An exploration of data-driven system identification and machine learning for plasma physics

Alan Kaptanoglu

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Steven Brunton, Chair

Uri Shumlak

Kai-Mei Fu

Gerald Seidler

Program Authorized to Offer Degree:

Physics

University of Washington

Abstract

An exploration of data-driven system identification
and machine learning for plasma physics

Alan Kaptanoglu

Chair of the Supervisory Committee:
Steven Brunton
Mechanical Engineering

Plasma is the most common state of visible matter in the universe and provides a myriad of scientific and engineering applications. However, the complexity of these systems poses a significant challenge for understanding and controlling plasmas. Fortunately, machine learning is increasingly used to handle complex, nonlinear systems, and the field of machine learning is advancing at an unprecedented pace, propelled forward by advances in sensing technology and computing power. This thesis summarizes work towards applying modern machine learning algorithms for fluid and plasma physics applications, with a focus on the understanding and control of magnetohydrodynamic (MHD) phenomena and fusion-relevant plasmas. Although this work is primarily focused on machine learning, first conventional numerical techniques are used to implement a two-temperature Hall-MHD model into the 3D PSI-Tet code, followed by an investigation of the plasma dynamics in the HIT-SI experiment. These simulations agree well with experimental measurements, and indicate that low-densities are required for significant closed flux surfaces − a recommendation that is now helping to guide the next generation of experimental design. Next, plasma modeling with machine learning is discussed in the context of the hierarchy of plasma models and it is illustrated that there is "plenty of room at the bottom" for physics-constrained reduced order models that approximate more complex MHD or kinetic plasma models. Variants

of the dynamic mode decomposition are explored on experimental data and simulations of the HIT-SI plasma device and indicate promise for magnetic mode spectroscopy and forecasting diagnostic measurements. Continuing, analytic reduced-order modeling methods are extended using techniques in system identification for extracting reduced-order models directly from data. In the process, new methods are invented to enforce physical constraints and stability in data-driven fluid and plasma models. For instance, the ability to build data-driven models that obey global conservation of energy or global conservation of cross-helicity is demonstrated, with promise for efficient simulations of ideal and resistive MHD turbulence. With the new functionality implemented into the open-source PySINDy code as part of this work, advanced system identification methods that can robustly extract dynamical equations from data are available to the larger scientific community. In total, this work illustrates that new machine learning methods can be directly tied with known physical laws in plasma physics, have promise to significantly impact much of the plasma physics and nonlinear systems fields, and can provide complementary, interpretable methods to the relatively black-box deep learning techniques that are frequently used in the plasma physics field for extracting diagnostic information, building reduced-order models, and performing real-time control.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

Λ_D: Charge separation distance, a measure of the important of electrical charge in a system.

Kn: Knudsen number defined through the ratio of the average mean free path to the typical experimental length scale (or length scale of interest). A measure of how well the system of interacting particles can be described as a fluid.

$\tau_{ss'}$: Rate of collisions between particle species s and s'. If $\tau_{ss} \ll 1$, particle species s is well-approximated by local Maxwellian distributions.

$\tau_{L/R}$: Characteristic time-scale for (global) resistive diffusion.

τ_A: Characteristic time-scale for an Alfvén wave transit. In a toroidal device, this is often the time scale for Alfvén waves to toroidally traverse the domain and the characteristic timescale for some unstable MHD modes.

r_L: The plasma Larmor radius, i.e. the characteristic radius of a charged particle undergoing helical motion around a magnetic field line.

λ_D: Debye length, the characteristic scale for the field of a charged particle to be "screened out" (exponentially damped) by the nearby charged particles of opposite sign.

ω_{ps}: The plasma frequency associated with particle species s.

ω_{cs}: The cyclotron frequency associated with charged particles of species s.

β: Ratio of the thermal and magnetic energies, and therefore a measure of the confinement quality in a fusion device.

(R, ϕ, Z): 3D coordinates in a cylindrical coordinate system.

θ: Poloidal angle defined from the magnetic axis of a plasma equilibrium in the (R, Z) plane.

ODE: Ordinary differential equation.

PDE: Partial differential equation.

NS: The Navier-Stokes equations describing a traditional fluid flow (not typically a conducting fluid).

DNS: Direct numerical simulation typically denotes a fluid or plasma simulation with high enough resolution to resolve all the experimental and turbulent scales.

MHD: Magnetohydrodynamics, a catch-all term for a variety of conducting single-fluid models.

Hall-MHD: A single-fluid magnetohydrodynamic model that retains some two-fluid effects through inclusion of the so-called "Hall" terms.

ROM: Reduced-order model, a low-dimensional model that approximates the evolution of a system that is typically modelled with a higher-dimensional model.

$\boldsymbol{X}$: A data matrix of size $\mathbb{R}^{N\times M}$, where N is the dimension of the state space (or number of different measurements) and M contains the rest of the dimensions (usually M is equal to the number of temporal snapshots M_t). For instance, for a coupled PDE system of two variables, $N = 2$ and $M = N_x \times N_y \times N_z \times M_t$ for a spatiotemporal grid with N_x points in the x-direction, and so on.

$\boldsymbol{u}$: The fluid flow field.

$\boldsymbol{B}$: The magnetic field.

$\boldsymbol{J}$: The electromagnetic current.

$\boldsymbol{E}$: The electric field.

ρ: The mass density, $\rho = m_i n$, where m_i is the mass of the ions and n is the number density.

T: The fluid temperature.

p: The plasma pressure.

POD: Proper orthogonal decomposition, more often called the biorthogonal decomposition (BOD) in plasma physics. Also known in other scientific fields as principal component analysis (PCA), Karhunen-Loève expansion, empirical orthogonal functions, and so on.

