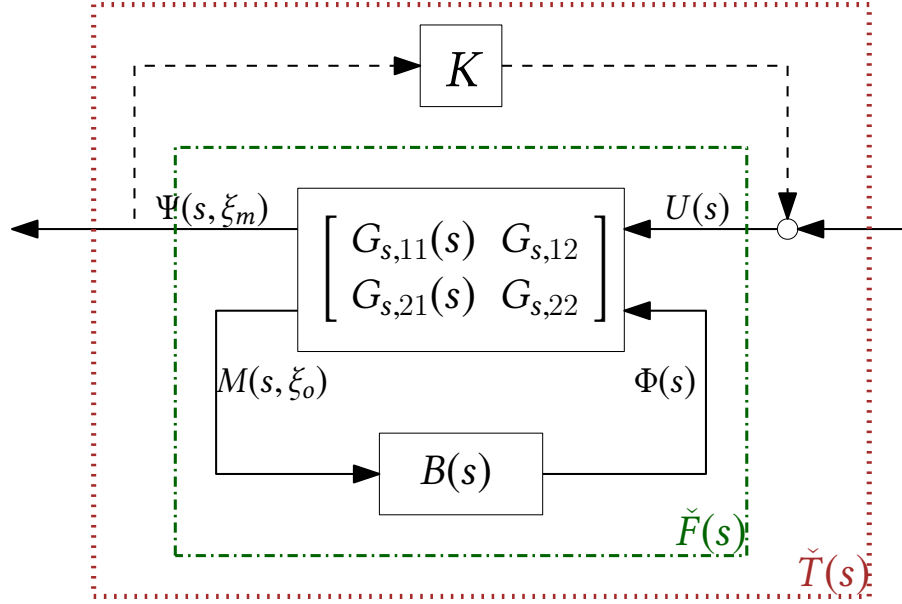
Figure 1.21 – *The modeled transfer function between speaker input and microphone output signals can be represented – with linearized heat release – as a lower linear fractional transformation $\check{F} = \mathscr{F}_\ell(G_s; B)$. In a stabilizing loop, these signals are connected through $\check{T}(s) = \frac{\check{F}(s)}{1-K\check{F}(s)}$.*

sponds to

$$\check{T}(s) = \frac{\check{F}(s)}{1 - K\check{F}(s)}.$$

For clarity, these different transfer functions are illustrated in Figure 1.21.

There remain however several unknown parameters if comparisons between $F(s)$ and $\check{F}(s)$ are to be made: an overall gain due to the unknown conversion factors of speaker and microphone, the microphone position x_m or ξ_m, and linear gain h_v and time constant t_{hr} or τ_{hr} of the heat release feedback $B(s)$. Additionally, the assumptions about friction, diffusive effects, and the steady upward flow that were made in this section — and had the advantage that the very simple model (1.27) and transfer function (1.31) could be obtained — also led to system poles moving onto the imaginary axis; in the root locus analysis, we argued that with damping, the poles will be moved slightly to the left, with higher damping at higher frequencies, and Figure 1.24 suggests that they lie on a parabola-shape.
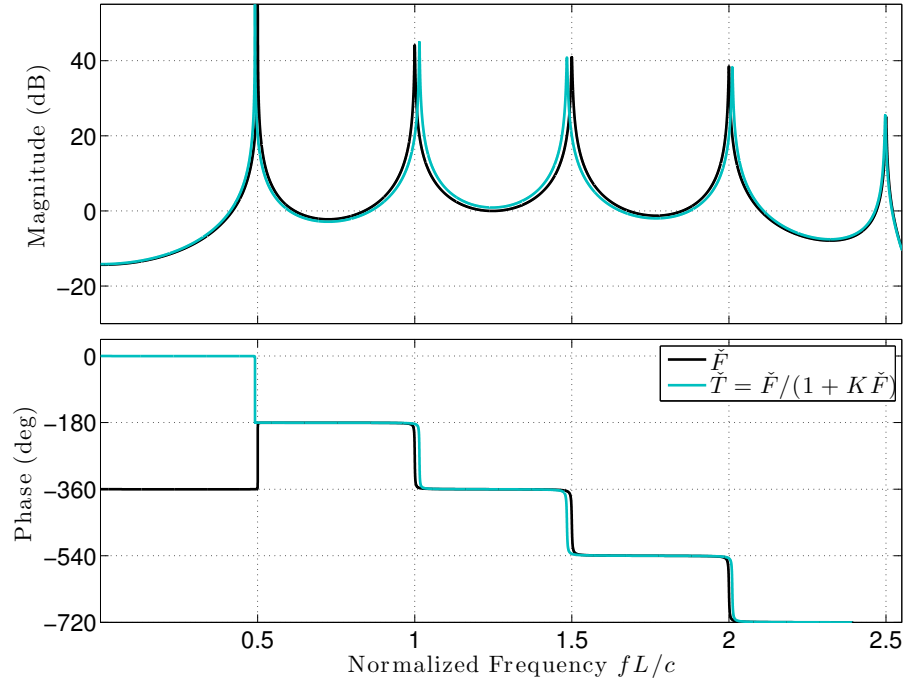
Figure 1.22 – *Bode plots of open-loop ($K = 0$) and closed-loop (with a stabilizing $K \neq 0$) frequency responses derived from the physical model of Figure 1.21. Note the opposite signs of the 180° phase increase near the fundamental mode indicating open-loop instability and closed-loop stability of that mode. The specific parameters used are $\tau_{hr} = 0.94$, $h_v = 0.034$, $K = -0.1$, $\varepsilon_0 = 5 \cdot 10^{-3}$, $\varepsilon_1 = 2 \cdot 10^{-5}$.*

If the damping effect is imitated by considering $j\omega + \varepsilon_0 + \varepsilon_1\omega^2$ instead of $j\omega$, and choices of the unknown heat release time constant, heat release gain and feedback gain are made judiciously, one obtains the responses shown in Figure 1.22; remarkably, despite employing very simple devices to treat the aforementioned difficulties, these responses qualitatively capture the most important features of Figures 1.7 and 1.8: resonant peaks, a phase increase of 180° at the first peak for the unstabilized open loop $\check{F}$, and phase drops of the same amount at the higher-order peaks for $\check{F}$ and at all peaks for the stabilized closed loop $\check{T} = \check{F}/(1 + K\check{F})$. That it is easy to find a wide range of parameter values that generate responses with these features after only a few manual iterations can be seen as evidence that the physical model indeed captures the important structures of the underlying physics. It should be emphasized however, that the real power of the model lies in explaining the thermoacoustic instability as a linearly destabilizing feedback of acoustic velocity and heat release, and the current section should be seen as more of a sanity check.

1.4.6 Neglecting Diffusive Effects Only

Inspecting Table 1.2, it appears the next effects to be included should be friction and non-zero Mach number, since Ma and f are of similar orders of magnitude, but $1/\mathrm{Pe} \ll \mathrm{Ma}$. It turns out to be necessary to use the (ψ, m, θ) form (1.24) of the gas dynamics, since the Neumann boundary condition (1.12) on velocity leads to an ill-posed problem – see also the remark after (1.47). By keeping all terms of first order in Ma and f, but still neglecting $1/\mathrm{Pe}$, (1.24) becomes

$$\frac{\partial}{\partial\tau}\begin{bmatrix}\psi\\m\\\theta\end{bmatrix} = \begin{bmatrix}0 & 0 & 0\\0 & -f & 0\\0 & 0 & 0\end{bmatrix}\begin{bmatrix}\psi\\m\\\theta\end{bmatrix} - \begin{bmatrix}\mathrm{Ma} & 1 & 0\\1 & \mathrm{Ma} & 0\\0 & 1 & \mathrm{Ma}\end{bmatrix}\frac{\partial}{\partial\xi}\begin{bmatrix}\psi\\m\\\theta\end{bmatrix} + \begin{bmatrix}1\\0\\\gamma/\bar{\gamma}\end{bmatrix}\phi,$$

together with the boundary conditions (1.28) and

$$\frac{\partial \theta}{\partial x}(\tau, 1) = 0,$$

which is derived from (1.13). We could have used the boundary condition on the bottom end $\xi = 0$ as well, the results are the same.

Putting this together in the form treated in Appendix 1.D, we obtain

$$\begin{bmatrix} \Psi' \\ M' \\ \Theta' \end{bmatrix}(s, \xi) = \begin{bmatrix} \frac{\text{Ma}s}{1-\text{Ma}^2} & \frac{f+s}{\text{Ma}^2-1} & 0 \\ \frac{s}{\text{Ma}^2-1} & -\frac{\text{Ma}(f+s)}{\text{Ma}^2-1} & 0 \\ \frac{s}{\text{Ma}-\text{Ma}^3} & \frac{f+s}{\text{Ma}^2-1} & -\frac{s}{\text{Ma}} \end{bmatrix} \begin{bmatrix} \Psi \\ M \\ \Theta \end{bmatrix}(s, \xi) + \begin{bmatrix} \frac{\text{Ma}}{\text{Ma}^2-1} \\ \frac{1}{1-\text{Ma}^2} \\ \frac{\frac{\gamma}{\gamma}+\frac{1}{\text{Ma}^2-1}}{\text{Ma}} \end{bmatrix} \delta(\xi - \xi_o) R(s)$$

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} U(s) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Psi \\ M \\ \Theta \end{bmatrix}(s, 0) + \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ \frac{s}{\text{Ma}(1-\text{Ma}^2)} & \frac{f+s}{\text{Ma}^2-1} & -\frac{s}{\text{Ma}} \end{bmatrix} \begin{bmatrix} \Psi \\ M \\ \Theta \end{bmatrix}(s, 1),$$

where Ψ, M, Θ, R, U are the Laplace transforms of ψ, m, θ, Φ, u. It is still possible to obtain an analytic transfer function

$$G_f : \begin{bmatrix} u(\cdot) \\ \Phi(\cdot) \end{bmatrix} \mapsto \left[\psi(\cdot, \xi_m) \right],$$

but it takes MATHEMATICA hours to compute it, and the result is much too complex to display here. However, neglecting all terms of order two or higher in Ma and f, the resulting transfer function is less ugly. In particular the transfer function from heat input to velocity takes the forms

$$G_{f,22}(s) = \frac{\text{Ma}\, s \cosh\left(\frac{3p(s)}{4}\right) + p(s) \sinh\left(\frac{p(s)}{4}\right)}{2p(s)\left(\sinh\left(p(s)/4\right) - \sinh\left(3p(s)/4\right)\right)} =$$

$$-\left(1 + \frac{\text{Ma}s \cosh\left(\frac{3p(s)}{4}\right)}{p(s)\sinh\left(p(s)/4\right)}\right) \frac{1}{4\cosh\left(p(s)/2\right)} = \left(1 + \frac{\text{Ma}s \cosh\left(\frac{3p(s)}{4}\right)}{p(s)\sinh\left(p(s)/4\right)}\right) G_{s,22}(p(s)), \quad (1.37)$$

where

$$p(s) = \sqrt{s(f + s)}.$$

This is interesting, because $p^{-1}(j\omega) = -f/2 \pm \sqrt{f^2/4 - \omega^2}$, i.e. the poles of $G_{s,22}(s)$ are mapped from the imaginary axis onto a line in the left-half plane with real part $-f/2$. The second form also shows that for $f, \mathrm{Ma} \to 0$, the transfer function reduces to the right bottom corner of (1.32) again. More generally, even the transfer function matrix obtained without neglecting the terms of second or higher order in f and Ma limits to (1.32) as $f, \mathrm{Ma} \to 0$, a good sanity check.

The first form of $G_{f,22}$ can be used to numerically compute its pole-zero pattern, which is interesting in regards to the comments made while interpreting Figure 1.19. Figure 1.23 shows the results of such computations and suggests that the intuition that friction would push the poles into the left-half plane was correct. There appear to be several additional poles – recall that $G_{s,22}$ had only poles at odd multiples of $j\pi$ – however each of those additional poles has a nearby zero making the effect on the root locus very small.

1.4.7   Full Linearized Model

In this section a numerical look is taken at the poles of the full model (1.23) or (1.24) and whether our intuition that the poles should lie on a parabola can be confirmed.

For reference, the full model (1.24) in pressure ψ, velocity m and temperature θ is

$$\frac{\partial}{\partial \tau} \begin{bmatrix} \psi \\ m \\ \theta \end{bmatrix} = f \begin{bmatrix} 0 & 2\bar{\gamma}\mathrm{Ma} & 0 \\ \gamma\mathrm{Ma} & -1 & -\bar{\gamma}\mathrm{Ma} \\ -(\gamma\mathrm{Ma})^2 & -2\gamma\mathrm{Ma} & \gamma\bar{\gamma}\mathrm{Ma}^2 \end{bmatrix} \begin{bmatrix} \psi \\ m \\ \theta \end{bmatrix} - \begin{bmatrix} \mathrm{Ma} & 1 & 0 \\ 1 & \mathrm{Ma} & 0 \\ 0 & 1 & \mathrm{Ma} \end{bmatrix} \frac{\partial}{\partial \xi} \begin{bmatrix} \psi \\ m \\ \theta \end{bmatrix}$$

$$+ \frac{1}{\mathrm{Pe}} \begin{bmatrix} 0 & 0 & \bar{\gamma} \\ 0 & \frac{4\mathrm{Pr}}{3} & 0 \\ 0 & 0 & \gamma \end{bmatrix} \frac{\partial^2}{\partial \xi^2} \begin{bmatrix} \psi \\ m \\ \theta \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ \gamma/\bar{\gamma} \end{bmatrix} \phi.$$

Figure 1.23 – *Numerically computed pole-zero pattern of the gas dynamics transfer function matrix $G_{f,22}$ for several orders of magnitude of the friction coefficient f. Poles are shown as solid blue symbols, zeros as hollow red ones. The Mach number Ma has barely any influence, hence a similar plot with varying Mach numbers is not shown here. Note that all the even-numbered poles have a zero nearby, negating their effect on the root locus.*

The complexity of this model is significantly increased compared to the model (1.27): while the latter is of first spatial order in two states, the former is of second order in three. We remark that we could as well use (1.23) with density and pressure as states and obtain very similar results (as done in [EBÅar, Sidebar 6]); however, using the temperature θ has the advantage that the zero heat flux boundary condition (1.13) can be incorporated seamlessly.[11]

Either way, the first-order form will have at least five states, hence the procedure of Appendix 1.D cannot be applied anymore. However, the poles of any derived transfer function will be a subset of the spectrum of the right-hand side operator, analogously to the situation

---

[11]On the other hand to derive an additional boundary condition on (1.23), the ideal gas law has to be applied to express $\frac{\partial \widetilde{T}}{\partial x}$ to first order in terms of density in pressure:

$$\frac{\partial \widetilde{T}}{\partial x} = \left(\frac{\partial \widetilde{p}}{\partial x} - \bar{p} \frac{\partial \widetilde{\rho}}{\partial x} / \bar{\rho} \right) / (\bar{\rho} R).$$

That yields an additional boundary condition for the (r, m, ψ) form (1.23).

in finite-dimensional LTI systems $\dot{x} = Ax + Bu$, where the eigenvalues of the A-matrix are a superset of the poles of the transfer function from u to x. Hence, instead of transfer function computations, we will compute the spectrum of the right-hand-side operator numerically using Chebyshev spectral methods.

Such methods basically amount to (spatial) discretization, however instead of equispaced discretization points, the Chebyshev points $\xi_k = \cos(k\pi/N)$, $k = 0, \ldots, N$ are used.[12] A differentiation matrix D then approximates the first spatial derivative on the Chebyshev grid, and the right-hand side operator can be rewritten in terms of the matrix. Its eigenvalues then approximate the spectrum of the right-hand side operator. For details on and more applications of such methods, see the book [Tre00].

Boundary conditions are treated slightly differently than in [Tre00]; the procedure is outlined in Appendix 1.F. The eigenvalues of the resulting $(3N - 3) \times (3N - 3)$ matrix are then computed using Arnoldi iterations (as implemented in MATLAB's `eigs` function) initialized at integer multiples of $j\pi$.

Figure 1.24 shows a part of the spectrum obtained using $N = 200$ grid points, along with the first 5 pressure modes for a specific set of the parameters Ma, Pr, f, and Pe as given in Table 1.2. As expected, the spectrum is located very close to, but left of the imaginary axis, and has an approximate parabola shape. The modes depicted on the left correspond to a standing half-wave and its harmonics; poles corresponding to higher frequencies are damped more.

The parameters Pe and f, which are known only to within maybe an order of magnitude, appear to have very well-defined effects. Increasing the amount of diffusion 1/Pe will "bend" the spectrum, i.e. the parabola shape will become narrower, while increasing the amount of friction f will shift the entire spectrum towards the left. These numerical computations again validate the claim made during the analysis of the root locus in Figure 1.19, namely that the

---

[12]The ξ_k are for the interval [0 1], simply shift and stretch the interval to obtain the Chebyshev points on an arbitrary interval.
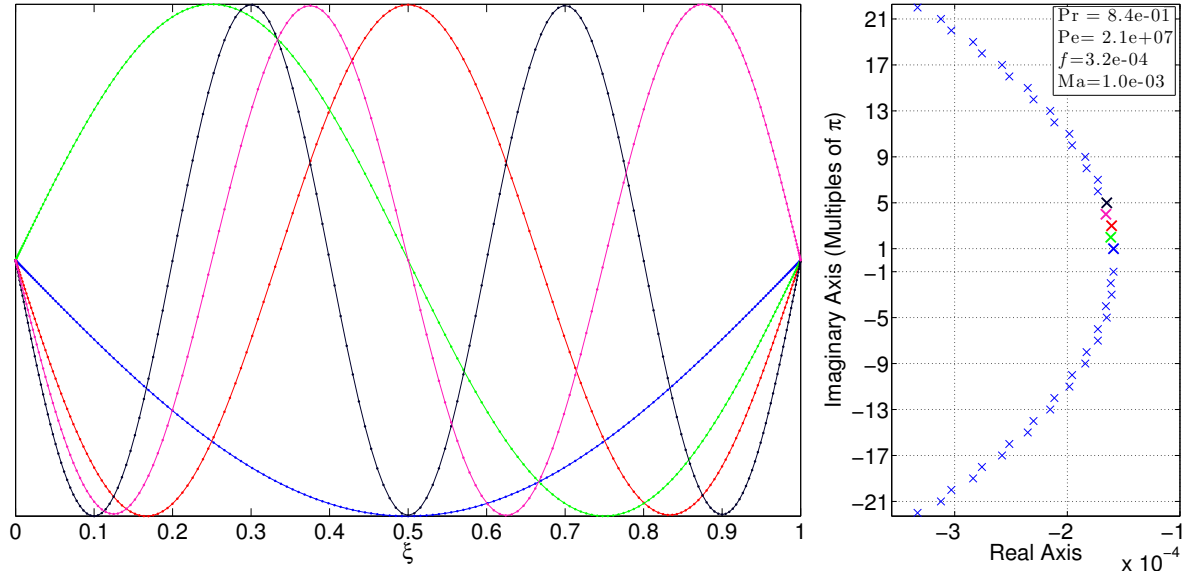
Figure 1.24 – *Numerical computation of the spectrum of the spatial differential operator in (1.24). The left plot shows the five pressure modes corresponding to the fundamental and the first four harmonics; they are indistinguishable from the pressure modes of the undamped model (1.32), corresponding to the resonances of a tube with two open ends. The right plot shows part of the numerically computed spectrum: poles located on the imaginary axis in the case without damping are now damped, i.e. shifted to the left; higher frequencies are damped more, yet still lightly (note the axis scales).*

poles located *on* the imaginary axis in the friction- and diffusion-less case would be shifted towards the left if diffusion and/or friction effects were retained.

It should also be mentioned that many of the $3N - 3$ eigenvalues of the discretization are in unexpected locations, including the right-half plane; however, it can be checked that the Chebyshev grid resolves the corresponding modes to less than a wave-length. In the terminology of e.g. [Tre00, Ch. 7], the points per wavelength (in the center of the interval – Chebyshev grids are denser near the boundaries and sparser in the center than equispaced grids) are less than 2, hence those poles and modes are numerical artifacts.

1.4.8   Future Directions

The strength of our model lies mainly in its simplicity and intuitive explanation of the thermo-acoustic instability, but due to its singular nature (poles on the imaginary axis) it is of limited utility for control and optimization. In Section 1.4.5 on page 58 we touched on this issue briefly and described a heuristic approach to address this issue, and in Sections 1.4.6 and 1.4.7 we have shown that a more complex model captures reality (as seen through the lens of system identification in Section 1.3) better, however those models suffer from increased complexity and reliance on poorly known parameters. Improving on the model and gaining better information on the involved parameters, by inclusion of new physical insights, by application of a more sophisticated identification procedure, or by using more advanced measurement devices than a microphone (e.g. a hotwire anemometer to measure the offset velocity $\bar{v}$ and fluctuation $\widetilde{v}$) would increase the usefulness of the model and hopefully ultimately allow for the formulation of an optimization problem for the energy conversion process.

The modeling approach taken here can also be applied to similar thermoacoustic systems. A natural candidate is the Sondhauss tube [Son70], also known more dramatically as the "acoustic laser" [GB00], which has one open and one closed end. An important new challenge arises

from the observation that a concentrated heat source, such as the hot wire in the Rijke tube, is not sufficient to generate the sound wave in a Sondhauss tube, but instead a so-called stack induces a temperature gradient over a finite length.

Finally, the recent Master's thesis [Min14] considers numerical implementation of the fully nonlinear thermoacoustic model equations, including also more general boundary conditions than simple pressure nodes. Analysis of the Rijke tube or similar thermoacoustic systems would benefit greatly from of a numerical testbed for control algorithms and optimized trajectories.

Appendix 1.A   Detailed Experimental Setup

All experiments were performed here at UCSB in a setup that shall be described next.

- The tube is a Pyrex glass tube with length = 4 ft ≈ 2.2 m, and internal diameter = 3 in ≈ 7.6 cm. Glass offers good heat resistance while having low electrical and thermal conductivity.

- The heater configuration changed, as we experimented with different ways of coiling the wire and mounting the coil. The wire is always Nickel-Chromium wire with diameters between 0.5 and 1 mm (corresponding to 24 to 19 gauge), and the length was chosen to obtain a coil resistance of roughly 10 Ω to match the impedance of the power supply.

- The microphone (RadioShack 33-3013) is a simple clip-on with built-in preamplifier.

- The speaker is a 20 W Auvio ceiling speaker (SKU 4000334).

- The audio amplifier for the speaker signal is a Techron 5515.

- The DC power supply is a Topward 6306D power supply with maximum power output of 360 W. The output power necessary to observe the humming was typically close to

maximum.

- Signal generation and collection, as well as the control algorithm are realized with Simulink Real-Time Windows Target.

- The Data Acquisition (DAQ) card to interface experiment and Simulink is a National Instruments 6053E card.

- The sampling period used is $T_s = 0.1$ ms (so $f_s = 10$ kHz), which is the highest the setup could sustain in real-time operation.

- To remove some of the high frequency content introduced by the sample-and-hold operation, the output of the DAQ card is filtered using an analog (passive) lowpass filter with bandwidth 1.9 kHz.

Refer also to Figure 1.1 for a photograph and schematic drawing of the setup.

