Faculty of Engineering and Built Environment

School of Engineering

Mechancial Engineering Project B

Semester 1 - 2018

PROJECT TITLE

Thermionic Amplifier System Identification

NAME

Alexander Fairclough

SUPERVISOR

Chris Renton

THE UNIVERSITY OF
NEWCASTLE
AUSTRALIA

FINAL YEAR PROJECT

Declaration

I declare that this thesis is my own work unless otherwise
acknowledged and is in accordance with the University's academic
integrity policy available from the Policy Library on the web at
http://www.newcastle.edu.au/policylibrary/000608.html

I certify that this assessment item has not been submitted previously
for academic credit in this or any other course.

I acknowledge that the Faculty of Engineering may, for the purpose
of assessing this thesis:
- Reproduce this thesis and provide a copy to another member of the
  Faculty; and/or
- Communicate a copy of this assessment item to a plagiarism
  checking service (which may then retain a copy of the item on its
  database for the purpose of future plagiarism checking).
- Submit the assessment item to other forms of plagiarism checking.
- Provide this thesis to future students.

I further acknowledge that the Faculty of Engineering may, for the
purpose of sector wide Quality Assurance:
- Reproduce this thesis to provide a copy to a member of another
  engineering faculty.

I certify that the electronic versions of this thesis I have submitted are
identical to this hardcopy version. Furthermore, I have checked that
the electronic version is complete and intact on the submitted
location.

Student Name: Alexander Fairclough

Signature

FYP Part B Report

8 Christian P. Robert. "Gibbs Samplers", Introducing Monte Carlo Methods with R, 2010
Publication

<1%

9 citeseerx.ist.psu.edu
Internet Source

<1%

10 Advances in Industrial Control, 2016.
Publication

<1%

11 Submitted to University of Sheffield
Student Paper

<1%

12 users.aalto.fi
Internet Source

<1%

13 edoc.hu-berlin.de
Internet Source

<1%

14 www.math.ohiou.edu
Internet Source

<1%

15 www.ijcse.com
Internet Source

<1%

16 aaltodoc.aalto.fi
Internet Source

<1%

17 bryant1.bryant.edu
Internet Source

<1%

18 Lindsten, Fredrik, Thomas B. Schön, and Michael I. Jordan. "Bayesian semiparametric Wiener system identification", Automatica,

<1%

2013.
Publication

19   Michael L. Stein. "Difference Filter Preconditioning for Large Covariance Matrices", SIAM Journal on Matrix Analysis and Applications, 2012
Publication

<1%

20   www.slideshare.net
Internet Source

<1%

21   www.gaussianprocess.org
Internet Source

<1%

22   www.cs.rpi.edu
Internet Source

<1%

23   auai.org
Internet Source

<1%

24   Olaf Menzer, Antje Maria Moffat, Wendy Meiring, Gitta Lasslop, Ernst Günter Schukat-Talamazzini, Markus Reichstein. "Random errors in carbon and water vapor fluxes assessed with Gaussian Processes", Agricultural and Forest Meteorology, 2013
Publication

<1%

25   Abotaleb, Ibrahim S., and Islam H. El-adaway. "Construction Bidding Markup Estimation Using a Multistage Decision Theory Approach", Journal of Construction Engineering and

<1%

Management, 2016.
Publication

26   Christian P. Robert, George Casella. "Monte Carlo Statistical Methods", Springer Nature, 2004
Publication

                                                           <1%

27   lib.dr.iastate.edu
Internet Source

  <1%

28   The Physical Basis of Electronics, 1964.
Publication

  <1%

29   www.research-collection.ethz.ch
Internet Source

  <1%

30   Submitted to Imperial College of Science, Technology and Medicine
Student Paper

  <1%

31   www.eurasip.org
Internet Source

  <1%

32   Submitted to University of Leeds
Student Paper

  <1%

33   archiv.ub.uni-heidelberg.de
Internet Source

  <1%

34   Nima Nonejad. "Particle Gibbs with ancestor sampling for stochastic volatility models with: heavy tails, in mean effects, leverage, serial dependence and structural breaks", Studies in

  <1%

Nonlinear Dynamics & Econometrics, 2015
Publication

35 d-nb.info
    Internet Source   <1%

36 Eric B. Ford. "TRANSIT TIMING
    OBSERVATIONS FROM *KEPLER*. II.   <1%
    CONFIRMATION OF TWO MULTIPLANET
    SYSTEMS VIA A NON-PARAMETRIC
    CORRELATION ANALYSIS", The Astrophysical
    Journal, 05/10/2012
    Publication

37 Christophe Andrieu. "Particle Markov chain
    Monte Carlo methods : Particle Markov Chain   <1%
    Monte Carlo Methods", Journal of the Royal
    Statistical Society Series B (Statistical
    Methodology), 06/2010
    Publication

38 etheses.whiterose.ac.uk
    Internet Source   <1%

39 Xuesong Zhou, Daichun Song, Youjie Ma,
    Deshu Cheng. "Grid-Connected Control and   <1%
    Simulation of Single-Phase Two-Level
    Photovoltaic Power Generation System Based
    on Repetitive Control", 2010 International
    Conference on Measuring Technology and
    Mechatronics Automation, 2010
    Publication

40 Submitted to The University of Manchester
Student Paper
<1%

41 Submitted to University of Reading
Student Paper
<1%

42 Springer Texts in Statistics, 1999.
Publication
<1%

43 130.243.105.49
Internet Source
<1%

44 jmlr.org
Internet Source
<1%

45 Jensen, Bjorn Sand, Jens Brehm Nielsen, and Jan Larsen. "Bounded Gaussian process regression", 2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP), 2013.
Publication
<1%

46 tesi.cab.unipd.it
Internet Source
<1%

47 www.jomb.org
Internet Source
<1%

48 codegur.com
Internet Source
<1%

49 Flora Yu-Hui Yeh, Marcus Gallagher. "An empirical study of the sample size variability of optimal active learning using Gaussian process
<1%

regression", 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), 2008
Publication

50   Submitted to University of Warwick
      Student Paper   <1%

51   www.ee.bgu.ac.il
      Internet Source   <1%

52   Tandon, R., S. Mohajer, and V. Poor. "On the
      Symmetric Feedback Capacity of the K-user
      Cyclic Z-Interference Channel", IEEE   <1%
      Transactions on Information Theory, 2012.
      Publication

53   Jan Neddermeyer. "Nonparametric particle
      filtering and smoothing with quasi-Monte Carlo
      sampling", Journal of Statistical Computation   <1%
      and Simulation, 2011
      Publication

54   The New Palgrave Dictionary of Economics,
      2008.   <1%
      Publication

55   openscholarship.wustl.edu
      Internet Source   <1%

56   Karl Berntorp, Stefano Di Cairano. "Particle
      Gibbs with Ancestor Sampling for Identification   <1%
      of Tire-Friction Parameters", IFAC-

PapersOnLine, 2017
Publication

57   Michael Webber. "Rates of Profit and
      Interregional Flows of Capital", Annals of the
      Association of American Geographers, 3/1987
      Publication   <1%

58   Andreas Svensson, Thomas B. Schön, Manon
      Kok. "Nonlinear State Space Smoothing Using
      the Conditional Particle Filter**This work was
      supported by the project Probabilistic modelling
      of dynamical systems (Contract number: 621-
      2013-5524) and CADICS, a Linnaeus Center,
      both funded by the Swedish Research Council
      (VR).", IFAC-PapersOnLine, 2015
      Publication   <1%

59   Hyun-Chul Kim, Daijin Kim, Z. Ghahramani,
      Sung Yang Bang. "Gender Classification with
      Bayesian Kernel Methods", The 2006 IEEE
      International Joint Conference on Neural
      Network Proceedings, 2006
      Publication   <1%

60   www.coursehero.com
      Internet Source   <1%

61   scholar.sun.ac.za
      Internet Source   <1%

62   vrossi.free.fr
      Internet Source   <1%

Internet Source

<1%

75 www.cs.ubc.ca
Internet Source

<1%

76 brage.bibsys.no
Internet Source

<1%

77 ubdc.ac.uk
Internet Source

<1%

78 Choi, . "Bayesian nonlinear regression with Gaussian process priors", Gaussian Process Regression Analysis for Functional Data, 2011.
Publication

<1%

79 eprints.lancs.ac.uk
Internet Source

<1%

80 Hans Petter Langtangen. "A Primer on Scientific Programming with Python", Springer Nature, 2014
Publication

<1%

81 "Fuel Cell Systems", Springer Nature, 1993
Publication

<1%

82 Jackman. "Front Matter", Bayesian Analysis for the Social Sciences, 10/23/2009
Publication

<1%

83 T. McKelvey, H. Akcay, L. Ljung. "Subspace-based multivariable system identification from

<1%

frequency response data", IEEE Transactions on Automatic Control, 1996
Publication

84   re.public.polimi.it
Internet Source
<1%

85   Campbell, Trevor, Sameera Ponda, Girish Chowdhary, and Jonathan How. "Planning under Uncertainty using Bayesian Nonparametric Models", AIAA Guidance Navigation and Control Conference, 2012.
Publication
<1%

86   documents.mx
Internet Source
<1%

87   eprints.gla.ac.uk
Internet Source
<1%

88   Dora Turk, Goele Pipeleers, Jan Swevers. "A Combined Global and Local Identification Approach for LPV Systems**This research is sponsored by the Fund for Scientific Research (FWOVlaanderen) through project G.0002.11, by the KU Leuven-BOF PFV/10/002 Center-of-Excellence Optimization in Engineering (OPTEC), and by the Belgian Program on Interuniversity Poles of Attraction, initiated by the Belgian State, Prime Ministers Office, Science Policy programming (IAP VII, DYSCO). Goele Pipeleers is a Postdoctoral Fellow of the
<1%

Research Foundation-Flanders (FWO - Vlaanderen).", IFAC-PapersOnLine, 2015
Publication

89   M. Briers. "Online Sensor Registration", 2005 IEEE Aerospace Conference, 2005
Publication   <1%

90   docslide.us
Internet Source   <1%

91   Introduction to Applied Bayesian Statistics and Estimation for Social Scientists, 2007.
Publication   <1%

92   www0.gsb.columbia.edu
Internet Source   <1%

93   web.cecs.pdx.edu
Internet Source   <1%

94   pdfs.semanticscholar.org
Internet Source   <1%

95   Zaytsev, A.A. Burnaev, E.V. Spokoiny, V.. "Properties of the Bayesian parameter estimation of a regression based on Gaussian processes.", Journal of Mathematical Sciences, Dec 9 2014 Issue
Publication   <1%

96   lasa.epfl.ch
Internet Source   <1%

97 P.M. Djuric. "Particle Filtering", IEEE Signal Processing Magazine, 9/2003

Publication

<1%

98 core.ac.uk

Internet Source

<1%

99 scholarbank.nus.edu.sg

Internet Source

<1%

100 ericschulz.github.io

Internet Source

<1%

101 Jose A. Castellanos. "Analysis of Particle Methods for Simultaneous Robot Localization and Mapping and a New Algorithm: Marginal-SLAM", Proceedings 2007 IEEE International Conference on Robotics and Automation, 04/2007

Publication

<1%

102 homepages.inf.ed.ac.uk

Internet Source

<1%

103 Jackman. "Markov Chain Monte Carlo", Bayesian Analysis for the Social Sciences, 10/23/2009

Publication

<1%

104 mlg.eng.cam.ac.uk

Internet Source

<1%

105 www.bcamath.org

Internet Source

<1%

106 Vasquez, J.R.. "Exploiting correlation effects within multiple-hypothesis tracking", Mathematical and Computer Modelling, 200605
Publication

<1%

107 www.theintertekgroup.com
Internet Source

<1%

108 Cowles, Mary Kathryn Carlin, Bradley P.. "Markov chain Monte Carlo convergence diagnostics: a comparative review.", Journal of the American Statistical Asso, June 1996 Issue
Publication

<1%

109 cis.temple.edu
Internet Source

<1%

110 www.cse.buffalo.edu
Internet Source

<1%

111 www.rt.isy.liu.se
Internet Source

<1%

112 L LUTES. "Direct Stochastic Analysis of Linear Systems", Random Vibrations, 2004
Publication

<1%

113 lib.tkk.fi
Internet Source

<1%

www.math.chalmers.se

114 Internet Source
<1%

115 aclweb.org
Internet Source
<1%

116 ir.canterbury.ac.nz
Internet Source
<1%

117 cdn.intechopen.com
Internet Source
<1%

118 Jyri Pakarinen. "A Review of Digital Techniques for Modeling Vacuum-Tube Guitar Amplifiers", Computer Music Journal, 06/2009
Publication
<1%

119 www.ece.mtu.edu
Internet Source
<1%

120 Yiteng Arden Huang, Jacob Benesty. "Adaptive multi-channel least mean square and Newton algorithms for blind channel identification", Signal Processing, 2002
Publication
<1%

121 "The New Palgrave: Dictionary of Economics, Volume 2 Command economy — epistemic game theory", The New Palgrave Dictionary of Economics, 2008.
Publication
<1%

122 Yu, Ye Sun, Lu. "Effect of overlay thickness,

overlay material, and pre-overlay treatment on evolution of asphalt con", Construction and Building Materials, Feb 20 2018 Issue
Publication

123   Nakano, Shin'ya, Kazue Suzuki, Kenji Kawamura, Frédéric Parrenin, and Tomoyuki Higuchi. "A sequential Bayesian approach for the estimation of the age–depth relationship of the Dome Fuji ice core", Nonlinear Processes in Geophysics, 2016.
Publication   <1%

124   A. Alessandri, C. Cervellera, A.F. Grassia, M. Sanguineti. "An approximate solution to optimal L/sub p/ state estimation problems", Proceedings of the 2005, American Control Conference, 2005., 2005
Publication   <1%

125   Wei, Yonghua, and Gabriel Huerta. "Dynamic Generalized Extreme Value Modeling via Particle Filters*", Communications in Statistics - Simulation and Computation, 2016.
Publication   <1%

126   Christian P. Robert. "Metropolis–Hastings Algorithms", Introducing Monte Carlo Methods with R, 2010
Publication   <1%

127   Lecture Notes in Mathematics, 2003.

Publication

<1%

128   Adrian Wills, Thomas B. Schön, Brett Ninness. "Parameter Estimation for Discrete-Time Nonlinear Systems Using EM", IFAC Proceedings Volumes, 2008
Publication

<1%

129   Schmidt, Paul, Volker J. Schmid, Christian Gaser, Dorothea Buck, Susanne Bührlen, Annette Förschler, and Mark Mühlau. "Fully Bayesian Inference for Structural MRI: Application to Segmentation and Statistical Analysis of T2-Hypointensities", PLoS ONE, 2013.
Publication

<1%

130   Svensson, Andreas, Thomas B. Schon, and Fredrik Lindsten. "Identification of jump Markov linear models using particle filters", 53rd IEEE Conference on Decision and Control, 2014.
Publication

<1%

131   Sun, Shiliang, Rongqing Huang, and Ya Gao. "Network-Scale Traffic Modeling and Forecasting with Graphical Lasso and Neural Networks", Journal of Transportation Engineering, 2012.
Publication

<1%

| Exclude quotes | Off | Exclude matches | Off |
| Exclude bibliography | On | | |

Project Summary

- I gained significant knowledge in the field of nonlinear system identification and modelling, probabalistic inference and regression, and estimation of intractable distributions.

- I performed research on the need for nonlinear modelling to emulate the sound of valve guitar amplifiers.

- I implemented a recently developed Particle Markov Chain Monte Carlo method to estimate linear dynamics and nonlinear saturation to fit a Wiener model.

- I used a Gaussian process for regression of the nonlinearity to place minimal structure on the inferred model using a nonparametric form.

- I implemented a Conditional Particle Filter with Ancestor Sampling to perform state estimation and ensure ergodicity of the PMCMC sampler

- I identified models for a series of test systems and a small tube amplifier using the PMCMC method and Gaussian white noise training data.

- I analysed the performance of the modelled systems using validation data sets and identified differences in their frequency response.

- I proposed plausible reasons for inconsistency between the true and identified models of the systems.

- I explore directions for further research into alternative system identification methods to solve these issues.

Bayesian Semiparametric System Identification of Thermionic Amplifiers

Final Year Project Report - MECH4841 Part B

2018

Alex Fairclough [1]

[1] *Student of Mechatronics and Mechanical Engineering,*
ES313 - Centre for Infinite Dimensional Non-Parametric Maps, Optical Flow, Particle Markov Chain Monte Carlo, SLAM and Stuff and Things and Possums
The University of Newcastle, Callaghan, NSW 2308, AUSTRALIA
Student Number: 3134480
E-mail: Alex.Fairclough@uon.edu.au

Abstract

The objective of this project is the research, development and characterisation of a nonlinear input-output model of a vacuum tube guitar amplifier. This is achieved by performing an analysis technique known as system identification.

Nonlinearities in vacuum tube guitar amplifiers have created much sought-after overdrive sounds that remain a keystone of a guitarists tone. Previous modelling techniques for tube amplifiers are either unfeasible for realtime embedded simulation, or make assumptions that discard salient details of valve saturation, causing a change in tonal character.

In this project, the feasibility of modelling a guitar tube amplifier with a semiparametric nonlinear Wiener system model is investigated. The Wiener model structure approximates the system using a combination of a linear state space model of system dynamics, and a static nonlinearity function modelled with Gaussian process regression. System identification using input-output measurements is performed within a flexible probabilistic framework to avoid placing assumptions and structure upon the model.

The method employs recently developed Particle Markov Chain Monte Carlo methods to identify a Wiener model of the system. Resulting model performance is measured by comparison of system frequency response and clipped waveform shape between the real and the modelled systems. The method performs well at identification of a nonlinear function for overdrive. The identification of the linear system struggles to capture low frequency dynamics.

Future researching is considered including development of a covariance function suited to modelling tube saturation, parametric model formulations to use Expectation Maximisation methods, and low latency real time simulations on an embedded hardware platform.

Acknowledgements

Contents

1. Introduction

Guitarists are always finding new ways to use and abuse electronic hardware for purposes in which they were never intended. Most audio amplifiers encountered are designed to produce a highly linear response. Instead, guitarists demand overdriven nonlinear response from their amplifiers. One of the cornerstones of rock guitar sounds is overdrive.