AE: Alfvén eigenmode, a special class of mixed MHD-kinetic instabilities that occur in some plasmas. There are many variations, including toroidal AE (TAE), reversed-shear AE (RSAE), beta-induced AE (BAE), and eccentricity AE (EAE).

TPR: True positive ratio, defined through the number of true positives divided by the number of true positives and false negatives. Also called "recall", "sensitivity", and "hit-rate".

FPR: False positive ratio, defined through the number of false positives divided by the number of false positives and true negatives. Also called "fall-out".

HIT-SI: The helicity injected torus with steady inductive helicity injection (HIT-SI) device at the University of Washington. See Sec. 2.1 for a detailed description.

NIMROD: A 3D finite element code for the solution of resistive and Hall-MHD equations, requiring the simulation domain to be toroidal or cylindrical.

PSI-Tet: A 3D finite element code for the solution of resistive and Hall-MHD equations, utilizing unstructured grids with no limitation on domain shape.

ACKNOWLEDGMENTS

I wish to express sincere appreciation to my advisors Professor Steven Brunton and Dr. Christopher Hansen, as well as my supervising committee. Thanks to all my coauthors, collaborators, and my fellow graduate students.

DEDICATION

To Bridget and my family

Chapter 1
INTRODUCTION

Plasmas are ionized gases that are ubiquitous in the universe. Most visible matter, including stars, interstellar gas clouds, and planetary magnetospheres, is plasma. Nuclear fusion powers the stars in the universe, and in the future, can hopefully help to provide sustainable power for societal needs, including carbon dioxide removal from the atmosphere. Beyond fusion, plama prediction and control are also central to numerous scientific, industrial, and technological applications, including materials processing (e.g. microchip fabrication), space physics (e.g. space weather, spacecraft propulsion), and other energy applications (e.g. plasma switches). Many plasmas of interest are highly nonlinear, providing serious challenges for the understanding and control of plasma phenomena. As a result, plasma simulations often take days to months of computer time to simulate experimentally relevant time-scales; this is many orders of magnitude longer than that needed by typical engineering applications, which aim to control plasma dynamics on the millisecond or even microsecond timescales. The overarching theme of this work is the discovery and exploration of new techniques in applied mathematics and machine learning that can facilitate the physical understanding, forecasting, and real-time control of complex plasma phenomena. There is a particular focus towards fusion-regime plasmas although much of this work remains relevant for general magnetohydrodynamic and neutral fluids.

1.1 Nuclear fusion via magnetic confinement

Nuclear fusion occurs when two atomic nuclei come close enough to fuse via the strong interaction. There are many types of nuclear fusion, but all require extremely high temperatures in order to overcome the repulsive Coulomb force between the nuclei. The sun fuses because

of its extraordinarily large mass; the mass in the center of the sun can fuse many different atomic elements because gravity confines the plasma and the temperatures can become immense. In the laboratory, magnetic confinement fusion relies instead on large magnetic fields which squeeze and confine the plasma to approximately fixed trajectories. There are other techniques, such as inertial confinement fusion (ICF), which produce an implosion (often driven by lasers) to rapidly heat and compress a small capsule − recent progress puts ICF at the cusp of break-even energy generation where enough fusion is produced to offset the input energy of the lasers [72]. By far the most accessible fusion reaction (the reaction with the largest cross-section) in the laboratory is with two different isotopes of hydrogen, deuterium and tritium, which produce a high-energy alpha particle and an even higher-energy neutron,

$$\mathrm{D} + \mathrm{T} \rightarrow {}^{4}\mathrm{He}(3.5\mathrm{MeV}) + \mathrm{neutron}(14.1\mathrm{MeV}). \tag{1.1}$$

In order for this reaction to occur, the plasma must be dramatically heated and then kept distant from material surfaces (which by necessity must be kept relatively cool to avoid damage and sputtering into the device). When enough fusion is occurring, these fusion products can "self-heat" the plasma, enabling a reduction in the amount of external heat needed to sustain the fusion processes. In order to harness the energy generated by fusion, scientists need a way to collect the energy from these fusion products. The conventional technique is to surround the plasma device with a lithium-containing blanket. Neutrons slow and deposit energy as heat into the blanket, and interact with lithium atoms, producing more tritium for the DT reaction in Eq. (1.1).

In the magnetic confinement regime of hot plasma and large magnetic fields, the plasma is often described well (at least at large scales) by various forms of magnetohydrodynamics (MHD). Dynamics associated with particle-dynamics (kinetic effects) are also often observed in these plasmas, but are usually first order corrections to MHD. In order to understand how to confine hot plasma, the principles of magnetic confinement will now be reviewed.

1.2 Magnetic confinement

Fundamentally, magnetic fields can confine charged particles because of the perpendicular nature of one of the Lorentz force terms. The Lorentz force acts on the component of a charged particle's velocity that is perpendicular to the field, causing the particle to exhibit helical orbits around magnetic field lines with a characteristic radius called the Larmor radius,

$$r_{Ls} \equiv \frac{u_{Ts}}{\omega_{cs}} = \frac{\sqrt{m_s T_s}}{q_s B}, \quad u_{Ts} \equiv \sqrt{\frac{T_s}{m_s}}, \quad \omega_{cs} \equiv \frac{q_s B}{m_s}. \tag{1.2}$$

Here the s subscript indicates that the definition is for particle species s, q_s is the particle charge, $B = |\boldsymbol{B}|$ is the magnitude of the magnetic field, T_s is the temperature of the fluid of species s, u_{Ts} is the particle thermal velocity of species s, and ω_{cs} is the cyclotron frequency of species s. The Larmor radius and the cyclotron frequency are fundamental length and time scales in a plasma. Unless specified otherwise, SI units are assumed for all variables except the temperature in units of eV.