Appendix 1.B   Estimating the Wire Temperature

Having an estimate of the temperature of the heating coil is useful regarding the choice of materials to be used and in the physical modeling process, see Section 1.4.4. It is straightforward to obtain a rough estimate from measurements of just voltage supplied to and current drawn by the heating coil by using the temperature dependence of the electrical resistance of the wire. The resistance R of a wire is commonly assumed to depend on its temperature T affinely:

$$R(T) = R_0\big(1 + \alpha(T - T_0)\big),$$

where R_0 is the resistance at T_0, a known point of reference, and α denotes the (linear) temperature coefficient, which is tabulated for different materials. Typically, $\alpha > 0$, i.e. the resistance increases as the material heats up; for the Nichrome wire used here, $\alpha \approx 1.76 \ 10^{-4}$ 1/K. The
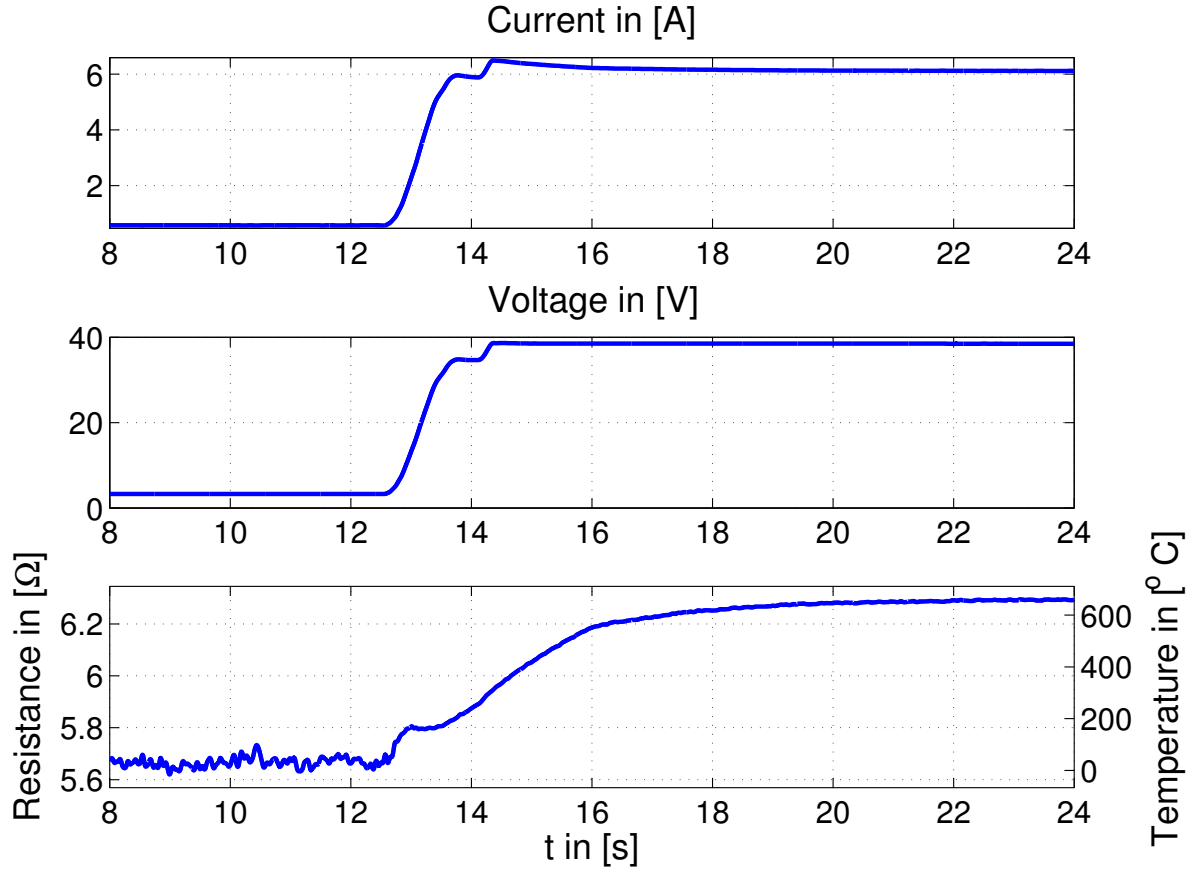
Figure 1.25 – *Smoothed time traces of voltage and current across the heating coil as the power is first set to a small value, and then increased until the humming sets in. The lower plot shows estimated resistance and temperature of the wire.*

relationship can be inverted to

$$T = T_0 + \frac{R - R_0}{\alpha R_0}.$$

For the reference point (R_0, T_0), room temperature $T_0 = 293$ K is assumed. The resistance R_0 at room temperature is estimated by setting the power supply to a small voltage and recording current and voltage across the coil; then the supplied power is increased until the instability sets in and current and voltage are again recorded. Since current measurements are very noisy, in particular at low voltages, where the signal-to-noise ratio is smaller, averages over relatively long times are taken. An example time trace is shown in Figure 1.25, and one arrives at an

estimate of

$$T_{\text{wire}} \approx 660^{\circ}\text{C} = 933 \text{ K}. \tag{1.38}$$

Appendix 1.C   Relationship of Acoustic Pressure and Velocity

Acoustic waves satisfy the linearized wave equation

$$\frac{\partial \widetilde{v}}{\partial t} = -\frac{1}{\overline{\rho}} \frac{\partial \widetilde{p}}{\partial x}$$
$$\frac{\partial \widetilde{p}}{\partial t} = -\gamma \overline{p} \frac{\partial \widetilde{v}}{\partial x},$$

and, formally, replacing $\frac{\partial}{\partial x}$ by jk and $\frac{\partial}{\partial t}$ by s, or applying temporal Laplace, spatial Fourier transforms, yields

$$s\widetilde{v} = -\frac{jk}{\overline{\rho}}\widetilde{p}$$

$$s\widetilde{p} = -\gamma\overline{p}jk\widetilde{v},$$

$$\implies \qquad\qquad\qquad s^2\widetilde{v} = \gamma\overline{p}/\overline{\rho}\,(-jk)^2\widetilde{v}$$

$$\text{using } c^2 = \gamma\overline{p}/\overline{\rho} \qquad \implies \qquad s^2/(jk)^2 = c^2 \tag{1.39}$$

$$\implies \widetilde{p}/\overline{p} = -j\gamma k/s\,\widetilde{v} \qquad \implies \qquad \widetilde{v}/c = -j\widetilde{p}/(\gamma\overline{p}), \tag{1.40}$$

where (1.39) is sometimes called a "dispersion relation," and it follows with (1.40), that in acoustic waves, the percentage change in pressure p is of the same order as the percentage change in Mach number (if $\overline{v}$ is very small).

Appendix 1.D   Transfer Function Computation for Two-Point Boundary Value Problems (TPBVPs)

Along similar lines as in [BD03, JB06], an exact formula for the frequency response of the class of spatially distributed systems that can be represented as linear TPBVPs is derived here. Inputs are restricted to be point-wise, either inducing a discontinuity, as r in (1.41), or as an inhomogeneous boundary condition like u in (1.42).

A unified way to treat linear TPBVPs with concentrated and boundary inputs can be described by using the state space realization

$$\frac{\mathrm{d}}{\mathrm{d}x}\Psi(x) = F\,\Psi(x) + G\,\delta(x - x_o)\,r, \qquad x \in [x_i, x_f] \tag{1.41}$$

$$N_u\,u = N_i\,\Psi(x_i) + N_f\,\Psi(x_f). \tag{1.42}$$

The first equation is the differential equation, while the second is a general form to express linear two-point boundary conditions.

The matrices F, G, N_u, N_i and N_f can depend on parameters; if (1.41) was obtained via Laplace transform of a PDE, this parameter will be the Laplace variable s. Similarly, r and u might have depended on time, and hence now depend on s, but they do not depend on x. The objective is to obtain a formula for the solution $\Psi(x)$ as a function of r and u; if the matrices depend on s, then so will the solution, and it can be interpreted as the transfer function of the original, spatio-temporal system.

The presence of $\delta(x - x_o)$ in the r input term implies that the solution $\Psi(x)$ could have a discontinuity at x_o, but is continuous everywhere else. The upper and lower limits at the jump point x_o

$$\Psi(x_o^-) := \lim_{x \nearrow x_o} \Psi(x), \qquad \Psi(x_o^+) := \lim_{x \searrow x_o} \Psi(x),$$
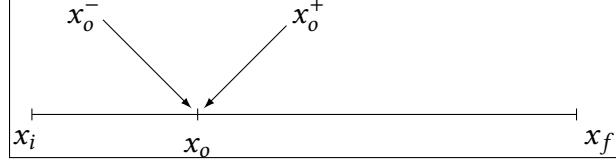
satisfy the following relation:

$$\Psi(x_o^+) = \Psi(x_o^-) + G\, r. \tag{1.43}$$

To see that:

$$\Psi(x_o^+) = \int_{x_i}^{x_o^+} \frac{\mathrm{d}}{\mathrm{d}x}\Psi(x)\mathrm{d}x = \int_{x_i}^{x_o^+} F\Psi(x) + G\delta(x - x_o)r\mathrm{d}x = \int_{x_i}^{x_o^+} F\Psi(x)\mathrm{d}x + G$$

$$\Psi(x_o^-) = \int_{x_i}^{x_o^-} \frac{\mathrm{d}}{\mathrm{d}x}\Psi(x)\mathrm{d}x = \int_{x_i}^{x_o^-} F\Psi(x) + G\delta(x - x_o)r\mathrm{d}x = \int_{x_i}^{x_o^-} F\Psi(x)\mathrm{d}x$$

$$\Psi(x_o^+) - \Psi(x_o^-) = G,$$

since the Dirac impulse has no effect if the interval of integration does not contain x_o^+.

The solution over each of the intervals $[x_i, x_o]$ and $[x_o, x_f]$ can be propagated from each end

$$\Psi(x) = \begin{cases} \Phi(x, x_i)\, \Psi(x_i) & x \in [x_i, x_o] \\ \Phi(x, x_f)\, \Psi(x_f) & x \in [x_o, x_f] \end{cases}, \tag{1.44}$$

where $\Phi(x_1, x_2)$ is the state transition matrix for the ODE (1.41). If the coefficients are constant in x as we have assumed here, then $\Phi(x_1, x_2) = e^{F(x_1 - x_2)}$. However, if (1.41) has space-varying coefficients, the analysis here remains unchanged, it is just not as simple anymore to find the state transition matrix.

The jump relation (1.43) can now be rewritten by first observing that (1.44) gives

$$\Psi(x_o^-) = \Phi(x_o, x_i)\, \Psi(x_i)$$

$$\Psi(x_f) = \Phi(x_f, x_o)\, \Psi(x_o^+),$$

which combined with (1.43) yields a relation between $\Psi(x_i)$ and $\Psi(x_f)$ as a function of the input r:

$$\Psi(x_f) = \Phi(x_f, x_o)\Big(\Phi(x_o, x_i)\Psi(x_i) + Gr\Big). \tag{1.45}$$

The given boundary conditions (1.42) can now be combined together with (1.45) in the following matrix-vector form

$$\begin{bmatrix} -\Phi(x_f, x_i) & I \\ N_i & N_f \end{bmatrix} \begin{bmatrix} \Psi(x_i) \\ \Psi(x_f) \end{bmatrix} = \begin{bmatrix} \Phi(x_f, x_o)\, G & 0 \\ 0 & N_u \end{bmatrix} \begin{bmatrix} r \\ u \end{bmatrix},$$

where I, respectively 0, always denote the identity, respectively zero, matrix of appropriate dimensions.

Finally, this permits rewriting the solution (1.44) in terms of the inputs r and u as

$$\Psi(x) = \begin{cases} \begin{bmatrix} \Phi(x, x_i) & 0 \end{bmatrix} \Gamma \begin{bmatrix} r \\ u \end{bmatrix} & x \in [x_i, x_o] \\[4mm] \begin{bmatrix} 0 & \Phi(x, x_f) \end{bmatrix} \Gamma \begin{bmatrix} r \\ u \end{bmatrix} & x \in [x_o, x_f], \end{cases} \tag{1.46}$$

where

$$\Gamma = \begin{bmatrix} -\Phi(x_f, x_i) & I \\ N_i & N_f \end{bmatrix}^{-1} \begin{bmatrix} \Phi(x_f, x_o)\, G & 0 \\ 0 & N_u \end{bmatrix}. \tag{1.47}$$

Remark. Γ exists if and only if $N_i + N_f\Phi(x_f, x_i)$ is invertible. That follows easily from the formula for the determinant of a block matrix

$$\det\left(\begin{bmatrix} -\Phi(x_f, x_i) & I \\ N_i & N_f \end{bmatrix}\right) = -\det\big(\Phi(x_f, x_i)\big)\det\big(N_f + N_i\Phi^{-1}(x_f, x_i)\big)$$

and the fact that transition matrices are always invertible, with $\Phi^{-1}(x_1, x_2) = \Phi(x_2, x_1)$, or from

the matrix inversion formula

$$
\begin{bmatrix} -\Phi(x_f,x_i) & I \\ N_i & N_f \end{bmatrix}^{-1} = \begin{bmatrix} -\Phi(x_i,x_f) + \Phi(x_i,x_f)SN_i\Phi(x_i,x_f) & \Phi(x_i,x_f)S \\ SN_i\Phi(x_i,x_f) & S \end{bmatrix},
$$

where $S := (N_f + N_i\Phi(x_i,x_f))^{-1}$ is the Schur complement of the block $-\Phi(x_f,x_i)$ in above block matrix. That also yields a form of Γ that requires only one inversion of a $(n \times n)$ matrix:

$$
\Gamma = \begin{bmatrix} -\Phi(x_i,x_f) + \Phi(x_i,x_f)SN_i\Phi(x_i,x_f) & \Phi(x_i,x_f)S \\ SN_i\Phi(x_i,x_f) & S \end{bmatrix} \begin{bmatrix} \Phi(x_f,x_o)\,G & 0 \\ 0 & N_u \end{bmatrix} =
$$

$$
\begin{bmatrix} -I + \Phi(x_i,x_f)SN_i & \Phi(x_i,x_f) \\ SN_i & I \end{bmatrix} \begin{bmatrix} \Phi(x_i,x_o)\,G & 0 \\ 0 & SN_u \end{bmatrix}.
$$

"$N_i + N_f\Phi(x_f,x_i)$ is invertible" should be interpreted as a well-posedness condition on the system (1.41) and (1.42), since else solutions are not unique:

Assume, it is violated. Then there exists a nonzero Z such that $\left(N_i + N_f\Phi(x_f,x_i)\right)Z = 0$, and if we have a solution $\Psi_0(x)$, then

$$
\Psi_Z(x) := \Psi_0(x) + \Phi(x,x_i)Z
$$

is another, since

$$
\frac{\mathrm{d}}{\mathrm{d}t}(\Psi_0 + \Phi(\cdot,x_i)Z) = F\Psi_0 + Gr + F\Phi(\cdot,x_i)Z = F\left(\Psi_0 + \Phi(\cdot,x_i)Z\right) + Gr
$$

$$
N_i\Big(\Psi_0(x_i)+\underbrace{\Phi(x_i,x_i)}_{=I}Z\Big) + N_f\Big(\Psi_0 + \Phi(x_f,x_i)Z\Big) = \underbrace{N_i\Psi_0(x_i)+N_f\Psi_0(x_f)}_{=N_u u} + \underbrace{\left(N_i+N_f\Phi(x_f,x_i)\right)Z}_{=0}.
$$

Unless $G \equiv 0$, the value of Ψ at x_o is discontinuous. One can choose to take either of the

one-sided limits, or the average value at the point x_o

$$\Psi_{\text{av}}(x_o) = \frac{1}{2} \begin{bmatrix} \Phi(x_o, x_i) & \Phi(x_o, x_f) \end{bmatrix} \Gamma \begin{bmatrix} r \\ u \end{bmatrix}. \tag{1.48}$$

In the case where any of the matrices F, G, N_u, N_i and N_f and inputs r, u depend on the Laplace transform variable s, we can use (1.46) to define a transfer function matrix $T(s)$. Say the output we are interested in is some combination $Y = C \cdot \Psi$ of the states at the location $x_m > x_o$; then from

$$\Psi(x_m) = \begin{bmatrix} 0 & \Phi(x_m, x_f) \end{bmatrix} \Gamma(s) \begin{bmatrix} r(s) \\ u(s) \end{bmatrix}$$

we have

$$Y(s) = \underbrace{C \begin{bmatrix} 0 & \Phi(x_m, x_f) \end{bmatrix} \Gamma(s)}_{=:T(s)} \begin{bmatrix} r(s) \\ u(s) \end{bmatrix}.$$

Note that the important computation is that of $\Phi(\cdot, \cdot)$. It is possible in general to compute Φ analytically (with computer algebra routines) for systems of dimension 4 or less, which corresponds to PDEs in which the spatial derivative order is at most 4. The transfer function T can be given analytically even for spatially varying systems and systems of order higher than 4, as long as Φ can be determined analytically. However, expressions quickly get so unwieldy that one is better off computing the frequency response numerically on a frequency grid.

Appendix 1.E   Departure Angles for Infinite-Dimensional Root Loci

The classical rule to compute the angle of departure φ_0 from an open-loop pole $s_0 \in \mathbb{C}$ states that

$$n_0 \varphi_0 = - \sum_{s_i \in \{\text{all poles} \neq s_0\}} \angle(s_0 - s_i) \; + \sum_{z_i \in \{\text{all zeros}\}} \angle(s_0 - z_i) - (2r - 1)\pi, \qquad (1.49)$$

where $\angle(s)$ represents the phase of the complex number s, $r = 1, 2, \ldots, n_0$, and n_0 is the multiplicity of the pole s_0 [FPEN02]. This rule is difficult to apply to a system with infinitely many poles (or zeros), so it is shown here, how a simple application of Taylor series yields an alternative expression that does not involve sums over the sets of all poles and zeros.

Single pole

Let $f(s) = \frac{h(s)}{g(s)}$ be a meromorphic function, i.e. $g(\cdot)$ and $h(\cdot)$ are both analytic. (It is sufficient for f to be meromorphic in an open set around the pole s_0 we are interested in.) The roots of g are then the open-loop poles (unless there are cancellations with h). Let us consider one point on the (positive) root locus, i.e. there is a $k_0 \geq 0$ such that $g(s_0) + k_0 h(s_0) = 0$, but for simplicity, assume also that $g'(s_0) + k_0 h'(s_0) \neq 0$ (if s_0 is a pole, that reduces to requiring it to be a single pole). To obtain the departure angle from s_0, let k_0 be perturbed by a small positive value δk and compute the perturbation δs needed to render $s_0 + \delta s$ a solution to $g(s_0 + \delta s) + (k_0 + \delta k) h(s_0 + \delta s) =$

0 to first order:

$$g(s_0 + \delta s) + (k_0 + \delta k)h(s_0 + \delta s) =$$

$$g(s_0) + g'(s_0)\delta s + (k_0 + \delta k)(h(s_0) + h'(s_0)\delta s) + \mathcal{O}\left(\delta s^2\right) =$$

$$g(s_0) + k_0 h(s_0) + (g'_0(s_0) + k_0 h'(s_0))\delta s + h(s_0)\delta k + \mathcal{O}\left(\delta s^2\right) \approx$$

$$(g'_0(s_0) + k_0 h'(s_0))\delta s + h(s_0)\delta k = 0$$

$$\Longleftrightarrow \delta s = -\frac{\delta k\, h(s_0)}{g'_0(s_0) + k_0 h'(s_0)},$$

where again $\mathcal{O}\left(\delta s^\nu\right)$ denotes terms of order ν or higher in δs.

For the phase of the departing branch that means

$$\angle(\delta s) = -\pi + \angle(h(s_0)) - \angle(g'(s_0) - k_0 h'(s_0)) \tag{1.50}$$

(since $\delta k > 0$, $\angle(\delta_k) = 0$), and if s_0 is an open-loop pole, then $k_0 = 0$ and

$$\angle(\delta s) = -\angle(g'(s_0)) + \angle(h(s_0)) - \pi, \tag{1.51}$$

which for a finite dimensional system reduces to (1.49) (with $n_0 = 1$).

Multiple poles

Now we consider a point on the (positive) root locus such that

$$\frac{\mathrm{d}^\nu g}{\mathrm{d}s^\nu}(s_0) + k_0 \frac{\mathrm{d}^\nu h}{\mathrm{d}s^\nu}(s_0) = 0$$

for $\nu = 0, \ldots, \mu - 1$, i.e. s_0 is pole with multiplicity μ. To obtain the departure angle from s_0 we again let k_0 be perturbed by a small positive value δk and compute the perturbation δs needed

to render $s_0 + \delta s$ a solution to $g(s_0 + \delta s) + (k_0 + \delta k)h(s_0 + \delta s) = 0$, however this time to μth order:

$$g(s_0 + \delta s) + (k_0 + \delta k)h(s_0 + \delta s) =$$

$$\sum_{\nu=0}^{\mu} \frac{\mathrm{d}^\nu g}{\mathrm{d}s^\nu}(s_0)\delta s^\nu + (k_0 + \delta k)\sum_{\nu=0}^{\mu} \frac{\mathrm{d}^\nu h}{\mathrm{d}s^\nu}(s_0)\delta s^\nu + \mathcal{O}\left(\delta s^{\mu+1}\right) =$$

$$\underbrace{\sum_{\nu=0}^{\mu-1}\left(\frac{\mathrm{d}^\nu g}{\mathrm{d}s^\nu}(s_0) + k_0\frac{\mathrm{d}^\nu h}{\mathrm{d}s^\nu}(s_0)\right)\delta s^\nu}_{=0} + \left(\frac{\mathrm{d}^\nu g}{\mathrm{d}s^\mu}(s_0) + k_0\frac{\mathrm{d}^\mu h}{\mathrm{d}s^\nu}(s_0)\right)\delta s^\mu + \delta k\, h(s_0) + \mathcal{O}\left(\delta s^{\mu+1}\right) \approx 0$$

$$\Longleftrightarrow \delta s = \sqrt[\mu]{-\frac{\delta k\, h(s_0)}{\frac{\mathrm{d}^\mu g}{\mathrm{d}s^\mu}(s_0) + k_0\frac{\mathrm{d}^\mu h}{\mathrm{d}s^\mu}(s_0)}},$$

and hence

$$\mu \cdot \angle(\delta s) = \angle(h(s_0)) - \angle\left(\frac{\mathrm{d}^\mu g}{\mathrm{d}s^\mu}(s_0) + k_0\frac{\mathrm{d}^\mu h}{\mathrm{d}s^\mu}(s_0)\right) - \pi + 2r\pi, \qquad r = 0,\ldots\mu-1 \quad (1.52)$$

and for $k_0 = 0$, i.e. s_0 a μth order open-loop pole

$$\mu \cdot \angle(\delta s) = \angle(h(s_0)) - \angle\left(\frac{\mathrm{d}^\mu g}{\mathrm{d}s^\mu}(s_0)\right) + (2r-1)\pi, \qquad r = 0,\ldots\mu-1, \quad (1.53)$$

which, for $\mu = n_0$, reduces again to (1.49).

Appendix 1.F   Boundary Conditions and Discretization of Linear Operators

Assume we discretized a linear operator, for instance on a Chebyshev grid, and obtained a $(n \times n)$ matrix M approximating the operator, and a $(b \times n)$ (typically very fat) matrix B such that the boundary conditions are given by $Bx = 0$. Note that $\ker(B)$ needs to be M-invariant,

i.e. $Bx = 0 \Rightarrow BMx = 0$, for the problem to be well-posed.

In summary, we can state: The eigenvalues of the operator M restricted by $Bx = 0$ are the eigenvalues of $V_k^* M V_k$, where V_k are the last $n - b$ columns of V in the singular value decomposition (SVD) of B because they form an orthonormal basis for $\ker(B)$.

To derive this result, we choose a basis adapted to $\ker(B)$, because that is where we want to restrict M to. A convenient choice is coming from the SVD of B:

$$
B = U \bar{\Sigma} V^* =
\begin{array}{c}
\quad\quad b \\
\left[\begin{array}{c|c}
\text{basis for} & \text{basis for} \\
\text{ran}(B) & \ker(B^*) \\
& \text{(likely } \emptyset)
\end{array}\right] b
\end{array}
\cdot
\begin{array}{cc}
\; b & n-b \\
\left[\begin{array}{c|c}
\Sigma & 0 \\
\hline
0 & 0
\end{array}\right] & \begin{array}{c} b \\ \text{likely } \emptyset \end{array}
\end{array}
$$

$$
\cdot \left(
\begin{array}{cc}
b & n-b \\
\left[\begin{array}{c|c}
\text{basis for} & \text{basis for} \\
\text{ran}(B^*) & \ker(B)
\end{array}\right] & n
\end{array}
\right)^*
$$

Let V_k be the matrix containing the last $n - b$ columns of V, i.e. the ones spanning the kernel $\ker(B)$, and V_r the other b rows (spanning the range of B^*, or equivalently, $\ker(B)^\perp$) and decompose the representation of M on the basis V as

$$
M_V = V^* M V = \left[\begin{array}{c|c}
V_r^* M V_r & V_r^* M V_k \\
\hline
V_k^* M V_r & V_k^* M V_k
\end{array}\right].
$$

Now the restriction of M_V to $\ker(B)$ is easily obtained:

$$
M_V\big|_{\ker(B)} = V_k^* M V_k,
$$

which is a $(n - b) \times (n - b)$ matrix, operating on vectors z_k that are the coordinates of a vector

79

in the V_k basis.

A representation of this operator in the original coordinates can be obtained as follows:

$$x_{\text{(original coords)}} = Vz = V_k z_k \text{(since } z_r = 0) =$$

$$V_k \quad \underbrace{(V_k^* M V_k)}_{\text{op. on } \ker(B) \text{ in } V\text{-coords.}} \quad \overbrace{V_k^* y_{\text{(original coords)}}}^{\text{Projection of } y \text{ onto } \ker(B) \textbf{ in } V\textbf{-coords.}} \quad = \left(V_k V_k^* M V_k V_k^*\right) y_{\text{(original coords)}}$$

Note that x is only the correct outcome of My if $y \in \ker(B)$. Due to the assumption that $\ker(B)$ is M-invariant, x will then be in $\ker(B)$, too.

Another interpretation: $P = V_k V_k^*$ is the orthogonal projection onto $\ker(B)$, and $V_k^* V_k V_k^* = V_k^*$ is the projection but expressed in V_k-coordinates (coordinates corresponding to V_r are 0). So $V_k V_k^* M V_k V_k^*$ really just projects onto $\ker(B)$, applies M and then projects again – since $\ker(B)$ is M-invariant, that second projection should be spurious, and $V_k V_k^* M V_k V_k^* = M V_k V_k^*$. $V_k V_k^* M V_k V_k^*$ has the same nonzero eigenvalues as $V_k^* M V_k$.

Appendix 1.G Transfer Functions in Dimensionless and Physical Variables

Since the identification results are necessarily in physical variables, whereas in the theoretical analysis we prefer dimensionless representations, how do the former relate to the latter? Consider two physical quantities, $f(x,t)$ and $g(x,t)$, and their normalized versions $\phi(\xi,\tau) := f(\xi L, \tau T)/\bar{f}$ and $\gamma(\xi,\tau) := g(\xi L, \tau T)/\bar{g}$ (hence, x is scaled by L to get $\xi = x/L$, similarly for τ).