Overdrive is the clipping of the peaks and troughs of a guitar signal caused by saturating the amplifier with an input signal whose magnitude exceeds the linear operating region of the amplifier they are using. While any signal amplifier imposes limits on the magnitude of the input, the nonlinearity which governs the transition to saturation can vary greatly. Vacuum tubes have a soft, asymmetric nonlinear transition to saturation creating a tone which is highly sought after. This is the primary reason they are still preferred for amplification of electric guitars.

Like many technologies that were once commonplace, valves have been depreciated and are rarely used in modern appliances. Tube amplifiers have significant size and weight due to the power transformers required for their operation. The mechanical properties of the construction and operation of tubes limits the lifespan of a vacuum tube. A guitar amplifier may require replacement of its vacuum tubes every year if used persistently, and often fail when subject to the tough handling they receive touring a musician. This creates incentives for the use of alternate modern technologies.

The invention of silicon transistors started an electronics revolution that allowed the creation of smaller, lighter, more efficient replacements for vacuum tubes in almost all of their applications. Guitar amplifiers are no exception, and transistor based *solid-state* amps are commonly available. While the high linearity of transistor amps allows for impressive clean performance, the sound of overdriven transistors is typically undesirable due to the amplifier characteristic exhibiting a sharp transition to saturation.

Many attempts have been made to achieve more desirable overdrive from transistor amplifiers. One of the most popular options is the use of an *effects pedal*, a small circuit placed inline between the guitar and amplifier to modify, modulate or distort the sound before it reaches the amplifier. Placing an overdrive pedal before a highly linear amplifier results in an acceptable approximation to the desired sound. As the nonlinearity is generated by the pedal and not the amplifier, a specific tone can be closely approximated when using many different amplifiers.

Among the disadvantages of analog effects pedals are a lack of customisation due to changes in the audio filters requiring internal components to be changed, and the inability to perfectly recreate a desired amplifier's sound. Each pedal design exhibits specific characteristics and limitations due to the choice of components in the filtering and clipping sections. A musician who wants a greater variety of sounds is required to purchase many different pedals. This creates the incentive for a flexible effects unit without the limitations of analog components, which can be completely modified or reprogrammed.

Digital Signal Processing (DSP) is the use of computer processing to perform a vast array of signal processing operations. Digital effects pedals use modern embedded microprocessors to perform high quality, real time DSP. Due to the limited computing power available on an embedded system, approximations must be made for the mathematical functions representing system processes, to ensure calculations complete within the audio sampling period. Computationally efficient approximations to overdrive originate with basic approximations to a smooth static nonlinearity (SNL) using a squashing function. This achieves a form of overdrive, but does not accurately recreate tube overdrive.

The first effects pedals based on DSP suffered from poor adoption in the guitar community, with complaints of the processing causing *tone suck*, a perceived loss of the quality of sound produced from the guitar. Among the error due to modelling approximations, early analog to digital converters captured data with a low bit depth and sample rate, adding quantisation error and reducing the maximum frequency accurately captured due to aliasing.

In recent years Analog to Digital Conversion (ADC) and Digital to Analog Conversion (DAC) has reached a level of quality more than sufficient for the perceived quality of a digital effect to be nearly indistinguishable when compared to an analog circuit. Due to this, the mind set of the average consumer guitarist is shifting towards acceptance of digital effects. The popularity of digital modelling effects offered by brands such as the Fractal Audio, Strymon, TC Electronics and Line 6 is increasing, and with it, the demand for new digital audio solutions.

Simulating an amplifier requires a mathematical model representing the phenomena and processes of the system. The model used must have sufficient complexity to capture the salient features of the system. For real time simulation on microprocessors, the model must also simplify the system enough to be computationally efficient while ensuring important characteristics of the system are not lost.

Many approaches to modelling require making assumptions about the structure and operation of the system such as linearity or the order of the model. These assumptions can be chosen to encode prior knowledge of the system into the model, or intentionally simplify the structure to make simulation easier and faster. Minimising assumptions while imposing enough structure on the model to make simulation possible is a challenge for all systems, particularly those with nonlinear behaviour. The most desired behavior of tube amps is in the nonlinear operating region of overdrive, therefore it is important that the model is capable of expressing this behaviour accurately and convincingly.

System identification is the process of identifying unknown parameters or features of a model using input and output data captured from a system. System identification is an ongoing field of research with no perfect solution, and can be performed with use of a wide range of estimation methods. Nonlinear systems with process and measurement noise pose a significant challenge to identify as they do not satisfy superposition. Therefore the estimation of unknown and potentially unobservable parameters and nonlinearities of a system requires methods robust to noise and uncertainty.

Bayesian inference is a probabilistic method which defines unknown variables which are not directly observable as stochastic quantities with an associated uncertainty. Modelling hypotheses express the subjective degree of belief in a quantity or process given the knowledge of the system. System identification can be performed using Bayesian inference by taking estimates from the probability distributions describing the values of model parameters likely to have generated the data we have measured.

The Gaussian (Normal) distribution describes the probability of real random variables as a function of their mean μ and covariance σ. Expanding multivariate Gaussian distributions to infinite dimensionality allows a probability distribution over *functions*, known as the Gaussian process. The Gaussian process characterises an unknown process as a nonparametric function $h(\cdot)$ completely specified by a mean function $m(x, x')$ and covariance function $K(x, x')$. A Gaussian process can be used to perform regression, identifying a function using input and output training data without a closed form expression for the function.

A useful, general class of nonlinear dynamical systems are those of the block-oriented system. Notably among this class are the Wiener and Hammerstein models, consisting of a linear dynamical system

followed by a memoryless static nonlinearity or visa-versa. A Wiener model is proposed, consisting of a state space model for the linear dynamics, and a nonparametric Gaussian process implemented as an approximation for the nonlinearity, combining to make a *semiparametric* system.

The Particle Filter (PF) is a Bayesian estimation method which employs Monte Carlo to approximate probability distributions using a finite number of random, discrete samples. However when the distribution is conditional on many variables, the Particle Filter alone can not estimate the posterior due The recently proposed method of particle Gibbs with ancestor sampling (PG-AS) employs a Particle Markov Chain Monte Carlo (PMCMC) framework to efficiently estimate the posterior distribution of the model parameters without introducing error. The PG-AS method has previously been demonstrated to accurately estimate highly nonlinear Wiener system models from input-output data (Lindsten et al., 2013).

The proposed semiparametric Wiener model aligns with the prior knowledge of the operation of tube amplifiers:

- Linear dynamical system which encompasses the frequency response of the system and dynamic behaviour of the amplifier stages

- Nonlinear behaviour exhibited due to the saturation or blocking of thermionic emissions in the valves

While the state space model is parametric, the approach taken is as 'grey-box' as possible. There are no strong assumptions imposed upon the system in the SSM. When the order of the system is unknown, a sparsity promoting prior is implemented to achieve automatic order determination. The lack of assumptions allows the inference method to generalise and capture any processes that are evident in the training data.

2. Vacuum Tube Amplification

Vacuum tubes are antiquated electronic components used to amplify small signal voltages. This is achieved through utilisation of the properties of *thermionic emission* of electrons from a heated electrode contained within an evacuated glass tube. Vacuum tubes, or 'valves', saw enormous use throughout the 20th century due to their applications in radio transmission/reception, telephone systems, defence applications, and computing. Vacuum tubes have been replaced by silicon transistors in most applications, due to the smaller size, improved linearity, increased efficiency and reduced supporting power electronics required to use transistors.

Long before their use in amplification, the first tube designs operated as simple diodes. The tube diode's construction is an evacuated glass tube enclosing two electrodes; a heated cathode element to emit electrons, and a plate/anode held at a positive voltage with respect to the cathode, to attract and collect the flow of electrons. The diode allows current to pass in only one direction from cathode to anode. Current is 'blocked' in the other direction as the anode is left unheated, preventing thermionic emission in this direction.

Amplification of a signal using vacuum tubes becomes possible by introducing more electrodes in their construction. The *triode* has a third electrode known as the *control grid*. The control grid is a wire mesh used to modulate the flow of electrons between the cathode and anode. When uncharged, the electrons pass uninhibited through the empty spaces of the grid however modulation of the electron flow is possible when a negative voltage relative to the cathode is applied to the grid, creating a negative field.

This field inhibits the flow of electrons from the cathode and reduces the current to the plate. Applying a relatively small voltage difference to the grid results in a much greater corresponding change in plate current, amplifying the signal applied to the grid. In this respect, valves operate as a *transconductance* amplifier, or a voltage controlled current source. The transconductance is defined as the ratio of the anode current to grid voltage in the linear operating region.The transconductance represents the relative increase of the signal from input to output and is also known as the 'gain' of the tube.
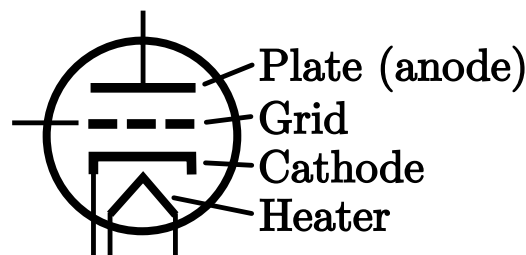

Figure 1: Triode Valve Schematic

A vacuum tube is typically identified by the number of electrodes it contains. In guitar and audio amplification, the most common tubes used are *triodes* containing three elements, and *pentodes* containing five. For guitar amplifiers, tubes are also typically classified by the intended use; as high gain pre-amplifier tubes suited for amplifying small signal inputs, and power-amplifier tubes that handle high power signals at the output stage of the amplifier which drives the loudspeaker.

An audio amplifier is usually designed for the amplified output to perfectly recreate the input signal with minimal change during the amplification. They achieve this by ensuring the relationship between
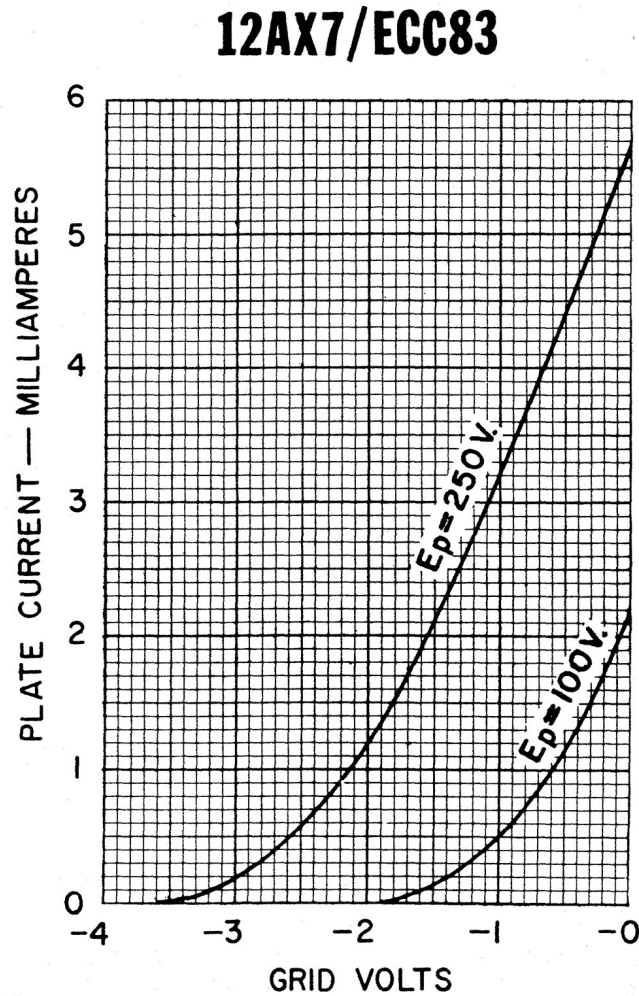
12AX7/ECC83

Figure 2: Grid Voltage to Plate Current Characteristic for 12AX7 Pre-amp Tube (Sylvania Electronic Products Inc., 1995)

input to output is as linear as possible. Vacuum tubes have a region of their voltage-current characteristic which is approximately linear, evident in Figure 2. This linear region is typically chosen as the operating point for the amplifier. This is achieved by *biasing* the grid with a negative voltage such that subsequent application of a small peak to peak signal is centred within the linear region.

When assuming linear operation, a change in input creates a corresponding change in output that is linearly scaled by the gain ratio of the amplifier. *Overdrive* is a phenomena that occurs when an input increases past the bounds of the linear region. The relationship between input and output becomes nonlinear, where increasing the amplitude of the input signal causes the output to taper off as it approaches a saturation limit. A metaphor to understand this saturation limit is the opening and closing of a tap. When a tap is opened completely, the flow of water will not change if you continue to attempt to turn the tap in the same direction.

Counter to intuition, this nonlinear operation is the reason valve amplifiers remain popular in modern

audio applications, due to the tonal differences perceived when listening to audio through an overdriven valve amplifier. Overdrive is often referred to as 'clipping' due to the visual effect it produces on the output waveform. The peaks of the signal which exceed the linear region appear to have been 'clipped' from the signal. A comparison between linear and overdriven amplification can be seen in Figure 3. The saturating nonlinearity of a vacuum tube exhibits a smooth transition to overdrive, whereas transistors exhibit overdrive with a fast transition as their saturation limit is imposed by the upper and lower bound of the supply voltage. This causes 'hard-clipping' which sounds harsh and unpleasant to the listener. This is a primary reason valves are considered superior to silicon transistors in electric guitar amplifiers.

The nonlinearity exhibited by tubes can be *asymmetric*, with a different saturation limit for positive or negative peaks of the signal. The asymmetry of the clipping is due to the difference between the limiting amount of electron emission possible when driven with a positive voltage, and the complete blocking of current to the plate when a large negative grid voltage is applied. Asymmetry has the effect of introducing odd order harmonic distortion into the signal, and is often partially attributed to the desired sound of tube amplifiers.
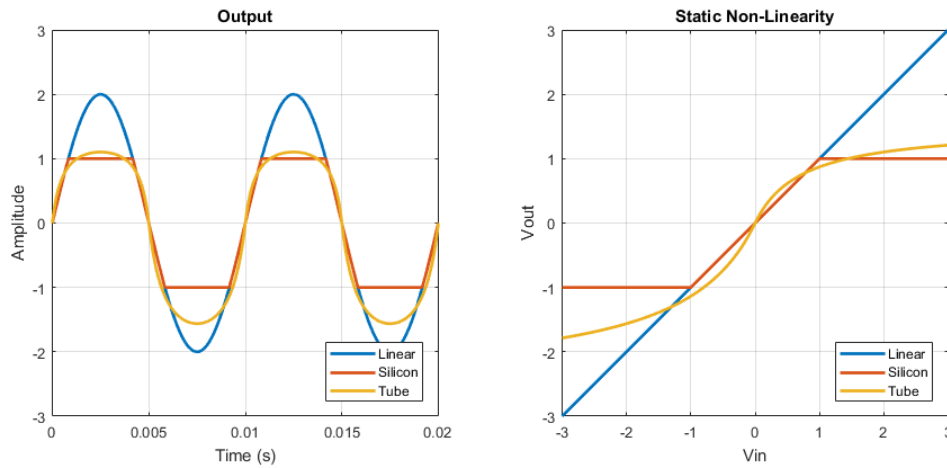

Figure 3: Comparison of Clipped Output Waveforms

3. Modelling & Simulation

Simulation is the process of imitating the behaviour and operation of an actual system by calculating the numerical solution to a model. A model is a collection of equations constructed to best represent the response of the system to specific environmental conditions and inputs (Franklin et al., 1998). Modelling and simulation often go hand in hand. A simulation can not be made without a model to calculate from, and a model cannot be verified without simulation. These are amongst the most useful tools an Engineer uses when designing solutions to problems based on real world phenomena as they allow work to be carried out in a controllable, repeatable mathematical environment.

Our understanding of many natural laws are simply the best models we have currently identified to fit expected behaviors of the universe around us. Models can vary tremendously in both structure and complexity; a trade off often exists when modifying either. Enforcing strong structure or assumptions upon a problem may encode knowledge that is believed to be very certain. Higher complexity may allow for increased accuracy, but can lead to *over fitting* of the data instead of correctly modelling true system dynamics.

A sensible choice of model for a tube amplifier is to use electrical circuit analysis. Electrical circuits are often designed and simulated with the aid of `SPICE` modelling. SPICE models of vacuum tubes have been developed for simulation based on historical research conducted during the boom of tube development. The primary focus of these models result in accurate approximations for the linear response of tubes, but few models have approximations which are valid and accurate in the nonlinear operating region of the valves. Modern research conducted by valve enthusiasts focuses on improving the nonlinear behaviour of SPICE simulations using experimental data. This approach utilises polynomial approximations and discrete look-up-table methods to calculate the model's nonlinear states, leading to improvements in the modelling of overdrive within SPICE simulations (Dempwolf and Zölzer, 2011; Koren, 1996).

However, SPICE simulation requires a complex, highly parametric model of the system created from first principles at the component level of the amplifier circuit. While the results can be accurate, the computation time required to evaluate the behaviour of the system makes this unfeasible for simulation in real time on embedded systems.

Block-oriented systems provide a useful framework for nonlinear modelling. They consist of interconnected subsystems representing linear dynamics and static nonlinearities. Conventional layouts are the Hammerstein (static nonlinearity followed by linear dynamical system), Wiener (linear dynamical system followed by static nonlinearity) and the combination of the two, the Hammerstein-Wiener model, which contains nonlinearities on the input *and* output of a linear dynamical system. Hammerstein and Wiener models have attracted significant attention for modelling and identification of nonlinear systems and techniques for system identification of these generalised models are a continuing area of research.