For brevity, the particles species subscript is often omitted going forward, especially when describing MHD models. Since charged particles follow magnetic field lines in helical orbits, one approach for confining plasma is to generate large magnetic fields with magnetic coils. As the magnetic field becomes very large, the Larmor radius shrinks inversely with the magnitude of the magnetic field. This causes a reduction in the heat conduction perpendicular to the magnetic field and therefore a reduction in the heat flux at the device walls. Typically the amount of confinement is measured by the amount of magnetic field energy required to contain a given amount of thermal energy, defined through the ratio

$$\beta \equiv \frac{\int p d^3\boldsymbol{x}}{\int \frac{B^2}{2\mu_0} d^3\boldsymbol{x}}. \tag{1.3}$$

Here p is the plasma pressure and μ_0 is the vacuum permeability. Since very hot plasmas are required for fusion, high β is desirable. Unfortunately, there are physical limits on the values of β, and different plasmas and device topologies have unique limits on the achievable

β [289].

Since plasma streams freely along field lines, heat conduction is far larger in the direction parallel to the magnetic field lines than in the perpendicular direction. For singly charged deuterons and $T_e = T_i$, the ratio of the parallel to perpendicular (classical) thermal conductivity is [52]

$$\frac{k_{\parallel,e}}{k_{\perp,i}} \approx \sqrt{\frac{m_i}{m_e}}(\omega_{ci}\tau_i)^2, \tag{1.4}$$

where m_i is the ion mass, m_e is the electron mass, τ_i is the ion-ion collision time for momentum exchange, and the parallel and perpendicular directions refer to directions relative to the local magnetic field lines. For typical parameters in a fusion device, such as a tokamak, one has temperatures $T_e = T_i = 3$ keV, magnetic field strength $B = 5$ T, plasma density $n = 10^{20}$ m^{-3}, and finds $k_{\parallel,e}/k_{\perp,i} \approx 10^{13}$! In reality, the perpendicular conductivity is much larger than the classical value because of turbulence, but the point remains that heat is rapidly lost along magnetic field lines, and much more slowly lost across magnetic field lines. To avoid large parallel heat loss, plasma devices are almost always closed magnetic configurations, and many are toroidal in shape.

1.3 Hierarchy of plasma models

To start modeling a dynamical plasma, it is worth considering the transition from a gas to a plasma. The gaseous state is typically modeled as a large collection of neutral atoms or molecules that only interact over very small distances. For the most part, particles in the gas fly around unperturbed by other particles, and occasionally, a collision occurs when two (or more) particles come exceedingly close to one another. When a gas is heated, this typically increases the average speed of the particles, but with sufficiently high heats, the collisions between particles can begin to knock the electrons off the atoms. When only a small percentage of the atoms have lost electrons, this system is often referred to as a "weakly ionized gas". When a somewhat larger amount of the atoms have lost electrons, the system is called a plasma.

Despite this relatively smooth transition from gas to plasma, plasmas behave far differently than neutral gases because plasmas are *ionized* gases. The charged particles (the positive charged ions and negatively charged electrons) can now feel electromagnetic forces over much larger spatial scales than the previous small scales associated with interatomic collisions of neutral atoms. The characteristic scale for the field of a charged particle to be "screened out" (exponentially damped) by the nearby charged particles of opposite sign is the Debye length λ_D. From this fundamental fact, that λ_D can be far larger than an interatomic spacing, emerges much of the incredible complexity and practicality of plasma dynamics.

To get a first handle on modeling the complex dynamics of plasmas, consider the most general view of a plasma as a collection of charge particles moving in an electromagnetic field. Classically, each particle has a well-defined position $\boldsymbol{x}$ and velocity $\boldsymbol{v}$. The density of a particular particle species s can then be written

$$N_s(\boldsymbol{x}, \boldsymbol{v}, t) = \sum_{i=1} \delta(\boldsymbol{x} - \boldsymbol{x}_i(t))\delta(\boldsymbol{v} - \boldsymbol{v}_i(t)), \tag{1.5}$$

where δ is a delta-function and the sum is over all particles of species s in the plasma. Assuming that the plasma can be described classically, if the position and velocity for every particle in the plasma are known at one time, the positions and velocities for all time can be determined with the Klimontovich equation (derived using Newton's second law). Newton's second law for a charge particle moving in an electromagnetic field is the Lorentz force law

$$\frac{d\boldsymbol{v}}{dt} = \frac{e}{m}(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}), \tag{1.6}$$

where e is a unit of electron charge, and $\boldsymbol{E}$ and $\boldsymbol{B}$ are the electric and magnetic fields respectively. Throughout this work, it is assumed that relativistic effects are negligible, although relativistic effects are straightforward to include. Of course, charged particles also *generate* electromagnetic fields, so self-consistent equations for the evolution of $\boldsymbol{E}$ and $\boldsymbol{B}$ are

also required − Maxwell's equations,

$$\nabla \cdot \boldsymbol{E} = \frac{\rho_q}{\epsilon_0}, \quad \nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t}, \tag{1.7}$$
$$\nabla \cdot \boldsymbol{B} = 0, \quad \nabla \times \boldsymbol{B} = \mu_0 \boldsymbol{J} + \frac{1}{c^2}\frac{\partial \boldsymbol{E}}{\partial t}.$$

Here ϵ_0 is the vacuum permittivity, c is the speed of light, and the charge density ρ_q and current density $\boldsymbol{J}$ must be computed from the plasma itself.