Let s be the physical, and σ be the dimensionless Laplace transform variable, so that we can

define Laplace transforms

$$F(x,s) = \int_0^\infty e^{-st} f(x,t) \mathrm{d}t \qquad\qquad G(x,s) = \int_0^\infty e^{-st} g(x,t) \mathrm{d}t$$

$$\Gamma(\xi,\sigma) = \int_0^\infty e^{-\sigma\tau} \gamma(\xi,\tau) \mathrm{d}\tau \qquad\qquad \Phi(\xi,\sigma) = \int_0^\infty e^{-\sigma\tau} \phi(\xi,\tau) \mathrm{d}\tau.$$

Then we can express the physical quantity in terms of the dimensionless quantity:

$$F(x,s) = \int_0^\infty e^{-st} f(x,t)\mathrm{d}t = \int_0^\infty e^{-st} \bar{f} \phi(x/L, t/T) \mathrm{d}t =$$

$$T\bar{f} \int_0^\infty e^{-s\tau T} \phi(x/L, \tau)\mathrm{d}\tau = T\bar{f}\Phi(x/L, sT)$$

Now let's say we *identified* a transfer function $P(s) = F(x_1,s)/G(x_2,s)$ and *derived* a transfer function $\Psi(\sigma) = \Phi(x_1/L,\sigma)/\Gamma(x_2/L,\sigma)$ (for some fixed x_i). Then, ideally,

$$P(s) = F(x_1,s)/G(x_2,s) = \frac{T\bar{f}\Phi(x_1/L,sT)}{T\bar{g}\Gamma(x_2/L,sT)} = \frac{\bar{f}}{\bar{g}} \Psi(sT). \tag{1.54}$$

Chapter 2

Frequency-Domain Methods for Optimal Periodic Control (OPC)

Optimal steady-state operation of industrial plants is a well-researched and well-understood way to choose control parameters; the next easiest operation condition is cycling, i.e. having the process follow an optimal finite-length trajectory over and over again. This problem of Optimal Periodic Control (OPC) has received attention in different decades and different communities. What is commonly regarded the first mention of this idea by Horn and Lin [HL67] is inspired by maximization of product concentration in chemical reactions, and for a simple example an improvement over the optimal steady state is obtained by periodically adjusting the reactant feed concentration. Later contributions include applications to vehicle cruise, e.g. [Gil76], where it is shown that for a simple model of a vehicle, fuel economy can be improved by periodically adjusting the thrust, and to production planning [MBF98].

Different approaches to the solution of the OPC problem have been made, mostly based on variational calculus. For example, in [HL67], it is shown that the first order necessary optimality conditions include that the co-state λ be periodic, too. Extensions by use of the maximum

principle were made soon after, as has been the application of relaxed steady-state analysis (which means, roughly speaking, infinitely fast switching between admissible values of the input). A comparison can be found in [BH71]. Second-variation methods [SE84] to derive sufficient conditions for optimality appeared later. A recent paper [VGD04] uses a flat output to parametrize all periodic solutions of a given system to transform the OPC to a Nonlinear Program. Early contributions of mainly Guardabassi et al. and Horn et al. are summarized in the Survey Paper [GLR74], a newer summary including also relaxed steady-state and quasi steady-state analysis is [Gil77].

A central question that is common in much of the above cited literature is whether steady-state operation could be improved by cycling. The Π-Test [BFG73] is a relatively effective method to answer this question. However, calculating the optimal periodic input and trajectory is in general a nontrivial problem. Apart from flatness, shooting methods [SE84] and Newton-Raphson techniques [HL67] have been employed to solve the Euler-Lagrange equations arising from variational calculus. A common disadvantage of those methods is the "curse of dimensionality:" the complexity of the problem increases with the state dimension of the plant, and infinite-dimensional plants can be handled only by discretization. A method that would be equally applicable to finite-dimensional and distributed-parameter systems is desirable.

The first part of this chapter, Section 2.2, expands on ideas formulated in [BGR72] by making use of harmonic balancing to find periodic orbits of the Hamiltonian system. Following this approach we obtain a system of polynomial equations in terms of the Fourier coefficients of the trajectories, the zeros of which coincide with trajectories satisfying the necessary conditions for optimality. The plant itself appears only with its frequency response as coefficients in the equations, thus making the complexity of the method independent of system order. This was the subject of [EB12].

In Section 2.3, the problem is approached by making use of the Fourier series representation

of periodic signals to rewrite the cost function as a polynomial in the Fourier coefficients of a particular signal. In doing so, the polynomial equations of the previous section result as a byproduct by setting the gradient of the cost with respect to the Fourier coefficients to zero. Yet a much more valuable outcome is that the polynomial form of the cost function allows for direct application of very efficient methods of polynomial minimization, namely Sum-of-Squares (SOS) methods, and treatment of more general system structures.

The organization is as follows. In Section 2.1, the form of the OPC problem considered here is defined before it is shown in Section 2.2 how the associated Hamiltonian system can be treated with harmonic balance techniques, resulting in the polynomial equations of Section 2.2.4. The results are demonstrated on two examples, one finite- and one infinite-dimensional. Section 2.3 then outlines the approach of formulating the cost as a polynomial and minimizing it using SOS programming.

In the Appendix it is shown how the Fourier coefficients of $y^p(t)$ are related to the Fourier coefficients of $y(t)$, and two technical results needed in developing the approach of Section 2.3 are given.

Related Publication:

[EB12] Jonathan P. Epperlein and Bassam Bamieh. A frequency domain method for optimal periodic control. In *American Control Conference (ACC), 2012*, pp. 5501–5506. IEEE, 2012.

2.1 Problem Setting

We are given a linear time-invariant (LTI) plant

$$\dot{x} = Ax + Bu, \qquad\qquad y = Cx, \qquad\qquad y \in \mathbb{R}, \, x \in \mathbb{R}^n, \, u \in \mathbb{R}^m \qquad (2.1)$$

and a cost function

$$J = \frac{1}{T} \int_0^T \frac{1}{2} x^T Q x + \frac{1}{2} u^T R u + \phi(y) \mathrm{d}t, \tag{2.2}$$

where $R > 0$, Q symmetric, but *not necessarily* definite and $\phi : \mathbb{R} \to \mathbb{R}$ a polynomial $\phi(y) = \sum_{q=1}^{p} \gamma_q y^q$ of *even* degree q. The task is to find a period T and a T-periodic solution (x, u) that minimize J:

$$\underset{x(0), u(\cdot), T}{\text{minimize}} \quad J(x(0), u, T) \tag{2.3a}$$

$$\text{subject to} \quad \begin{aligned} \dot{x} &= Ax + Bu \\ x(0) &= x(T). \end{aligned} \tag{2.3b}$$

Note that in contrast to a common optimal control problem initial and final state are neither given nor free, but have to be equal. Thus, in addition to the period T and the control signal u, $x(0)$ is a variable with respect to which the minimization is performed.

Also note that ϕ would typically include only even powers, reflecting a cost that is symmetric. If $\phi(\cdot) \geq 0$, then it is crucial that Q is not positive semidefinite, since $Q \geq 0$ would imply $J \geq 0$, rendering the zero solution $x(0) = 0$, $x \equiv 0$, and $u \equiv 0$ a global optimum. Nonconvexity of J should thus be assumed.

2.2 The Variational/Harmonic Balance Approach to OPC

In this section, the OPC problem formulated in Section 2.1 is approached with variational calculus and its Hamiltonian system (2.5) is derived. The tools for treating it – harmonic balance and homotopy continuation – are introduced in Sections 2.2.2 and 2.2.3, respectively, setting the stage for the method proposed in Section 2.2.4.
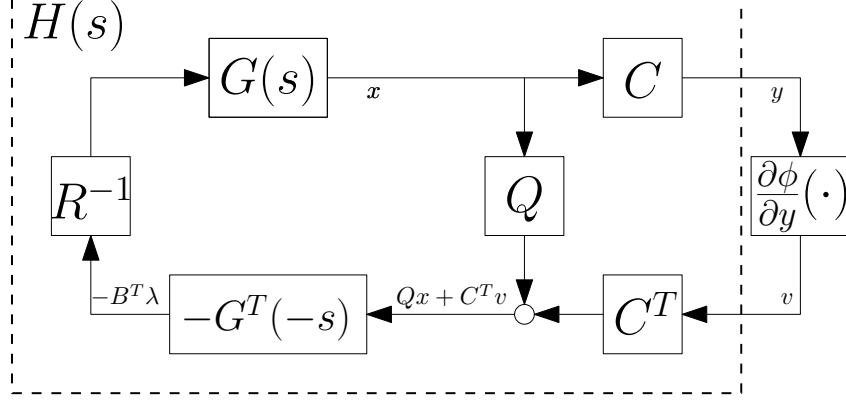
Figure 2.1 – *Signal flow of the Hamiltonian system (2.5) associated with the optimization problem (2.3). $G(s) := (sI - A)^{-1}B$ is the transfer function from input u to state x.*

2.2.1 The Hamiltonian System

It can be shown via variational calculus [HL67] that with the definition of the Hamiltonian as

$$\mathcal{H}(x,u,\lambda) = \frac{1}{2}(x^T Q x + u^T R u) + \phi(Cx) + \lambda^T(Ax + Bu)$$

the first order conditions for optimality in addition to the constraints (2.3b) are

$$\dot{\lambda} = -A^T\lambda - Qx - C^T\frac{\partial\phi}{\partial y}(Cx) = -\frac{\partial\mathcal{H}}{\partial x} \tag{2.4a}$$

$$0 = Ru + B^T\lambda = \frac{\partial\mathcal{H}}{\partial u} \tag{2.4b}$$

$$\lambda(0) = \lambda(T) \tag{2.4c}$$

$$0 = J(u,x,T) - \mathcal{H}(x(T),u(T),\lambda(T)). \tag{2.4d}$$

The transversality condition (2.4d) originates from setting the coefficient of the variation δT to zero, thus it is dropped if the period is fixed. The boundary condition (2.4c) on λ indicates that any trajectory optimal for (2.3) necessarily is (when extended over $\mathbb{R}_+$) a periodic solution of
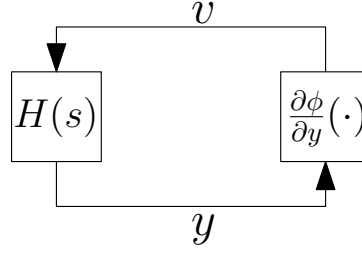
Figure 2.2 – *Lur'e structure*

the Hamiltonian system

$$\dot{x} = Ax + Bu \tag{2.5a}$$

$$y = Cx \tag{2.5b}$$

$$\dot{\lambda} = -A^T\lambda - Qx - C^T\frac{\partial\phi}{\partial y}(y) \tag{2.5c}$$

$$0 = Ru + B^T\lambda. \tag{2.5d}$$

The signal flow of the Hamiltonian system is shown in Figure 2.1, from where it is apparent that the transfer function from v to y is linear and single-input single-output, which allows us to recast the system as the classical Lur'e structure of Figure 2.2.

2.2.2 Harmonic Balance

Harmonic balance is a tool for finding oscillatory solutions of linear systems with a memoryless feedback nonlinearity such as in Figure 2.2 with

$$y = Cx, \qquad \dot{x} = Ax + Bv, \qquad v = \psi(y), \tag{2.6}$$

where ψ plays the role of $\frac{\partial\phi}{\partial y}$ and $H(s)$ is the transfer function from v to y. Since every periodic signal can be represented by its Fourier series, any periodic solution of system (2.6) can be

written as

$$y(t) = \sum_{k=-\infty}^{+\infty} \alpha_k e^{jk\omega t} \tag{2.7}$$

and, since the static nonlinearity $\psi(y)$ does not change the periodicity or period of y,

$$v(t) = \sum_{k=-\infty}^{+\infty} \beta_k e^{jk\omega t}. \tag{2.8}$$

The coefficients β_k are going to depend on the coefficients α, and *not* on the chosen fundamental frequency ω so for now we just write them as $\beta_k(\alpha)$.[1] For the signal to constitute a periodic solution of (2.6) it has to satisfy a self-excitation condition, i.e. if the loop is cut at the y-signal, then the signal has to be unaltered after passing through the now open loop:

$$\sum_{k=-\infty}^{+\infty} \alpha_k e^{jk\omega t} = \sum_{k=-\infty}^{+\infty} H(jk\omega t)\beta_k(\alpha)e^{jk\omega t}, \tag{2.9}$$

and since the functions $e^{jk\omega t}$ are mutually orthogonal, this equation can be interpreted component-wise to get

$$\alpha_k - H(jk\omega t)\beta_k(\alpha) = 0 \qquad \forall k \in \mathbb{Z}. \tag{2.10}$$

Any solution of (2.10), consisting of ω and $\alpha = (\dots, \alpha_{-1}, \alpha_0, \alpha_1, \dots)$, then corresponds to a periodic solution. However, to make the problem tractable, restriction to a finite number N of harmonics is necessary; this is not unreasonable, since H will typically have lowpass character, and thus the component equations tend to $\alpha_k + 0 = 0$ for large enough k, which justifies considering only a finite number of harmonics. Still, as we are dealing with an approximation, one has to be careful in interpreting the results. Once a solution to (2.10) is found, there are lengthy but checkable conditions under which existence of a nearby oscillatory solution of

---

[1]To see that β_k indeed does not depend on ω, scale the time in the equation for β_k and note that ω disappears:

$$\beta_k = \frac{\omega}{2\pi} \int_0^{2\pi/\omega} \psi\left(\sum_{\ell=-\infty}^{\infty} \alpha_\ell e^{j\ell\omega t}\right) e^{jk\omega t} \mathrm{d}t \overset{\tau=\omega t}{=} \frac{1}{2\pi} \int_0^{2\pi} \psi\left(\sum_{\ell=-\infty}^{\infty} \alpha_\ell e^{j\ell\tau}\right) e^{jk\tau} \mathrm{d}\tau$$

([2.5](#)) is guaranteed [Mee81]. It has also been shown [Mee72] that, under conditions involving a Lipschitz condition on $\beta_k(\alpha)$, for every oscillatory solution, there exists a finite N, for which the harmonic balance ([2.10](#)) with N components will predict it. It is not possible though to obtain N a priori, and a converse result [Fit66] states that for a given ψ and $N < \infty$, a H can be given for which a periodic solution does exist, but Equations ([2.10](#)) have no solution. For $N = 1$ one obtains the classical describing function method covered in virtually every basic controls textbook, e.g. [GL00].

A technicality that needs to be addressed is the non-uniqueness of solutions to ([2.10](#)) due to the time-invariance of H and ψ. If a periodic function $\xi(t)$ is an oscillatory solution of ([2.6](#)), then so is $\xi(t + \tau)$ for any $\tau \in \mathbb{R}$, in other words the phase of the solution is undefined. In terms of the Fourier expansions that means that the relative phase of the α_k is defined uniquely, but the absolute phase is not, and thus one of the α_k can be picked to be real and positive, customarily that is $\alpha_1 = \alpha_{-1}$. Of course, $\overline{\alpha_k} = \alpha_{-k}$ is required, too, to make the resulting signals real.

2.2.3 Homotopy Continuation

Homotopy continuation methods solve systems of polynomial equations by embedding them into a system with an additional parameter t, the so-called *homtopy*. The homotopy is chosen so that for $t = 0$, the zeros of the resulting system are very easy to find, and for $t = 1$, one obtains the original system.

Concretely: Let $F(x) = 0$ be a square polynomial system, i.e. there are n indeterminates x_i and

n polynomial equations:

$$f_1(x_1, \ldots, x_n) = 0$$

$$f_2(x_1, \ldots, x_n) = 0$$

$$F(x) = 0 \quad \Leftrightarrow \qquad \vdots \qquad \vdots$$

$$f_n(x_1, \ldots, x_n) = 0.$$

Then,

$$\Phi(x, t) = tF(x) + (1 - t)F_0(x), \qquad t \in [0, 1] \tag{2.11}$$

is such a homotopy. The solutions of $F(x) = 0$ are obtained by solving the *starting system* $F_0(x) = 0$ and continuing all isolated solutions from $t = 0$ to $t = 1$, where they correspond to solutions of the original systems $F(x) = \Phi(x, 1) = 0$. The ingenuity goes into two different parts. The first is *root counting*, i.e. choosing the starting system such that it has at least as many roots as the original system (and of course such that those roots are easy to compute). The second is the continuation from $t = 0$ to $t = 1$, avoiding or dealing with discontinuities such as branches splitting, merging, or escaping to infinity. Homotopy continuation methods are not restricted to quadratic systems, positive exponents or the particular form of (2.11); for details on root counting, the choice of homotopy and other issues, see e.g. [Ver99b] and the references therein. Of note is that the complexity of such methods depends mainly on the number of branches that have to be tracked, in other words, on how many solutions there are and how tightly their number can be bounded.

2.2.4 The Polynomial Equations of Harmonic Balance

To the best of our knowledge, the only application of harmonic balance to optimal periodic control is in [BGR72], where describing function analysis is used to find periodic solutions

of (2.5). We build on ideas described there. The transfer function H in Figure 2.2 has the state-space representation

$$
\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} A & -BR^{-1}B^T \\ -Q & -A^T \end{bmatrix}\begin{bmatrix} x \\ \lambda \end{bmatrix} + \begin{bmatrix} 0 \\ -C \end{bmatrix} v
$$
$$
y = \begin{bmatrix} C^T & 0 \end{bmatrix}\begin{bmatrix} x \\ \lambda \end{bmatrix}
$$

(2.12)

and the transfer function

$$
H(s) = -C\, G(s)R^{-1}\, (I + G^T(-s)Q\, G(s)R^{-1})^{-1}\, G^T(-s)C^T, \tag{2.13}
$$

where

$$
G(s) = (sI - A)^{-1}B.
$$

This transfer behavior retains a Hamiltonian structure, in particular it holds that

$$
H(s) = H^T(-s),
$$

i.e. the poles and zeros are symmetric with respect to the imaginary axis. Since $H(s)$ has only real coefficients, $H(-j\omega) = \overline{H(j\omega)}$ is true, too, and so for every diagonal element h_{kk}: $\overline{h_{kk}(j\omega)} = h_{kk}(j\omega)$, which means they are real for all frequencies ω. For single-output plants: $H(j\omega) \in \mathbb{R}$. Note also, that for infinite-dimensional systems with transcendental transfer functions the above statements about $H(s)$ still hold, whereas the state-space representation is restricted to the finite-dimensional case.

For sake of readability, we will restrict ourselves to scalar signals y and v. Then let α_k and β_k be their Fourier coefficients like in (2.7) and (2.8), and for ease of notation define also real

numbers $H_k := H(j\omega k)$. Equations (2.10) then become

$$\alpha_k - H_k \beta_k(\alpha) = 0 \qquad \forall k \in \mathbb{Z}. \tag{2.14}$$

What will allow us to write this as a system of *polynomial* equations amenable to numerical solution is that the explicit form of the functions $\beta_k(\alpha)$ can be generated automatically:

$$\beta_k(\alpha) = \sum_{q=0}^{p-1} (q+1)\gamma_{q+1}\Psi_{k,q}(\alpha),$$

where the functions $\Psi_{k,q}$ are derived in Appendix 2.A, important here is only that they are of the form

$$\Psi_{k,q}(\alpha) = \sum M \cdot \alpha_{\ell_1}\alpha_{\ell_2}\cdots\alpha_{\ell_q},$$

i.e. a sum of monomials of total degree q. The harmonic balance equations become

$$\alpha_k - H_k \sum_{q=0}^{p-1} (q+1)\gamma_{q+1}\Psi_{k,q}(\alpha) = 0, \tag{2.15}$$

that is they now take the form of a polynomial system of equations with degree $p-1$, for the numerical solution of which we turn to homotopy methods as outlined in Section 2.2.2; the solver we are employing is the homotopy continuation tool PHCpack [Ver99a] and its Matlab interface PHClab [GV08].

As for the complexity of the resulting numerical computations: For a given N and p, system (2.15) will consist of $2N$ equations in $2N$ variables. This is because it is not possible (to our knowledge) to express the constraint $x = \overline{y}$ as polynomials involving only x and y (no conjugates). Thus, α_k and α_{-k} have to be treated as different variables and solutions with $\alpha_{-k} \neq \overline{\alpha_k}$ to be elminated afterwards. Since we picked α_1 to be real, α_{-1} can be eliminated from the equations altogether by replacing it with α_1, and of course α_0 has to be real.

While this possibly introduces a lot of overhead by tracking spurious branches corresponding to solutions for which $\alpha_{-k} \neq \overline{\alpha_k}$, it is worth repeating that the complexity of Equations (2.15) does *not* increase with the order n of the plant but only with the degree p of $\phi(\cdot)$ and the number of considered harmonics N.

2.2.5 Examples

2.2.5.1 Double Integrator

This example is taken from [SE84], where a shooting method is employed, and also treated in [VGD04] with help of the flatness-based approach. The plant is a double integrator

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u, \qquad y = \begin{bmatrix} 0 & 1 \end{bmatrix} x. \tag{2.16}$$

and the cost function is

$$J(x(0), u, T) = \frac{1}{T} \int_0^T \frac{x_1^2 - x_2^2}{2} + \frac{y^4}{4} + \frac{u^2}{10} dt, \tag{2.17}$$

so in accordance with Equations (2.2) and (2.13)

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad R = 0.2, \quad p = 4, \quad \gamma_4 = 0.25,$$

$$H(s) = \frac{10s^2}{s^4 + 10s^2 + 10}.$$

Since $H(0) = 0$, the optimal y cannot have a DC-Term, as the harmonic balance for α_0 reduces to $\alpha_0 - 0 = 0$.

The describing function of $\frac{\partial \phi}{\partial y}(y) = y^3$ is

$$\mathcal{N}(\alpha_1) = \frac{1}{2\alpha_1 \pi} \int_{-\pi}^{\pi} \left(\alpha_1 \cos(\tau)\right)^3 \cos(\tau) \mathrm{d}\tau = \frac{3\alpha_1^2}{4}$$

and so the case of $N = 1$ can be solved explicitly:

$$\alpha_1 - H(j\omega)\frac{3\alpha_1^3}{4} = 0$$

$$\Leftrightarrow \qquad \frac{3\alpha_1^2}{4} \frac{-10\omega^2}{\omega^4 - 10\omega^2 + 10} = 1$$

$$\Leftrightarrow \qquad \alpha_1(\omega) = \sqrt{\frac{2}{15}} \frac{\sqrt{-\omega^4 + 10\omega^2 - 10}}{\omega}.$$

For $N = 3$, the harmonic balance equations are already a lot more complicated (see also Appendix 2.A), and explicit solution is not feasible anymore.

Figure 2.3 shows the achieved cost functions over a range of ω for $N \in \{1, 3, 5\}$, and Figure 2.4 shows the corresponding optimal trajectories of y. Finally Figure 2.5 shows the values of optimal Fourier coefficients α_k over ω for $N = 3$. Noteworthy are the following points:

1. Our results, in particular the optimal cost and frequency, $J^* \approx -0.0233$ and $\omega^* \approx 1.782$, are in agreement with the results from [SE84] and [VGD04].

2. The describing function analysis yields a very good estimate already, which is to be expected since H has low-pass character.

3. For $N = 1$, real solutions α_1 only exist where $H(j\omega) > 0$, which is the case between the poles at $\omega_l = \sqrt{5 - \sqrt{15}}$ and $\omega_h = \sqrt{5 + \sqrt{15}}$. For $N = 3, 5$, no solutions could be found outside this range, either. This phase criterion for feasibility of first order conditions seems intuitive; however, this idea is not justified rigorously as of yet.

4. For higher N, there can be multiple solutions at some frequencies beyond the ambiguity through time-invariance. In Figure 2.5, only the solutions corresponding to the minimum

94

cost function are shown.