The *state space model* is often used to represent and model linear time-invariant dynamical systems. The model can be described in a parametric form as

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} + \omega_t, \qquad\qquad \omega_t \sim \mathcal{N}(0, Q), \qquad\qquad (3.1)$$

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u} + e_t, \qquad\qquad e_t \sim \mathcal{N}(0, R). \qquad\qquad (3.2)$$

Any linear transfer function, or combinations thereof, can be represented with an equivalent state space model. A difficulty in representing a transfer function in state space form is that there are

many equivalent state space models for a unique transfer function. *Observer canonical form* is a representation where the first state is chosen as the output of the system. To make sure the output is equal to the first state the $\mathbf{C}$ matrix is fixed to the value

$$\mathbf{C} = [1, 0, \ldots, 0].$$

Consider a 3rd order input-output transfer function in the frequency domain

$$H(s) = \frac{Y(s)}{U(s)} = \frac{b_0 s^3 + b_1 s^2 + b_2 s + b_3}{s^3 + a_1 s^2 + a_2 s + a_3} \tag{3.3}$$

The observer canonical form of the state space model of this transfer function is

$$sX(s) = \mathbf{A}\mathbf{X}(s) + \mathbf{B}U(s) = \begin{bmatrix} -a_1 & 1 & 0 \\ -a_2 & 0 & 1 \\ -a_3 & 0 & 0 \end{bmatrix} \mathbf{X}(s) + \begin{bmatrix} b_1 - a_1 b_0 \\ b_2 - a_2 b_0 \\ b_3 - a_3 b_0 \end{bmatrix} U(s), \tag{3.4}$$

$$Y(s) = \mathbf{C}\mathbf{X}(s) + \mathbf{D}U(s) = [1, 0, 0]\mathbf{X}(s) + b_0 U(s), \tag{3.5}$$

and this observer canonical form can be generalised to any state space model of order n_x. Any form of state space model can be converted to observer canonical form with the use of a similarity transform, preserving the input-output behavior no matter how it is described (Cheever, 2015)

The Wiener model is ideal for the purpose of modelling valve amplifiers. When driven with a low input voltage, the amplifier behaves approximately linearly, and the frequency response of the system is shaped by linear analogue filtering stages. This behaviour is captured in the linear dynamical system block $\mathcal{G}$. When the input voltage overloads the tube, the input/output voltage characteristic begins to saturate smoothly, exhibiting the nonlinearity responsible for overdrive (King, 1923). This overdrive effect is captured by the static nonlinearity block $h(\cdot)$ of the Wiener model.
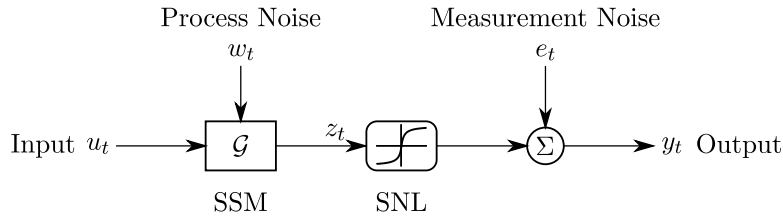

Figure 4: Block diagram of a Wiener system with linear dynamical system $\mathcal{G}$ followed by static non-linearity $h(\cdot)$ with process noise w_t and measurement noise e_t.

An approximate static nonlinearity function to represent amplifier overdrive is the class of squashing functions. One such form of squashing function is

$$y = \frac{x}{1 + |x|}. \tag{3.6}$$

This function is smooth, continuously differentiable, and is extremely efficient to compute with the limited processing power of a microprocessor. Some issues with the use of this function for modelling overdrive is the lack of a linear region, never reaching asymptotic saturation and neglecting asymmetry. Other function choices are suitable such as $y = tanh(x)$, which is asymptotic at ± 1, but suffers from the prohibitive computational cost of evaluating $tanh(\cdot)$ for each sample.

Gaussian process regression is a tool strongly suited for the purpose of modelling the function of a static nonlinearity. Sampling from the GP prediction directly approximates the continuous unknown function $h(z)$, mapping the output of the linear dynamic system z to the final system output y. The fundamentals of Gaussian processes are covered in Section 7.

4. Bayesian Inference

System identification often represents an inverse problem. Knowledge of the system's behaviour is extracted from input/output measurements to decide on a model and/or its parameters that adequately explains the creation of the data. The Bayesian view of probability provides a set of logical rules as a means to measure uncertainty, degrees of belief, or states of knowledge. These expression can then be used to *infer* a degree of plausibility of a previously unknown quantity or function using knowledge or measurement of its consequences (Perez, 2014).

Bayes's Rule is key to probabilistic inference, and often takes the form of an expression for the *posterior* distribution. In the context of system identification the following form is used.

$$p(\theta|D, I) = \frac{p(D|\theta, I)p(\theta|I)}{p(D|I)} = \frac{p(D|\theta, I)p(\theta|I)}{\int_\Theta p(D|\theta, I)p(\theta|I)d\theta}, \tag{4.1}$$

where,

- θ is the continuously variable parameter to be estimated,

- I is a proposition that represents background information,

- D is a proposition representing data or evidence,

- $p(\theta|D, I)$ is the posterior probability of the parameter,

- $p(\theta|I)$ is the prior probability of the parameter,

- $p(D|\theta, I)$ is the likelihood function of θ, the probability of obtaining the data given the parameters and prior information,

- $p(D|I)$ is a normalising constant known as the *marginal likelihood*, with integral over the domain Θ, the set of all values of θ.

It is important to note that in this context, the probability distributions $p(\theta|D, I)$ and $p(\theta|I)$ represent what we know *about* θ, not any property of θ. The parameter itself is not random; instead the distributions indicate *"the range of values that are consistent with the data and our prior information. What is distributed is not the parameter but the probability."* (Jaynes, 1986)

The posterior distribution is the target of Bayesian inference, giving an updated measure of belief in the parameter's possible value using the combination of all information at hand. A point estimate of the parameters of the system $\hat{\theta}$ can be sampled from this distribution in a multitude of ways. Among them are the *maximum a posteriori* θ_{MAP}, the mean or *expected value* of the posterior distribution, the median, or the maximum likelihood estimate.

Marginalisation is a key ability of Bayesian methods, allowing unwanted parameters to be eliminated from the estimation problem. Let γ represent a *nuisance parameter* appearing in an estimation problem alongside the parameter of interest θ. Applying Bayes's Theorem can yield the joint posterior of both variables $p(\theta, \gamma|D, I)$ which must be marginalised to yield the posterior of θ alone. Marginalisation is achieved by integrating the posterior over γ

$$p(\theta|DI) = \int_\Gamma p(\theta, \gamma|D, I)d\gamma, \tag{4.2}$$

where $\gamma \in \Gamma$ represents the set of all values of γ (Perez, 2014).

Specification of a prior function is one of the challenges of properly utilising Bayesian inference. The prior is typically taken from a general purpose *toolbox* of priors, such as a Uniform or Normal distribution with known mean and covariance. The accuracy of the inferred parameters will suffer if the prior is poorly specified. For instance, if the prior has a very low covariance, it indicates strong prior belief of the parameter's value. If the true value of the parameter is distant from the peak of the sharp prior distribution, the prior will outweigh any efforts to change the belief of the posterior towards the true value indicated by the data.

5. Sequential Monte Carlo

5.1. Monte Carlo

The implementation of Bayesian inference methods often encounters a significant road block. When a necessary distribution has no closed form solution or is analytically intractable, it can not be sampled from directly. Overcoming this road block can often be achieved by using an approximation method.

Monte Carlo is one methodology for approximating a distribution without introducing significant error to the solution. Monte Carlo approximations are useful as substitutes due to their *convergence* properties, which can be used to prove the accuracy of using a Monte Carlo method as an estimator.

A random variable X, distributed by the probability density function p_X, has an *expected value $E\{X\}$*. This value can be estimated by computing its expectation integral, that is,

$$E\{X\} = \int_{-\infty}^{\infty} x p_X(x) dx. \tag{5.1}$$

Evaluating this integral can often prove to be difficult. Instead of computing the integral directly, estimating it using a large amount of discrete random variables can yield a solution that can be proved to converge almost surely to the correct answer. These estimates are often useful to substitute an exact calculation inside of a Bayesian estimator.

Definition 5.1 (Strong Law of Large Numbers) *If $\{X_i, i \geq 1\}$ is a sequence of independent identically distributed (i.i.d.) random variables with $E\{|X_i|\} < \infty$, then the sequence of random variables S_n defined as*

$$S_n(\omega) \triangleq \frac{1}{N} \sum_{i=1}^{N} X_i(\omega), \tag{5.2}$$

converges almost surely to $\mu = E\{X_i\}$ for any $i \geq 1$, that is,

$$P\left(\left\{\omega \,\Big|\, \lim_{N \to \infty} S_n(\omega) = \mu\right\}\right) = 1. \tag{5.3}$$

The Strong Law of Large Numbers Theorem dictates that the sequence of discrete samples $S_n(\omega)$ converges almost surely to the expected value μ as $n \to \infty$ (Wills, 2018).

Consider the sequence $S_n(\omega)$ given in Equation 5.2 where the N i.i.d. random variables X_i are drawn from the uniform distribution $X_i \sim \mathcal{U}(0, 1)$. For the case where $N = 1$ the cumulative distribution function (CDF) of the sequence simply resembles the uniform distribution. As the number of random variables N increases, the CDF of the sequence visibly converges to the expected value of the function $E\{X\} = 0.5$.

Monte Carlo integration is a key application of the SLLN. In Bayesian analysis, Monte Carlo is often used for the computation of expectation integrals. Taking the sequence of i.i.d. random variables $X, X_1, X_2, ...$ then

$$\frac{1}{N} \sum_{i=1}^{N} X_i \overset{a.s.}{\to} \int x p_X(x) dx, \quad N \to \infty. \tag{5.4}$$

Often, we are not interested in the expectation of X itself, but a function of X which produces a measured output Y. Let f be a function that maps the i.i.d. random variables X_i to $Y_i = f(X_i)$, then the samples Y_i are also i.i.d. The SLLN theorem applied to Y_i can be stated as

$$\frac{1}{N}\sum_{i=1}^{N} Y_i \overset{a.s.}{\to} \int y p_Y(y) dy, \quad N \to \infty, \tag{5.5}$$

and using the Law of the Unconscious Statistician this result can be restated as

$$\frac{1}{N}\sum_{i=1}^{N} f(X_i) \overset{a.s.}{\to} \int f(x) p_X(x) dx, \quad N \to \infty. \tag{5.6}$$

This states that the expectation integral can be solved by taking the mean of the random variables $f(X_i)$, assuming an infinite number of variables are used (Wills, 2018).
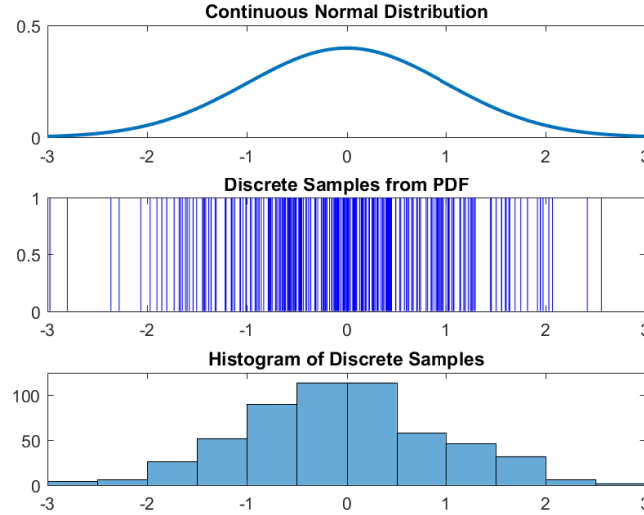

Figure 5: Monte Carlo Approximation of $\mathcal{N}(0,1)$.

Figure 5 demonstrates a simple example of Monte Carlo approximation. A sequence of random samples $X_1, X_2, ...X_n$ are drawn from a normal distribution $\mathcal{N}(0,1)$. Plotting the values of the variables X_i generates a 'discrete' pdf, where samples are more densely packed where they have a higher likelihood of being drawn. Taking the histogram of these samples sorts values into bins and counts how many times the values of the variables are located within each bin. This clearly shows the similarity between the mass distribution of the discretised pdf and the true pdf of the normal distribution.

The expectation integral can be approximated by recalling Equation 5.4 by taking the mean of the samples,

$$\hat{\mu} = E\{X_i\} \approx \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{5.7}$$

In the example shown in Figure 5, using $n = 250$ samples, the estimated mean was calculated to be $\hat{\mu} \approx -0.0038$. While close to the true mean $\mu = 0$, this estimate contains error that cannot feasibly

be eliminated outside of the limiting case of $n \to \infty$, making it impossible as computers are finite arithmetic machines.

5.2. Particle Filtering

The Particle Filter (PF) is a class of Sequential Monte Carlo method which generates approximations of probability density functions (pdfs). A particle filter is typically used when solving Bayesian estimation problems wherein a desired pdf is intractable (has no closed form solution), or is non-Gaussian/multi-modal. These types of problems occur in many areas of open-ended research such as, in this case, nonlinear system identification.

The principle behind the particle filter is to perform Monte Carlo integration of a pdf using a large amount of randomly distributed *particles*. The particles are simply discrete random variables whose probability distribution resembles the desired distribution. If an infinite number of particles are used, the Strong Law of Large Numbers (Equation 5.6) dictates that the particle filter approximation converges almost surely to the desired distribution.

State estimation is the process of estimating the state vector $\mathbf{x}$ for a state space model. State estimation is required when the true value of the states are unknown and can not be measured directly. The recursive Bayesian estimator is the basis of several methods that perform state estimation in a probabilistic framework such as Kalman Filters, Particle Filters ans Expectation Maximisation algoritms.

In general, the recursive Bayesian state estimator takes the form of Algorithm 1.

Algorithm 1: Recursive Bayesian Estimator

---

1. Initialisation: Choose the prior for the state (initial uncertainty)

$$p(\mathbf{x}_0|\mathbf{y}_0) = p(\mathbf{x}_0) \tag{5.8}$$

2. **for** $t = 1, 2, ..., T$ **do**
    a. Correction

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})} \tag{5.9}$$

    b. Prediction

$$p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}) = \int p(\mathbf{x}_{t+1}|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t})d\mathbf{x}_t \tag{5.10}$$

The algorithm is initialised with an estimate for the state based on the prior, Equation 5.8. The correction step then uses information extracted from the latest measurement y_t using the process model, and updates the current state. The prediction step then generates an estimate for the state at time $t+1$ using the process model.

The resulting estimate is the combination of two sets of data; our belief of how the system will behave, and the measurement of its actual behavior. The merging of this data is often referred to as *data fusion*.

The prediction and correction can be combined into a single step by substituting Equation 5.9 into Equation 5.10, yielding a single step update (Wills, 2018),

$$p(\mathbf{x}_{t+1}|\mathbf{y}_t) = \int \frac{p(\mathbf{x}_{t+1}|\mathbf{x}_t)p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})}d\mathbf{x}_t. \tag{5.11}$$

A particle filter aims to approximate the integral in Equation 5.11 using Sequential Monte Carlo. This is advantageous compared to using other implementations of the Bayes filter such as the Kalman Filter (or the EKF/UFK), as the target distribution can be nonlinear, non-Gaussian or analytically intractable without breaking the method.

The single step prediction integral can be written in a form to accentuate the application of the SLLN Equation 5.6, such that it can be solved as an expectation integral. Samples drawn from the predicted state pdf $p(\cdot)$ are simply mapped through a function $f(\cdot)$ (Renton, 2017),

$$p(\mathbf{x}_{t+1}|\mathbf{y}_t) = \int \underbrace{\frac{p(\mathbf{x}_{t+1}|\mathbf{x}_t)p(\mathbf{y}_t|\mathbf{x}_t)}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})}}_{f(\cdot)}\underbrace{p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}_{p(\cdot)}d\mathbf{x}_t. \tag{5.12}$$

Applying the SLLN yields,

$$p(\mathbf{x}_{t+1}|\mathbf{y}_t) = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\frac{p(\mathbf{x}_{t+1}|\mathbf{x}_t^{(i)})p(\mathbf{y}_t|\mathbf{x}_t^{(i)})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})}, \qquad \mathbf{x}_t^{(i)} \sim p(\mathbf{x}_t|\mathbf{y}_{1:t-1}), \tag{5.13}$$

$$p(\mathbf{x}_{t+1}|\mathbf{y}_t) = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\underbrace{\frac{p(\mathbf{y}_t|\mathbf{x}_t^{(i)})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})}}_{w_t^{(i)}}p(\mathbf{x}_{t+1}|\mathbf{x}_t^{(i)}), \qquad \mathbf{x}_t^{(i)} \sim p(\mathbf{x}_t|\mathbf{y}_{1:t-1}). \tag{5.14}$$

The *likelihood weights* $w_t^{(i)}$ are a sequence of random variables representing the relative importance of each particle. In this sense the approximation can be viewed as a mixture distribution,

$$p(\mathbf{x}_{t+1}|\mathbf{y}_t) = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}w_t^{(i)}p(\mathbf{x}_{t+1}|\mathbf{x}_t^{(i)}), \qquad \mathbf{x}_t^{(i)} \sim p(\mathbf{x}_t|\mathbf{y}_{1:t-1}). \tag{5.15}$$

The weights define a categorical distribution $\mathcal{C}(\{w_t^i\}_{i=1}^N)$ from which integers can be sampled. These integers correspond to an index for the set of particles from which the updated particle samples at the next time-step can be drawn. Drawing from this mixture distribution is equivalent to what is commonly known as a *re-sampling* step.

In practice, there will always be a finite number of N particles. This modifies our estimate to a finite sum

$$p(\mathbf{x}_{t+1}|\mathbf{y}_t) \approx \frac{1}{N}\sum_{i=1}^{N}w_t^{(i)}p(\mathbf{x}_{t+1}|\mathbf{x}_t^{(i)}), \tag{5.16}$$

$$\approx \sum_{i=1}^{N}\tilde{w}_t^{(i)}p(\mathbf{x}_{t+1}|\mathbf{x}_t^{(i)}), \tag{5.17}$$

where the weights have been normalised to equal one, such that,

$$\sum_{i=1}^{N} \tilde{w}_t^{(i)} = 1. \tag{5.18}$$

This is achieved by dividing each individual weight by the sum of all N weights,

$$\tilde{w}_t^{(i)} = \frac{w_t^{(i)}}{\sum_{j=1}^{N} w_t^{(j)}}. \tag{5.19}$$

Any proposal distribution from which it is possible to draw samples can be used as the prediction density when using the mixture distribution approach due to the use of the likelihood weights. The weights encode the likelihood of those particles relative to the true prediction density. The use of these weights is often referred to as *importance sampling*, and the resulting particle filter is known as the *conditional particle filter*.