However, the Klimontovich equation is impractical because it requires an infeasible amount of knowledge − the position and velocity of every particle in the plasma (often there are greater than 10^{20} particles in the plasma) at some initial time t. In order to obtain a useful model for plasma behavior, the *single particle* distribution function for species s, $f_s(\boldsymbol{x}, \boldsymbol{v}, t)$, is defined as the probability that a number of species s particles occupies a volume of phase space $(\boldsymbol{x}, \boldsymbol{v}, t)$. The evolution equation for $f_s(\boldsymbol{x}, \boldsymbol{v}, t)$ can be derived directly from the Klimontovich equation with an ensemble average and by assuming that the mean-free-path time is much larger than the time during a collision, $\tau_{\text{mfp}} \gg \tau_{\text{coll}}$, so that collisions can be treated as isolated events that can be entirely packaged into a single "collision term" on the right-hand side (RHS) of the Boltzmann equation below,

$$\frac{\partial f_s}{\partial t} + \boldsymbol{v} \cdot \frac{\partial f_s}{\partial \boldsymbol{x}} + \frac{q_s}{m_s}(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B})\frac{\partial f_s}{\partial \boldsymbol{v}} = \frac{\partial f_s}{\partial t}|_{\text{coll}}. \tag{1.8}$$

The collision term on the RHS is a sum of all the types of collisions that occur in the plasmas − Coulomb collisions, nuclear fusion collisions, neutral-ion collisions, and so forth. Some of these collisions, such as ionization and recombination, transform particles between species. Setting the collision term to zero results in the *Vlasov equation*, which is useful for collisionless plasmas including many space plasmas. Now the particle number density n_s for particle species s and the corresponding fluid velocity $\boldsymbol{u}_s$ are defined through the zeroth and

first moments of the distribution function,

$$n_s = \int f_s d\boldsymbol{v}, \quad \boldsymbol{u}_s = \int \boldsymbol{v} f_s d\boldsymbol{v}, \tag{1.9}$$

and the definitions for the charge density ρ_q and electromagnetic current $\boldsymbol{J}$ follow,

$$\rho_q = \sum_s \rho_{q,s} n_s, \quad \boldsymbol{J} = \sum_s \rho_{q,s} n_s \boldsymbol{u}_s. \tag{1.10}$$

Particle-in-cell (PIC) and other kinetic methods use the Boltzmann or Vlasov equations from Eq. (1.8) by discretizing space and time, and using a small number of discrete "super-particles" that represent averages over a much larger set of plasma particles. This is still an attempt to track all the particles and their interactions in a 7-dimensional phase space, so these models are computationally expensive and emergent simulation dynamics can be difficult to understand analytically. Further reductions are often made for realistic plasmas.

This equation can be reduced further, to the familiar 4-dimensional space-time $(\boldsymbol{x}, t)$, by taking moments of the Boltzmann equation in velocity space. Typically the zeroth, first, and second moments are taken to produce the 5-moment model for each particle species (dropping the subscripts from now on) [232],

$$\rho = m \int f(\boldsymbol{v}) d\boldsymbol{v}, \tag{1.11}$$

$$\rho \boldsymbol{u} = m \int \boldsymbol{v} f(\boldsymbol{v}) d\boldsymbol{v}, \tag{1.12}$$

$$\hat{\boldsymbol{P}} = m \int (\boldsymbol{v} - \boldsymbol{u})(\boldsymbol{v} - \boldsymbol{u}) f(\boldsymbol{v}) d\boldsymbol{v}, \tag{1.13}$$

$$\boldsymbol{\Pi} = \hat{\boldsymbol{P}} - p\boldsymbol{\delta}, \tag{1.14}$$

$$p = \frac{m}{3} \int |\boldsymbol{v} - \boldsymbol{u}|^2 f(\boldsymbol{v}) d\boldsymbol{v}, \tag{1.15}$$

$$\boldsymbol{h} = \frac{m}{2} \int |\boldsymbol{v} - \boldsymbol{u}|^2 (\boldsymbol{v} - \boldsymbol{u}) f(\boldsymbol{v}) d\boldsymbol{v}, \tag{1.16}$$

$$\mathcal{E} = \frac{p}{\gamma - 1} + \frac{1}{2} \rho u^2. \tag{1.17}$$

Here $\rho = mn$ is the fluid (mass) density, $\boldsymbol{u}$ is the fluid velocity, $\hat{\boldsymbol{P}}$ is the pressure tensor, p is the scalar pressure, $\mathcal{E}$ is the energy density, γ is the adiabatic index, $\boldsymbol{\Pi}$ is the viscous stress tensor, and $\boldsymbol{h}$ is the heat flux. Now a 5-moment equation can be defined for each of the fluids corresponding to each particle species,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{u}) = \frac{\partial \rho}{\partial t}|_{\text{coll}}, \tag{1.18}$$

$$\frac{\partial \rho \boldsymbol{u}}{\partial t} + \nabla \cdot (\rho \boldsymbol{u}\boldsymbol{u} + p\boldsymbol{\delta} + \boldsymbol{\Pi}) = \frac{\rho e}{m}(\boldsymbol{E} + \boldsymbol{u} \times \boldsymbol{B}) + \frac{\partial (\rho \boldsymbol{u})}{\partial t}|_{\text{coll}}, \tag{1.19}$$

$$\frac{\partial \mathcal{E}}{\partial t} + \nabla \cdot ((\mathcal{E} + p)\boldsymbol{u} + \boldsymbol{u} \cdot \boldsymbol{\Pi} + \boldsymbol{h}) = \frac{\rho e}{m}\boldsymbol{u} \cdot \boldsymbol{E} + \frac{\partial \mathcal{E}}{\partial t}|_{\text{coll}}. \tag{1.20}$$