2.2.5.2 Vibrating String

This example demonstrates our method on an infinite-dimensional example, namely a damped wave equation with boundary control

$$\frac{\partial^2 h}{\partial t^2} = \frac{\partial^2 h}{\partial z^2} - D\frac{\partial h}{\partial t} \qquad \begin{aligned} h(L, t) &\equiv 0 \\ \frac{\partial h}{\partial t}(0, t) &= u(t), \end{aligned} \tag{2.18}$$

which could be interpreted as a string with one loose and one fixed end where the control input is the vertical velocity of the loose end. Then, h denotes the vertical displacement of the string, z is the spatial coordinate, D is the damping coefficient, L is the length of the string; see

Figure 2.6. As first state and output we define the velocity at $L/2$

$$y(t) = x_1(t) := \frac{\partial h}{\partial t}(L/2, t). \tag{2.19}$$

The transfer function from input u to output y can be obtained by standard techniques, see also Appendix 1.D, and is

$$G_1(s) = \frac{1}{2\cosh\left(L\sqrt{s^2 + Ds}/2\right)}. \tag{2.20}$$

As the second state we pick the vertical displacement at the same location, thus $G_2(s) = G_1(s)/s$. For the cost function we assume that we want the string to have high average velocity in the middle, while keeping the displacement at bay and minimizing control effort. The cost function

$$J(x(0), u, T) = \frac{1}{T}\int_0^T \frac{x_2^2 - x_1^2}{2} + \frac{x_1^4}{4} + \frac{u^2}{2}\,\mathrm{d}t \tag{2.21}$$
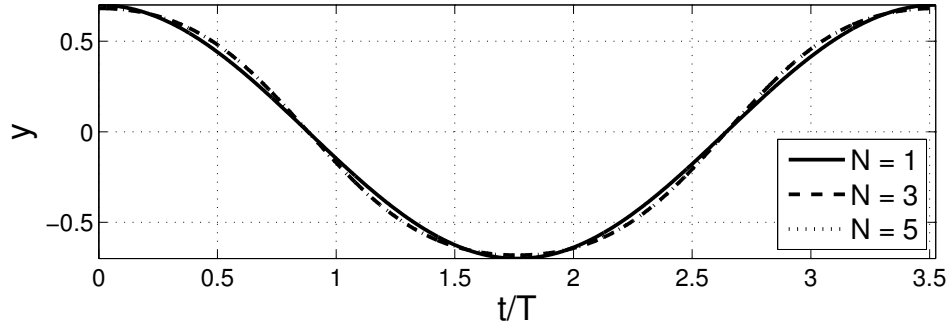
Figure 2.4 – *The optimal output trajectories $y(t)$ for Example 2.2.5.1. The fundamental frequency is $\omega^* \approx 1.782$. As is the case for the cost function, there is virtually no difference between $N = 3$ and $N = 5$.*

96
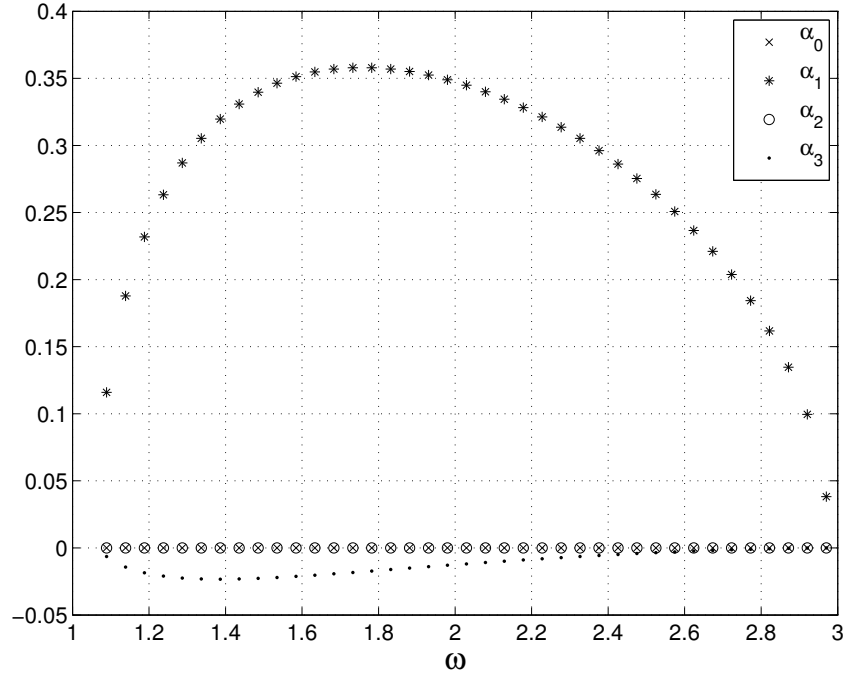
Figure 2.5 – *The solution of the harmonic balance equations for Example 2.2.5.1 yields the Fourier coefficients of the optimal output trajectory $y(t)$ at every frequency ω.*
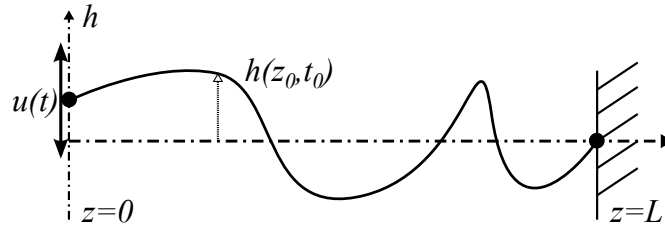
Figure 2.6 – *Interpretation of the damped wave equation (2.18) as an actuated string attached to a wall.*
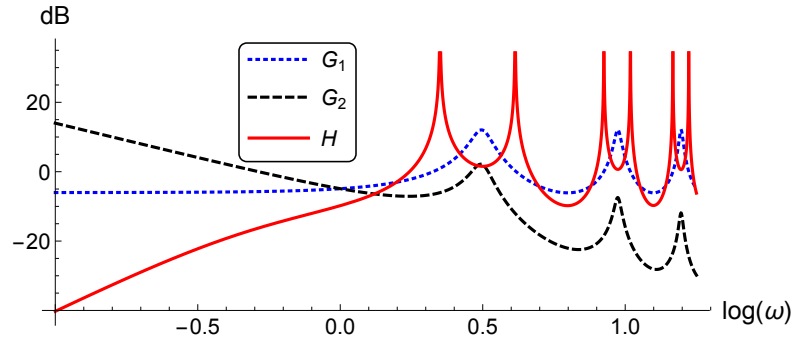
Figure 2.7 – *Bode magnitude plot of the transfer functions G_1, G_2 and H for Example 2.2.5.2. As can be seen clearly, G_1 and in particular H do not roll off.*
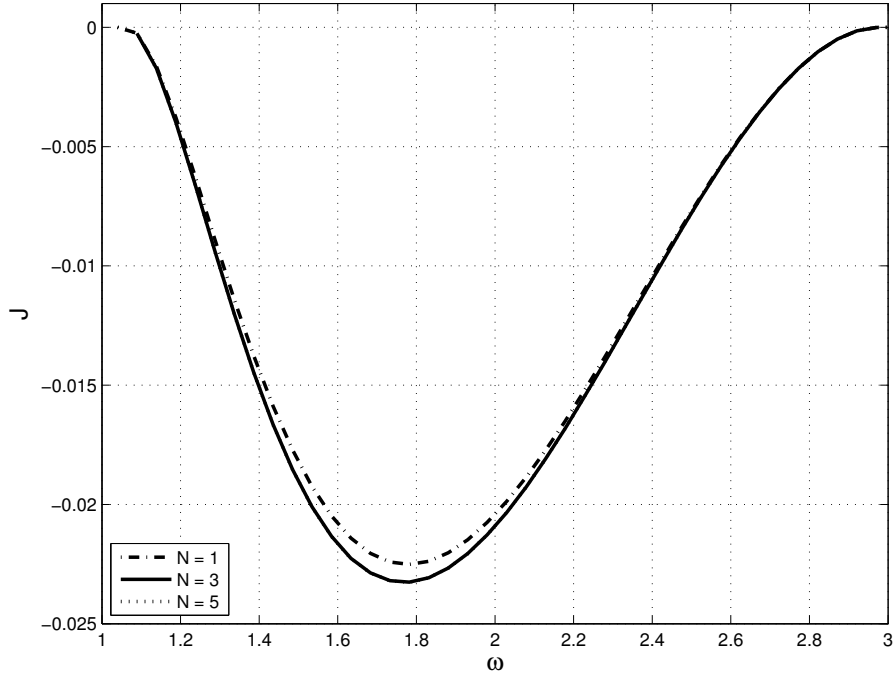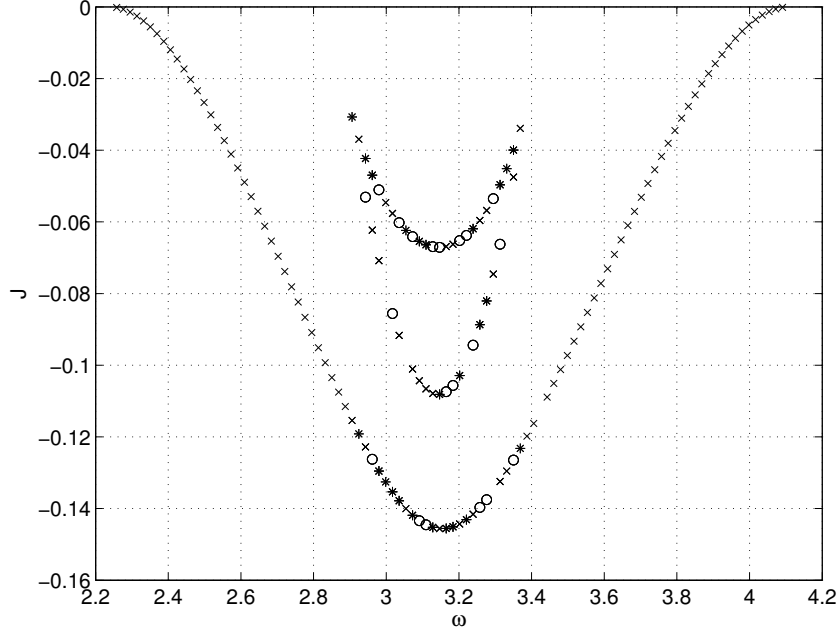
Figure 2.8 – *Solving the harmonic balance equations of Example 2.2.5.2 for $N = 3$ and computing the cost function yields above picture. Note the multiple solutions at some frequencies around the optimal frequency; they correspond to other local minima. Saddle points are also present, but they correspond to complex Fourier coefficients and are not included here.*

reflects those requirements. Again, in accordance with Equations (2.2) and (2.13)

$$Q = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \quad R = 1, \quad p = 4, \quad \gamma_4 = 0.25,$$

$$H(s) = \frac{-s^2}{\left(4\cosh\left(\frac{L\sqrt{s^2+Ds}}{2}\right)\cosh\left(\frac{L\sqrt{s^2-Ds}}{2}\right) - 1\right)s^2 - 1}.$$

Two major differences to the first example need to be pointed out. The first is, that H has infinitely many imaginary poles, and also infinitely many intervals with $H(j\omega) > 0$.

The second difference is that the example unfortunately is not a well-constructed one, because $G_1(s)$ does *not* have the lowpass character that would be desired to justify the harmonic balance, and consequently, neither does $H(s)$, as is illustrated by the Bode magnitude plots in Figure 2.7. The results on the first of the intervals with $H(j\omega) > 0$ ($\omega \in [\omega_1, \omega_2]$, where $\omega_1 \approx 2.24$ and
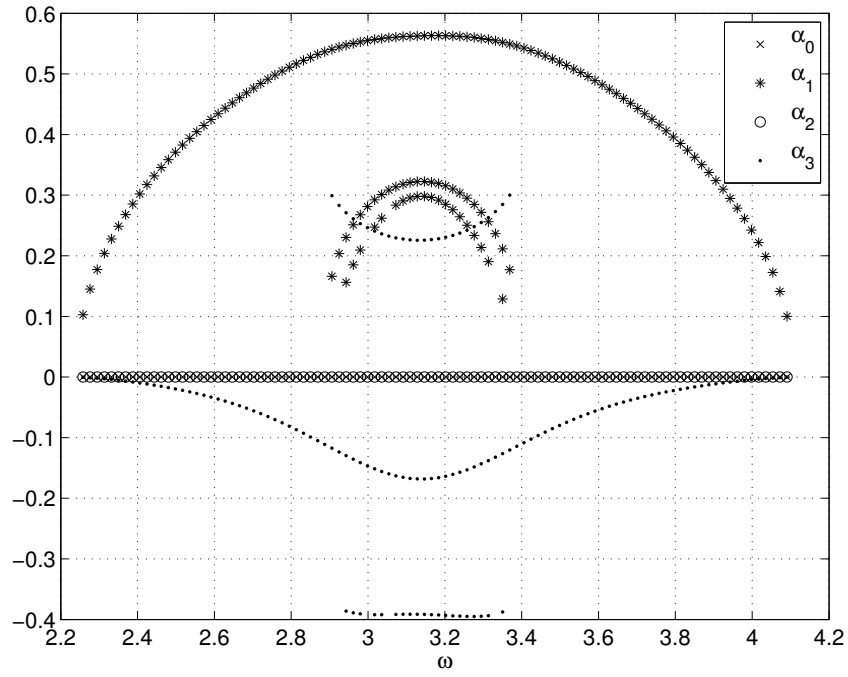
Figure 2.9 – *The solution of the harmonic balance equations for Example 2.2.5.2 with N = 3 yields the Fourier coefficients of the optimal output trajectory $y(t)$ at every frequency ω. Note again the multiple solutions and that there are no even harmonics present.*
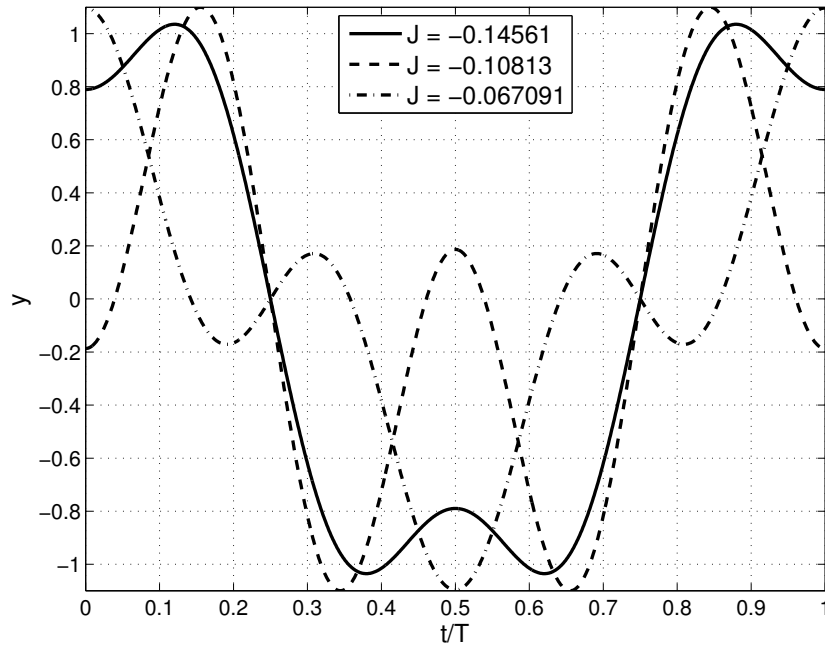
Figure 2.10 – *The output trajectories corresponding to solutions of the harmonic balance equation $y(t)$ for Example 2.2.5.2 with N = 3. The fundamental frequency is $\omega^* \approx 3.1466$.*

$\omega_2 \approx 4.12$) are shown in Figures 2.8, 2.9 and 2.10 for the values $L = 1$ and $D = 0.5$. As can be seen, the higher harmonics have a more significant effect in this example, which is likely due to the absence of roll-off in H.

To elaborate on this, Figure 2.11 shows how the cost function develops as we consider the intervals of higher frequency. The optimal cost decreases as ω increases, and this can be explained intuitively: G_2 rolls off, hence there is a "discount" on the position penalty $x_2^2/2$ as we increase the frequency, whereas G_1 does not decrease as $\omega \to \infty$, and so even at higher frequencies, the ratio of velocity to input, i.e. $|G_1|$, stays the same. Hence, the optimal periodic solution will have an infinite frequency, which minimizes the penalty on x_2.

2.3 A Different View: Polynomial Minimization

In this section, the OPC problem (2.3) is approached in a slightly different way: instead of obtaining polynomial equations representing the first-order necessary conditions (2.4) from variational calculus, the Fourier-series representation is used to directly transform the cost function J (2.2) into a polynomial. From the condition for an extremum, namely $\mathrm{grad}\ J = 0$, we then recover the same polynomial equations (2.14) as in the prior section; the more important contribution however is that it is much more efficient to minimize the polynomial form of the cost function directly using for instance the Sum-of-Squares (SOS) method [PS03], which allows reformulation as a semidefinite program.

2.3.1 Polynomial Form of Cost J

For ease of notation, we consider a single-input single-output system

$$\dot{x} = Ax + Bu \qquad y = Cx \qquad x \in \mathbb{R}^n$$
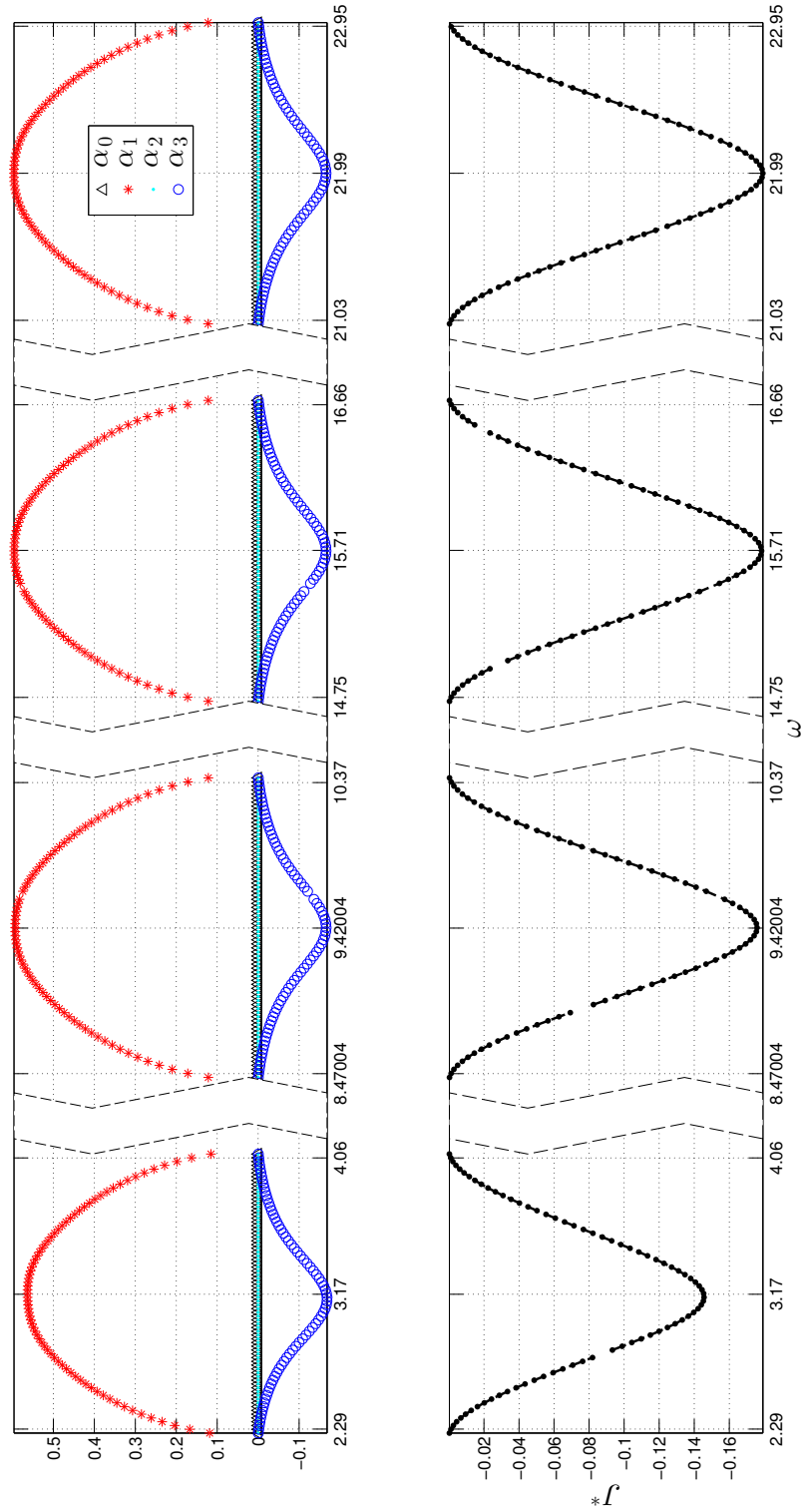
Figure 2.11 – *Optimal Fourier coefficients and cost function for Example 2.2.5.2 at increasing frequencies. As the frequency increases, the achievable cost function decreases. Only intervals with $H(j\omega) > 0$ are shown.*
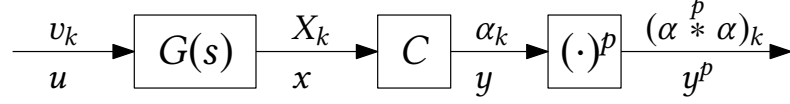
Figure 2.12 – *Relationship between input, state and output, and their Fourier coefficients.*

with

$$G(s) := (sI - A)^{-1}B \qquad G(s) \in \mathbb{C}^{n \times 1}$$

the transfer function from input u to state x, and the cost function

$$J = \frac{1}{T} \int_0^T \begin{bmatrix} x^T & u \end{bmatrix} \begin{bmatrix} Q & S \\ 0 & R \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} + y^p \, \mathrm{d}t, \tag{2.22}$$

where $p = 2\bar{p}$ is an even number. This J is slightly more general than (2.2) by allowing cross-terms, but less so by restricting the non-quadratic part to a single power. We note that it should not require much work to extend the results to allow for polynomials in y and multi-input multi-output systems.

Assume a fixed T and $\omega := 2\pi/T$. Then we denote by α_k, v_k and X_k the Fourier coefficients of y, u and x, respectively, and an application of Parseval's identity yields

$$\frac{1}{T} \int_0^T x^T Q x \, \mathrm{d}t = \sum_{k=-\infty}^{\infty} X_k^* Q X_k$$

and so on. From the system structure 2.12 we can read off

$$v_k = \frac{\alpha_k}{C G_k} \qquad \text{and} \qquad X_k = \frac{G_k \alpha_k}{C G_k},$$

where $G_k := G(jk\omega)$, and so we can rewrite the quadratic part of J as

$$J_Q = \frac{1}{T} \int_0^T \begin{bmatrix} x^T & u \end{bmatrix} \begin{bmatrix} Q & S \\ 0 & R \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} dt = \sum_{k=-\infty}^{\infty} X_k^* Q X_k + X_k^* S v_k + R \overline{v_k} v_k =$$

$$\sum_{k=-\infty}^{\infty} \left(\frac{\alpha_k}{CG_k} \right)^* \left(G_k^* Q G_k + G_k^* S v_k + R \right) \frac{\alpha_k}{CG_k}. \quad (2.23)$$

The last expression looks slightly familiar, and indeed, defining

$$H_k := CG_k \left(I + R^{-1} G_k^*(S + QG_k) \right)^{-1} R^{-1}(CG_k)^*, \quad (2.24)$$

we see that, for $S = 0$, H_k is equal to the frequency response (2.13) encountered in the previous section. Using H_k, we can rewrite

$$J_Q = \sum_{k=-\infty}^{\infty} \alpha_k^* \frac{1}{H_k} \alpha_k = \alpha^* \operatorname{diag}\{H_k\}^{-1} \alpha, \quad (2.25)$$

where $\alpha = (\ldots, \alpha_{-1}, \alpha_0, \alpha_1, \ldots)$ and

$$\operatorname{diag}\{H_k\} := \begin{bmatrix} \ddots & \ddots & & & \\ \ddots & H_{-1} & 0 & & \\ & 0 & H_0 & 0 & \\ & & 0 & H_1 & \ddots \\ & & & \ddots & \ddots \end{bmatrix}$$

are a doubly infinite vector and matrix.

The non-quadratic part requires slightly more work. It is well-known that convolution in the time-domain corresponds to simple multiplication in the frequency domain, and vice-versa. Define $\alpha \overset{q}{*} \alpha$ to be the q-fold convolution of α with itself (see also (2.29) and Appendix 2.A for

103

further nomenclature), and observe that we can equivalently write

$$\alpha \overset{q}{*} \alpha = M_\alpha^{q-1} \alpha,$$

where

$$M_\alpha := \begin{bmatrix} \ddots & \ddots & \ddots & & & \\ & \ddots & \alpha_0 & \alpha_{-1} & \cdots & \\ & \cdots & \alpha_1 & \alpha_0 & \alpha_{-1} & \cdots \\ & & \cdots & \alpha_1 & \alpha_0 & \ddots \\ & & & & \ddots & \ddots \end{bmatrix} \tag{2.26}$$

is a doubly infinite Toeplitz matrix. If the Fourier coefficients α_k correspond to real signals, then $\alpha_{-k} = \overline{\alpha_k}$ and we also have $M_\alpha^* = M_\alpha$. Using these definitions and Parseval's identity, we can write the non-quadratic part of J as

$$J_p := \frac{1}{T} \int_0^T y^p \mathrm{d}t = \frac{1}{T} \int_0^T y^{\bar{p}} \cdot y^{\bar{p}} \mathrm{d}t = \sum_{k=-\infty}^{\infty} (\alpha \overset{\bar{p}}{*} \alpha)_k^* \cdot (\alpha \overset{\bar{p}}{*} \alpha)_k =$$

$$(M_\alpha^{\bar{p}-1} \alpha)^* M_\alpha^{\bar{p}-1} \alpha = \alpha^* M_\alpha^{*\,\bar{p}-1} M_\alpha^{\bar{p}-1} \alpha = \alpha^* M_\alpha^{p-2} \alpha,$$

and

$$J(\alpha) = J_Q(\alpha) + J_P(\alpha) = \alpha^* \left(\mathrm{diag}\left\{ \frac{1}{H_k} \right\} + M_\alpha^{p-2} \right) \alpha. \tag{2.27}$$

While it might not be obvious, (2.27) is a compactly written polynomial in the coefficients α_k, $k \in \mathbb{Z}$: H_k depends on the problem parameters (system and cost) and the chosen frequency ω only, and $\alpha^* M_\alpha^{p-2} \alpha$ yields a pth-degree polynomial. Using results from Appendix 2.B, we recover the polynomial equations of harmonic balance (2.15) as

$$\frac{\mathrm{d}J}{\mathrm{d}\alpha_{-k}}(\alpha) = \frac{2\alpha_k}{H_k} + p\,(\alpha \overset{p-1}{*} \alpha)_k = 0.$$

2.3.2 Sum-of-Squares Programming (SOS)

Sum Of Squares (SOS) Programming [PS03] is a very efficient way of minimizing a multivariate polynomial $f \in \mathbb{R}[x_1, \ldots, x_n]$ over $\mathbb{R}^n$.