A state space model can be written in the form of the *standard proposal* for a particle filter, encoding the process and measurement models and their associated noise as the mean and covariance of normal distributions,

$$p(x_{t+1}|\theta, x_t) = \mathcal{N}(x_{t+1}; \ Ax_t + Bu_t, Q), \tag{5.20}$$
$$p(y_t|\theta, x_t) = \mathcal{N}(y_t; \ Cx_t + Du_t, r). \tag{5.21}$$

These distributions are a natural choice for the prediction and measurement densities of the particle filter.

Particles located in areas of low relevance will quickly have a negligible weight after a few iterations. This results in the filter having few effective particles and is known as *degeneracy*. Avoiding degeneracy is desirable to ensure the particles used are most effective and computational power is not wasted. A common technique to achieve this is *resampling*.

Resampling is the process of replacing ineffective particles with new particles belonging to areas of more relevance as determined by the strength of the weights. The particles which have higher importance weights are more likely to survive, while those with low weights are more likely to perish. This leads to the set naturally distributing themselves to important areas of the distribution.

Drawing a sample from the mixture distribution in Equation 5.17 simplifies the process of resampling in the particle filter to the following process (Renton, 2017),

1. Draw an integer j such that $\qquad\qquad\qquad P(j = i) = \tilde{w}^{(i)}, \tag{5.22}$

2. Generate a sample according to $\qquad\qquad\qquad \mathbf{x}^{(i)} \sim p(\mathbf{x}_{t+1}|\mathbf{x}_t^{(j)}), \tag{5.23}$

3. Repeat $\tag{5.24}$

This leads to the implementation of the Particle Filter in Algorithm 2 (Renton, 2017).

Algorithm 2: Particle Filter

---

1. Initialisation: Choose the prior for the state (initial uncertainty)

$$\mathbf{x}_1^{(i)} \sim p(\mathbf{x}_1|\mathbf{y}_0) = \sum_{i=1}^{N} \tilde{w}_t^{(i)} p(\mathbf{x}_1|\mathbf{x}_0^{(i)}), \qquad \sum_{i=1}^{N} \tilde{w}_t^{(i)} = 1. \tag{5.25}$$

2. **for** $t = 1, 2, ..., T$ **do**
   a. Update normalised weights according to the measurement likelihood

$$\tilde{w}_t^{(i)} = \frac{p(\mathbf{y}_t|\mathbf{x}_t^{(i)})}{\sum_{j=1}^{N} p(\mathbf{y}_t|\mathbf{x}_t^{(j)})} \tag{5.26}$$

   b. Draw samples from the prediction density

$$\mathbf{x}_{t+1} \sim p(\mathbf{x}_{t+1}|\mathbf{y}_t) \approx \sum_{i=1}^{N} \tilde{w}_t^{(i)} p(\mathbf{x}_{t+1}|\mathbf{x}_t^{(i)}) \tag{5.27}$$

---

6. Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) is a sampling method which characterises a distribution without knowledge of all the distribution's mathematical properties. This is achieved using Monte Carlo techniques to draw random samples from the distribution and using the SLLN to approximate the expected value. The principle behind Markov chain Monte Carlo is to simulate an ergodic Markov chain such that the limiting distribution π converges to a target distribution of interest as t approaches infinity, regardless of the initial state (Robert and Casella, 2013).

Definition 6.1 *A Markov chain Monte Carlo method which aims to simulate a distribution π is any method producing an ergodic Markov chain (X^t), such that the stationary distribution of the Markov chain is π.*

6.1. Markov Chains

A Markov chain can be thought of as a sequence of dependent random variables evolving over time that does not exhibit a memory effect. This lack of memory means the future of the process is independent of all previous values, and the probability of any given transition depends only on its present value (Meyn and Tweedie, 2012). This is referred to as the *Markov property*.

The *transition kernel* or *Markov kernel* is the conditional probability distribution determining the probability of the value of the next random variable in the process. Literature on the study of Markov chains often uses notation that defines a Markov transition kernel $K(X_n, X_{n+1})$, however this report notates a Markov transition kernel as a conditional distribution $k(x_{t+1}|x_t)$ due to familiarity in the context of Bayesian probability.

Definition 6.2 *The sequence of random variables $x_0, x_1, ...$ is called a Markov Chain if for every index $k > 0$, and for all t*

$$k(x_{t+k}|x_{1:t}) = k(x_{t+k}|x_t) \tag{6.1}$$

A simple illustration of a Markov chain is a *random walk*. The Markov chain satisfies

$$x_{t+1} = x_t + \epsilon_t \tag{6.2}$$

where $\epsilon_t \sim \mathcal{N}(0,1)$ is i.i.d noise. As this noise is independent, the transition kernel $k(x_{t+1}|x_t)$ corresponds to a normal distribution $\mathcal{N}(x_t, 1)$. The random walk kernel satisfies the Markov property as the transition probability at any time step is conditional only on the current state x_t.

Markov chains can have very strong stability properties that are exploited in the design of MCMC methods. A *stationary* distribution has exists by construction in MCMC. That is, the MCMC sampler constructs a stationary distribution $\pi(x)$ such that if a sample from the method is distributed according to $X_n \sim \pi$, then another sample will be similarly distributed giving $X_{n+1} \sim \pi$.

Formally the kernel and stationary distribution then satisfy the following definition:

Definition 6.3 *Let $k(y|x)$ be a Markovian transition kernel, and let $\pi(x)$ be some probability density function. k is said to leave π invariant if*

$$\int k(y|x)\pi(x)dx = \pi(y) \tag{6.3}$$

A stationary distribution imposes the constraint that the kernel must satisfy *irreducibility*, requiring the kernel to allow free movement over the entire state space. The existence of a stationary distribution also ensures that most chains involved in MCMC are *recurrent*, visiting an arbitrary set A an infinite amount of times. A recurrent chain the stationary distribution is also the *limiting* distribution. This has the important consequence that no matter the starting value x_0, the sequence x_t has a positive probability of eventually reaching any region of the state space (Robert et al., 2010).

A Markov chain that is stationary, irreducible and recurrent is *ergodic*. The ergodic theorem is the key property that allows MCMC to operate. If a given kernel k produces an ergodic Markov chain with stationary distribution π, sampling random variables X from this kernel k will eventually produce samples distributed according to π (Robert et al., 2010). In particular, this allows the SLLN to apply where,

$$\frac{1}{T}\sum_{t=1}^{T} h(X^t) \to E[h(X)], \tag{6.4}$$

and is often called the *Ergodic Theorem*.

6.2. Gibbs Sampling

Gibbs sampling is a class of MCMC algorithms that samples a multi-dimensional target distribution by separating it into a series of smaller problems of lower dimensionality. Suppose the random variable $\mathbf{X}$ has dimensionality $k > 1$, such that each sample can be split such that $\mathbf{x}^{(t)} = [x_1^{(t)}, \ldots, x_k^{(t)}]$. Gibbs sampling draws separate samples for each component of $\mathbf{x}$, alternately conditioning each sample on the remaining $k - 1$ dimensions.

New information gained from drawing a new sample is used immediately as the algorithm proceeds in its iterations. The next component to be sampled is conditioned on the latest values available for all other dimensions. Algorithm 3 demonstrates operation of a Gibbs sampler applicable to any dimensionality $k > 1$. The second step

---

Algorithm 3: Generalised Gibbs Sampler

Given $\mathbf{x}^{(t)}$, generate samples
for $t = 1, \ldots, T$ **do**
$\quad X_1^{t+1} \sim p_1(x_1 | x_2^{(t)}, \ldots, x_k^{(t)})$
$\quad X_2^{t+1} \sim p_2(x_2 | x_1^{(t+1)}, x_3^{(t)}, \ldots, x_k^{(t)})$
$\quad \ldots$
$\quad X_k^{t+1} \sim p_k(x_k | x_1^{(t+1)}, \ldots, x_{k-1}^{(t+1)})$
return $\mathbf{x}^{1:T}$

To illustrate, first consider a two stage Gibbs sampler. If two random variables X and Y have joint density $f(x, y)$, with conditional densities $f_{Y|X}$ and $f_{X|Y}$, the two stage Gibbs sampler generates a Markov chain (X_t, Y_t) through following the steps in Algorithm 4.

Algorithm 4: Two Stage Gibbs sampler

Initialise $X_0 \leftarrow x_0$
for $t = 1, \ldots, T$ **do**
$\quad$ Draw $Y_t \sim f_{Y|X}(\cdot | x_{t-1})$
$\quad$ Draw $X_t \sim f_{X|Y}(\cdot | y_t)$

Implementing this Gibbs sampler is simple with the assumption that it is possible to draw samples from both true conditional distributions. When each sample is drawn from the true conditionals, the sequence (X_t, Y_t) and (X_{t+1}, Y_{t+1}) create an ergodic Markov chain as they are both distributed from f. Therefore convergence of the method is ensured if the supports of the conditionals are connected (Robert et al., 2010).

The Bivariate normal distribution has a two dimensional joint distribution $p(x, y)$ which can be written in the form:

$$p(x, y) \sim \mathcal{N}\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right) \tag{6.5}$$

The conditionals for this matrix are available in closed form and the Gibbs sampler targeting the joint distribution $f(x, y)$ is constructed in Algorithm 5.

Algorithm 5: Bivariate Normal Gibbs Sampler

Given y_t, generate samples
for $t = 1, \ldots, T$ **do**
$\quad X_{t+1} | y_t \sim \mathcal{N}(\rho y_t, 1 - \rho^2)$
$\quad Y_{t+1} | x_t \sim \mathcal{N}(\rho x_{t+1}, 1 - \rho^2)$
return $X_{1:t}, Y_{1:T}$

The resulting approximation of the Bivariate distribution with a cross-covariance of $\rho = 0.8$ generated by the Gibbs sampler of Algorithm 5 is shown in Figure 6 with $t = 3,000$ samples.
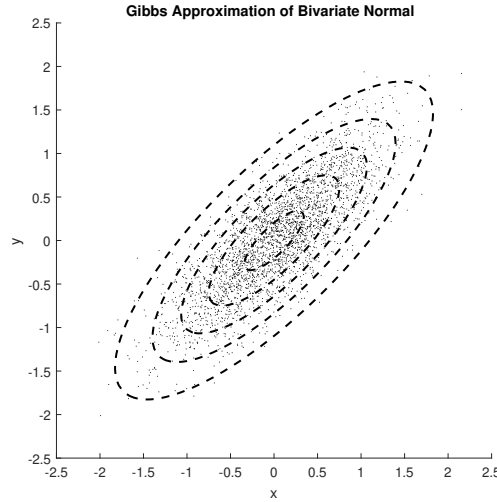
Figure 6: Gibbs Approximation of Bivariate

7. Gaussian Processes

Regression is the task of estimating the relationship between variables in a set of data and making predictions of continuous quantities. A familiar example of regression is computing the *line of best fit* to a noisy data set. Regression is a suitable method for identifying a possible function $y = h(z)$, describing the static nonlinearity of a Wiener model.

The Gaussian process (GP) in a one dimensional regression problem describes a distribution over functions. The strength of using a Gaussian process over a parametric model is the ability to elegantly deal with uncertainty about the structure of the function and perform regression without imposing structure. In a parametric model, an identified function $f(x)$ can only ever be as accurate as the assumed parametric structure, which may be too restrictive. Instead, with the Gaussian process model the function can be inferred directly, allowing it to let the data speak for itself and represent the underlying function.

A Gaussian process is defined by its mean function $m(\mathbf{x})$ and covariance function $k(\mathbf{x}, \mathbf{x}')$ which completely specifies the processes' behaviour. For a process $f(\mathbf{x})$ the Gaussian process is written as

$$f(x) \sim \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')\right). \tag{7.1}$$

While an ordinary probability distribution describes stochastic variables, a Gaussian process governs the properties of functions. A Gaussian process is a collection of random variables, any finite number of which have a multivariate Gaussian distribution (Rasmussen and Williams, 2006). In the regression case, the random variables represent the value of the function $f(x)$ at location x. A Gaussian process is a generalization of a multivariate Gaussian distribution to infinite dimensionality.

A convenient view is to think of the GP as a distribution over functions, where inference occurs in the *function space*. The *prior* distribution in a GP gives a prior probability to every possible function, where functions are considered more likely due to the display of desired characteristics such as being smooth, stationary, periodic, etc. The desired properties represent prior knowledge or assumptions

20

about the physical process being learned. The covariance function and its hyperparameters encode the prior.

The GP calculates a function mapping inputs $\mathbf{x} \in \mathcal{X}$ to an output $y = f(\mathbf{x})$ using observed pairs of data $\mathcal{D} = \{(y_i, \mathbf{x}_i) | i = 1, ..., m\}$ where m is the number of samples. Let $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_m]^T$ and $\mathbf{Y} = [y_1, ..., y_m]^T$ represent pairwise inputs of data.

Let the likelihood (sampling distribution) describing random aspects of the data-generating process be of the form

$$p(y|f(\mathbf{x}), \theta), \tag{7.2}$$

where θ holds any additional parameters which condition the distribution.

The Gaussian process model places a GP prior on the generative process f, putting a prior on function values directly. Each input position $x \in \mathcal{X}$ has an associated random variable $f(x)$. The GP prior on f means the prior joint distribution of a collection of function values $\mathbf{f} = [f(x_1), ..., f(x_2)]^T$ associated with the collection of inputs is a multivariate Gaussian

$$p(\mathbf{f}|\mathbf{X}) = \mathcal{N}(\mathbf{f}|\mathbf{m}, \mathbf{K}), \tag{7.3}$$

with mean $\mathbf{m}$ and covariance matrix $\mathbf{K}$. $\mathbf{K}$ is assembled such that $K_{ij} = k(\mathbf{x_i}, \mathbf{x_j})$ and $\mathbf{m} = [m(\mathbf{x_1}), ..., m(\mathbf{x_m})]^T$.

The likelihood of f given $\mathcal{D}$ factorises, and depends on f only through its value $\mathbf{f}$ at the observed inputs giving

$$p(\mathbf{y}|f, \theta) = \prod_{i=1}^{m} p(y_i | f(\mathbf{x_i}), \theta) = p(\mathbf{y}|\mathbf{f}, \theta) \tag{7.4}$$

Conditioning the likelihood on observed inputs $\mathbf{f}$ is equivalent to conditioning over the full function f, and is the key to Gaussian processes possessing the ability to capture an infinite dimensional process while only requiring inference over finite dimensional quantities. The posterior distribution of the function values $\mathbf{f}$ is given by Bayes's Rule

$$p(\mathbf{f}|\mathcal{D}, \theta) = \frac{p(\mathbf{y}|\mathbf{f}, \theta)p(\mathbf{f}|\mathbf{X})}{p(\mathcal{D}|\theta)} = \frac{\mathcal{N}(\mathbf{f}|\mathbf{m}, \mathbf{K})}{p(\mathcal{D}|\theta)} \prod_{i=1}^{m} p(y_i | f(\mathbf{x_i}), \theta). \tag{7.5}$$

The result of the Gaussian process does not yield a closed form expression for a single function, but instead requires computation of the posterior predictive distribution $f(\mathbf{x}_*)$ point-wise for *test inputs* denoted $\mathbf{x}_*$. The vector of test inputs is therefore denoted $\mathbf{X}_*$, and the vector of function values for test inputs $\mathbf{f}_*$.

To obtain the predictive distribution $\mathbf{f}_*$, integration is performed over the posterior distribution

$$p(\mathbf{f}_*|\mathcal{D}, \mathbf{X}_*, \theta) = \int p(\mathbf{f}_*|\mathbf{f}, \mathbf{X}, \mathbf{X}_*)p(\mathbf{f}|\mathcal{D}, \theta)d\mathbf{f}, \tag{7.6}$$

where $p(\mathbf{f}_*|\mathbf{f}, \mathbf{X}, \mathbf{X}_*)$ describes the dependency of $\mathbf{f}_*$ on $\mathbf{f}$ due to the GP prior.

The joint prior distribution of $\mathbf{f}$ and $\mathbf{f}_*$ is multivariate normal

$$p(\mathbf{f}, \mathbf{f}_*|\mathbf{X}, \mathbf{X}_*) = \mathcal{N}\left(\begin{bmatrix} \mathbf{f} \\ \mathbf{f}_* \end{bmatrix} \middle| \begin{bmatrix} \mathbf{m} \\ \mathbf{m}_* \end{bmatrix}, \begin{bmatrix} \mathbf{K}(\mathbf{X}, \mathbf{X}) & \mathbf{K}(\mathbf{X}, \mathbf{X}_*) \\ \mathbf{K}(\mathbf{X}_*, \mathbf{X}) & \mathbf{K}(\mathbf{X}_*, \mathbf{X}_*) \end{bmatrix}\right). \tag{7.7}$$

For simplicity, the notation for each component of the covariance matrix is compressed where

$$[K]_{ij} = k(x_i, x_j), \qquad\qquad i, j = 1, \ldots, T, \qquad\qquad (7.8)$$

$$[K_*]_{ij} = k(x_i, x_j^*), \qquad\qquad i = 1, \ldots, T, j = 1, \ldots, M, \qquad\qquad (7.9)$$

$$[K_{**}]_{ij} = k(x_i^*, x_j^*), \qquad\qquad i, j = 1, \ldots, M. \qquad\qquad (7.10)$$

Due to the requirement of the covariance functions to be positive semi-definite, the covariance matrix K is also positive semi-definite. As such, $K_*^T = K(X_*, X)$, and the prior can be written as

$$p(\mathbf{f}, \mathbf{f}_* | \mathbf{X}, \mathbf{X}_*) = \mathcal{N}\left(\begin{bmatrix} \mathbf{f} \\ \mathbf{f}_* \end{bmatrix} \middle| \begin{bmatrix} \mathbf{m} \\ \mathbf{m}_* \end{bmatrix}, \begin{bmatrix} \mathbf{K} & \mathbf{K}_* \\ \mathbf{K}_*^T & \mathbf{K}_{**} \end{bmatrix} \right). \qquad (7.11)$$

To obtain the posterior distribution over functions, the joint prior must be restricted only to functions which agree with observed data points. Due to the probabilistic framework, this is very efficient, and corresponds to the formulaic conditioning of the joint prior on the observations. This yields the conditional distribution which can be used for prediction,

$$p(\mathbf{f}_* | \mathbf{f}, \mathbf{X}_*, \mathbf{X}) = \mathcal{N}(\mathbf{f}_* | \mu_*, \mathbf{\Sigma}_*), \qquad (7.12)$$

where the mean $\mu*$ and covariance Σ_* are given by

$$\mu_* = \mathbf{m}_* + \mathbf{K}_*^T \mathbf{K}^{-1}(\mathbf{f} - \mathbf{m}), \qquad (7.13)$$

$$\mathbf{\Sigma}_* = \mathbf{K}_{**} - \mathbf{K}_*^T \mathbf{K}^{-1} \mathbf{K}_*. \qquad (7.14)$$

Generating samples from the mean and covariance of the conditional distribution in Equation (7.12) allows for sampling of function values $\mathbf{f}_*$ for test inputs $\mathbf{X}_*$. This result is essential for the application of the Gaussian process for performing nonparametric function regression.