Maxwell's equations in Eq. (1.7) are still required for the time evolutions of $\boldsymbol{E}$ and $\boldsymbol{B}$. Throughout this work a plasma of singly charged deuterons and electrons is assumed, so that the charges of the particles will be everywhere assumed equal or opposite to a unit of electron charge, e. Note that terms like $\boldsymbol{u}\boldsymbol{u}$ are shorthand for the tensor in index notation $u_i u_j$ (not to be confused with the species index), $\boldsymbol{\Pi} = \Pi_{ij}$, and similarly for the pressure tensor $\boldsymbol{P}$ and delta function $\boldsymbol{\delta}$. The model is complete once closures are introduced specifying the form of the stress tensor $\boldsymbol{\Pi}$ and the heat flux $\boldsymbol{h}$.

More complicated 10-moment [157] and 13-moment [401] models can be calculated by taking additional moments of the Boltzmann equation, although these models tend to be significantly harder to interpret and numerically implement [297]. Many of the plasma model varieties are summarized in Fig. 1.1, which plots the different plasma models as a function of the Knudsen number Kn (ratio of the mean free path to the smallest length scale of interest) and charge separation Λ_d (the degree to which electrical charge is relevant). For instance, when $Kn \gg 1$, $\Lambda_d \ll 1$, electrical charge is fairly unimportant and mean free paths are very large, typifying the collisionless kinetic regime, which is described by the Vlasov equation.

There are a very large number of models that stem from the 5-moment model, depending on the assumptions made, the composition of the plasma, and so on. If τ denotes the smallest timescale of experimental interest, the primary assumption from the Boltzmann equation to

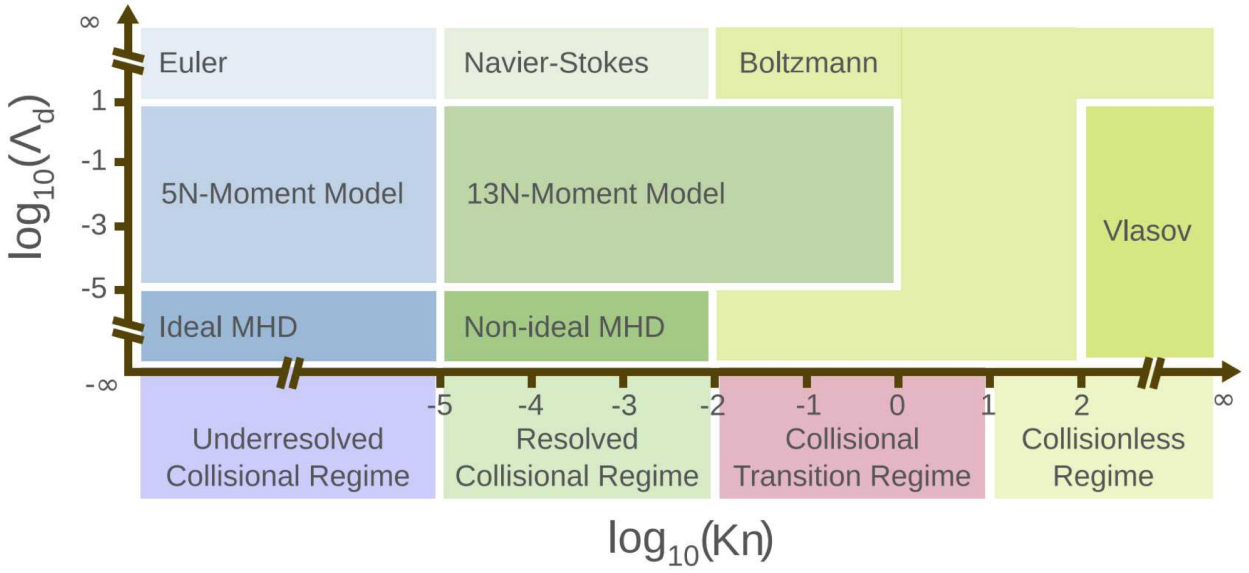


Figure 1.1: Hierarchy of plasma models as a function of Knudsen number (Kn) and charge separation distance (Λ_d). Reproduced with permission from Dr. Uri Shumlak.

the 5-moment model was that the electron-electron, ion-ion, neutral-neutral, etc. collision rates satisfy $\tau_{ee} \ll \tau$, $\tau_{ii} \ll \tau$, $\tau_{nn} \ll \tau$, so that each species is in separate local equilibrium. If there is a single species of ion and single species of neutral particle, the system can be represented with a coupled three-fluid model with ionization, recombination, and other atomic processes. If the plasma is assumed to be fully ionized in the domain of interest, the system can be reduced significantly to the two-fluid model by removing the neutral terms from the equations. In this limit, the ion and electron species are in separate local thermal equilibria, and the equations are often closed with the Braginskii closure [52]. With additional assumptions, a whole host of multi-fluid and single-fluid models can be defined, summarized in Fig. 1.2. With enough assumptions, single-fluid conducting models can be derived and such models are traditionally called MHD. For instance, if the dynamic velocities satisfy $v \ll c$, the dynamic frequencies satisfy $\omega \ll \omega_{pe}$, and the smallest experimental length scale satisfies $L \gg \lambda_D$, the system is well-described by a set of equations for a single conducting MHD fluid. Ideal MHD is obtained with even further assumptions, including zero dissipation

from resistive or viscous effects.