A polynomial $f(x_1, \ldots, x_n)$ is an SOS polynomial if there are polynomials $f_i(x_1, \ldots, x_n)$ such that $f(x_1, \ldots, x_n) = \sum_i f_i^2(x_1, \ldots, x_n)$. Obviously, if f is a SOS polynomial, then $f(x) \geq 0$ for all $x \in \mathbb{R}^n$. It is appealing but wrong to conclude that the converse holds, too, namely that f being a SOS polynomial follows from $f \geq 0$. Since that is not the case[2], going from "Is $f(x)$ positive $\forall x$?" to "Is f an SOS polynomial?" is a *relaxation*.

It follows that "$\min_{\mathbb{R}^n} f(x)$" can be relaxed to "$\max_{\mathbb{R}} \lambda$ such that $f(x) - \lambda$ is an SOS polynomial." In [PS03] it is claimed that this relaxation is often tight; in the examples considered here, this is found to be mostly true.

The value of the relaxation lies in the fact that the relaxed problem can be recast as a semidefinite program (SDP) and thus solved in polynomial time; more precisely, computation time is polynomial in the number of variables n if the degree d is fixed and in d if n is fixed. The optimal point $p^\star \in \mathbb{R}^n : f(p^\star) = \lambda_{\max}$ can be simultaneously obtained through SDP duality.

Conveniently, the MATLAB toolbox SOStools [PAV+13] provides a function `findbound` that can be called with a polynomial as input and optional inequality and equality constraints, and provides a lower bound along with a point at which it is reached (if an additional duality condition is satisfied). Internally, the SOS program is recast as an SDP and solved with an SDP solver (e.g. SDPT3 [TTT03] or SeDuMi [SRPT09]). For us that means that once our problem is formulated as a minimization of a polynomial (possibly with polynomial constraints) it can be solved numerically without further work on our end.

---

[2]The converse does hold for $n = 1$, for quadratic polynomials with any n, and for at most quartic polynomials with $n = 2$ [PAV+13]. That the converse does not hold in general is shown by the counterexample of Motzkin's polynomial $m(x, y) = -1 + x^4 y^2 + x^2 y^4 - 3x^2 y^2$, for which $m(x, y) \geq 0$, but it cannot be written as a SOS.

2.3.3 Minimizing OPC Cost with SOS

We are already close to applying SOS: (2.27) is a very compact way of writing the cost J as a polynomial, but it is not yet useful for implementation, since it involves infinite matrices and vectors and will lead to a polynomial of degree p with infinitely many terms in infinitely many variables. Truncation is straightforward for the quadratic part J_Q, but care needs to be taken with the non-quadractic part. The reason is that nonlinear operations add higher harmonics to a band-limited signal — e.g. $2\cos^2(\omega_0 t) = 1 + \cos(2\omega_0 t)$, so while the original signal consisted only of the fundamental, its square now contains the second harmonic.

It should be clear that

$$
\begin{bmatrix} \cdots & 0 & \alpha_{-N} & \alpha_{-N+1} & \cdots & \alpha_N & 0 & \cdots \end{bmatrix} M \begin{bmatrix} \vdots \\ 0 \\ \alpha_{-N} \\ \alpha_{-N+1} \\ \vdots \\ \alpha_N \\ 0 \\ \vdots \end{bmatrix}
$$

$$
= \begin{bmatrix} \alpha_{-N} & \alpha_{-N+1} & \cdots & \alpha_N \end{bmatrix} M_{[-N:N,-N:N]} \begin{bmatrix} \alpha_{-N} \\ \alpha_{-N+1} \\ \vdots \\ \alpha_N \end{bmatrix},
$$

where $M_{[-N:N,-N:N]}$ denotes the block of the doubly infinite matrix M with row and column indices ranging from $-N$ to N. However,

$$
\left(M^p\right)_{[-N:N,-N:N]} \neq \left(M_{[-N:N,-N:N]}\right)^p,
$$

but instead, sufficiently many elements of M must be kept when computing the pth power. For a general doubly infinite matrix, "sufficiently many" means "all," but in the case of M_α with $\alpha_k = \alpha_{-k} = 0$ for all $k > N$, there is a finite answer:

$$\left(M_\alpha^p\right)_{[-N:N,-N:N]} = \left(\left(M_{\alpha,[-pN:pN,-pN:pN]}\right)^p\right)_{[-N:N,-N:N]},$$

which is derived as (2.39) in Appendix 2.C. $M_{\alpha,[-pN:pN,-pN:pN]}$ is the *finite* Toeplitz matrix uniquely defined (for given N) by its first row r_α^T and first column c_α:

$$
\begin{aligned}
r_\alpha^T &= \begin{bmatrix} \alpha_0 & \alpha_{-1} & \cdots & \alpha_{-N} & \overbrace{0 \quad \cdots \quad \cdots \quad 0}^{(2p-1)N \text{ elements}} \end{bmatrix} \\
c_\alpha^T &= \begin{bmatrix} \alpha_0 & \alpha_1 & \cdots & \alpha_N & \underbrace{0 \quad \cdots \quad \cdots \quad 0}_{(2p-1)N \text{ elements}} \end{bmatrix}
\end{aligned}
$$

and so the cost function J for a finite vector of Fourier coefficients $\alpha = \begin{bmatrix} \alpha_{-N} & \cdots & \alpha_N \end{bmatrix}$ can be generated as

$$
J_N = \alpha^* \left(\begin{bmatrix} \frac{1}{H_{-N}} & 0 & \cdots & \cdots & 0 \\ 0 & \frac{1}{H_{-N+1}} & 0 & \cdots & 0 \\ 0 & & \ddots & & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & \cdots & & \cdots & \frac{1}{H_N} \end{bmatrix} + \left(\left(M_{\alpha,[-qN:qN,-qN:qN]}\right)^q\right)_{[-N:N,-N:N]} \right) \alpha,
$$

where $q := p - 2$; see also (2.39).

Two small implementation issues have to be addressed: First, since SOS methods minimize over $\mathbb{R}$, we write $\alpha_{\pm k} = x_k \pm j y_k$ and obtain a multivariate polynomial in the $2N$ variables $x_0, x_1, ..., x_N$ and $y_0, y_1, ... y_N$. Secondly, to enforce a real signal, we set $y_0 = 0$ and to fix the phase of the optimal trajectory (see also the end of Section 2.2.2), we also set $y_1 = 0$ and require $x_1 \geq 0$.

Application to Examples 2.2.5.1 and 2.2.5.2 yields virtually identical results, hence we refer to

Figures 2.3 through 2.10 and instead discuss the important differences between the methods to obtain them:

Computational effort The formulation as an SDP allows for very efficient algorithms to be applied, and furthermore explicitly enforcing the constraint $\alpha_k = \overline{\alpha_{-k}}$ (which was not possible in Section 2.2.4) means that no spurious solutions have to be computed and then discarded. This leads to a large advantage in complexity for the SOS approach, and computation times from Example 2.2.5.1 illustrate the vastly superior efficiency of SOS: Solution on a grid of 100 frequencies for $N = 3$ harmonics takes about 88 s using SOStools,[3] and about 210 s using PHCpack to solve the first-order conditions as in Section 2.2.4. For $N = 5$ harmonics, SOS takes about 200 s, but PHCpack will not finish for hours.

Relaxation As mentioned in Section 2.3.2, the SOS approach involves relaxations, and so it finds upper bounds (on the optimal cost) and returns a point at which this bound is achieved only if a duality condition holds (see [PAV$^+$13, Sec. 4.3]). In Example 2.2.5.1, this leads to no solutions being found for frequencies between $\omega_l = \sqrt{5 - \sqrt{15}}$ and $\omega_{\text{SOS}} \approx 1.25$. In the example, that does not influence the global optimum, however it generally cannot be determined a priori nor a posteriori whether the global optimum lies in the range where SOStools fails to return a solution.

Other extrema Of course, SOS has no way of finding extrema beside the global minimum. In order to obtain those, the first-order conditions have to be solved, hence if (for whatever reason) local extrema and saddle points are of interest, one would have to resort to the method of Section 2.2.

---

[3]using SDPT3 and a degree for the relaxation (see [PAV$^+$13, Sec. 4.3]) of 5

2.4 Conclusions and Future Directions

We have first shown how the OPC problem (2.3) for a linear plant with finite or infinite state dimension can be treated by means of variational calculus and harmonic balance, resulting in a system of multinomial equations whose zeros are the Fourier coefficients of trajectories satisfying first-order necessary conditions for optimality. The approach is demonstrated on two examples, the first of which has been treated in the OPC literature previously, whereas the second involves an infinite-dimensional system.

In the second part we have shown that by using the Fourier series representation of the output signal y to parametrize all involved signals and rewriting the cost function J in terms of the Fourier coefficients of y, we obtain J as a single polynomial in the – infinitely many – Fourier coefficients. The same polynomial equations as in the first part can be derived by setting the gradient of J with respect to the Fourier coefficients to zero. More importantly, there are much more computationally effective methods to minimize a polynomial than finding the zeros of its gradient, and we outlined how Sum-of-Squares (SOS) programming can be used to obtain the same results for the treated examples. Some technical issues encountered when going from infinitely many to finitely many Fourier coefficients are addressed along the way and in the appendix.

Since this approach of rewriting the cost function as a polynomial does not hinge on a compact form for the Hamiltonian system anymore – in fact, no variational calculus is required at all – it lends itself to more general structures. The approach made use of two properties of the particular OPC problem treated here: there is a signal on whose Fourier coefficients all other signals' Fourier coefficients depend polynomially, and the cost function can be rewritten as a polynomial in the Fourier coefficients of the signals it depends on.

As for the second requirement, the cost function J can be rewritten using Parseval's identity, whenever it is quadratic in a polynomial function of the involved signals. To formalize the first,
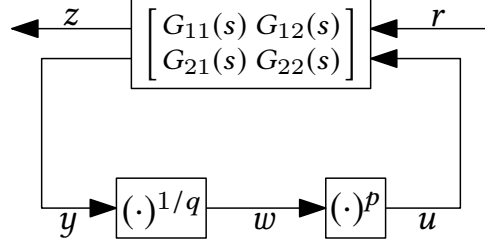
Figure 2.13 – *LLFT-like structure for which OPC problems can be approached similarly as in Section 2.3*

let us call a signal "parametrizing" if its Fourier coefficients determine all other signals' Fourier coefficients through polynomial dependencies. This definition of a parametrizing signal bears resemblance to the so-called flat output, yet they are not equivalent: For $\dot{x} = u$, u is clearly a parametrizing signal, but not a flat output. On the other hand, for $\dot{x} = |x| + u$, x is a flat output (since $u = f(x, \dot{x}) = \dot{x} - |x|$), yet there is no polynomial relationship between the Fourier coefficients of u and x.

From what we have derived in the previous sections, we see that it is sufficient for a signal $a = \sum_k \alpha_k e^{jk\omega t}$ to be parametrizing, if every other signal can be reached by:

- going through a linear system $G(s)$ in either direction. If $b = G(s)a$ (or $a = G(s)b$) Then, $\beta_k = G(jk\omega)\alpha_k$ (or $\beta_k = \alpha_k/G(jk\omega)$).

- going forward through a polynomial. If $b = \sum_\ell c_\ell a^\ell$, then $\beta_k = \sum_\ell c_\ell \Psi_{k,\ell}(\alpha)$, where $\Psi_{k,\ell}(\cdot)$ is defined in (2.30).

- going backward through an odd inverse power. If $a = b^{1/q}$, then $b = a^q$ and we can apply the polynomial case.

Combining those findings, we see that we can treat the fairly general structure shown in Figure 2.13, since the signal w is parametrizing. This structure is promisingly close to that of the nonlinear Rijke tube model shown in Figure 1.15, but we are not quite there yet, since the nonlinearity $\sqrt{|\bar{v} + \cdot|}$ does not allow straightforward parametrization. Expansion of the concept to

even more general systems, and a constructive method to find a parametrizing signal are thus very interesting questions for future work in this direction.

Appendix 2.A   The Fourier Coefficients of y^q.

We derive the functions $\Psi_{k,q}(\alpha)$ so that

$$\left(\sum_{k=-N}^{N} \alpha_k e^{jk\omega t} \right)^q = \sum_{k=-N}^{N} \Psi_{k,q}(\alpha) e^{jk\omega t}, \tag{2.28}$$

where $\alpha = (\alpha_{-N}, \ldots, \alpha_0, \alpha_1, \ldots, \alpha_N)$. Note that raising a signal with N harmonics to the qth power yields a signal with Nq harmonics, but in view of the harmonic balance equation (2.15) the additional harmonics will have to be ignored. Convolution of time signals corresponds to multiplication of their Fourier coefficients, and, by symmetry, multiplication in the time-domain corresponds to a convolution of the Fourier coefficients. Let us denote the sequence obtained by convolutions of sequences x and α by $x * \alpha$, and define the m-fold convolution of α with itself as

$$\alpha \overset{m}{*} \alpha := \underbrace{\alpha * \alpha * \cdots * \alpha}_{m \text{ terms}}, \tag{2.29}$$

then we can identify

$$\Psi_{k,q}(\alpha) = (\alpha \overset{q}{*} \alpha)_k,$$

111

where $(\cdots)_k$ denotes the kth element of a sequence. Writing down $\alpha \overset{m}{*} \alpha$ for some m

$$(\alpha * \alpha)_k = \sum_{i=-N}^{N} \alpha_{k-j}\alpha_j$$

$$(\alpha * \alpha * \alpha)_k = \sum_i \alpha_{k-i} \sum_j \alpha_{i-j}\alpha_j = \sum_{i,j} \alpha_{k-i}\alpha_{i-j}\alpha_j$$

$$\vdots \qquad\qquad \vdots$$

$$(\alpha \overset{m}{*} \alpha)_k = \sum_{i_1,i_1,\ldots,i_{m-1}} \alpha_{k-i_1}\alpha_{i_1-i_2} \cdots \alpha_{i_{m-1}}$$

we note that the sum of the coefficients on the right hand is always k and there are always m of them, thus the summation could be run over all length-m integer partitions of k instead of $i_1, i_2, \ldots$. However, some monomials appear more often than others, i.e. a factor has to be added to every summand. For illustrative purposes, let $m = 3$, $N = 2$, $k = 0$. The element $(\alpha \overset{3}{*} \alpha)_0$ is then

$$\alpha_0^3 + 6(\alpha_{-2}\alpha_0\alpha_2 + \alpha_{-1}\alpha_0\alpha_1) + 3(\alpha_{-2}\alpha_1^2 + \alpha_{-1}^2\alpha_2),$$

so where there is only one index as in α_0^3, the coefficient is 1, where there are 2 as in $\alpha_{-2}\alpha_1^2$ and $\alpha_{-1}^2\alpha_2$ it is 3, and for 3 different indices it is 6.

Combinatorics help to find those coefficients generally: the coefficient for each monomial should be the number of ways of rearranging the monomial $\alpha_{\ell_1}\alpha_{\ell_2} \cdots \alpha_{\ell_q}$, which in turn depends only on how many of the ℓ_j are equal. For the limiting case where all of them are equal, there is only one way, and if they all are distinct, then there are $q!$ ways. For the in-between cases, let $c_i = \mathrm{card}\{j \,|\, \ell_j = i\}$, i.e. the number of factors with index i. Then, there are

$$\binom{q}{c_{-N}, c_{-N+1}, \cdots, c_N} := \frac{q!}{c_{-N}! \, c_{-N+1}! \cdots c_N!}$$

ways of rearranging. This so-called multinomial coefficient counts the ways of putting q distinct objects ("Factor in ith position") into r boxes ("Factors with index j"), where c_j objects

112

go into the jth container. Note also that $0! = 1$. The above findings can be combined to

$$\Psi_{k,q}(\alpha) = \sum_{\substack{\ell_1+\ell_2+\ldots+\ell_q=k \\ \ell_i \in [-N,N] \; \forall i=1,\ldots,q}} \binom{q}{c_{-N}, \cdots, c_N} \alpha_{\ell_1} \alpha_{\ell_2} \cdots \alpha_{\ell_q}, \tag{2.30}$$

where c_i denotes the frequency of index i in the current partition, e.g. if a partition of 1 is $\ell_1 = -1$, $\ell_2 = -1$, $\ell_3 = 3$, then $c_{-1} = 2$, $c_3 = 1$, $c_{-N\cdots-2,0,1,2,4,\ldots N} = 0$. As an example, consider $\Psi_{k,3}$ with $N = 3$ (Note that $\alpha_1 = \alpha_{-1} > 0$ has been included already):

$$\Psi_{-3,3}(\alpha) = \alpha_1^3 + 6\alpha_0\alpha_1\alpha_{-2} + 3\alpha_0^2\alpha_{-3} + 3\alpha_1\alpha_{-2}^2 + 6\alpha_1^2\alpha_{-3} + 6\alpha_2\alpha_{-2}\alpha_{-3} + 3\alpha_3\alpha_{-3}^2$$

$$\Psi_{0,3}(\alpha) = \alpha_0^3 + 6\alpha_1^2\alpha_0 + 3\alpha_1^2\alpha_{-2} + 3\alpha_2\alpha_1^2 + 6\alpha_2\alpha_0\alpha_{-2} + 6\alpha_2\alpha_1\alpha_{-3} + 6\alpha_3\alpha_1\alpha_{-2} + 6\alpha_3\alpha_0\alpha_{-3}$$

$$\Psi_{1,3}(\alpha) = 3\alpha_1\alpha_0^2 + 3\alpha_1^3 + 6\alpha_2\alpha_0\alpha_1 + 6\alpha_2\alpha_1\alpha_{-2} + 3\alpha_2^2\alpha_{-3} + 3\alpha_3\alpha_1^2 + 6\alpha_3\alpha_0\alpha_{-2} + 6\alpha_3\alpha_1\alpha_{-3}$$

$$\Psi_{2,3}(\alpha) = 3\alpha_1^2\alpha_0 + 3\alpha_2\alpha_0^2 + 6\alpha_2\alpha_1^2 + 3\alpha_2^2\alpha_{-2} + 6\alpha_3\alpha_0\alpha_1 + 6\alpha_3\alpha_1\alpha_{-2} + 6\alpha_3\alpha_2\alpha_{-3}.$$

Appendix 2.B   The Gradient of $\alpha^* M_\alpha^p \alpha$

Let $\alpha^T = (\ldots, \alpha_{-1}, \alpha_0, \alpha_1, \ldots)$ be a doubly infinite vector of Fourier coefficients such that $\alpha_{-k} = \overline{\alpha_k}$, M_α be defined as in (2.26), and Λ be a doubly infinite diagonal matrix with diagonal entries λ_k, $k \in \mathbb{Z}$.

Also, let e_k be the kth elementary vector with elements equal to 1 at the kth position and equal to zero everywhere else. Then we have $E_k := M_{e_k}$ as a doubly infinite matrix with elements on the kth subdiagonal equal to 1, and all others equal to 0, i.e. $[E_k]_{i,j} = \delta_{k,(i-j)}$.[4] Alternatively, we can identify E_k to be the symbol of the right-shift operator.

---

[4]Here, $[\cdot]_{i,j}$ denotes the element in the ith row, jth column, and $\delta_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{else} \end{cases}$ is the Kronecker symbol.

Then, checking the following properties is rather straightforward:

$$\frac{\mathrm{d}\alpha^*}{\mathrm{d}\alpha_k} = e_{-k}^T \tag{2.31}$$

$$\frac{\mathrm{d}M_\alpha}{\mathrm{d}\alpha_k} = E_k \tag{2.32}$$

$$\frac{\mathrm{d}(\alpha^*\Lambda\alpha)}{\mathrm{d}\alpha_k} = \alpha_{-k}(\lambda_k + \lambda_{-k}) \tag{2.33}$$

$$E_k^T = E_k^* = E_{-k} \tag{2.34}$$

$$M_\alpha e_k = E_k \alpha \tag{2.35}$$

$$e_{-k}^T M_\alpha = \left(M_\alpha^* e_{-k}\right)^* = \alpha^* E_k \tag{2.36}$$

$$M_\alpha E_k = E_k M_\alpha. \tag{2.37}$$

It follows by the product rule that

$$\frac{\mathrm{d}\left(\alpha^* M_\alpha^p \alpha\right)}{\mathrm{d}\alpha_k} = e_{-k}^T M_\alpha^p \alpha + \alpha^* \left(p M_\alpha^{p-1} E_k\right)\alpha + \alpha^* M_\alpha^p e_k =$$

$$\alpha^* E_k M_\alpha^{p-1}\alpha + p\alpha^* M_\alpha^{p-1} E_k \alpha + \alpha^* M_\alpha^{p-1} E_k \alpha =$$

$$(p+2)\alpha^* M_\alpha^{p-1} E_k \alpha = (p+2)\alpha^* M_\alpha^p e_k = (p+2)\left(M_\alpha^p \alpha\right)^* e_k$$

and finally, since $M_\alpha^p \alpha = \alpha^{\overset{p+1}{*}} \alpha$:

$$\frac{\mathrm{d}\left(\alpha^* M_\alpha^p \alpha\right)}{\mathrm{d}\alpha_k} = (p+2)\left(e_k^T (\alpha^{\overset{p+1}{*}}\alpha)\right)^* = (p+2)\overline{\left((\alpha^{\overset{p+1}{*}}\alpha)_k\right)} = (p+2)(\alpha^{\overset{p+1}{*}}\alpha)_{-k}. \tag{2.38}$$

Appendix 2.C   Computing $\left(M_\alpha^p\right)_{[-N:N,-N:N]}$

Let $\alpha = (\ldots, 0, \alpha_{-N}, \ldots, \alpha_N, 0, \ldots)$ and $\beta = (\ldots, 0, \beta_{-M}, \ldots, \beta_M, 0, \ldots)$ be infinite vectors with $2N+1$ (respectively $2M+1$) nonzero elements, and consider the product of the infinite Toeplitz

114

matrices M_α and M_β. For the element in the ith row, jth column, we have

$$[M_\alpha M_\beta]_{i,j} = \sum_{\ell \in \mathbb{Z}} [M_\alpha]_{i,\ell} [M_\beta]_{\ell,j} = \sum_{\ell \in \mathbb{Z}} \alpha_{i-\ell} \beta_{\ell-j} = \sum_{k \in \mathbb{Z}} \alpha_k \beta_{(i-j)-k}.$$

The range for which α_k is nonzero is $-N < k < N$, and the range for which $\beta_{(i-j)-k}$ is nonzero is $(i-j) - M < k < (i-j) + M$, and both overlap only if $|i - j| < M + N$. Hence, the product of matrices with zeros outside the $\pm N$th and $\pm M$th diagonal will have zeros outside the $\pm(M+N)$th diagonal, but generally will have nonzero elements up to the $\pm(M+N)$th diagonal.

Combined with the observation that *all* nonzero elements of α and β go into the computation of every diagonal element of $M_\alpha M_\beta$, it follows that in order to obtain the correct truncation of M_α^p to the inner N bands, we have to consider that per multiplication, the matrix M_α with $2N + 1$ nonzero bands adds N bands on either side, and so the matrix to truncate from needs to be of size $(2pN + 1) \times (2pN + 1)$. In other words, the initial matrix that will be taken to the pth power can be constructed as the finite $(2pN + 1) \times (2pN + 1)$ Toeplitz matrix generated from the nonzero part of α, padded with $(p - 1)N$ zeros on either end:

$$\left(M_\alpha^p \right)_{[-N:N,-N:N]} = \left(\mathring{M}_{\mathring{\alpha}}^p \right)_{[-N:N,-N:N]} = \left(\left(M_{\alpha,[-pN:pN,-pN:pN]} \right)^p \right)_{[-N:N,-N:N]}, \tag{2.39}$$

where

$$\mathring{\alpha} = \begin{bmatrix} o_{-pN} & \cdots & o_{-N-1} & \alpha_{-N} & \cdots & \alpha_N & o_{N+1} & \cdots & o_{pN} \end{bmatrix}$$

$$\mathring{M}_{\mathring{\alpha}} = \begin{bmatrix}
\alpha_0 & \alpha_1 & \cdots & & & o_{pN} & 0 & \cdots & & \cdots & 0 \\
\alpha_{-1} & \alpha_0 & \ddots & & & & \ddots & & & & \vdots \\
\vdots & \ddots & \ddots & & & & & & & & \\
 & & & \ddots & & & & & & & \vdots \\
 & & & & \ddots & & & & & & \vdots \\
o_{-pN+1} & & \ddots & \ddots & & \ddots & & & \ddots & & 0 \\
o_{-pN} & o_{-pN+1} & \cdots & o_{-N-1} & \alpha_{-N} & \cdots & \alpha_0 & \cdots & \alpha_N & o_{N+1} & \cdots & o_{pN} \\
0 & \ddots & \ddots & & & & & & & & o_{pN-1} \\
\vdots & & & & & & & & & & \\
 & & & & \ddots & & & & \ddots & \vdots \\
 & & & & & \ddots & & & \ddots & \ddots & \alpha_1 \\
0 & & & & & o_{-pN} & & \cdots & \alpha_{-1} & \alpha_0
\end{bmatrix}$$

are a finite vector and a finite matrix, with the zero-padding elements $o_k \equiv 0$ added to illustrate their generation.

116

Chapter 3

Stability, Stabilizability and Linear-Quadratic Control of Spatially Invariant Systems over $\mathcal{H}_k$ Spaces

The control of infinite-dimensional, spatially distributed and large-scale systems has been the subject of much work and research for many decades now. Classical references include [LM71, Ban83, BDPDM07] and many others. In the past decade or so, more detailed research questions have emerged related to distributed and cooperative control that involve questions of distributed control architecture, communication requirements and multi-agent systems. Amongst these, the results of this chapter are closely related to [BPD02, DD03].