Addition of noise to the inference of the function in GP regression is analytically tractable under the assumption that the noise is additive independent Gaussian such that

$$y = f(x) + \epsilon, \qquad\qquad \epsilon \sim \mathcal{N}(0, \sigma_n^2). \qquad (7.15)$$

A convenient method of handling this noise is to introduce it into the covariance matrix such that

$$cov(\mathbf{y}) = \mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma_n \mathbf{I}, \qquad (7.16)$$

where I is the identity matrix, adding the noise independently to each measurement. Rewriting the joint distribution in terms of the observed output and the testing inputs gives

$$p(\mathbf{y}, \mathbf{f}_* | \mathbf{X}, \mathbf{X}_*) = \mathcal{N}\left(\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \middle| \begin{bmatrix} \mathbf{m} \\ \mathbf{m}_* \end{bmatrix}, \begin{bmatrix} \mathbf{K} + \sigma_n^2 \mathbf{I} & \mathbf{K}_* \\ \mathbf{K}_*^T & \mathbf{K}_{**} \end{bmatrix} \right). \qquad (7.17)$$

Finally, the resultant predictive distribution is found through conditioning of the joint prior, yielding addition of noise in the predicted mean and covariance

$$p(\mathbf{f}_* | \mathbf{f}, \mathbf{X}_*, \mathbf{X}) = \mathcal{N}(\mathbf{f}_* | \mu_*, \mathbf{\Sigma}_*) \qquad (7.18)$$

where the mean μ_* and covariance Σ_* are given by

$$\mu_* = \mathbf{m}_* + \mathbf{K}_*^T(\mathbf{K} + \sigma_n^2\mathbf{I})^{-1}(\mathbf{f} - \mathbf{m}), \tag{7.19}$$

$$\mathbf{\Sigma}_* = \mathbf{K}_{**} - \mathbf{K}_*^T(\mathbf{K} + \sigma_n^2\mathbf{I})^{-1}\mathbf{K}_* \tag{7.20}$$

Restating Equation 7.18 with the expansion of μ_* and Σ_* yields the final predictive distribution

$$p(\mathbf{f}_*|\mathbf{f}, \mathbf{X}_*, \mathcal{D}) = \mathcal{N}(\mathbf{f}_*|\mathbf{K}_*^T(\mathbf{K} + \sigma_n^2\mathbf{I})^{-1}\mathbf{y}, \mathbf{K}_{**} - \mathbf{K}_*^T(\mathbf{K} + \sigma_n^2\mathbf{I})^{-1}\mathbf{K}_*) \tag{7.21}$$

7.1. Covariance Functions (Kernels)

The covariance function of a Gaussian process represents the dependencies between function values for any pair of inputs $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$, describing the strength and direction of linear dependence, similarity, or mutual informativeness of the function values $f(x)$ and $f(x')$ as a function of the corresponding inputs (Kuss, 2006). Any valid covariance function must produce a positive semi-definite matrix to ensure the existence of all finite-dimensional distributions. A function is said to be positive semi-definite on $\mathcal{X} \times \mathcal{X}$ if

$$\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j k(\mathbf{x_i}, \mathbf{x_j}) = \alpha^{\mathbf{T}}\mathbf{K}\alpha \geq 0 \tag{7.22}$$

holds for any choice of $m, \alpha \in \mathbb{R}^m$, and $\mathbf{x_1}, ..., \mathbf{x_m} \in \mathcal{X}$.

Addition of function or *process noise* can be included in the covariance function through use of the form

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 c(\mathbf{x}, \mathbf{x}') \tag{7.23}$$

where $|c(\mathbf{x}, \mathbf{x}')| \leq 1$ is a positive semi-definite covariance function and $\sigma_f^2 > 0$ is the *signal variance*. This value represents the variance of the noise-free signal, not the noise present on measurements of the output σ_n, introduced in Equation (7.16).

Covariance functions are *stationary* if they are invariant under any translation $\mathbf{t} \in \mathcal{X}$ such that $k(\mathbf{x}, \mathbf{x}') = k(\mathbf{x} + \mathbf{t}, \mathbf{x}' + \mathbf{t})$. Thus a GP is described as stationary if its covariance function is stationary with a constant mean function. The use of stationary covariance functions is often reasonable, and makes the assumption that the covariance between function values depends only on the difference between inputs $x - x'$, and decays with distance.

Naive use of arbitrary covariance functions can yield impressive results, and there is reasonable justification for the use of the squared exponential kernel and the Matérn kernel (Rasmussen and Williams, 2006; Stein, 2012). However, if dynamics or underlying physical processes of the system such as periodicity are discerned prior to construction of the GP, a covariance function can be designed or chosen to best utilise this knowledge to fit the function distribution.

There are many useful covariance functions commonly used in the context of GP regression, forming a general purpose toolbox to chose from. Of particular interest are the squared exponential kernel, and the Matérn kernel, two of the most popular choices for general purpose GP regression.

Definition 7.1 (Squared Exponential)

$$cov(f(\mathbf{x}), f(\mathbf{x}')) = k_{SE}(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{|\mathbf{x} - \mathbf{x}'|^2}{2}\right) \tag{7.24}$$

The squared exponential kernel can be shown to correspond to a Bayesian linear regression model which expands the input into a feature space defined by an infinite number of Gaussian basis functions due to Mercer's theorem, or an infinite number of Gaussian-shaped basis functions (Rasmussen and Williams, 2006). This draws parallels to the operation of the Support Vector Machine (SVM) where the higher dimensional kernel space is no longer only n-dimensional, but infinite-dimensional. The squared exponential kernel is infinitely differentiable, causing the predicted functions of a GP using the SE kernel to be smooth, as the GP will be mean-square differentiable for all orders (Rasmussen and Williams, 2006).

The squared exponential kernel is often found in the augmented form shown in Equation 7.25, with the addition of *hyperparameters* which allow tuning of its characteristics.

$$cov(f(\mathbf{x}), f(\mathbf{x}')) = k_{SE}(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{|\mathbf{x} - \mathbf{x}'|^2}{2l^2}\right) \tag{7.25}$$

The smoothness of predicted functions can be tuned using parameter l, *the characteristic length scale*, determining the mean number of zero crossings. An intuitive method of thinking about the length scale is as a correlation length, which controls the distance at which two observations are independent and no longer informative. Signal variance σ_f represents the variance of the process which generates the data used.

Definition 7.2 (Matérn kernel)

$$cov(f(\mathbf{x}), f(\mathbf{x}')) = k_{Matern}(x, x') = \alpha \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}\Delta x}{l}\right)^\nu K_\nu\left(\frac{\sqrt{2\nu}\Delta x}{l}\right), \tag{7.26}$$

where $\Delta x = |x - x'|$, Γ is the gamma function and K_ν is the modified Bessel function of the second kind of order $\nu > 0$.

The parameter ν controls the differentiability of the covariance function. Particular choice of ν can create other covariance functions as special cases. For instance, the squared exponential kernel can be shown to be calculated from the Matérn kernel as ν approaches ∞.

7.2. Hyperparameters

The choice of hyperparameters heavily influences the quality of the regression. If the parameters are chosen poorly or without consideration, the output will not best reflect the system due to an inherent tradeoff between discarding information from the data, or by introducing unnecessary complexity and over-fitting to the data. In this way, the covariance function and the hyperparameters both impose modelling assumptions.

A large focus of previous work has been placed on the characteristic length scale due to its presence in many kernels and the strong impact it has on generalisation. The length scale must allow the GP to capture sufficient complexity without overfitting of input data. There is also a relationship between small generalisation error and computational practicality due to numeric conditioning of the required kernel matrices depending on the scaling of the kernel (Schaback and Wendland, 2006).

:: Figure demonstrating the effect of different hyper parameters on function regression with noisy data ::

Many procedures have been investigated for determination of suitable hyperparameters in both numeric and probabilistic frameworks. A common approach is the use of optimisation methods to minimise a cost function such as squared error between the original output and proposed function (Rasmussen and Williams, 2006). Additional techniques have also been investigated such as leave-one-out cross validation (LOOCV) to help avoid over-fitting (Rippa, 1999). In a probabilistic sense, the desired goal is to maximise the likelihood of the output, conditioned on the hyper parameters $p(y|\eta, r, x_{1:T})$. This is can be achieved using an estimator such as a Metropolis Hastings method as described in Section 8.2.

8. Bayesian Semiparametric Model

A novel approach to the Wiener identification problem is proposed in the paper (Lindsten et al., 2013), contributing a methodology that utilises a Bayesian model of the Wiener system with an inference algorithm based on a PMCMC method named *particle Gibbs with ancestor sampling* (PG-AS). The Wiener system model is proposed in a *semiparametric* form, coined due to the mixture of a parametric state space model as the linear dynamical block, and a nonparametric Gaussian process for the static nonlinearity block of the Wiener system. The semiparametric Wiener model and PG-AS method described is implemented to perform nonlinear system identification of guitar amplifiers.

The block diagram of the Wiener system was introduced in Section 3, and is repeated here in Figure 7 for clarity.
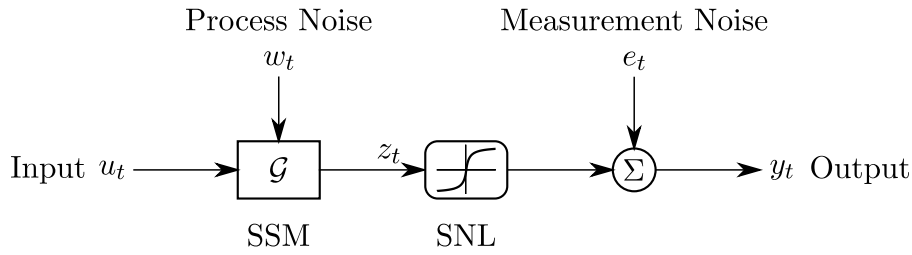

Figure 7: Block diagram of a Wiener system with linear dynamical system $\mathcal{G}$ followed by static non-linearity $h(\cdot)$ with process noise w_t and measurement noise e_t.

Let the target of inference be $\theta = \{G; h(\cdot)\}$, consisting of the parameters of the state space model G, and the nonlinear mapping function modelled by a Gaussian Process $h(\cdot)$. To perform system identification, the posterior probability of the model parameters $p(\theta|y_{1:t})$ is targeted by a multi stage Gibbs sampler.

A key challenge to inference of this system is drawing from the posterior state density $p(x_{1:T}|\theta, y_{1:T})$, which is not available in closed form due to the presence of the nonlinear mapping $h(\cdot)$ in the conditional parameters θ. Since sampling from the exact state posterior is not possible, an estimation method must be used.

PMCMC uses a particle filter to construct a proposal Markov kernel as part of a Gibbs sampling sequence. The use of the particle filter leaves the target posterior distribution invariant. The resultant Markov kernel is used in the Gibbs sampler without introducing systematic error into the result of the inference (Lindsten et al., 2013).

Calculation of the posterior pdf $p(\theta|y, u)$ is possible for the proposed Wiener model when solved with the PG-AS algorithm. The paper (Lindsten et al., 2013) was the first known to provide a posterior parameter density for the Wiener identification problem.

The state space form of the proposed system including nonlinearity is given by

$$x_{t+1} = Ax_t + Bu_t + w_t, \qquad\qquad w_t \sim \mathcal{N}(0, Q), \qquad\qquad (8.1)$$
$$z_t = Cx_t, \qquad\qquad\qquad\qquad\qquad\qquad (8.2)$$
$$y_t = h(z_t) + e_t, \qquad\qquad\qquad e_t \sim \mathcal{N}(0, r), \qquad\qquad (8.3)$$

where $x_t \in \mathbb{R}^{n_x}$ is the state of the dynamical system, $w_t \in \mathbb{R}^{n_x}$ is the process noise, $u_t \in \mathbb{R}^{n_u}$ is the input signal, $z_t \in \mathbb{R}$ is the output from the linear system, and $y_t \in \mathbb{R}$ is the measured final output of the system mapped through the static nonlinearity $h(z_t)$ with measurement noise e_t.

The guitar amplifier is a single input, single output system, with an unknown amount of states in the linear system. Observability of the linear system is assumed, and as such the matrix $C = [1 \ 0 \ \cdots \ 0]$ is fixed without loss of generality. Let $\Gamma = [A, B]$ denote the parameter matrices of the state space system. The identification problem is to identify the parameters Γ, Q, r, the nonlinear mapping $h(\cdot)$ and suitable hyperparameters.

8.1. Prior Distributions

If the order of the state space model n_x is known, *conjugate priors* are a suitable choice of prior on the matrices of the state space model and its noise covariance. For the parameters of the state space model Γ and process noise Q, a suitable conjugate prior is the matrix normal, inverse Wishart (MNIW) distribution. Let the prior $p(\Gamma, Q) = p(\Gamma|Q)p(Q)$ with distributions given by

$$p(\Gamma|Q) = \mathcal{MN}(\Gamma; M, Q, L), \tag{8.4}$$

$$p(Q) = \mathcal{IW}(Q; n_0, S_0), \tag{8.5}$$

where, $\mathcal{MN}()$ is a matrix normal density with mean M and left and right covariances L^{-1} and V, and $\mathcal{IW}()$ is an inverse Wishart distribution with n degrees of freedom and scale matrix S. The measurement noise r is also given a conjugate IW prior

$$p(r) = \mathcal{IW}(r; m_0, R_0). \tag{8.6}$$

When the exact order n_x of the system is unknown, it can be inferred as part of the system identification problem through careful construction of the prior distributions. Automatic order determination is achieved by over parameterisation the model and the choice of a *sparsity promoting prior* which 'switches off' the extra states which show no supporting evidence (Lindsten et al., 2013). The class of generalised hyperbolic (GH) distributions have been successfully used as sparsity-promoting priors for different purposes. The Laplace, normal inverse-Gaussian, and Student's t-distribution all fall under this category. The generalised inverse-Gaussian (GIG) belongs to the class of generalized hyperbolic (GH) distributions, proven to be effective for automatic order determination in the Wiener system identification problem (Lindsten et al., 2013). The density function for the GIG distribution is given by

$$\mathcal{GIG}(\tau, \nu, a, b) = \frac{(a/b)^{\nu/2}}{2K_\nu(\sqrt{ab})} \tau^{\nu-1} \exp(-\frac{1}{2}(a\tau + b\tau^{-1})), \tag{8.7}$$

where K_ν is a modified Bessel function of the 2nd kind. The hyper parameters must be set such that $a \geq 0, b \geq 0$ and $\nu \in \mathbb{R}$ to ensure the distribution is defined.

To achieve automatic order determination, the prior for the SSM parameters Γ is defined by placing independent, zero-mean Gaussian priors on the columns $\{\gamma_j\}_{j=1}^{n_x+n_u}$ such that

$$p(\Gamma|\bar{\tau}) = \prod_{j=1}^{n_x+n_u} \mathcal{N}(\gamma_j; 0, \tau_j \mathbf{I}_{n_x}), \tag{8.8}$$

where $\mathbf{I}_{n_x}$ is an $n_x \times n_x$ identity matrix and the vector $\bar{\tau} = \{\tau_j\}_{j=1}^{n_x+n_u}$ contains hyperparameters governing the variances of each of the columns. Each column is assigned an independent GIG prior

$$p(\tau_j) = \mathcal{GIG}(\tau_j; \nu, a, b), \tag{8.9}$$

for $j = 1, \ldots, n_x + n_u$.

With this construction, the resulting distributions of the columns of Γ will have peaks at the origin. If there is insufficient evidence to indicate a particular state/input is non-zero, the variance parameter τ corresponding to that state will be driven to zero, and in turn, drive the corresponding column of Γ to zero. This results in a matrix with a sparse column structure, where components which have been driven to zero can be discarded from the final model. These discarded components are unobservable from the perspective of the model and are not modes of the true system but are present in the model structure only (Lindsten et al., 2013).

The hyper parameters of these priors can be chosen empirically using the process outlined by (Fox, 2009), where the observed empirical covariance of the output measurements $y_{1:T}$ determines the means n_0, m_0 and covariances S_0, R_0.

8.2. Particle Gibbs with Ancestral Sampling

Data collected from experiments performed on a system are stored as pairwise input/output samples. These act as training pairs for inference, allowing comparison between the real system and the predicted estimate from the model. Let $\Pi \triangleq \{\Gamma, Q, r, \bar{\tau}\}$ if performing automatic order determination, or $\Pi \triangleq \{\Gamma, Q, r\}$ if the model order is known. The variables to be predicted include all parameters of the model such that $\theta \triangleq \{\Pi, \eta, h(\cdot)\}$. The GP $h(\cdot)$ is included as although it is not a parameter, the result of the GP function regression is a necessary piece of the model to be estimated.

The desired target of the PMCMC estimator is the posterior probability $p(\theta|y)$. This is obtained by applying the PG-AS inference method, targeted on the intractable joint posterior distribution of the parameters and states given the measurements of the true system,

$$p(\theta, x|y) = p(x|\theta, y)p(\theta|y), \tag{8.10}$$

and marginalising over the states. Conditioning on the input data u is implicit and is omitted from the notation for simplicity.