1.4 Magnetohydrodynamics

Magnetohydrodynamic models are the most common types of models used to simulate the macroscopic dynamics of fusion-relevant plasmas because magnetic fields are often very large (Larmor radii are very small) and MHD equilibria are far better understood than other forms of equilibria. After all, most future fusion reactors would prefer to operate in steady-state operation, although there are pulsed-pulse designs as well [126]. This work focuses on a single-fluid, two-temperature Hall-MHD model consisting of a single ion species and electrons, which includes resistive dissipation, allows for some two-fluid effects, and reduces to ideal MHD and many other MHD variants in the appropriate limits. Switching from evolution equations for $\mathcal{E}_e$ and $\mathcal{E}_i$ to the electron and ion temperatures $T_e = m_e p_e/2\rho$ and $T_i = m_i p_i/2\rho$ produces,

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \boldsymbol{u}) + D\nabla^2 \rho, \tag{1.21}$$

$$\frac{\partial \boldsymbol{u}}{\partial t} = -\boldsymbol{u} \cdot \nabla \boldsymbol{u} + \frac{1}{m_i n}\left[\boldsymbol{J} \times \boldsymbol{B} - \nabla\left(n(T_i + T_e)\right) - \nabla \cdot \boldsymbol{\Pi}\right], \tag{1.22}$$

$$\frac{\partial T_i}{\partial t} = -\boldsymbol{u} \cdot \nabla T_i + (\gamma - 1)\left[-T_i \nabla \cdot \boldsymbol{u} - \frac{1}{n}\left(\nabla \cdot \boldsymbol{h}_i - Q_i\right)\right], \tag{1.23}$$

$$\frac{\partial T_e}{\partial t} = -\boldsymbol{u}_e \cdot \nabla T_e + (\gamma - 1)\left[-T_e \nabla \cdot \boldsymbol{u}_e - \frac{1}{n}\left(\nabla \cdot \boldsymbol{h}_e - Q_e\right)\right], \tag{1.24}$$

$$\frac{\partial \boldsymbol{B}}{\partial t} = -\nabla \times \boldsymbol{E}. \tag{1.25}$$

Note that there is a single density evolution equation and a single velocity field evolution equation. Quasi-neutrality $n = n_e \approx n_i$ ($L \gg \lambda_D$) has been assumed, the fluid velocity satisfies $\boldsymbol{u} \approx \boldsymbol{u}_i$, and $\boldsymbol{u}_e$ is determined through