The setting in [BPD02] is for systems and signals defined over a spatial domain that forms a group. This setting provides an idealization where many system-theoretic properties can be easily and intuitively investigated. The group structure allows for considering systems whose action is translation-invariant, so-called spatially invariant systems. These are idealizations in an analogous manner that time-invariant systems are commonly used as useful idealizations of

117

systems that might in reality be slowly time-varying or operate over finite time horizons.

The spatially invariant setting allows for simple "pointwise" methods for testing stability, stabilizability and solving optimal linear-quadratic control problems, and more importantly, optimal linear-quadratic controllers (even centralized ones) inherit the spatial invariance of the plant as well as some degree of locality. While the proof of the former property, namely that the optimal spatially invariant controller achieves the same performance as any optimal spatially varying one, is valid regardless of the underlying (Hilbert) state space [BPD02, Thm. 9], the "localized" property of quadratically optimal controller kernels is shown only for the special case of $\mathfrak{L}_2$ state spaces, and the question arises as to whether these results are valid for more general Hilbert space settings such as Sobolev spaces. Sobolev spaces commonly arise as the state space for systems that have wave-like dynamics, or more generally partial differential equations (PDEs) with higher order temporal derivatives.

It should be remarked that the recent paper [Cur12] treats the situation of second-order temporal derivatives in great detail, and in this setting even provides conditions for well-posedness of solutions to the underlying PDE problem. The results of [Cur12, Lemma A.1] are essentially equivalent to most of what follows, and the author of that work should be thanked for pointing this out. However, the exposition here differs considerably. We show how the spatially invariant systems theory from [BPD02], in particular the "pointwise" analysis, can be generalized *directly* from the setting of $\mathfrak{L}_2$-space. Our generalization is based on a version of Plancherel's theorem (Theorem 3.6) and a careful definition of the symbols of adjoints (Lemma 3.5) with the necessary accounting of the spatial frequency weights related to the Sobolev norms; we think that this approach is more intuitive and transparent.

Tests for stability, stabilizability and constructions of optimal solutions to linear-quadratic control problems can then be done in the spatial frequency domain, and we show how in Sections 3.1 and 3.2. Section 3.3 uses these results to treat wave and beam equation examples

118

in detail, and investigates properties of the optimal LQR feedback gains. We will see that the optimal controllers satisfy a localization property, which is an important contribution in light of [Cur11, Sec. IV] claiming the opposite.

This chapter is an extended version of the

Related Publication

[EB14] Jonathan P. Epperlein and Bassam Bamieh. Distributed control of spatially invariant systems over Sobolev spaces. In *European Control Conference (ECC), 2014*, pp. 2133–2138. IEEE, 2014.

3.1   Preliminaries

This section introduces some definitions and nomenclature and provides the results necessary to generalize the treatment of the $\mathfrak{L}_2$ setting to $\mathcal{H}_k$ state spaces which are defined in Definition 3.2. Theorem 3.6 presents the correct frequency weighting to retain certain properties of the spatial Fourier transformation $\mathscr{F}$ that enable us to rewrite operator-valued Riccati and Lyapunov equations in terms of families of matrix-valued ones in Section 3.2; special care needs to be taken with adjoint operators in such equations, and Lemma 3.5 quantifies what "care" means.

3.1.1   Translations and Invariance

The results in [BPD02] generally apply if the the state space is a function space whose spatial domain is a locally compact Abelian group. In what follows, $\mathbb{G}$ will denote such a group. The commonly encountered instances are

- the real line $\mathbb{G} = \mathbb{R}$,

- the unit circle $\mathbb{G} = \partial \mathbb{D}$,

- the integers $\mathbb{G} = \mathbb{Z}$, and

- the integers modulo $N \in \mathbb{Z}$: $\mathbb{G} = \mathbb{Z}_N$.

and Cartesian products of these, e.g. the infinite cylinder $\mathbb{R} \times \partial \mathbb{D}$ or d-dimensional Euclidean space $\mathbb{R}^d$.

The issues addressed here arise, roughly, from the fact that the spatial derivative operator $\frac{\partial}{\partial x}$ is unbounded, and the necessity to restrict the domain of such right-hand side operators to ensure well-posedness; since there is no spatial derivative on $\mathbb{Z}$ and $\mathbb{Z}_N$ (or – if we consider the difference operator $\Delta f(x) = f(x) - f(x - 1)$ as the analogue – since Δ is bounded), in the sequel $\mathbb{G}$ is either the real line $\mathbb{R}$, or the unit circle $\partial \mathbb{D}$. If the spatial domain is a Cartesian product it is useful to keep in mind though that the analysis also applies to $\mathbb{Z}$ and $\mathbb{Z}_N$.

The *group operation* on $\mathbb{G}$ is denoted by $+$ and corresponds to simple addition for $\mathbb{G} = \mathbb{R}$ and $\mathbb{G} = \mathbb{Z}$, and addition modulo 2π (respectively N) for $\mathbb{G} = \partial \mathbb{D}$ (respectively $\mathbb{G} = \mathbb{Z}_N$). For Cartesian products, it is applied elementwise. The group operation induces a *translation operator* T_ξ on functions on $\mathbb{G}$:

$$[T_\xi f](x) := f(x - \xi).$$

Definition 3.1. *An operator K is* translation invariant *(or* shift invariant*), if its action commutes with T_ξ:*

$$KT_\xi = T_\xi K \quad \forall \xi \in \mathbb{G}.$$

and its domain $\mathfrak{D}(K)$ is invariant under T_ξ, too: $T_\xi \mathfrak{D}(K) = \mathfrak{D}(K)$..

Some classes of spatially invariant operators of particular interest are

- shift operators themselves,

- the differential operator $\frac{\partial}{\partial x}$,

- spatial convolutions, which we will write as

$$Kf = k * f = \int_{\mathbb{G}} k(x - \xi) f(\xi) \, \mathrm{d}\xi,$$

where $\mathrm{d}x$ is the Lebesgue measure on $\mathbb{G}$, and we have

$$\int_{\mathbb{G}} f(x)\mathrm{d}x = \begin{cases} \int_{-\infty}^{\infty} f(x)\mathrm{d}x & \text{if } \mathbb{G} = \mathbb{R} \\ \int_{0}^{2\pi} f(x)\mathrm{d}x & \text{if } \mathbb{G} = \partial\mathbb{D}, \end{cases}$$

- compositions and linear combinations of spatially invariant operators.

3.1.2 Hilbert Spaces $\mathfrak{L}_2(\mathbb{G})$, $\mathcal{H}_k(\mathbb{G})$, $\mathfrak{L}_W(\mathbb{G})$

A *Hilbert space H* is a Banach space equipped with an *inner product*, which we shall denote by $\langle x, y \rangle_H$, and a corresponding *norm* $\|x\|_H^2 := \langle x, x \rangle_H$; the subscript is optional if we need to be explicit about the space. The Hilbert spaces of interest here are

Definition 3.2.

- *The space of square-integrable functions*

$$\mathfrak{L}_2^n(\mathbb{G}) := \left\{ f : \mathbb{G} \to \mathbb{C}^n \mid \|f\|_{\mathfrak{L}_2^n} < \infty \right\}$$

$$\langle f, g \rangle_{\mathfrak{L}_2^n} := \int_{\mathbb{G}} g^*(x) f(x) \, \mathrm{d}x$$

- *The weighted space with weight $W : \mathbb{G} \to \mathbb{C}^{n \times n}$, which is positive definite on $\mathbb{G}$ and*

121

grows no more than polynomially in its argument

$$\mathfrak{L}_W(\mathbb{G}) := \left\{ f : \mathbb{G} \to \mathbb{C}^n \mid \|f\|_{\mathfrak{L}_W} < \infty \right\}$$

$$\langle f, g \rangle_{\mathfrak{L}_W} := \int_{\mathbb{G}} g^*(\lambda) W(\lambda) f(\lambda) \, d\lambda$$

- *The Sobolev space*

$$\mathcal{H}_k^n(\mathbb{G}) := \left\{ f : \mathbb{G} \to \mathbb{C}^n \mid \|f\|_{\mathcal{H}_k^n} < \infty \right\}$$

$$\langle f, g \rangle_{\mathcal{H}_k^n} := \sum_{\ell=0}^{k} \left\langle \frac{\partial^\ell f}{\partial x^\ell}, \frac{\partial^\ell g}{\partial x^\ell} \right\rangle_{\mathfrak{L}_2^n}$$

Equivalently, this space consists of all functions whose derivatives up to kth order are in the respective $\mathfrak{L}_2$-space.

Hence, weighted spaces and Sobolev spaces are different generalizations of $\mathfrak{L}_2^n$, for example we have

$$\mathfrak{L}_2^2(\mathbb{G}) = \mathcal{H}_0^2(\mathbb{G}) = \mathfrak{L}_{\left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]}^2(\mathbb{G}).$$

It will turn out (Theorem 3.6), that $\mathcal{H}_k$ spaces are the adequate generalization in the physical domain, while weighted spaces are introduced in the frequency domain.

Remark. In many cases it is overly restrictive to require *all* derivatives up to order k of the elements of the state space to live in $\mathfrak{L}_2^n$. Requiring this of only some derivatives (and in particular not the function itself) yields so-called *homogenous Sobolev spaces*. To introduce a concrete example, we have

$$\mathring{\mathcal{H}}_2(\mathbb{G}) := \left\{ f : \mathbb{G} \to \mathbb{C} \mid \left\| \frac{\partial^2 f}{\partial x^2} \right\|_{\mathfrak{L}_2} < \infty \right\}$$

$$\langle f, g \rangle_{\mathring{\mathcal{H}}_2} := \left\langle \frac{\partial^2 f}{\partial x^2}, \frac{\partial^2 g}{\partial x^2} \right\rangle_{\mathfrak{L}_2}.$$

The corresponding relaxation for the weighted spaces is to require the weight $W(\lambda)$ to be positive *semi*definite only. Ramifications include the non-invertibility of $W(\lambda)$ at some λ and that with respect to $\|\cdot\|_{\hat{\mathcal{H}}_2}$, all functions whose difference is linear in x are equal. Even though the results should apply with at most minor modifications, we avoid any such complications by restricting ourselves to the spaces defined in Definition 3.2.

Operators can not necessarily be defined on the whole space, and we denote the *domain* of an operator $K : H_1 \to H_2$ by $\mathfrak{D}(K) \subseteq H_1$. An operator K is *densely defined*, if $\mathfrak{D}(K)$ is dense in H_1. With every operator we associate its adjoint:

Definition 3.3. *For an operator $K : H_1 \supset \mathfrak{D}(K) \to H_2$, the* adjoint *$K^\dagger : H_2 \to H_1$ is defined by the identity*

$$\langle u, Kv \rangle_{H_2} = \left\langle K^\dagger u, v \right\rangle_{H_1} \quad \forall v \in \mathfrak{D}(K) \tag{3.1}$$

and for all u for which $\langle u, K\cdot \rangle_{H_2}$ defines a bounded (or, equivalently, continuous) linear functional on $\mathfrak{D}(K)$. $\mathfrak{D}(K^\dagger)$ is defined as the set of such u.

Note that we are using the somewhat nonstandard dagger $\cdot^\dagger$ to avoid confusion with the *complex conjugate transpose* M^* of a matrix $M \in \mathbb{C}^{n \times n}$.

If we consider one of the spaces outlined in Definition 3.2, a very important class of linear operators can be defined:

Definition 3.4. *A multiplication operator is a linear operator K with $\mathfrak{D}(K) \subseteq \mathfrak{L}_W(\mathbb{G})$ for which there exists a measurable function $K : \mathbb{G} \to \mathbb{C}^{m \times n}$ such that*

$$\left[Kf \right](\lambda) = K(\lambda) \cdot f(\lambda) \quad \forall f \in \mathfrak{D}(K).$$

The function $K(\cdot)$ is the symbol *of K. We take the liberty to denote operator and symbol with the same letter and always write the symbol with an argument, either as $K(\lambda)$ or, for the sake of readability, as K_λ.*

Many operations on a multiplication operator can then be done "pointwise," e.g. the symbol of the inverse of K is $(K_\lambda)^{-1}$; however care has to be taken with the adjoint.

Lemma 3.5 (Symbols of adjoints). *Let K be a multiplication operator with dense domain $\mathfrak{D}(K) \subseteq \mathfrak{L}_V$ and range in $\mathfrak{L}_W$. Then the symbol of the adjoint $K^\dagger$ is given by*

$$K^\dagger(\lambda) = V^{-1}(\lambda)K^*(\lambda)W(\lambda) \tag{3.2}$$

Proof. Since the weights are positive definite, we have $W_\lambda = W_\lambda^*\ \forall \lambda$. Then:

$$\langle Kf, g\rangle_{\mathfrak{L}_W} = \int_{\mathbb{G}} g^*(\lambda)W_\lambda K(\lambda)V_\lambda^{-1}V_\lambda f(\lambda)\,\mathrm{d}\lambda = \int_{\mathbb{G}}\left[V_\lambda^{-1}K_\lambda^*W_\lambda g(\lambda)\right]^* V_\lambda f(\lambda)\,\mathrm{d}\lambda$$
$$=: \left\langle f, K^\dagger g\right\rangle_{\mathfrak{L}_V}.$$

$\square$

3.1.3 Fourier Analysis on Groups

The generalized Fourier transform $\mathscr{F}$ maps functions on $\mathbb{G}$ to functions on the *dual group* $\widehat{\mathbb{G}}$ via

$$[\mathscr{F}f](\lambda) = \widehat{f}(\lambda) := \int_{\mathbb{G}} f(x)\,\lambda(-x)\,\mathrm{d}x \quad \forall \lambda \in \widehat{\mathbb{G}}, \tag{3.3}$$

where, in full generality, λ is a character of $\mathbb{G}$, i.e. a homomorphism from $\mathbb{G}$ to $\partial\mathbb{D}$, and $\widehat{\mathbb{G}}$ is the set of all characters of $\mathbb{G}$. The dual group of $\mathbb{R}$ is $\mathbb{R}$ itself, whereas the dual group of $\partial\mathbb{D}$ is $\mathbb{Z}$, the set of all integers, see also Table 3.1. For a full exposition of Fourier analysis on groups, see [Rud62].

Hence, when we consider $\mathbb{G} = \mathbb{R}$, then $\mathscr{F}$ is the common Fourier transform, and when we consider $\mathbb{G} = \partial\mathbb{D}$, then $\mathscr{F}$ denotes Fourier series expansion. When considering functions on dual groups we will talk about the *transform* or *frequency domain*, and the entity in the

Table 3.1 – *Common groups $\mathbb{G}$ and their dual groups $\widehat{\mathbb{G}}$*

$\mathbb{G}$	$\mathbb{R}$	$\partial\mathbb{D}$	$\mathbb{Z}$	$\mathbb{Z}_N$	$\mathbb{G}_1 \times \mathbb{G}_2$
$\widehat{\mathbb{G}}$	$\mathbb{R}$	$\mathbb{Z}$	$\partial\mathbb{D}$	$\mathbb{Z}_N$	$\widehat{\mathbb{G}}_1 \times \widehat{\mathbb{G}}_2$

transform domain corresponding to $\bullet$ will be denoted $\widehat{\bullet}$.

The generalized Fourier transform as an operator from $\mathcal{H}^n_k(\mathbb{G})$ into $\mathfrak{L}_W(\widehat{\mathbb{G}})$ (for appropriate choice of W, see Theorem 3.6) has appealing properties, in particular

1. $\mathcal{F}$ is linear;

2. $\mathcal{F}$ transforms many translation invariant operators – including the ones mentioned in 3.1.1 – into multiplication operators;[1] and

3. with appropriate normalization of measures $\mathrm{d}x$ and $\mathrm{d}\lambda$, $\mathcal{F}$ is an isometric isomorphism, i.e. the norm is preserved: $\|f\| = \|\widehat{f}\|$.

The last point above follows from what is known as Plancherel's theorem, an extension of which to the case of Sobolev and weighted spaces we state next:

Theorem 3.6 (Plancherel's Theorem)**.** *Let $H = \mathcal{H}_{\ell_1} \times \mathcal{H}_{\ell_2} \times \cdots \times \mathcal{H}_{\ell_n}$ be a Cartesian product of Sobolev spaces and associate with it a weight matrix W*

$$W := \mathrm{diag}\{w_{\ell_1}, \ldots, w_{\ell_n}\},$$

where the individual weights are

$$w_m(\lambda) := \sum_{\nu=0}^{m} \lambda^{2\nu}.$$

(Note that $j\lambda$ is the symbol of $\widehat{\frac{\partial}{\partial x}}$, the Fourier transform of the differential operator.) Now define

---

[1]This generally holds for *all* translation invariant *bounded* operators, see e.g. [BC49, Thm.74] and [CZ95, Thm. A.6.28], for the case of unbounded operators, however, we know of no general result.

$\widehat{H} := \mathfrak{L}_W(\widehat{\mathbb{G}})$ *and we have*

$$\langle f, g \rangle_H = \left\langle \widehat{f}, \widehat{g} \right\rangle_{\widehat{H}} \tag{3.4}$$

Proof. The case for $\mathfrak{L}_2$ spaces is well-known, see e.g. [Rud62, 1.6.1 & 1.6.2], so it is enough to reduce the present case to $\mathfrak{L}_2$. First, we note that the Cartesian product of Hilbert spaces is again a Hilbert space, if the inner product is defined as the sum of the inner products of each individual space. For us that means that

$$\langle f, g \rangle_H = \sum_{\nu=1}^n \langle f_\nu, g_\nu \rangle_{\mathcal{H}_{\ell_\nu}}, \tag{3.5}$$

if we write elements of H as $f = (f_1, \dots, f_n)$. Thus, proving the statement for $n = 1$ generalizes easily to the general result. By Definition 3.2 and Plancherel's theorem for $\mathfrak{L}_2$ spaces:

$$
\begin{aligned}
\langle f, g \rangle_{\mathcal{H}_\ell} &= \sum_{\nu=0}^{\ell} \left\langle \frac{\partial^\nu f}{\partial x^\nu}, \frac{\partial^\nu g}{\partial x^\nu} \right\rangle_{\mathfrak{L}_2} \\
&= \sum_{\nu=0}^{\ell} \left\langle \widehat{\frac{\partial^\nu f}{\partial x^\nu}}, \widehat{\frac{\partial^\nu g}{\partial x^\nu}} \right\rangle_{\mathfrak{L}_2} \\
&= \sum_{\nu=0}^{\ell} \left\langle (j\lambda)^\nu \widehat{f}, (j\lambda)^\nu \widehat{g} \right\rangle_{\mathfrak{L}_2} \\
&= \sum_{\nu=0}^{\ell} \int_{\mathbb{G}} \widehat{g}_\lambda^* (-j\lambda)^\nu (j\lambda)^\nu \widehat{f}_\lambda \, d\lambda \\
&= \int_{\mathbb{G}} \widehat{g}_\lambda^* \left(\sum_{\nu=0}^{\ell} \lambda^{2\nu} \right) \widehat{f}_\lambda \, d\lambda \qquad\qquad = \left\langle \widehat{f}, \widehat{g} \right\rangle_{\mathfrak{L}_{w_\ell}}.
\end{aligned}
$$

$\square$

To obtain an interpretation of this theorem, consider that Plancherel's theorem in its classical form (in this context also know as Parseval's identity) ensures that $f \in \mathfrak{L}_2$ implies $\widehat{f} \in \mathfrak{L}_2$ and that $\|f\|_{\mathfrak{L}_2} = \left\| \widehat{f} \right\|_{\mathfrak{L}_2}$. If f more generally lives in a Sobolev space, then the space in which its Fourier transform $\widehat{f}$ lives needs to be chosen properly to preserve this isometry property of the Fourier transform; Theorem 3.6 establishes this choice.

126

$$
\begin{array}{ccc}
\boxed{f(x)} & \xrightarrow{\quad K \quad} & [Kf](x) \\[4pt]
\mathscr{F} \downarrow & & \downarrow \mathscr{F} \\[4pt]
\widehat{f}(\lambda) & \xrightarrow{\quad \widehat{K} \quad} & \boxed{\left[\widehat{K}\,\widehat{f}\,\right](\lambda) = \widehat{Kf}(\lambda)}
\end{array}
$$

Figure 3.1 – *The definition of the operator $\widehat{K} := \mathscr{F}K\mathscr{F}^{-1}$ implies that the Fourier transform of Kf can be found by either applying K and then the Fourier transform, or applying the Fourier transform first, followed by $\widehat{K}$.*

We can then associate with every operator K on a Sobolev space H an operator $\widehat{K}$ on the transform space $\widehat{H}$ by $\widehat{K} := \mathscr{F}K\mathscr{F}^{-1}$, see Figure 3.1 for an illustration. The isometry property gained by choosing $\widehat{H}$ according to Theorem 3.6 then ensures the induced norms of both operators are equal, i.e.

$$
\|K\|_{H \to H} = \sup_{\|f\|_H = 1} \|Kf\|_H = \sup_{\|\widehat{f}\|_{\widehat{H}} = 1} \left\|\widehat{K}\,\widehat{f}\,\right\|_{\widehat{H}} = \left\|\widehat{K}\right\|_{\widehat{H} \to \widehat{H}} .
$$

For translation invariant operators that correspond to multiplication operators in the transform domain we can say even more, and the following corollary provides a characterization of the induced operator norm in terms of its symbol's spectral norm:

Corollary 3.7. *Let H and W be defined as in Theorem 3.6 and let $K : H \to H$ be a linear, spatially invariant operator from H into itself so that $\widehat{K} = \mathscr{F}K\mathscr{F}^{-1}$ is a multiplication operator. Then*

$$
\|K\|_{H \to H} = \sup_{\lambda \in \widehat{\mathbb{G}}} \sqrt{\sigma_{max}\left(W_\lambda^{1/2}\widehat{K}(\lambda)W_\lambda^{-1/2}\right)}. \tag{3.6}
$$

Here, $\sigma_{max}(\cdot)$ denotes the maximal singular value, and $W^{1/2}$ corresponds to the upper triangular matrix that satisfies $W = (W^{1/2})^*W^{1/2}$. This Cholesky decomposition exists for all positive definite matrices W.

Proof. We write the Cholesky decomposition as $W_\lambda = R_\lambda^* R_\lambda$. Then

$$
\begin{aligned}
\|K\|_{H\to H}^2 &= \sup_{\|f\|=1} \langle Kf, Kf\rangle_H = \sup_{\|\widehat{f}\|=1} \left\langle \widehat{K}\widehat{f}, \widehat{K}\widehat{f}\right\rangle_{\widehat{H}} \\
&= \sup \int_{\widehat{\mathbb{G}}} \widehat{f}_\lambda^* R_\lambda^* R_\lambda^{-*} \widehat{K}_\lambda^* W_\lambda \widehat{K}_\lambda R_\lambda^{-1} R_\lambda \widehat{f}_\lambda \, \mathrm{d}\lambda \\
&= \sup \int_{\widehat{\mathbb{G}}} \widehat{f}_\lambda^* R_\lambda^* \left(R_\lambda \widehat{K}_\lambda R_\lambda^{-1}\right)^* \left(R_\lambda \widehat{K}_\lambda R_\lambda^{-1}\right) R_\lambda \widehat{f}_\lambda \, \mathrm{d}\lambda \\
&\leq \sup \sigma_{\max}\left(R_\lambda \widehat{K}_\lambda R_\lambda^{-1}\right) \|\widehat{f}\|_{\widehat{H}}^2.
\end{aligned}
$$

It then is a standard argument to show that for any $\varepsilon > 0$, there is a function $\widehat{g}_\varepsilon$ with $\|g_\varepsilon\|_H = 1$ which achieves above bound to within ε. $\qquad\square$

3.1.4 Spatial Invariance, State-Space Description

The class of problems considered here is that of linear spatio-temporal systems, which will be represented in their *state-space* or *evolution form*

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} \Psi(x,t) &= [A\Psi](x,t) + [Bu](x,t) \\
\Psi(\cdot, 0) &= \Psi_0,
\end{aligned}
\tag{3.7}
$$

where $\Psi(\cdot, t)$, $u(\cdot, t)$ live on the Sobolev spaces H and $\mathcal{U}$, respectively; W and V will be the spatial weights on their transform spaces $\widehat{H}$ and $\widehat{\mathcal{U}}$ according to Theorem 3.6. Time is denoted by t, while the variable $x \in \mathbb{G}$ collects all the spatial coordinates; hence, "translation invariance with respect to $\mathbb{G}$" and *spatial invariance* can be (and are) used synonymously.

Since we are dealing with infinite-dimensional state and input spaces, linear operators are not necessarily bounded, and so the notion of "solution" to (3.7) requires some care. It shall suffice here to say that a common notion of "well-posedness" is to require that A generate a so-called C_0 semigroup $T(t)$. To contrast that with the finite-dimensional case: matrices $A, B \in \mathbb{R}^{\bullet \times \bullet}$

are always bounded linear operators, whose domain is the entire space, and the generated semigroup is $T(t) = e^{At}$. To collect all assumptions on involved operators, spaces and the like in one place, we state

Assumption 3.8.