The basis of Gibbs sampling is to divide the components of a joint density into its conditionals and update each component element-wise using samples from a conditional distribution, excluding the active element of each sample (Kuss, 2006). To target the posterior density $p(\theta, x|y)$, the multi-stage Gibbs sampler iterates through the sampling process of:

$$\text{Draw } \Pi^* \sim p(\Pi^*|h, x_{1:T}, y_{1:T}), \tag{8.11}$$

$$\text{Draw } \eta^* \sim p(\eta^*|\Pi^*, x_{1:T}, y_{1:T}), \tag{8.12}$$

$$\text{Draw } h^* \sim p(h^*|\eta^*, \Pi^*, x_{1:T}, y_{1:T}), \tag{8.13}$$

$$\text{Draw } x_{1:T}^* \sim p(x_{1:T}^*|\Pi^*, \eta^*, h^*, y_{1:T}). \tag{8.14}$$

The last step of the sampler, Equation 8.14, requires sampling a state trajectory from the joint smoothing density $p(x_{1:T}^*|\theta, y_{1:T})$. This distribution is not available in closed form as it is conditional on θ, containing the GP nonlinearity $h(\cdot)$. This renders the inference problem intractable, even if the

parameters of the model were fixed. To overcome this, a particle filter is employed to estimate the expected value of the state trajectory. The use of a particle filter estimate inside the Gibbs sampling steps means the method becomes a *particle Gibbs* sampler.

The particle Gibbs sampler used is an extension on the ideal Gibbs sampler targeting the desired intractable joint density, employing a particle filter to construct a valid Markov kernel for the sampler. The family of Markov kernels desired is

$$\{M_\theta : \theta \in \Theta\}, \tag{8.15}$$

such that for each θ, $M_\theta(x_{1:T}|x'_{1:T})$ leaves the joint smoothing density $p(x_{1:T}|\theta, y_{1:T})$ invariant (Lindsten et al., 2013).

A conditional particle filter with ancestor sampling (CPF-AS) is chosen as it leaves the target distribution invariant for any number of particles $N \geq 2$, defining an irreducible and aperiodic Markov kernel (Lindsten et al., 2014).

In a CPF-AS, the Nth particle is specified a-priori at each time step from a reference trajectory denoted $x'_{1:T} = (x'_1, \ldots, x'_T)$ with a corresponding ancestral path $a^N_{1:T}$. This serves the purpose of ensuring distribution of particles in a relevant region of the state space and ensuring it defines an invariant distribution within the MCMC sampler (Lindsten et al., 2013). The CPF-AS can be viewed with the mixture distribution perspective of the particle filter discussed in Section 5.2.

To implement ancestor sampling, instead of drawing N independent samples from the proposal density

$$x^i_t \sim q(x_t|\theta, x^{a^i_t}_{t-1}), \tag{8.16}$$

samples are drawn for only $i = 1, \ldots, N-1$. The Nth particle is then set deterministically by directly using the particle from the reference trajectory $x^N_t = x'_t$.

To construct an artificial history for the Nth particle, the particle must be associated with an ancestor from time $t - 1$. This is achieved with the *ancestor sampling* step, setting the ancestor index a^N_t by drawing a sample from the likelihood of the previous i'th particle x^i_{t-1} becoming x'_t as it moved forward in time. This is equivalent to performing a single backwards simulation step to sample from

$$\hat{p}(x_{t-1}|\theta, x'_t, y_{1:T}) = \sum_{i=1}^N w^{a^N_t}_{t-1|t} p(x'_t|\theta, x^{a^N_t}_{t-1}).$$

The ancestor index a^N_t is then drawn from the categorical distribution

$$a^N_t \sim (P(a^N_t) = w^{a^N_t}_{t-1|t}).$$

The implication of this is such that after a complete pass, the entire reference trajectory is contained in the Nth sample, ensuring $x^N_{1:T} = x'_{1:T}$. By pre specifying this collection of particles, the CPF-AS "defines an *irreducible and aperiodic* Markov kernel M^N_θ with invariant distribution $p(x_{1:T}|\theta, y_{1:T})$" (Lindsten et al., 2013).

The use of ancestor sampling can be thought of as conceptually similar to the use of an entire backward simulation pass, such as done in a particle smoother. Ancestor sampling is advantageous compared to using a backwards smoother as it lowers the computational complexity of the algorithm

After a complete pass of the particle filter, a full trajectory $x^*_{1:T}$ is chosen from amongst the particles as the state estimate for the Gibbs sampling loop. This is done at the final time step $t = T$ drawing an index J from the final particle weights where

$$J \sim (P(j = i) = w^i_T), \tag{8.17}$$

and returning the chosen trajectory by taking the particles corresponding to that index,

$$x^*_{1:T} = x^J_{1:T}. \tag{8.18}$$

The procedure for the CPF-AS is given in Algorithm 6, as outlined in (Lindsten et al., 2013).

Algorithm 6: Conditional Particle Filter with Ancestor Sampling (CPF-AS)

---

1. Initialisation:
a. Draw $x^i_1 \sim q(x_1 | \theta y_1)$ for $i = 1, \ldots, N - 1$.
b. Set $x^N_1 = x'_1$.
c. **for** $i = 1, \ldots, N$ **do**

$$w^i_1 = \frac{p(y_1 | \theta, x^i_1) p(x^i_1)}{q(x^i_1 | \theta, y_1)}. \tag{8.19}$$

d. Normalise the weights such that $\sum_{i=1}^{N} w^i_1 = 1$.
2. **for** $t = 2, \ldots, T$ **do**
   a. Draw an integer a^i_t such that $P(a^i_t = j) = w^j_{t-1}$ for $i = 1, \ldots, N - 1$.
   b. Draw $x^i_t \sim q(x_t | \theta, x^{a^i_t}_{t-1}, y_t)$ for $i = 1, \ldots, N - 1$.
   c. Draw $a^N_t \sim P(a^N_t = j) = w^j_{t-1} p(x'_t | \theta, x^j_{t-1})$.
   d. Set $x^N_t = x'_t$.
   e. **for** $i = 1, \ldots, N$ **do**

$$w^i_t = \frac{p(y_t | \theta, x^i_t) p(x^i_t | \theta, x^{a^i_t}_{t-1})}{q(x^i_t | \theta, x^{a^N_t}_{t-1}, y_t)}. \tag{8.20}$$

   f. Normalise the weights such that $\sum_{i=1}^{N} w^i_t = 1$.
3. Draw J such that $P(J = i) = w^i_T$.
return $x^*_{1:T} = x^J_{1:T}$.

8.3. Posterior Distributions

If the order of the model is unknown and automatic order determination is used with the associated GH prior, there is no closed-form expression for the posterior distribution of Γ, Q and $\bar{\tau}$. The problem of finding the posterior is split into the appropriate sampling problems,

$$\text{Draw } \Gamma^* \sim p(\Gamma | Q, \bar{\tau}, x_{1:T}), \tag{8.21}$$
$$\text{Draw } Q^* \sim p(Q | \Gamma^*, x_{1:T}), \tag{8.22}$$
$$\text{Draw } \bar{\tau}^* \sim p(\bar{\tau} | \Gamma^*). \tag{8.23}$$

Finding the posterior of Γ requires re-arranging the state-update into a vector operation where the states and inputs are concatenated, and the relation is described as

$$x_{t+1} = \left[\bar{x}_{t,1} I_{n_x}, \ldots, \bar{x}_{t,(n_x+n_u)} I_{n_x} \right] \text{vec}(\Gamma) + w_t,$$

where $\bar{x}_t = [x_t; u_t]$. The vec($\cdot$) operator stacks the columns of a matrix into a vector. This can be achieved in MATLAB using the `reshape()` function.

The state update equation of a state space model can be written in the following vector form,

Let the matrices,

$$X = [x_2, \ldots, x_T], \quad \bar{X} = \begin{bmatrix} x_1, \ldots, x_{T-1} \\ u_1, \ldots, u_{T-1} \end{bmatrix}, \quad W = [w_1, \ldots, w_T], \tag{8.24}$$

then the state space update can be written

$$vec(X) = (\bar{X}^T \otimes I_{n_x})\text{vec}(\Gamma) + \text{vec}(W), \tag{8.25}$$

where $\otimes$ denotes the Kronecker product.

Hence, using the GH prior, the posterior of Γ can be sampled from the multivariate normal distribution

$$p(\Gamma|Q, \bar{\tau}, x_{1:T}) = \mathcal{N}(\text{vec}(\Gamma); \mu_\Gamma, \Sigma_\Gamma), \tag{8.26}$$

where,

$$\mu_\Gamma = \Sigma_\Gamma((\bar{X}^T \otimes Q^{-1})\text{vec}(X)), \tag{8.27}$$

$$\Sigma_\Gamma = (\text{diag}(\bar{\tau})^{-1} \otimes I_{n_x} + (\bar{X}^T \otimes Q^{-1})(\bar{X}^T \otimes I_{n_x}))^{-1}. \tag{8.28}$$

Computing the posterior of Q relies on the previous choice of the IW conjugate prior and is conditioned on the fixed sample of Γ from the previous step of the Gibbs sampler. the posterior is given by

$$p(Q|\Gamma, x_{1:T}) = \mathcal{IW}(Q; T - 1 + n_0, S_{GH} + S_0), \tag{8.29}$$

where $S_{GH} = (X - \Gamma\bar{X})(X - \Gamma\bar{X})^T$.

The posterior of the variance parameters τ of the GH prior are conjugate to the independent column priors,

$$p(\bar{\tau}|\Gamma) = \prod_{j=1}^{n_x+n_u} \mathcal{GIG}(\tau_j; \nu - \frac{n_x}{2}, a, b + \gamma_j^T \gamma_j), \tag{8.30}$$

where $\{\gamma_j\}_{j=1}^{n_x+n_u}$ are the columns of Γ.

When the model order is known and the associated MNIW prior is used, the posterior density of Γ and Q are available in closed form. Following the form of X, $\bar{X}$, and W given in Equation 8.24, then the state update distribution can be written

$$X = \Gamma\bar{X} + W. \tag{8.31}$$

Due to the choice of conjugate priors, the posterior of Γ and Q are then given by the MN and IW distributions,

$$p(\Gamma, Q|x_{1:T}) = \mathcal{MN}(\Gamma; S_{X\bar{X}}S_{\bar{X}\bar{X}}^{-1}, Q, S_{\bar{X}\bar{X}}) \times \mathbb{IW}(Q; T - 1 + n_0, S_{X|\bar{X}} + S_0), \tag{8.32}$$

where

$$S_{\bar{X}\bar{X}} = \bar{X}\bar{X} + L, \tag{8.33}$$

$$S_{X\bar{X}} = X\bar{X}^T + ML, \tag{8.34}$$

$$S_{XX} = XX^T + MLM^T, \tag{8.35}$$

$$S_{X|\bar{X}} = S_{XX} - S_{X\bar{X}}S_{\bar{X}\bar{X}}^{-1}S_{X\bar{X}}^T. \tag{8.36}$$

Regardless of if the model order is known, the posterior of the output noise covariance r is computed using a vector based construction. Let

$$S_r = (\mathbf{y} - \mathbf{h})^T(\mathbf{y} - \mathbf{h}), \tag{8.37}$$

where $\mathbf{y} = [y_1, \ldots, y_t]^T$ and $\mathbf{h} = [h(Cx_1), \ldots, h(Cx_t)]^T$.

The posterior is then given by an IW distribution

$$p(r|h, x_{1:T}, y_{1:T}) = \mathcal{IW}(r; T + m_0, S_r + R_0). \tag{8.38}$$

The posterior of $h(\cdot)$ is given by a Gaussian process. As the GP describes a distribution of functions, a sample function must be drawn from the posterior in a way that can represented and used effectively. The simplest method is to evaluate the GP on a fixed number of test points evenly distributed on a grid only once per iteration of the MCMC sampler. The points are then linearly interpolated to calculate the function at any arbitrary value $h(z_t)$. This significantly reduces the complexity of the computations performed inside the particle filter compared to performing inference using every datapoint z_t as a test input to the GP.

Let the test points be held in the vector $\mathbf{z}^* = [z^{(1)}, \ldots, z^{(M)}]^T$ and the corresponding function outputs $\mathbf{h}^* = [h(z^{(1)}), \ldots, h(z^{(M)})]^T$. Using the equation for the predictive posterior distribution of the GP from Section 7,

$$p(\mathbf{h}_*|\eta, r, x_{1:T}, y_{1:T}) = \mathcal{N}(\mathbf{h}_*|\mathbf{m}^* + K_*^T(K + \sigma_n^2 I_T)^{-1}(\mathbf{y} - \mathbf{m}), K_{**} - K_*^T(K + \sigma_n^2 I_T)^{-1}K_*) \tag{8.39}$$

where $\sigma_n^2 = r$ and I_T is an identity matrix with dimensionality of the length T of the samples.

The posterior distribution of the hyper parameters η is not available in closed form for most valid covariance functions. Due to this, sampling from the posterior requires an approximation method. A Metropolis-Hastings accept/reject step is proposed by (Lindsten et al., 2013) using a random walk kernel to sample a proposed value such that $\eta' \sim \mathcal{N}(\eta, \epsilon)$. The proposed sample is then accepted if it is more likely, with probability

$$1 \wedge \frac{p(y_{1:T}|\eta', r, x_{1:T})}{p(y_{1:T}|\eta, r, x_{1:T})} \frac{p(\eta')}{p(\eta)} \frac{v(\eta|\eta')}{v(\eta'|\eta)}, \tag{8.40}$$

else the previous value is kept.

The probability of the hyperparameters P_η is given by the covariance matrix K, and the distribution can be written

$$p(y_{1:T}|\eta, r, x_{1:T}) = \mathcal{N}(\mathbf{y}; \mathbf{m}, P_\eta + \sigma_n^2 I_T). \tag{8.41}$$

The Cholesky factorisation of $P_\eta + \sigma_n^2 I_T$ can be used to compute the acceptance probability such that

$$\frac{p(y_{1:T}|\eta', r, x_{1:T})}{p(y_{1:T}|\eta, r, x_{1:T})} = \frac{\det(R_\eta)}{\det(R_{\eta'})} \exp(\frac{1}{2}s_\eta^T s_\eta - \frac{1}{2}s_{\eta'}^T s_{\eta'}), \tag{8.42}$$

where $s_\eta = R_\eta^{-T}(\mathbf{y} - \mathbf{m})$.

Psuedo-code of the system identification procedure used to implement the PG-AS method and perform system identification on the semiparametric Wiener model is given in Algorithm 7.

Algorithm 7: Wiener System Identification with Particle Gibbs with Ancestor Sampling

---

1. Initialisation:
    - a. Initialise the state space system using a sub-space method.
    - b. Set $h(\cdot) = 1$.
    - c. Set $x_{1:T}[0] = \mathbf{0}$
2. **for** $k = 1, \ldots, K$ **do**
    - a. Sample $\Pi[k]$ from:
    - **if** *MNIW prior* **then**
        - Sample $\{\Gamma[k], Q[k]\}$ $p(\Gamma, Q | x_{1:T}[k-1])$.
        - Sample $r[k] \sim p(r | \mathbf{h}_*[k-1], x_{1:T}[k-1], y_{1:T})$.
    - **if** *GH prior* **then**
        - Sample $\Gamma[k] \sim p(\Gamma | Q[k-1], \bar{\tau}[k-1], x_{1:T}[k-1])$.
        - Sample $Q[k] \sim p(Q | \Gamma[k], x_{1:T}[k-1])$.
        - Sample $\bar{\tau} \sim p(\bar{\tau} | \Gamma[k])$.
        - Sample $r[k] \sim p(r | \mathbf{h}_*[k-1], x_{1:T}[k-1], y_{1:T})$.
    - b. Sample $\eta[k] | r[k], x_{1:T}[k-1], y_{1:T}$ using MH accept/reject algorithm.
    - c. Sample $\mathbf{h}_*[k]$ from the GP posterior $p(\mathbf{h}_* | \eta, r, x_{1:T}, y_{1:T})$.
    - d. Set $\theta[k] = \{\Pi[k], \eta[k], \mathbf{h}_*[k]\}$.
    - e. Sample $x_{1:T}^*$ by running the CPF-AS targeting $p(x_{1:T} | \theta[k], y_{1:T})$ conditioned on previous state trajectory $x_{1:T}[k-1]$.

---

9. Results

Quantifying the accuracy of guitar amplifier models is a challenge due to the subjective nature of the listener's hearing. Listeners often describe qualities of the way it feels to play through an amplifier or a model, where non-measurable characteristics such as 'feel', 'touch response' or 'warmth' are desired.

The frequency content of a discrete time-domain signal can be analysed using the Fast Fourier Transform (FFT). The FFT samples a signal over a fixed window, dividing it into its frequency components. The most common use for the FFT is graphing the one-sided Power Spectral Density (PSD) of a signal, describing the relative magnitude of the signal's frequency content in $\frac{dB}{Hz}$.

A *spectrograph* describes the frequency content of a signal as it moves through time by performing multiple FFTs over small windows and concatenating the results. This is useful for analysing a signal whose frequency content changes with time, such as that of an electric guitar being played. Comparison of the frequency response of the amplifier and identified model with the spectrograph done visually.

The input signal used to generate the training data for the models is zero-mean Gaussian white noise. Gaussian noise is an ideal input signal as it has equal power distribution over the entire frequency domain. This is verified for interest in Appendix C.

The performance of the system identification method can be tested by applying it to a synthetic benchmark system. This allows comparison of the model output with knowledge of the true system frequency response and static nonlinearity.

9.1. LPF Test System

The first synthetic Wiener system is composed of a 2nd order Butterworth low-pass filter with a cut off frequency of 2.8 kHz. The static nonlinearity used for the test system is a form of smooth squashing function,

$$h(z) = \frac{z}{g + 1/\kappa z} \qquad\qquad z \geq 0,$$
$$h(z) = \frac{z}{g - z} \qquad\qquad z < 0,$$

where g is a gain parameter modifying the smoothness of the function and rate of saturation, and κ adjusts the asymmetry of the function. The squashing function is also used as the mean function $m(z)$ for the Gaussian process set with parameters $g = 1$ and $\kappa = 1$, giving a smooth symmetric mean function.