$$\boldsymbol{u}_e = \boldsymbol{u} - \frac{\boldsymbol{J}}{ne}. \tag{1.26}$$

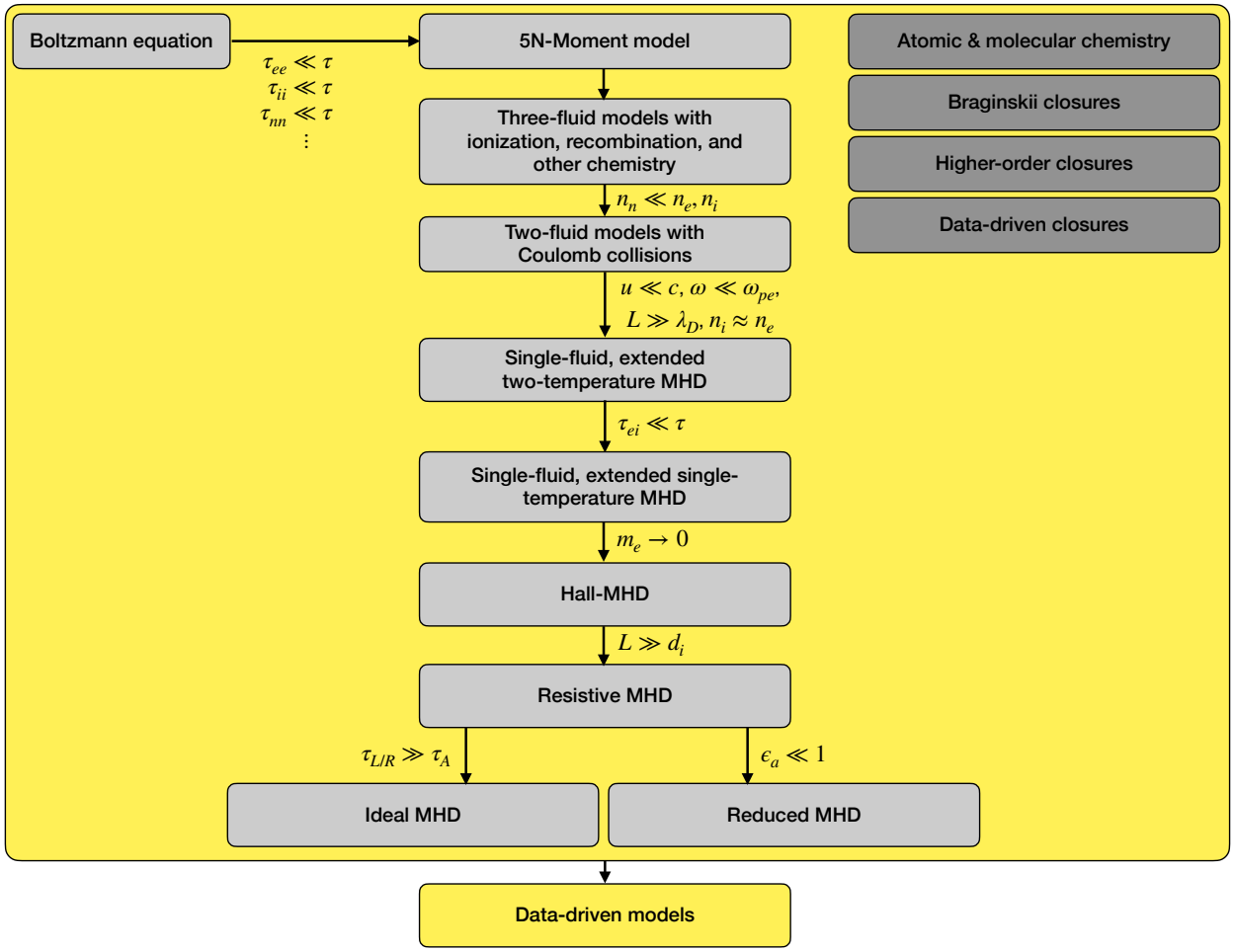


Figure 1.2: Hierarchy of models coming from the 5-moment fluid model. There is one 5-moment model per fluid, plus Maxwell's equations. Here τ_{ei} is the electron-ion collision rate, ϵ_a is the inverse aspect ratio of a device, $\tau_{L/R}$ is the resistive diffusion time, and τ_A is the characteristic timescale for an Alfvén wave transit. Data-driven methods can be applied to any place in the hierarchy for discovering reduced-order models and model closures, although this work is primarily focused with data-driven models for MHD.

For a quasi-neutral plasma with equal electron and ion temperatures, λ_D can be defined through

$$\lambda_D = \sqrt{\frac{\epsilon_0 T}{ne^2}}. \tag{1.27}$$

$T = T_e = T_i$ is the plasma temperature. The model is completed with the following closures:

$$\mu_0 \boldsymbol{J} = \nabla \times \boldsymbol{B}, \tag{1.28}$$

$$\boldsymbol{E} = -\boldsymbol{u} \times \boldsymbol{B} + \eta \boldsymbol{J} + \frac{1}{ne}\left(\boldsymbol{J} \times \boldsymbol{B} - \nabla(nT_e)\right) + f_{m_e}\frac{m_e}{ne^2}\frac{\partial \boldsymbol{J}}{\partial t}, \tag{1.29}$$

$$\boldsymbol{h}_s = -n\left[\chi_{\parallel,s}\hat{\boldsymbol{b}}\hat{\boldsymbol{b}} + \chi_{\perp,s}\left(\boldsymbol{I} - \hat{\boldsymbol{b}}\hat{\boldsymbol{b}}\right)\right] \cdot \nabla T_s, \tag{1.30}$$

$$Q_i = -\left(\nabla \boldsymbol{u}\right)^T : \boldsymbol{\Pi} - Q_{\text{coll}}, \tag{1.31}$$

$$Q_e = \eta \boldsymbol{J}^2 + Q_{\text{coll}}, \tag{1.32}$$

$$Q_{\text{coll}} = 2 \times 10^{-14}\frac{n}{T_e^{\frac{3}{2}}}(T_i - T_e), \tag{1.33}$$

$$\boldsymbol{\Pi} = -\nu(\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T - \frac{2}{3}\boldsymbol{I}\nabla \cdot \boldsymbol{u}), \tag{1.34}$$

Here $\hat{\boldsymbol{b}}$ is a unit vector in the direction of the magnetic field, η is the plasma resistivity, ν is the plasma viscosity, $\chi_\parallel$ ($\chi_\perp$) is the heat conductivity parallel (perpendicular) to the magnetic field, and $\boldsymbol{I}$ is the identity matrix. Strictly, Eq. 1.21 should not contain a diffusive term. However, an artificial particle diffusivity $D = 250$ m^2/s is used to prevent numerical issues with small scale density oscillations characteristic of the dynamics in HIT-SI simulations that are discussed in Sec. 2.1; this technique is common in MHD simulations of dynamic plasma systems [178, 327] and in other contexts such as the analytic work in Chapter 3 it should be understood that this term is dropped. Similarly, with the assumptions made to derive this model, the last term in Eq. (1.29) should be omitted; the term is included here for completeness because it provides a stabilizing effect on the magnetic field evolution for

HIT-SI simulations that are described in Chapter 2.

In the ideal, single-temperature Hall-MHD limit with zero-velocity (no-slip) boundary conditions, there are several global conserved quantities within a bounded volume: the total energy

$$W \equiv \int \left(\frac{B^2}{2\mu_0} + \frac{1}{2}\rho u^2 + \frac{p}{\gamma - 1} \right) d^3\boldsymbol{x}, \tag{1.35}$$

total cross-helicity (not an ideal invariant of Hall-MHD, but an ideal invariant of standard MHD assuming $\boldsymbol{B} \cdot \hat{\boldsymbol{n}} = 0$ and $\boldsymbol{u} \cdot \hat{\boldsymbol{n}} = 0$),

$$H_c \equiv \int \boldsymbol{u} \cdot \boldsymbol{B} d^3\boldsymbol{x}, \tag{1.36}$$

total magnetic-helicity (requires a perfectly conducting boundary condition to be conserved),

$$H_m \equiv \int \boldsymbol{A} \cdot \boldsymbol{B} d^3\boldsymbol{x}, \tag{1.37}$$

and total generalized-helicity [133],

$$H_g \equiv \int (\boldsymbol{A} + d_i \boldsymbol{u}) \cdot (\boldsymbol{B} + d_i \nabla \times \boldsymbol{u}) d^3\boldsymbol{x}. \tag{1.38}$$

Note that H_m and H_g are not gauge-invariant except with a perfectly conducting boundary or other slight modifications to fix this ambiguity [31]. Here the ion inertial length d_i is the characteristic length associated with the Hall-term and represents the scale at which ions decouple from electrons. The ion inertial length is defined through the ion plasma frequency ω_{pi},

$$d_i \equiv \frac{c}{\omega_{pi}} = \frac{m_i}{e}\sqrt{\frac{1}{\mu_0 \rho}}, \quad \omega_{pi}^2 \equiv \frac{n_i e^2}{m_i \epsilon_0}. \tag{1.39}$$

For the no-slip boundary, mass is also conserved. Momentum is not conserved (unless $\boldsymbol{B} = 0$ everywhere on the boundary), even in the inviscid limit with $\boldsymbol{B} \cdot \hat{\boldsymbol{n}} = 0$, because momentum is transported along magnetic field lines. Note that an inviscid MHD model is one with no

dissipation (resisitivity and viscosity are negligible), while ideal MHD refers to an inviscid model with additional assumptions such as $d_i = 0$. These global invariants are important constraints on the underlying structure of the dynamics, and Chapter 3 will show that one can build this extra structure directly into data-driven models.