1. *All operators are* spatially invariant*;*

2. *all spatially invariant operators correspond to* multiplication operators *in the transform domain;*

3. *all corresponding multiplication operators have* continuous symbols*;*

4. *in systems of the form* (3.7)*, A generates a C_0* semigroup *on the chosen state-space;*

5. *additionally, in* (3.7)*, B is a* bounded operator*; and*

6. *all function spaces are (Cartesian products of)* Sobolev spaces *in the spatial domain and consequently are* weighted spaces *in the transform domain, chosen according to Theorem 3.6.*

3.2 Stability, Stabilizability and Linear-Quadratic Regulation (LQR)

Restricting our treatment to problems as laid out in Section 3.1.4 enables us to make several statements about linear systems on infinite-dimensional Sobolev spaces that parallel well-known results for finite-dimensional systems; it will turn out, however, that instead of dealing with a single matrix equation like in the finite-dimensional case, we will have to consider a whole family of such equations, indexed by the dual group $\widehat{\mathbb{G}}$ of the spatial coordinate system.

3.2.1 Exponential Stability

Stability is a property of the autonomous system

$$\frac{\mathrm{d}}{\mathrm{d}t}\Psi(x,t) = A\Psi(x,t), \quad \Psi(\cdot,0) = \Psi_0. \tag{3.8}$$

While for systems on a finite-dimensional state space most notions of stability are equivalent, the case is different for the infinite-dimensional case, and one has to pick a certain definition of stability. The following notion of stability is one of the most useful ones, and the one we will be considering here:

Definition 3.9 (Exponential stability)**.** *The autonomous system* (3.8) *is* exponentially stable*, if there exist constants $M, \alpha > 0$ such that*

$$\|T(t)\| \leq Me^{-\alpha t} \qquad \forall t \geq 0,$$

where $T(t)$ is the semigroup generated by A.

In the present case of spatially invariant operators, exponential stability can be checked by considering a family of Lyapunov equations:

Theorem 3.10 (Lyapunov test for stability)**.** *Assume that A in* (3.8) *generates a C_0 semigroup $T(t)$ on the Sobolev Space H. Then, $T(t)$ is exponentially stable, if and only if for all $\lambda \in \widehat{\mathbb{G}}$, $\widehat{A}(\lambda)$ is stable and for the positive definite solution $\widehat{P}_\lambda$ of the Lyapunov equation*

$$\widehat{P}_\lambda \widehat{A}_\lambda + \widehat{A}_\lambda^* \widehat{P}_\lambda + W_\lambda = 0, \tag{3.9}$$

we have

$$\sup_{\lambda \in \widehat{\mathbb{G}}} \ \sigma_{max}\left(W^{-1/2}(\lambda)\widehat{P}_\lambda W^{-1/2}(\lambda)\right) < \infty, \tag{3.10}$$

where $W(\lambda)$ is as in Theorem 3.6.

Proof. It is well established, see e.g. [Ban83, Lemma 4.3.3] or [PZ81, Thm. 3.4], that exponential stability of $T(t)$ is equivalent to existence of a bounded positive definite linear operator Π solving the Lyapunov equation

$$\langle A\Psi, \Pi\Psi \rangle_H + \langle \Pi\Psi, A\Psi \rangle_H = -\langle \Psi, \Psi \rangle_H \ \ \forall \Psi \in \mathfrak{D}(A). \tag{3.11}$$

It should also be clear that such a Π has to be spatially invariant itself.[2] Then, (3.11) can be rewritten as

$$0 = \left\langle \widehat{A\Psi}, \widehat{\Pi\Psi} \right\rangle_{\widehat{H}} + \left\langle \widehat{\Pi\Psi}, \widehat{A\Psi} \right\rangle_{\widehat{H}} + \left\langle \widehat{\Psi}, \widehat{\Psi} \right\rangle_{\widehat{H}} = \int_{\widehat{\mathbb{G}}} \widehat{\Psi}_\lambda^* \left(\widehat{A}_\lambda^* W_\lambda \widehat{\Pi}_\lambda + \widehat{\Pi}_\lambda^* W_\lambda \widehat{A}_\lambda + W_\lambda \right) \widehat{\Psi}_\lambda \, \mathrm{d}\lambda,$$

which, since $\mathfrak{D}(A)$ is dense, is equivalent to

$$\widehat{A}_\lambda^* W_\lambda \widehat{\Pi}_\lambda + \widehat{\Pi}_\lambda^* W_\lambda \widehat{A}_\lambda + W_\lambda = 0 \quad \forall \lambda \in \widehat{\mathbb{G}}.$$

Now define $\widehat{P}_\lambda := W_\lambda \widehat{\Pi}_\lambda$ and note that $\Pi^\dagger = \Pi$ iff $\widehat{\Pi}^\dagger = \widehat{\Pi}$, hence by Lemma 3.5 it follows that $\widehat{P}_\lambda^* = \widehat{P}_\lambda$ iff $\Pi^\dagger = \Pi$ and we obtain (3.9). From Corollary 3.7 we see that (3.10) is equivalent to boundedness of Π. $\qquad\square$

3.2.2 Exponential Stabilizability

A closely related question to that answered above is whether a system with input (3.7) can be made into an exponentially stable autonomous one by a spatially invariant feedback law of the form $u(x, t) = -F\Psi(x, t)$.

Definition 3.11 (Exponential stabilizability)**.** *The system* (3.7) *is* exponentially stabilizable, *if there exists a bounded operator* $F : H \to \mathfrak{U}$ *such that* $A - BF$ *generates an exponentially stable* C_0 *semigroup on* H.

---

[2]It can be shown that, if it exists, Π has the form $\Pi\Psi = \int_0^\infty T^\dagger(t)T(t)\Psi \, dt$.

Analogous to Theorem 3.10, a family of equations can be checked to ascertain exponential stabilizability:

Theorem 3.12 (Stabilizability)**.** *Assume that in a system* (3.7)*, A is the generator of a C_0 semigroup $T(t)$ on some Sobolev space H. Then, the system is exponentially stabilizable, if and only if for all $\lambda \in \widehat{\mathbb{G}}$, the pair $(\widehat{A}_\lambda, \widehat{B}_\lambda)$ is stabilizable and the positive definite solution $\widehat{P}_\lambda$ of the Riccati equation*

$$\widehat{P}_\lambda \widehat{A}_\lambda + \widehat{A}_\lambda^* \widehat{P}_\lambda + W_\lambda - \widehat{P}_\lambda \widehat{B}_\lambda V_\lambda^{-1} \widehat{B}_\lambda^* \widehat{P}_\lambda = 0, \tag{3.12}$$

where W and V are the weights on the spaces $\widehat{H}$ and $\widehat{\mathcal{U}}$ and defined as in Theorem 3.6, respectively, satisfies (3.10)*.*

Proof. Sufficiency: Assume, a $\widehat{P}_\lambda$ as required exists. Then let $\widehat{F}_\lambda := V_\lambda^{-1} \widehat{B}_\lambda^* \widehat{P}_\lambda / 2$ and do some algebra to obtain

$$(\widehat{A}_\lambda - \widehat{B}_\lambda \widehat{F}_\lambda)^* \widehat{P}_\lambda + \widehat{P}_\lambda (\widehat{A}_\lambda - \widehat{B}_\lambda \widehat{F}_\lambda) + W_\lambda = 0,$$

i.e. by Theorem 3.10, $A - BF$ generates an exponentially stable C_0 semigroup.

Necessity: This follows e.g. from [Zab75, Thm. 2^0], or as a corollary to Theorem 3.13 below, with Q and R the identity and the observation that $(A^\dagger, I)$ is always exponentially stabilizable.

$\square$

3.2.3 LQR Control

Lastly, if we find that a system is indeed stabilizable, which particular feedback operator F should we pick? A possible answer to that is: the operator F that minimizes a quadratic cost functional

$$J(\Psi_0, u) = \int_0^\infty \langle Q\Psi, \Psi \rangle_H + \langle Ru, u \rangle_{\mathcal{U}} \, dt, \tag{3.13}$$

where R is a positive definite operator, i.e. $\langle Ru, u \rangle_{\mathcal{U}} > 0$ for all nonzero u, and Q is a positive semi-definite operator, i.e. $\langle Q\Psi, \Psi \rangle_H \geq 0$. Analogous to finite dimensions, the optimal F

corresponds to a solution of a certain Riccati equation, only now it depends on the parameter $\lambda \in \widehat{\mathbb{G}}$.

Theorem 3.13 (LQR). *Consider system (3.7), where A is the generator of a C_0 semigroup $T(t)$ on a Sobolev space H, and the cost functional (3.13). Then we have:*

If (A, B) is exponentially stabilizable, then the Riccati equation

$$\widehat{P}_\lambda \widehat{A}_\lambda + \widehat{A}_\lambda^* \widehat{P}_\lambda + W_\lambda \widehat{Q}_\lambda - \widehat{P}_\lambda \widehat{B}_\lambda \widehat{R}_\lambda^{-1} V_\lambda^{-1} \widehat{B}_\lambda^* \widehat{P}_\lambda = 0 \qquad (3.14)$$

has a unique solution $\widehat{P}_\lambda = \widehat{P}_\lambda^*$. $\widehat{P}_\lambda$ satisfies (3.10), i.e. it defines a bounded operator Π with Fourier symbol $\widehat{\Pi}_\lambda = W_\lambda^{-1} \widehat{P}_\lambda$ which is such that the feedback law $u = -R^{-1} B^\dagger \Pi \Psi$ minimizes J for every Ψ_0.

If additionally $(A^\dagger, Q^{1/2})$ is exponentially stabilizable, then $A - BR^{-1}B^\dagger\Pi$ generates an exponentially stable C_0 semigroup, i.e. the minimizing feedback law is also stabilizing.

Proof. The first part follows from e.g. [Gib79, Thm. 4.11], where the operator Π is obtained as the solution of the operator Riccati equation

$$A^\dagger \Pi + \Pi A - \Pi B R^{-1} B^\dagger \Pi + Q = 0.$$

Associating every operator with its Fourier transform yields

$$\mathscr{F}^{-1} \left(\widehat{A^\dagger} \widehat{\Pi} + \widehat{\Pi} \widehat{A} - \widehat{\Pi} \widehat{BR^{-1}} \widehat{B^\dagger} \widehat{\Pi} + \widehat{Q} \right) \mathscr{F} = 0,$$

and after setting $\widehat{P}_\lambda = W_\lambda \widehat{\Pi}_\lambda$ this is equivalent to (3.14). See also the proof of Theorem 3.10. The second part is essentially in [CP78, Cor. 4.17], but it should be intuitively clear that any observable[3] unstable modes will lead to J diverging to infinity, whereas unstable but unobservable modes will allow for finite cost, even if the state diverges. $\qquad \square$

---

[3]Stabilizability of $(A^\dagger, Q^{1/2})$ is equivalent to detectability of $(A, Q^{1/2})$.

Remark. Once the symbol $\widehat{F}(\lambda)$ of the optimal feedback operator is computed, the question of implementing the optimal controller F in the spatial domain arises. The answer lies in the fact that all multiplication operators we encounter here correspond to spatial convolutions,[4] and the spatial convolution kernel k of an operator $K = \mathscr{F}^{-1}\widehat{K}\mathscr{F}$ with $\widehat{K}$ a multiplication operator corresponds to the inverse Fourier transform of the symbol $\widehat{K}(\lambda)$: $k(x) = [\mathscr{F}^{-1}\widehat{K}(\cdot)](x)$.
To see that, consider for simplicity operators K and $\widehat{K}$ with spatial and transform domain both just $\mathfrak{L}_2(\mathbb{R})$. Then, for any $g \in \mathfrak{L}_2(\mathbb{R})$:

$$[Kg](x) = [\mathscr{F}^{-1}\widehat{K}\mathscr{F}g](x) = \int_{\mathbb{R}} e^{\lambda x}\, \widehat{K}(\lambda) \int_{\mathbb{R}} e^{-\lambda\xi} g(\xi)\,\mathrm{d}\xi\,\mathrm{d}\lambda =$$

$$\iint e^{\lambda(x-\xi)}\widehat{K}(\lambda)\,\mathrm{d}\lambda\, g(\xi)\,\mathrm{d}\xi = \int \overbrace{[\mathscr{F}^{-1}\widehat{K}(\lambda)](x-\xi)}^{=k(x-\xi)} g(\xi)\,\mathrm{d}\xi = [k(\cdot) * g(\cdot)](x),$$

thus K is a convolution operator with convolution kernel $k(\cdot)$ equal to the inverse Fourier transform of $\widehat{K}(\cdot)$.

3.3   Examples

In this final section, two examples of wave and beam equations are presented and analyzed using the framework developed in the prequel. For these examples we find that optimal LQR feedback gains are indeed spatially localized in the sense that their convolution kernels decay exponentially, contradicting claims made in the literature [Cur11, Sec. IV] about these examples. This indicates that the results on the locality of quadratically optimal controllers stated in [BPD02] for systems over $\mathfrak{L}_2$ may indeed be valid for the more general setting of Sobolev spaces. These generalizations will be reported elsewhere.

---

[4]We have to allow for distributions in the convolution kernel, e.g. $[T_\xi f](x) = \int_{\mathbb{G}} \delta(x - \xi - y)f(y)\,\mathrm{d}y$, hence $T_\xi f = \delta(\cdot - \xi) * f$.

3.3.1 Wave Equation

This example is taken from [Cur11]. Consider the following damped wave equation on the real line, i.e. $\mathbb{G} = \mathbb{R}$, with fully distributed input:

$$\frac{\partial^2 \phi}{\partial t^2} + \alpha \frac{\partial \phi}{\partial t} - \frac{\partial^2 \phi}{\partial x^2} = u(x,t). \tag{3.15}$$

We write it in a state-space form like (3.7) by defining $\Psi^T := [\phi \quad \frac{\partial \phi}{\partial t}]$, which leads to

$$\frac{\mathrm{d}}{\mathrm{d}t} \Psi = \begin{bmatrix} 0 & 1 \\ \frac{\partial^2}{\partial x^2} & -\alpha \end{bmatrix} \Psi + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u =: A\Psi + Bu. \tag{3.16}$$

It is well-known (see e.g. [Cur11]) that, if one chooses this realization, A generates a C_0 semigroup not on $\mathfrak{L}_2^2(\mathbb{R})$ but on $H := \mathcal{H}_1(\mathbb{R}) \times \mathfrak{L}_2(\mathbb{R})$.[5] The input space $\mathcal{U}$ is $\mathfrak{L}_2(\mathbb{G})$. Since $\frac{\partial}{\partial x}$ is a spatially invariant operator with symbol $j\lambda$, after taking Fourier transforms the system takes the form of an infinite family of decoupled ordinary differential equations

$$\frac{\mathrm{d}}{\mathrm{d}t} \widehat{\Psi} = \begin{bmatrix} 0 & 1 \\ -\lambda^2 & -\alpha \end{bmatrix} \widehat{\Psi} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \widehat{u} =: \widehat{A}\widehat{\Psi} + \widehat{B}\widehat{u}. \tag{3.17}$$

From Theorem 3.6 we know that the proper choice of state-space for (3.17) is $\widehat{H} := \mathfrak{L}_{1+\lambda^2}(\mathbb{R}) \times \mathfrak{L}_2(\mathbb{R}) = \mathfrak{L}_W^2(\mathbb{R})$, where $W(\lambda) = \mathrm{diag}\{1 + \lambda^2, 1\}$.

It is easy to see that $\widehat{A}$ does not generate an exponentially stable C_0 semigroup, since $\widehat{A}(0)$ is not stable, so it makes sense to solve an LQR problem. Like in [Cur11], we choose the matrices Q and R in (3.13) to be the identity on the respective space. Then, applying Theorem 3.13, we need to solve the Riccati equation

$$\widehat{P}_\lambda \widehat{A}_\lambda + \widehat{A}_\lambda^* \widehat{P}_\lambda + W_\lambda - \widehat{P}_\lambda \widehat{B}\widehat{B}^* \widehat{P}_\lambda = 0, \tag{3.18}$$

---

[5]Instead of $\mathcal{H}_1$, the slightly less restrictive $\mathring{\mathcal{H}}_1$ can be chosen, see the remark made below Definition 3.2.

which has the solution

$$\widehat{P}_\lambda = \begin{bmatrix} f(\lambda)g(\lambda) - \alpha\lambda^2 & f(\lambda) - \lambda^2 \\ f(\lambda) - \lambda^2 & g(\lambda) - \alpha \end{bmatrix},$$ (3.19)

where

$$f(\lambda) := \sqrt{1 + \lambda^2 + \lambda^4}$$

$$g(\lambda) := \sqrt{\alpha^2 - 2\lambda^2 + 2f(\lambda) + 1}.$$

In order to check that Π with $\widehat{\Pi}_\lambda = W_\lambda^{-1}\widehat{P}_\lambda$ defines a bounded operator on H, according to Corollary 3.7 we have to make sure the supremum over $\widehat{\mathbb{G}} = \mathbb{R}$ of the largest singular value of the matrix

$$W_\lambda^{1/2}\widehat{\Pi}_\lambda W_\lambda^{-1/2} = W_\lambda^{-1/2}\widehat{P}_\lambda W_\lambda^{-1/2} = \begin{bmatrix} \frac{-\alpha\lambda^2 + f(\lambda)g(\lambda)}{\lambda^2 + 1} & \frac{f(\lambda) - \lambda^2}{\sqrt{\lambda^2 + 1}} \\ \frac{f(\lambda) - \lambda^2}{\sqrt{\lambda^2 + 1}} & -\alpha + g(\lambda) \end{bmatrix}$$

is finite. With a little algebra it can be seen that the elements are all bounded – e.g. the off-diagonal elements are equal to

$$\frac{\lambda^2 + 1}{(\lambda^2 + f(\lambda))/\sqrt{\lambda^2 + 1}},$$

which is clearly bounded. Since from boundedness of the elements, it follows that e.g. the Frobenius norm is bounded, and since all matrix norms are equivalent, we can conclude that $\|\Pi\| = \left\|\widehat{\Pi}\right\| = < \infty$.

We can now compute the symbol of the optimal feedback operator $F = R^{-1}B^\dagger\Pi$:

$$\widehat{F}_\lambda = \widehat{B}_\lambda^*\widehat{P}_\lambda = \begin{bmatrix} -\lambda^2 + f(\lambda) & -\alpha + g(\lambda) \end{bmatrix} = \begin{bmatrix} \frac{\lambda^2 + 1}{\lambda^2 + f(\lambda)} & \frac{1}{\alpha + g(\lambda)}\left(1 + 2\frac{\lambda^2 + 1}{\lambda^2 + f(\lambda)}\right) \end{bmatrix}.$$

Lastly, we can confirm that Theorem 3.10 is valid here, too. By construction, and since $(A^\dagger, I)$ is always exponentially stabilizable, we know that the closed loop operator $A_{cl} = A - BF$ must

generate an exponentially stable semigroup. Its symbol is

$$\widehat{A}_{cl}(\lambda) = \begin{bmatrix} 0 & 1 \\ -f(\lambda) & -g(\lambda) \end{bmatrix}$$

and we seek a positive definite solution $\widehat{Q}(\lambda)$ of

$$\widehat{Q}_\lambda \widehat{A}_{cl}(\lambda) + \widehat{A}_{cl}^*(\lambda)\widehat{Q}_\lambda + W_\lambda = 0,$$

which corresponds to (3.9). The solution is

$$\widehat{Q}(\lambda) = \begin{bmatrix} q_1(\lambda) & p(\lambda) \\ p(\lambda) & q_2(\lambda) \end{bmatrix}$$

$$p(\lambda) = \frac{\lambda^2 + 1}{2f(\lambda)}$$

$$q_2(\lambda) = \frac{1 + 2p(\lambda)}{2g(\lambda)}$$

$$q_1(\lambda) = f(\lambda)q_2(\lambda) + g(\lambda)p(\lambda).$$

Since for all $\lambda \in \mathbb{R}$ we have $p(\lambda) > 0$ and

$$\det\left(\widehat{Q}(\lambda)\right) = f(\lambda)q_2^2(\lambda) + \frac{p(\lambda)}{2} > 0,$$

all leading principal minors of $\widehat{Q}$ are positive, which implies positive definiteness. With similar algebra as above, it can also be seen that the largest singular value of $W_\lambda^{-1/2}\widehat{Q}_\lambda W_\lambda^{-1/2}$ is bounded, and so Theorem 3.10 is confirmed.

In order to see whether the statements about locality from [BPD02] hold in this case, too, we will need to obtain the convolution kernel $F(\cdot)$ to represent F in the form

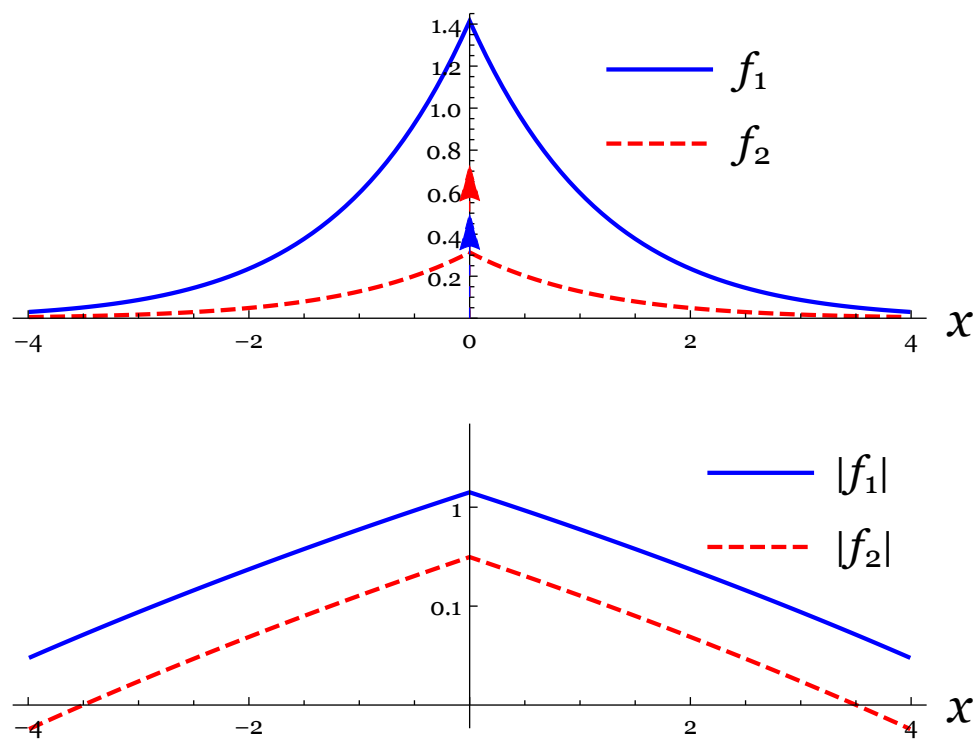$$[F\Psi](x) = \int_{\mathbb{R}} F(x - \xi)\Psi(\xi)\,d\xi.$$

Figure 3.2 – *The elements of the convolution kernel F(x) for the wave equation example in linear and logarithmic plots. In the linear plot, the arrows represent Dirac δ functions; note the exponential decay, seen clearly in the logarithmic plot.*

Following the remark below Theorem 3.13, $F(x)$ is given by the inverse Fourier transform of $\widehat{F}_\lambda$

$$F(x) = \int_{\widehat{\mathbb{G}}} \widehat{F}_\lambda e^{j\lambda x} \, \mathrm{d}\lambda. \tag{3.20}$$

For the example at hand, (3.20) needs to be evaluated numerically for every x. Doing so yields what is shown in Figure 3.2, where it is clearly seen that the absolute value of both components of the convolution kernel decay rapidly with the size of the argument x. The interpretation of this is that to compute the control input at a location x_o, information from other locations becomes exponentially less important as the distance to x_o increases; it then would make sense to truncate the kernel to a function with finite support, which amounts to discarding the information from locations farther than a distance Δ from x_o.

3.3.2 Beam Equation

In [CR10], the following beam equation is considered:

$$\frac{\partial^2 \phi}{\partial t^2} - \gamma \frac{\partial^3 \phi}{\partial^2 x \partial t} + \frac{\partial^4 \phi}{\partial x^4} = \int_{\mathbb{R}} b(x - \xi) u(\xi, t) \, \mathrm{d}\xi, \tag{3.21}$$

where we have to require that $\int_{\mathbb{R}} |b(x)| \, \mathrm{d}x < \infty$. For this system, an appropriate state space is $H = \mathcal{H}_2(\mathbb{R}) \times \mathcal{L}_2(\mathbb{R})$, see e.g. [CR10], and the input space is chosen to be $\mathcal{U} = \mathcal{L}_2(\mathbb{R})$. This again is easily rewritten as a state-space system

$$\frac{\mathrm{d}\Psi}{\mathrm{d}t} = \begin{bmatrix} 0 & 1 \\ -\frac{\partial^4}{\partial x^4} & \gamma \frac{\partial^2}{\partial x^2} \end{bmatrix} \Psi + Bu,$$

where

$$[Bu](x) := \begin{bmatrix} 0 \\ [b(\cdot) * u(\cdot)](x) \end{bmatrix},$$

139

and Fourier transformed to a system on $\widehat{H} = \mathfrak{L}_W(\mathbb{R})$ with $W = \mathrm{diag}\{1 + \lambda^2 + \lambda^4, 1\}$:

$$\frac{\mathrm{d}\widehat{\Psi}}{\mathrm{d}t} = \begin{bmatrix} 0 & 1 \\ -\lambda^4 & -\gamma\lambda^2 \end{bmatrix} \widehat{\Psi} + \begin{bmatrix} 0 \\ \widehat{b}(\lambda) \end{bmatrix} \widehat{u} =: \widehat{A}\widehat{\Psi} + \widehat{B}\widehat{u}. \tag{3.22}$$

As in the prior example, $\widehat{A}$ is not stable for $\lambda = 0$; in addition it is obvious in the transform domain description, that $\left(\widehat{A}(0), \widehat{B}(0)\right)$ also is not stabilizable, unless $\widehat{b}(0) \neq 0$. This requirement corresponds to the convolution kernel $b(\cdot)$ having nonzero mean, which is not at all obvious if one considers only (3.21).