A -40dB noise floor was chosen, setting the measurement noise $r = 10^{-40/20} = 0.01$. The process noise was set equally across all states with no cross-covariance by scaling the identity matrix such that $Q = 0.1^2[\mathbf{I}_{n_x}]$. The MNIW prior for the state space model was used, with the true system order $n_x = 2$. Output data was generated from the test system using $t = 15,000$ samples and analysed with the PMCMC method for $k = 10,000$ iterations with $N_p = 5$ particles in the particle filter. The bode plot of the true system and identified model is shown in Figure 8.

The estimated frequency response matches closely with the true system, accurately estimating the 8.2 kHz cut off frequency. The estimated rate of attenuation is a constant $-40\frac{dB}{dec}$, consistent with a
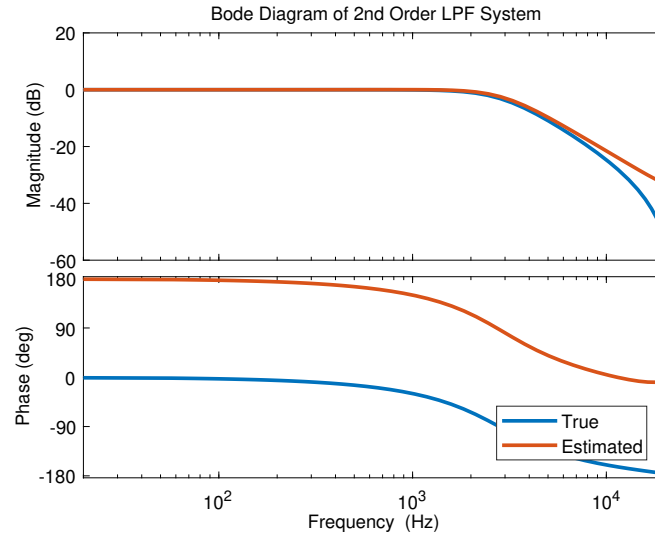
Figure 8: 2nd Order System Bode Plot

2nd order low pass filter with two poles in the same location. The true Butterworth filter response displays a sharper rate of attenuation that can be attributed to the non-zero input pass-through term $\mathbf{D}$ of the state space model. In this method the term is fixed to $\mathbf{D} = 0$ and as such cannot characterise this behaviour.

The phase response of the estimated system is $180°$ out of phase with the true system. This caused the SNL to be mirrored across the x-axis of Figure 9, and as such the function estimate shown was mirrored to account for this.
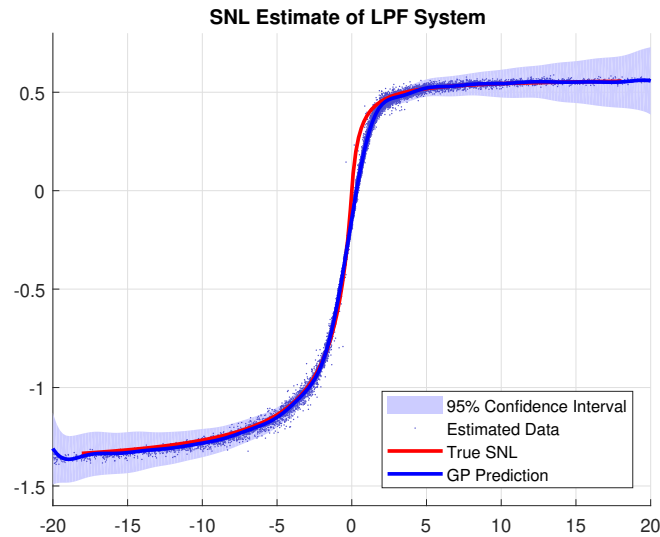
Figure 9: 2nd Order System SNL

The estimate of the SNL function produced by the GP regression very closely matches the true

function. The GP used a subset of $m = 100$ inputs split into three areas of focus:

- 30 points between $z = \pm 4$,

- 35 points between $-25 < z < -4$,

- 35 points between $4 < z < 25$.

The inducing inputs and kernel approximation were calculated using the sparse VFE method made available from the `GPML MATLAB toolbox` (Rasmussen and Nickisch, 2018). The square exponential kernel function was used with length scale and noise hyperparameters determined by the Metropolis Hastings step of the PG-AS sampler. This allows effective use of the $15,000$ samples of input data in the particle filter while reducing the computational complexity of the GP from $\mathcal{O}(n^3)$ to $\mathcal{O}(nm^2)$.

The Markov Chain Monte Carlo estimator creates estimates of the probability distribution of each parameter. As the system is 2nd order, it is feasible to plot histograms for each parameter to visualise the estimates produced by the MCMC method. The resulting histograms are shown in Figure 10. The expected value of each parameter is plotted as a dotted line. The first 5000 samples were omitted from the samples to account for the *burn-in* of the sampler and ensure the MCMC method had converged to a stationary distribution. It is notable that the first two parameters of the $\mathbf{A}$ matrix appear to have a slightly multi-modal distribution, however when the expected value is taken the mean is located in the centre of the two distributions more closely resembling a Gaussian with a higher covariance. Allowing more iterations in the MCMC method may have smoothed out the multi-modal 'lumps' of these distributions.
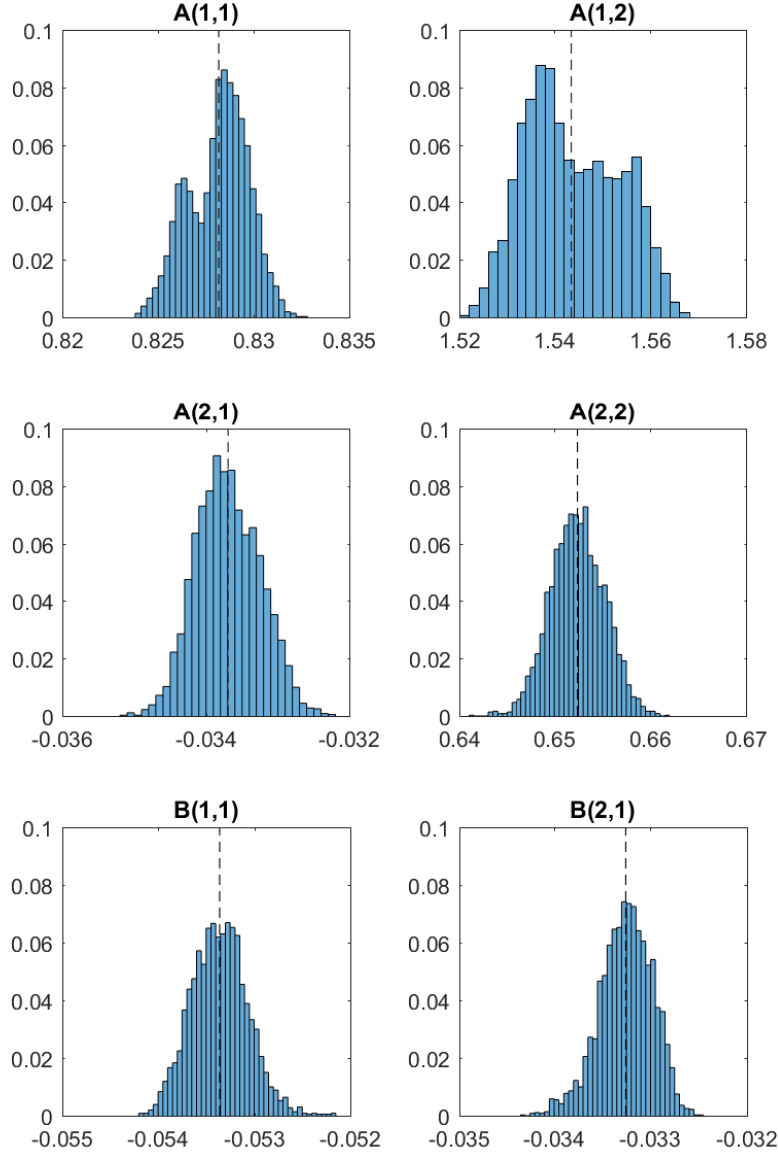
Figure 10: 2nd Order System SSM Parameter Distributions

9.2. Peaking Filter Test System

The second synthetic Wiener system is composed of an 8th order state space model built from cascaded 2nd order peaking filters centered at approximately 120 Hz, 800 Hz, 1.8 kHz, and 4 kHz with varying amplitude and Q factors.

Output data was generated from the test system using $t = 1000$ samples and analysed with the PM-

CMC method for $k = 20,000$ iterations with $N_p = 15$ particles in the particle filter. The measurement and process noise were chosen to be the same as in Section 9.1 giving $r = 10^{-40/20} = 0.01$ and $Q = 0.1^2[\mathbf{I}_{n_x}]$. The GH prior was used for the state space model, with an estimated system order $n_x = 12$. The Gaussian process used the Matérn kernel function and was given the full set of data as inputs. The bode plot of the true system and identified model is shown in Figure 11.
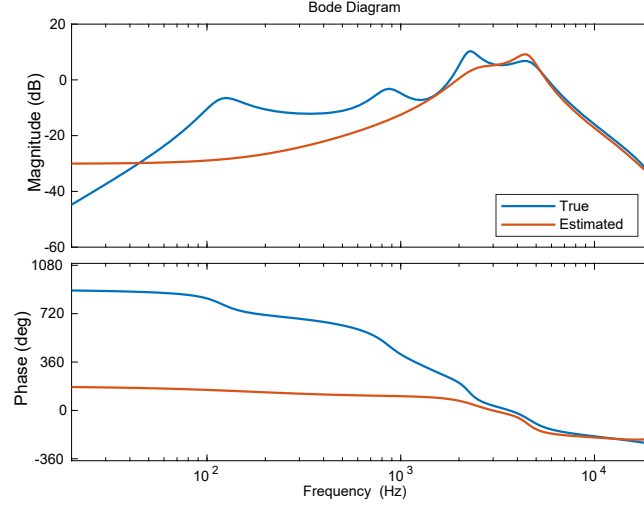
Figure 11: 8th Order System Bode Plot
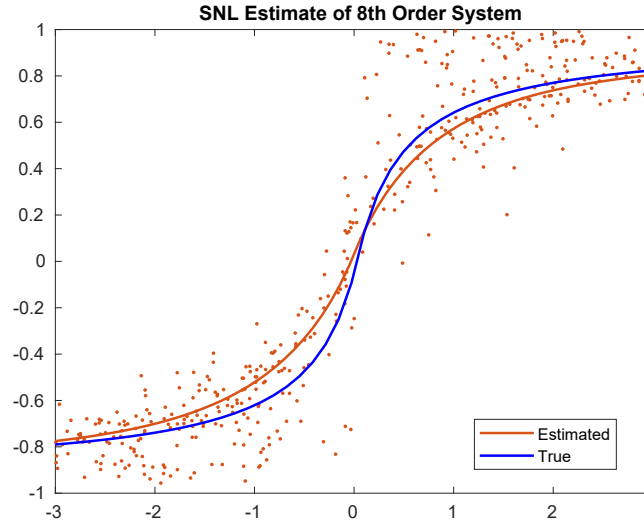
Figure 12: 8th Order System SNL

Significant deviation from the true frequency response is seen in the identified model. The error is most prominent in the bass frequencies where peaks of the filter at 120 Hz and 800 Hz aren't captured. The method performed better at identifying high frequency components. The identified model has peaks at the correct frequencies of 1.8 kHz and 4 kHz. The high frequency roll-off behavior is also characterised well, beginning at 4 kHz and decreasing at the correct rate of 60 $\frac{dB}{dec}$.

The lack of characterisation of the lower frequencies indicates that the dynamics of the system were unobservable at these frequencies. This could be caused from lack of sufficient excitation at these frequencies due to the low number of samples, or the significant phase delay caused by the cascaded filter systems making this appear as a fractional time delay. The GH prior appears to have determined these modes of operation unnecessary from the model.

The static nonlinearity displays a good fit to the true function, but does not fit exactly due to increased noise in the data z_t originating from model inaccuracy. This is an advantage to the use of a GP, as it elegantly handles noise on the input data, providing a good fit.

9.3. 5W Tube Amplifier

The Blackheart 'Little Giant' BH5H 5W tube amplifier has a minimal circuit structure with a low RMS power output. This allows for overdrive of the both the pre-amp and power-amp stages of the amplifier at sound levels suitable for use in a home or bedroom. The amplifier uses two 12AX7 triode tubes in the pre-amplifier and a single EL84/6BQ5 pentode tube in the power amplifier. It has a solid state rectifier that supplies DC voltages to bias the tubes to their linear operating region.

Regardless of the system identification method, the salient features of the system can only be inferred from what is evident and observable in the data. Any experiment performed to capture data from a system must sufficiently excite the dynamics which are to be identified. As it is desired to identify the overdrive characteristic of the tube amp in the model, it is necessary for the gain and volume settings of the amplifier to be set high enough to cause significant saturation while recording the training data.

The method of recording the output influences the observable dynamics of the system. External sources of dynamics can be a source of error in the output measurement which the system identification method will attempt to model as though it was the behaviour of the amplifier. Use of a calibrated microphone ensures the frequency response of the microphone does not effect the measurements. Performing the experiment in a heavily damped studio environment with approximately flat frequency response aims to remove the influence of acoustic effects of the room such as reverberation or resonant peaks.
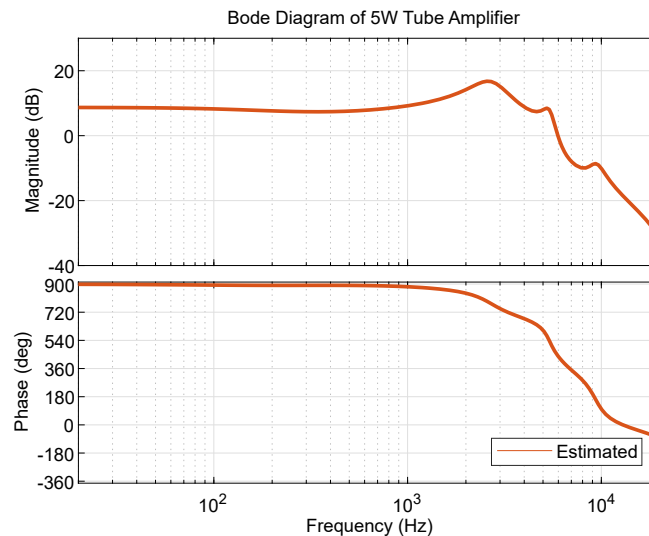

Figure 13: Blackheart 5W Tube Amp Identified Bode Plot

The output waveform of the identified model of the BH-5H amplifier matches the true output exceptionally when emulating the white noise training data shown in Figure 15. The bode plot in Figure 13 shows that the model has many high-frequency characteristics, however the model does not identify poles in the lower frequencies. The low frequency response is expected to show a high-pass effect centered at approximately 80 Hz typical of a guitar cabinet.

The error in frequency response is easily in the spectrograph of Figure 17. The identified model is applied to a validation dataset of 16 seconds of electric guitar. The low-frequencies of the model show
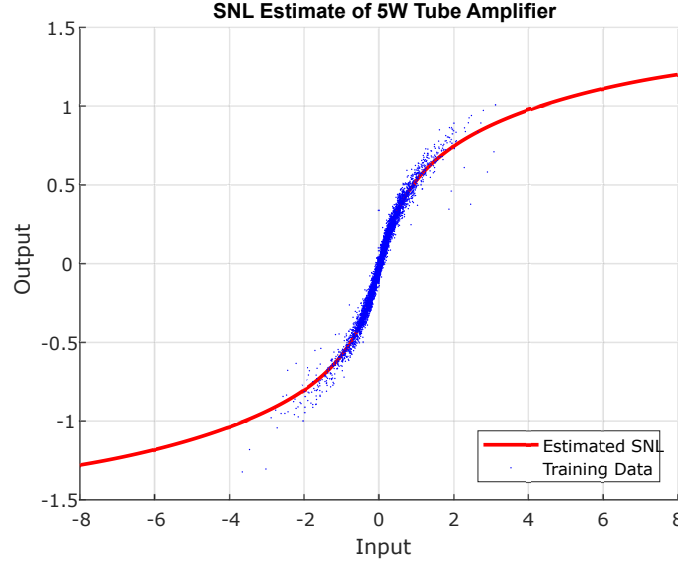
Figure 14: Blackheart 5W Tube Amp Identified Static Nonlinearity
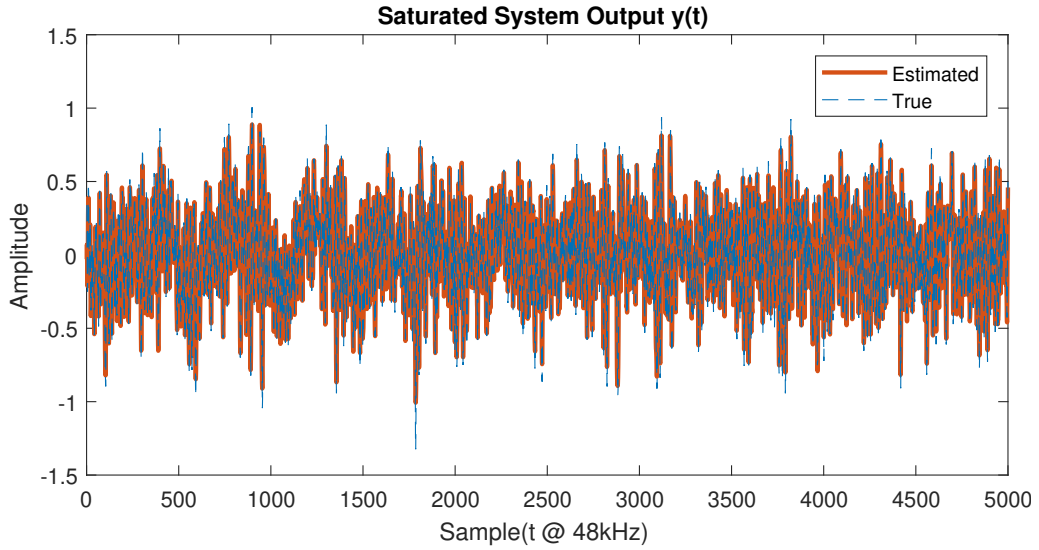

Figure 15: Blackheart Amp Real vs Estimated Output Waveforms

considerable error, with increased intensity of frequencies below 80 Hz relative to the true frequency response. The high frequencies of the model also suffer from error, primarily in the region above 10 kHz.

The low-frequency error of this model may by fixed by manually including a high pass filter in the system. Enforcing this structure on the model would invalidate some of the motivation for using the grey box structure of the proposed method as assumptions about the expected model behavior are being manually placed on the system instead of being automatically identified. Similar to the error seen in Section 9.1, the low frequency errors seen may be a result of fixing the input pass-through term of the output to $\mathbf{D} = 0$, or poor numerical properties of enforcing the observer canonical form
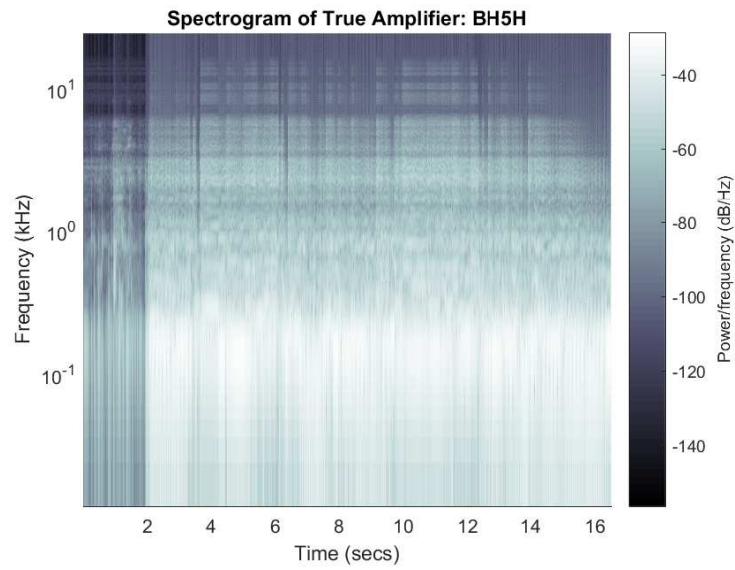
by fixing $C = [1, 0, 0, \ldots]$.

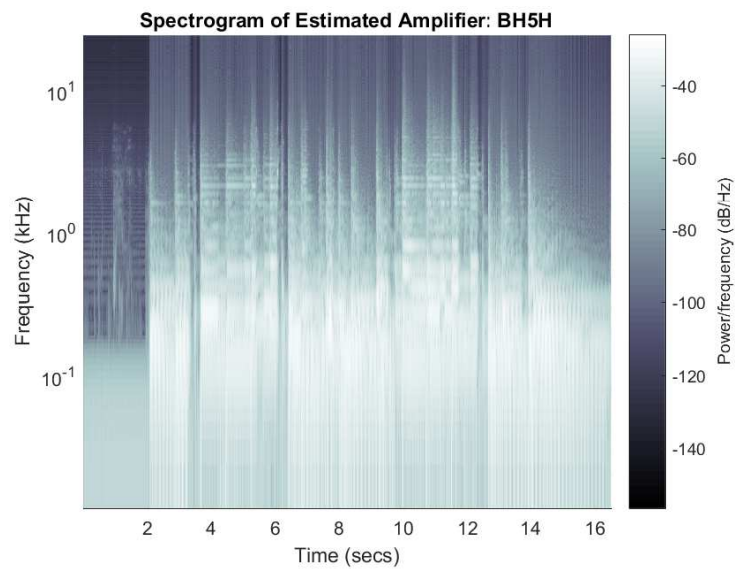Figure 16: True Spectrograph from Guitar Dataset: BH5H

Figure 17: Estimated Spectrograph from Guitar Dataset: BH5H

10. Conclusion

Nonlinear system identification is explored for the purpose of accurately simulating guitar valve amplifiers. The Wiener model is proposed as a suitable grey-box model, defining the system as a linear state space model followed by a static nonlinearity. Gaussian process regression is shown to be suitable for modelling of the one dimensional memory-less static nonlinearity of the system, reducing imposed structure on the model. A Particle Markov Chain Monte Carlo estimator is applied to achieve nonlinear system identification within a Bayesian framework.

The probabilistic framework permits use of a PMCMC method of particle Gibbs with ancestor sampling, targeting the intractable posterior $p(\theta|y)$ to generate estimates of the model parameters. The Gibbs sampler splits the task of estimating the posterior into a series of iterative samples from lower dimensional distributions. Estimates of the parametric state space model, and its associated noise terms, are sampled using conjugate priors. The hyperparameters of the Gaussian process are estimated using a Metropolis Hastings accept/reject method. A conditional particle filter with ancestor sampling is used to sample the state prediction density. The use of the CPF-AS estimate for the state prediction generates an invariant Markov chain and ensures ergodicity of the target distribution of the Gibbs sampler. Sparse input approximations are made for the Gaussian process kernel, allowing the method to handle larger training data sets using a subset of inputs.

The method is applied to two synthetic test systems, and a 5W tube amplifier. The method shows promise for identification of the nonlinear operation of a guitar valve amplifier as a Wiener model. The static nonlinearity is successfully identified in all test systems, capturing both smooth and asymmetric properties of the functions. The method of inference for the state space model captures high frequency dynamics of the system, but omits relevant low-frequency characteristics of the amplifier.

To improve the Gaussian process regression, it may be possible to construct a covariance function specific to the problem of modelling the static nonlinearity of a saturated guitar amplifier. Instead of using the squared exponential or Matérn kernel, the underlying knowledge of tube saturation could be implemented within the covariance function. A desirable characteristic for a general purpose overdrive kernel would be an adaptive length scale, allowing the function to vary rapidly when close to the origin using a small length scale, and become smoother when highly saturated using a long length scale to ensure the tails of the function are asymptotic. Constructing a valid symmetric positive definite kernel function to achieve this is as a direction for further research of modelling of tube overdrive with a Gaussian process.

The consistent error of the method in modelling the response of the amplifier at low frequencies is a substantial roadblock to the implementation of this system identification method for emulating guitar amplifiers. The exact source of the error is difficult to determine.

A potential source of this error could be due to the constraints placed on the state space model, such as the restriction to observer canonical form with $C = [1, 0, 0, \ldots]$. This limits the search space of the A matrix of the state space model. The canonical form is chosen to solve the problem of non-uniqueness, that is, multiple solutions to the matrices of the state space model can produce a system with the same transfer function. Methods exist for system identification of fully-parametric nonlinear state space models that overcome this non-uniqueness without enforcing observer canonical form by using a gradient projection method to solve the identification of the system using a gradient-based optimisation method to compute the maximiser for a maximum likelihood method (Verdult et al., 2003). Estimating the input-output pass-through term D may also be necessary to achieve accurate

frequency response estimation.

Particle methods may be applicable to the use of maximum likelihood identification of a fully-parametric Wiener system. Particle smoothing can be used to compute an estimate for the expectation integral of an expectation-maximisation method as demonstrated in (Wills et al., 2013), however there is no guarantee for convergence to the true solution due to the possibility of getting stuck in local optima. Their method also offers the possibility of a Hammerstein-Wiener model, which may more accurately model the coupled overdrive characteristics of the pre-amp and power-amp of a guitar amplifier.

While fully-parametric methods offer a promising direction to system identification of the Wiener system, they have the downside of requiring a parametric equation for the static nonlinearity. The use of a Gaussian process is desirable due to its nonparametric nature and regression over infinite basis functions. A possible research direction may be to implement a subset of regressors using a finite set of basis functions, adhering to a parametric form which could be solved with EM while maintaining some of the flexibility of the Gaussian process nonlinearity.

To more accurately model the accuracy of the *perceived* frequency response of a system, future methods may make use of a cost function based on psycho-acoustic masking. This is an effect wherein peaks of a certain frequency will block the perception of signals below a *masking* threshold around the peak frequency. These masking thresholds decay asymmetrically around the frequency peak, however a Gaussian assumption is a suitably close approximation (Smith et al., 2011). Evaluating the perceived accuracy of the model in terms of the frequency response also opens the possibility of system identification performed by regression in the frequency domain.

A critical assumption made throughout the modelling of the Wiener system is that the nonlinearity is static and memoryless. Recent advances in industry leading guitar amplifier modelling products have used dynamic nonlinearities. This is achieved by interpolating between multiple SNL functions stored in an N-dimensional look-up table. A state estimator is used to determine which functions should be selected for the current time steps (Yeh, 2012). This may be necessary to capture effects seen in valve amplifiers such as magnetic saturation of the transformer causing a dynamic hysteresis effect.

Implementation of the resulting model for real time simulation on a microprocessor is considered a desirable long term goal. The model is proposed to be ported to C for execution on an STM32F407 microprocessor. The state-space model can be implemented discretely, and the nonlinear mapping of the Gaussian process can be evaluated on a fine grid and stored in memory as a look-up-table (LUT). The goal of real time emulation of a guitar amplifier on an embedded system means that future research should aim to be feasible with the limited computational power available on microprocessors.

11. References and Citations

References

Adams, R., 2013. Computing log-sum-exp.
URL https://hips.seas.harvard.edu/blog/2013/01/09/computing-log-sum-exp/

Cheever, E., 2015. Transfer function to state space.
URL http://lpsa.swarthmore.edu/Representations/SysRepTransformations/TF2SS.html

Dempwolf, K., Zölzer, U., 2011. A physically-motivated triode model for circuit simulations. In: 14th international conference on Digital Audio Effects DAFx. Vol. 11.

Fox, E. B., 2009. Bayesian nonparametric learning of complex dynamical phenomena. Ph.D. thesis, Massachusetts Institute of Technology.

Franklin, G. F., Powell, J. D., Workman, M. L., 1998. Digital control of dynamic systems. Vol. 3. Addison-wesley Menlo Park, CA.

Jaynes, E. T., 1986. Bayesian methods: General background.

King, R. W., 1923. Thermionic vacuum tubes and their applications. Bell Labs Technical Journal 2 (4), 31–100.

Koren, N., 1996. Improved vacuum tube models for spice simulations. Glass Audio 8 (5), 18–27.

Kuss, M., 2006. Gaussian process models for robust regression, classification, and reinforcement learning. Ph.D. thesis, Technische Universität.

Lindsten, F., Jordan, M. I., Schön, T. B., 2014. Particle gibbs with ancestor sampling. Journal of Machine Learning Research 15 (1), 2145–2184.

Lindsten, F., Schön, T. B., Jordan, M. I., 2013. Bayesian semiparametric wiener system identification. Automatica 49 (7), 2053–2063.

Meyn, S. P., Tweedie, R. L., 2012. Markov chains and stochastic stability. Springer Science & Business Media.

Perez, T., 2014. Reasoning under Uncertainty - A Bayesian Tutorial. University of Newcastle.

Rasmussen, C., Nickisch, H., 2018. Gaussian process regression and classification toolbox.
URL http://gaussianprocess.org/gpml/code

Rasmussen, C. E., Williams, C. K., 2006. Gaussian processes for machine learning. Vol. 1. MIT press Cambridge.

Renton, C., 2017. Course Notes: Mechatronic System Design 2. University of Newcastle.

Renton, C., 2018. Slack message providing help with log pdfs.

Rippa, S., Nov 1999. An algorithm for selecting a good value for the parameter c in radial basis function interpolation. Advances in Computational Mathematics 11 (2), 193–210.
URL https://doi.org/10.1023/A:1018975909870

Robert, C., Casella, G., 2013. Monte Carlo statistical methods. Springer Science & Business Media.

Robert, C. P., Casella, G., Casella, G., 2010. Introducing monte carlo methods with r. Vol. 18. Springer.

Schaback, R., Wendland, H., 2006. Kernel techniques: from machine learning to meshless methods. Acta numerica 15, 543–639.

Smith, J. O., et al., 2011. Spectral audio signal processing. Vol. 1334027739. W3K.

Stein, M. L., 2012. Interpolation of spatial data: some theory for kriging. Springer Science & Business Media.

Sylvania Electronic Products Inc., 1995. Engineering data service sheet. [Online; accessed September 28, 2017].
URL http://www.philbrickarchive.org/12ax7_sylvania.pdf

Verdult, V., Bergboer, N., Verhaegen, M., 2003. Identification of fully parameterized linear and non-linear state-space systems by projected gradient search.

Wills, A., 2018. Course Notes: Advanced Estimation. University of Newcastle.

Wills, A., Schön, T. B., Ljung, L., Ninness, B., 2013. Identification of hammerstein–wiener models. Automatica 49 (1), 70–81.

Yeh, D. T., 2012. Automated physical modeling of nonlinear audio circuits for real-time audio effectspart ii: Bjt and vacuum tube examples. IEEE Transactions on Audio, Speech, and Language Processing 20 (4), 1207–1216.

A. Engineering Australia Competencies

Professional and Personal Attributes

Solving problems of great complexity requires the drive to self-manage my time while being part of a wider team. Organisation holds strong importance and the ability to prioritise critical tasks and processes is key to being an effective member of any professional environment.

There has been a strong emphasis on the need of professionalism in every moment of my education and further into my career. The importance of communication when delivering presentations and reports has given me a vision to always strive to confidently and clearly communicate technical information and processes in a manner that is understandable to someone from any discipline.

Application of Engineering Abilities

Throughout all areas of my degree, completion of assignments and projects have helped me develop a strong skill set in problem solving. Approaching a problem may first seem insurmountable, but with systematic analysis any project can be effectively managed and divided into a set of realistic, time based and achievable tasks. When the steps to be completed are identified the technical specifications of each problem can be approached to ensure that the synthesis of a solution to any engineering design problem is solved optimally.

Core Knowledge and Skill Base

Completing a combined double-degree of Mechanical Engineering alongside Mechatronics Engineering has given me the opportunity to develop a vast amount of knowledge spread across many topic areas. While I have found a strong interest in embedded systems engineering, automation and signal processing which are strongly mechatronics based, I feel that the interdisciplinary links I have made with the mechanical engineering discipline are invaluable.

The projects I have completed throughout my degree have required the use of the entire knowledge base I have acquired while at University. I have an appreciation for bridging the link between technical knowledge and real world implementation for idealised systems. Completing mechatronics projects involves the use of advanced modelling, estimation, control and evaluation of results. In researching and applying many techniques from the field of mechatronics engineering, I have developed a strong passion for seeking out new technical knowledge on the cutting edge of research to apply in innovative ways. This report is the culmination of that passionate spark.

B. Log Distribution Calculations

Calculation of pdfs can suffer from numeric instability due to the use of exp(), leading to overflow and underflow. Using log relationships, the pdfs can be calculated with increased speed, precision and numeric stability.

The typical formula for the pdf of a multivariate Gaussian distribution is given by:

$$p(x; \mu, \Sigma) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu)\right)}{\sqrt{|2\pi\Sigma|}} \tag{B.1}$$

Applying the natural logarithm yields:

$$\ln(p(x; \mu, \Sigma)) = -\frac{1}{2}\ln(|2\pi\Sigma|) - \frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu) \tag{B.2}$$

Computing the determinant of Σ inside the logarithm is undesirable. The determinant becomes inaccurate as the dimension of Σ increases, or eigenvalues of Σ are close to zero. This can cause numeric over/underflow before the log is computed. Cholesky factorisation is performed to avoid this as follows (Renton, 2018):

Let $A =$chol$(2\pi\Sigma)$.

$$\ln(p(x; \mu, \Sigma)) = -sum(\ln(diag(A))) - \frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu) \tag{B.3}$$

The log form of a distribution is advantageous as it allows the calculation of the weights in the particle filter $\frac{p(x)}{q(x)}$ to be performed as a subtraction $ln(p(x)) - ln(q(x))$.

The distributions and weights are stored directly in log form to retain the numerical and computational benefit of this method.

As such, the weight normalisation can also be performed as a log operation where

$$\tilde{\omega} = \frac{\omega}{\sum_{i=1}^{N_p} \omega^{(i)}},$$

is instead calculated using the log-sum-exponential trick (Adams, 2013),

$$\log(\tilde{\omega}) = \log(\omega) - \log\left(\sum_{i=1}^{N_p} \exp\left(\log(\omega^{(i)})\right)\right), \tag{B.4}$$

$$= \log(\omega) - \max(\log(\omega) - \log\left(\sum_{i=1}^{N_p} \exp\left(\omega^{(i)} - \max(\log(\omega))\right)\right). \tag{B.5}$$

C. Flat Power Spectral Density of White Noise

Figure 18 shows the FFT of the average power spectral density of 50,000 realisations of a sequence of 2048 draws from $u \sim \mathcal{N}(0, 2)$. The sharp cut-off seen on the right of the figure occurs at the Nyquist sampling frequency of $\frac{\text{Fs}}{2}$ where Fs = 48 kHz in this example.
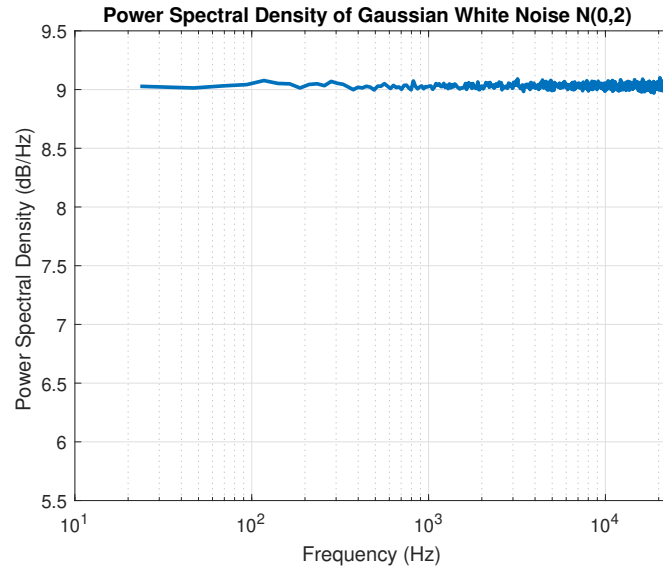

Figure 18: PSD of Zero Mean Gaussian Noise $\mathcal{N}(0, 2)$