Under this constraint, we can attempt an LQR controller design. For simplicity, Q and R are chosen to be the identities on H and $\mathcal{U}$. Then, solving (3.14) yields

$$\widehat{P}(\lambda) = \frac{1}{\widehat{b}^2(\lambda)} \begin{bmatrix} h(\lambda)e(\lambda) - \gamma\lambda^6 & h(\lambda) - \lambda^4 \\ h(\lambda) - \lambda^4 & e(\lambda) - \gamma\lambda^2 \end{bmatrix}, \tag{3.23}$$

where

$$h(\lambda) := \sqrt{\left(\lambda^4 + \lambda^2 + 1\right)\widehat{b}^2(\lambda) + \lambda^8}$$

$$e(\lambda) := \sqrt{\left(\gamma^2 - 2\right)\lambda^4 + \widehat{b}^2(\lambda) + 2h(\lambda)}.$$

By Theorem 3.13, this defines a stabilizing feedback if the system is stabilizable. To establish stabilizability, on the other hand, we have to return to Theorem 3.12. Due to our choice of weights Q and R, the Riccati equations from Theorems 3.12 and 3.13, (3.12) and (3.14), are equal, so applying Theorem 3.12 means checking boundedness of

$$\sup_{\lambda \in \widehat{\mathbb{G}}} \sigma_{\max}\left(W^{-1/2}(\lambda)\widehat{P}(\lambda)W^{-1/2}(\lambda)\right) < \infty,$$

which is more challenging than in the previous example, and the algebra has been moved to Appendix 3.A. There, we show that above supremum is indeed finite, and thus the system
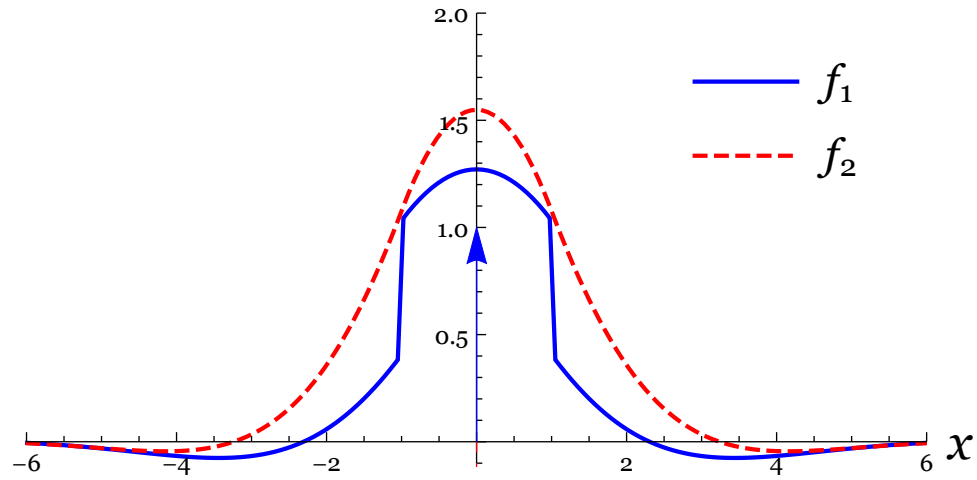
140

Figure 3.3 – *Linear plot of convolution kernel $F(x)$ for the beam equation example; $b(x)$ is a simple rectangular window, i.e. $\widehat{b}(\lambda) = \text{sinc}(\lambda)$. As before, the arrows denote Dirac δ functions. Note the discontinuity at $x = \pm 1$, the edges of the window.*
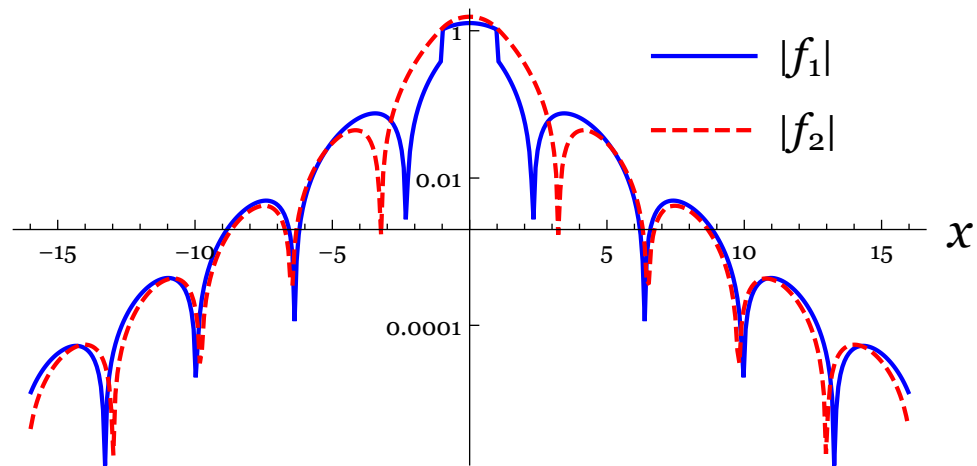

Figure 3.4 – *Logarithmic plot of the convolution kernel $F(x)$ for the beam equation example. Note again the exponential decay.*

is stabilizable, and thus the operator with the symbol $\widehat{\Pi}_\lambda = W_\lambda^{-1} \widehat{P}_\lambda$ yields an exponentially stabilizing feedback. The optimal feedback operator has the symbol

$$\widehat{F}_\lambda = \widehat{B}_\lambda^* \widehat{P}_\lambda = \frac{1}{\widehat{b}(\lambda)} \left[\lambda^4 - h(\lambda) \quad \gamma \lambda^2 - e(\lambda) \right],$$

and as before, we would like to check the locality of its kernel. In order to do so, we pick the convolution kernel b to be a rectangular window, i.e. $b(x) = 1$ if $x \in [-1, 1]$, and $b(x) = 0$ else. Two things stand out: In Fig. 3.3, it appears that the convolution kernel of the feedback operator "inherits" the discontinuities of the convolution kernel of the input operator B; in Fig. 3.4, it can be seen that the kernel can be bounded from above by a straight line, which indicates exponential or even more rapid decay. These observations are qualitatively the same if other windows are considered; we have looked at e.g. $b(x) = \delta(x)$, which corresponds to $B = [0 \ 1]^T$, at cosine, and at Tukey windows.

3.4 Summary and Future Directions

This chapter has shown how spatially invariant systems theory can be readily generalized to situations where the distributed systems realizations are chosen over a Sobolev space. The proper statements of conditions for stability, stabilizability, and solvability of LQR problems are made in terms of standard conditions on parametrized families of finite-dimensional systems along with a boundedness condition, analogously to [BPD02].

The distinction between the case of the standard $\mathfrak{L}_2$ state space treated there and the present Sobolev case is that additional proper accounting of the spatial frequency weights needs to be done. The main contribution here is the derivation of simple procedures for doing so.

Analysis of wave and beam equation examples using the developed framework shows that their optimal LQR feedback gains are indeed spatially localized in the sense that their convolution

kernels decay exponentially, contradicting claims made in [Cur11, Sec. IV] about these examples.

The examples indicate that the results on the locality of optimal quadratic controllers stated in [BPD02] for systems over $\mathfrak{L}_2$ may indeed be valid for the more general setting of Sobolev spaces, and proving a general statement about localization of controller kernels is the subject of ongoing research.

The requirement for the spatial domain to be spatially invariant restricts the approach to systems on either infinite ($\mathbb{R}$ or $\mathbb{Z}$) or periodic ($\partial \mathbb{D}$ or $\mathbb{Z}_N$) domains and hence severely limits the number of practical applications. Efforts have been made to overcome such restrictions e.g. by embedding a system on a finite spatial domain into a spatially invariant domain [DD03], and more research in this direction is certainly warranted.

Appendix 3.A   Algebra Pertinent to Section 3.3.2

We will prove here, that $W_\lambda^{-1/2}\widehat{P}_\lambda W_\lambda^{-1/2}$ has bounded entries. For convenience, we let $f(\lambda) :=$ $1 + \lambda^2 + \lambda^4$ and omit the argument λ on $\widehat{b}, h, e$, and f. Then we get

$$W_\lambda^{-1/2}\widehat{P}_\lambda W_\lambda^{-1/2} = \begin{bmatrix} \frac{eh-\gamma\lambda^6}{f\,\widehat{b}^2} & \frac{h-\lambda^4}{\sqrt{f}\,\widehat{b}^2} \\ \frac{h-\lambda^4}{\sqrt{f}\,\widehat{b}^2} & \frac{e-\gamma\lambda^2}{\widehat{b}^2} \end{bmatrix}.$$

If we denote the elements by $\begin{bmatrix} a & b \\ b & d \end{bmatrix}$, we can rewrite them as

$$a = \frac{1}{f(eh + \gamma\lambda^6)}\left((\gamma^2 - 2)\lambda^4 f + 2hf + e^2 + \frac{2f\lambda^8}{\lambda^4 + h}\right)$$

$$b = \sqrt{f}/(h + \lambda^4)$$

$$d = \left(1 + 2\frac{f}{h + \lambda^4}\right)/(e + \gamma\lambda^2).$$

143

We note that for $\lambda \neq 0$, the denominators are all positive, hence there are no poles except for possibly at zero. At $\lambda = 0$, all denominators are $\widehat{b}(0)$, but we have required $\widehat{b}(0) \neq 0$, see the remark after (3.22). To assess the behavior as $\lambda \to \pm\infty$, we collect the highest powers of λ in each term and obtain

$$a \longrightarrow \frac{\gamma^2 + 1}{2\gamma\lambda^2}$$
$$b \longrightarrow \frac{1}{2\lambda^2}$$
$$c \longrightarrow \frac{1}{\gamma\lambda^2}.$$

Thus, all elements are bounded, from which it follows easily that e.g. the Frobenius norm of the matrix is bounded. Since all matrix norms are equivalent, it follows that $\widehat{\Pi}_\lambda = W_\lambda^{-1}\widehat{P}_\lambda$ does indeed define a bounded operator.

Bibliography

[AFR+00]   A. M. Annaswamy, M. Fleifil, J. W. Rumsey, R. Prasanth, J. P. Hathout, and A. F. Ghoniem. Thermoacoustic instability: Model-based optimal control designs and experimental validation. *IEEE Transactions on Control Systems Technology*, 8(6):905–918, 2000.

[BAKJ04]   Andrzej Banaszuk, Kartik B. Ariyur, Miroslav Krstić, and Clas A. Jacobson. An adaptive algorithm for control of combustion instability. *Automatica*, 40(11):1965–1972, 2004.

[Ban83]   Stephen P. Banks. *State-Space and Frequency-Domain Methods in the Control of Distributed Parameter Systems*. Institution of Electrical Engineers, Stevenage, UK, 1983.

[Bay93]   D. S. Bayard. Statistical plant set estimation using Schroeder-phased multisinusoidal input design. *Applied Mathematics and Computation*, 58(2):169–198, 1993.

[BC49]   S. Bochner and K. Chandrasekharan. *Fourier Transforms*. Annals of mathematics studies. Princeton University Press, 1949.

[BD03]   Bassam Bamieh and Mohammed Dahleh. Exact computation of traces and $\mathcal{H}^2$ norms for a class of infinite-dimensional problems. *Automatic Control, IEEE Transactions on*, 48(4):646–649, 2003.

[BDPDM07] Alain Bensoussan, Giuseppe Da Prato, Michel C. Delfour, and Sanjoy K. Mitter. *Representation and control of infinite dimensional systems*. Birkhäuser Boston, 2007.

[BFG73]   S. Bittanti, G. Fronza, and G. Guardabassi. Periodic control: A frequency domain approach. *IEEE Transactions on Automatic Control*, 18(1):33–39, 1973.

[BGR72]   U. Bertelè, G. Guardabassi, and S. Ricci. Suboptimal periodic control: A describing function approach. *IEEE Transactions on Automatic Control*, 17(3):368–370, 1972.

[BH71]   J. E. Bailey and F. J. M. Horn. Comparison between two sufficient conditions for improvement of an optimal steady-state process by periodic operation. *Journal of Optimization Theory and Applications*, 7(5), 1971.

[BPD02]   Bassam Bamieh, Fernando Paganini, and Munther A. Dahleh. Distributed control of spatially invariant systems. *Automatic Control, IEEE Transactions on*, 47(7):1091–1107, 2002.

[BS00]   S. Backhaus and G. W. Swift. A thermoacoustic-Stirling heat engine: Detailed study. *The Journal of the Acoustical Society of America*, 107(6):3148–3166, 2000.

[Bur92]   J. C. Burgess. Chirp design for acoustical system identification. *The Journal of the Acoustical Society of America*, 91(3):1525, 1992.

[Can02]   Sébastien Candel. Combustion dynamics and control: Progress and challenges. *Proceedings of the Combustion Institute*, 29(1):1–28, 2002.

[CP78]   Ruth F. Curtain and Anthony J. Pritchard. *Infinite Dimensional Linear Systems Theory*, volume 8. Springer-Verlag New York, 1978.

[CR10]   Ruth F. Curtain and Leiba Rodman. Analytic solutions of matrix riccati equations with analytic coefficients. *SIAM Journal on Matrix Analysis and Applications*, 31(4):2075–2092, July 2010.

[Cur11]   Ruth F. Curtain. Comments on "Distributed control of spatially invariant systems". *Automatic Control, IEEE Transactions on*, 56(3):707 –710, March 2011.

[Cur12]   Ruth F. Curtain. Riccati equations for second order spatially invariant partial differential systems. *Automatica*, 48(1):139 – 144, 2012.

[CZ95]   Ruth F. Curtain and Hans Zwart. *An introduction to infinite-dimensional linear systems theory*, volume 21. Springer, 1995.

[DCVdH96]   R. A. De Callafon and Paul M. J. Van den Hof. FREQID-frequency domain identification toolbox for use with Matlab. *Selected Topics in Identification, Modelling and Control*, 9:129–134, 1996.

[DD03]   Raffaello D'Andrea and Geir E. Dullerud. Distributed control design for spatially interconnected systems. *Automatic Control, IEEE Transactions on*, 48(9):1478–1495, 2003.

[DM05]   Ann P. Dowling and Aimee S. Morgans. Feedback control of combustion oscillations. *Annu. Rev. Fluid Mech.*, 37:151–182, 2005.

[Dow97]   Ann P. Dowling. Nonlinear self-excited oscillations of a ducted flame. *Journal of Fluid Mechanics*, 346:271–290, 1997.

[EB12]   Jonathan P. Epperlein and Bassam Bamieh. A frequency domain method for optimal periodic control. In *American Control Conference (ACC), 2012*, pages 5501–5506. IEEE, 2012.

[EBÅar]   Jonathan P. Epperlein, Bassam Bamieh, and Karl J. Åström. ThermoAcoustics and the Rijke tube: Experiments, identification and modeling. *Control Systems Magazine*, to appear.

[Fin64]   Bernard S. Finn. Laplace and the speed of sound. *Isis*, pages 7–19, 1964.

[Fit66]   R.E. Fitts. *Linearization of nonlinear feedback systems*. PhD thesis, MIT, 1966.

[FL99]   Urban Forssell and Lennart Ljung. Closed-loop identification revisited. *Automatica*, 35(7):1215–1241, 1999.

[FPEN02]   G. F. Franklin, J. D. Powell, and A. Emami-Naeini. *Feedback Control of Dynamic Systems*. Prentice Hall, 4 edition, 2002.

[GB00]   S. L. Garrett and S. Backhaus. The power of sound. *American Scientist*, 88(6):516–525, 2000.

[Gev93]   Michel Gevers. Towards a joint design of identification and control? In *Essays on Control*, pages 111–151. Springer, 1993.

[Gib79]   J. S. Gibson. The Riccati integral equations for optimal control problems on Hilbert spaces. *SIAM Journal on Control and Optimization*, 17(4):537–565, 1979.

[Gil76]   E. G. Gilbert. Vehicle cruise: Improved fuel economy by periodic control. *Automatica*, 12:159–166, 1976.

[Gil77]   E. G. Gilbert. Optimal periodic control: A general theory of necessary conditions. *SIAM J.Control and Optimization*, 15(5):717–746, 1977.

[GL00]   T. Glad and L. Ljung. *Control Theory: Multivariable and Nonlinear Methods*. Taylor & Francis, New York, 2000.

[GLR74]   G. Guardabassi, A. Locatelli, and S. Rinaldi. Status of periodic optimization of dynamical systems. *Journal of Optimization Theory and Applications*, 14(1):1–20, 1974.

[GLS77]   I. Gustavsson, L. Ljung, and T. Söderström. Identification of processes in closed loop—identifiability and accuracy aspects. *Automatica*, 13(1):59 – 75, 1977.

[GV08]   Y. Guan and J. Verschelde. PHClab: A MATLAB/Octave Interface to PHCpack. In Michael Stillman, Jan Verschelde, Nobuki Takayama, Douglas N. Arnold, and Fadil Santosa, editors, *Software for Algebraic Geometry*, volume 148 of *The IMA Volumes in Mathematics and its Applications*, pages 15–32. Springer New York, 2008.

[HAFG98]   J. P. Hathout, A. M. Annaswamy, M. Fleifil, and A. F. Ghoniem. A model-based active control design for thermoacoustic instability. *Combustion Science and Technology*, 132(1):99–138, 1998.

[Hec88]   M. A. Heckl. Active control of the noise from a Rijke tube. *Journal of Sound and Vibration*, 124(1):117–133, 1988.

[HL67]   F. J. M. Horn and R. C. Lin. Periodic processes: A variational approach. *I&EC Process Design and Development*, 6(1):21–30, 1967.

[IMR11]   Simon J. Illingworth, Aimee S. Morgans, and Clarence W. Rowley. Feedback control of flow resonances using balanced reduced-order models. *Journal of Sound and Vibration*, 330(8):1567 – 1581, 2011.

[JB06]   Mihailo R. Jovanović and Bassam Bamieh. A formula for frequency responses of distributed systems with one spatial variable. *Systems & control letters*, 55(1):27–37, 2006.

[Kin14]   L. V. King. On the convection of heat from small cylinders in a stream of fluid: determination of the convection constants of small platinum wires with applications to hot-wire anemometry. *Philos. Trans. Royal Soc.*, 214:373–432, 1914.

[Lig54]   M. J. Lighthill. The response of laminar skin friction and heat transfer to fluctuations in the stream velocity. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 224(1156):1–23, 1954.

[Lju99]   Lennart Ljung. *System Identification: Theory for the User*. PTR Prentice Hall, Upper Saddle River, NJ, 1999.

[Lju07]   Lennart Ljung. *System Identification Toolbox for Use with MATLAB*. The Math-Works, Inc., 2007.

[LM71]   Jacques L. Lions and Sanjoy K. Mitter. *Optimal Control of Systems Governed by Partial Differential Equations*, volume 1200. Springer Berlin, 1971.

[LS48]   Harold Levine and Julian Schwinger. On the radiation of sound from an un-flanged circular pipe. *Phys. Rev.*, 73:383–406, Feb 1948.

[Mat03]   Konstantin I. Matveev. *Thermoacoustic Instabilities in the Rijke Tube: Experiments and Modeling*. PhD thesis, California Inistitute of Technology, 2003.

[MBF98]   H. Maurer, Ch. Büskens, and G. Feichtinger. Solution techniques for periodic control problems: A case study in production planning. *Optimal Control Applications and Methods*, 19:185–203, 1998.

[McI86]   A.C. McIntosh. Flame resonance and acoustics in the presence of heat loss. *Lectures in Applied Mathematics*, 24(Part 1):269–301, 1986.

[McI90]   A. C. McIntosh. On flame resonance in tubes. *Combustion Science and Technology*, 69(4-6):147–152, 1990.

[MD07]   A. S. Morgans and A. P. Dowling. Model-based control of combustion instabilities. *Journal of Sound and Vibration*, 299(1–2):261 – 282, 2007.

[Mee72]   A. I. Mees. The describing function matrix. *IMA J Appl Math*, 10(1):49–67, 1972.

[Mee81]   A. I. Mees. *Dynamics of Feedback Systems*. Wiley, New York, 1981.

[Min14]   Lawrence Minnetian. Rijke tube simulation: Modeling, simulation and identification of thermoacoustic phenomena in the Rijke tube. Master's thesis, ETH Zürich, 2014.

[OCGZK13]   Nejat Olgac, Rudy Cepega-Gomez, Umut Zalluhoglu, and Ayhan S. Kammer. An unconventional perspective in thermo-acoustic instability (TAI) using a novel mathematical tool. Technical report, University of Connecticut, Storrs, CT, 2013.

[PAV+13]   A. Papachristodoulou, J. Anderson, G. Valmorbida, S. Prajna, P. Seiler, and P. A. Parrilo. *SOSTOOLS: Sum of squares optimization toolbox for MATLAB*. http://arxiv.org/abs/1310.4716, 2013. Available from http://www.eng.ox.ac.uk/control/sostools, http://www.cds.caltech.edu/sostools and http://www.mit.edu/˜parrilo/sostools.

[PS03]   Pablo A. Parrilo and Bernd Sturmfels. *Minimizing Polynomial Functions*, volume 60 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 83–99. AMS, 2003.

[PW02]   M. C. Potter and D. C. Wiggert. *Mechanics of Fluids*. Cole, Pacific Grove, CA, USA, 3rd edition, 2002.

[PZ81]   Anthony J. Pritchard and Jerzy Zabczyk. Stability and stabilizability of infinite-dimensional systems. *SIAM Review*, 23(1):25–52, 1981.

[Ray78]   John William Strutt Rayleigh. The explanation of certain acoustical phenomena. *Nature*, 18(455):319–321, 1878.

[Ray96]   John William Strutt Rayleigh. *The Theory of Sound*, volume 2. Macmillan, 1896.

[RBFB93]   R. L. Raun, M. W. Beckstead, J. C. Finlinson, and K. P. Brooks. A review of Rijke tubes, Rijke burners and related devices. *Progress in Energy and Combustion science*, 19(4):313–364, 1993.

[Rij59]   P. L. Rijke. Über eine neue Art, die in einer an beiden Enden offenen Röhre enthaltene Luft in Schwingungen zu versetzen. *Annalen der Physik*, 183:339–343, 1859.

[Rud62]   Walter Rudin. *Fourier analysis on groups*. Wiley Online Library, 1962.

[SC67]   D. A. Saville and S. W. Churchill. Laminar free convection in boundary layers near horizontal cylinders and vertical axisymmetric bodies. *J. Fluid Mech*, 29(2):391–399, 1967.

[SE84]   J. L. Speyer and R. T. Evans. A second variational theory for optimal periodic processes. *IEEE Transactions on Automatic Control*, 29, Feb. 1984.

[SG00]   Herrmann Schlichting and Klaus Gersten. *Boundary Layer Theory*. Springer, 8 edition, 2000.

[SHIK10]   De Wet Swanepoel, James W. Hall III, and Dirk Koekemoer. Vuvuzela: good for your team, bad for your ears. *South African Medical Journal*, 100(2):99–100, 2010.

[Son70]   Carl Sondhauss. Ueber das Tönen erhitzter Röhren und die Schwingungen der Luft in Pfeifen von verschiedener Gestalt. *Annalen der Physik*, 216(5):53–76, 1870.

[SP07]   Sigurd Skogestad and Ian Postlethwaite. *Multivariable Feedback Control: Analysis and Design*, volume 2. Wiley New York, 2007.

[SRPT09]   Jos F. Sturm, Oleksandr Romanko, Imre Polik, and Tamas Terlaky. SeDuMi, 2009. http://mloss.org/software/view/202/.

[Swi88]   G. W. Swift. Thermoacoustic engines. *J. Acoustic Society Am.*, 84(4), 1988.

[TD92]   Jorge E. Tierno and John C. Doyle. Multimode active stabilization of a Rijke tube. *Active Control of Noise and Vibration, 1992: presented at the ASME Winter Annual Meet,, vol. DSC*, 38, 1992.

[Tre00]   Lloyd N. Trefethen. *Spectral Methods in MATLAB*. Society for Industrial and Applied Mathematics, 2000.

[TTT03]   R. H. Tütüncü, K. C. Toh, and M. J. Todd. Solving semidefinite-quadratic-linear programs using SDPT3. *Mathematical programming*, 95(2):189–217, 2003.

[Ver99a]   Jan Verschelde. Algorithm 795: PHCPACK: A general-purpose solver for polynomial systems by homotopy continuation. *ACM Transactions on Mathematical Software*, 25(2):251–276, Jun 1999.

[Ver99b]   Jan Verschelde. Polynomial Homotopies for Dense, Sparse and Determinantal Systems. *ArXiv Mathematics e-prints*, July 1999.

[VGD04]   S. Varigonda, T. T. Georgiou, and P. Daoutidis. Numerical solution of the optimal periodic control problem using differential flatness. *Automatic Control, IEEE Transactions on*, 49(2):271–275, Feb. 2004.

[Zab75]   J. Zabczyk. Remarks on the algebraic Riccati equation in Hilbert space. *Applied Mathematics & Optimization*, 2(3):251–258, 1975.

[ZN97]   B. T. Zinn and Y. Neumeier. An overview of active control of combustion instabilities. *35th Aerospace Sciences Meeting and Exhibit*, 461:1997, 1997.