1.4.1 MHD equilibria

Most future fusion reactor concepts require a device that operates in steady-state. Many other plasmas are also quasi-steady-state or otherwise exhibit equilibrium features. Therefore, understanding basic MHD equilibria is important for designing machine learning methods and many other tasks. The ideal MHD equilibrium equations can be obtained by neglecting dissipation, setting Eqs. (1.21) − (1.25) equal to zero, and assuming $\boldsymbol{u} = 0$, i.e. it is a static equilibrium. This gives

$$\begin{aligned} \boldsymbol{J} \times \boldsymbol{B} &= \nabla p, \\ \nabla \times \boldsymbol{B} &= \mu_0 \boldsymbol{J}, \\ \nabla \cdot \boldsymbol{B} &= 0. \end{aligned} \tag{1.40}$$

Many fusion devices are toroidal and exhibit toroidal axisymmetry. Therefore these equations only depend on the coordinates (R, Z) in a cylindrical coordinate system (R, Z, ϕ). In this case, the poloidal flux stream function ψ is defined such that

$$\boldsymbol{B} = B_\phi \boldsymbol{e}_\phi + \frac{1}{R} \nabla \psi \times \boldsymbol{e}_\phi, \tag{1.41}$$

where $\boldsymbol{e}_\phi$ is a unit vector in the toroidal direction. Rewriting the MHD equilibrium equations in terms of ψ leads to the Grad-Shafranov equation [148, 395],

$$\Delta^* \psi = -\mu_0 R^2 \frac{dp}{d\psi} - \frac{1}{2}\frac{d(RB_\phi)^2}{d\psi}. \tag{1.42}$$
$$\Delta^* \equiv R\frac{\partial}{\partial R}\frac{1}{R}\frac{\partial}{\partial R} + \frac{\partial^2}{\partial Z^2}$$

Here $p(\psi)$ is the pressure, which is now a function of the poloidal flux. By specifying the two free functions, p and RB_ϕ, one can solve the Grad-Shafranov equation numerically for the magnetic field and current density in the MHD equilibrium. This is a very challenging partial differential equation (PDE) to solve analytically, and there are many different ways to solve it numerically [194, 159]. Once the equilibrium is found, there are a number of important quantities to compute that characterize the quality of the equilibrium, including the plasma β (confinement quality), plasma safety factor (a metric of device stability), and more general stability criteria.

1.5 Data-driven modeling

This introduction has illustrated that there are a tremendous number of known plasma models of varying model complexity, from MHD to the Klimontovich equations. However, a large gap exists in the lower levels of this hierarchy between simple circuit models and the many MHD variants. These low-level models are motivated because higher fidelity models typically require computationally intensive and high-dimensional simulations [73, 328, 150], obfuscating the dynamics and precluding model-based real-time control. Moreover, many high-dimensional nonlinear systems tend to evolve on low-dimensional attractors [414]. Plasmas across a large range of parameter regimes, geometry, and degree of nonlinearity exhibit this feature [196, 432, 341, 434, 409, 69, 152, 207]. In these cases, the evolution of only a few coherent structures, obtained from model-reduction techniques [35, 36], can closely approximate the evolution of the high-dimensional physical system. Additionally, recent progress in theoretical, data-driven, and machine learning methods are revolutionizing the analysis,

modeling, and control of high-dimensional, nonlinear systems, especially in the field of fluid mechanics [320, 323, 414, 77, 374, 63, 65].

The applications of reduced-order models (ROMs), i.e. low-dimensional models, include understanding reduced physical mechanisms [261, 153], computationally efficient simulations [185], digital twins (virtual time-dependent models for a dynamic system, constantly updating with sensor measurements) [213], and real-time control [362, 131, 246]. For example, acceleration of ICF simulations and digital twins can facilitate an exploration of the implosion parameter space [185], surrogate closure models can lead to more accurate and efficient fluid simulations [436], surrogate gyrokinetic transport models can speed up tokamak simulations by orders of magnitude [91, 431], and steady-state tokamak operation will require the active avoidance or mitigation of disruptions, which can seriously damage components of the device [255]. For these real-time control challenges, there are a wealth of model-based control techniques such as model predictive control [7] that can be leveraged for plasma systems. However, existing models like MHD or PIC simulations are often too high-dimensional and computationally expensive to operate in real-time, and heuristic models are often too low-fidelity to be useful for control. Deep learning methods are increasingly making significant contributions in plasma physics, but it remains difficult to interpret these models or build in guarantees regarding model stability or performance.

In addition to being computational efficient, reduced-order models can help uncover key mechanisms that govern the evolution of the dominant coherent structures. This aspect of reduced-order modeling has a rich history, from the famous Lorenz model in 1963 [261], through the present era, including the low-order mechanistic model of the cylinder wake in 2003 by Noack et al. [320]. Recently, data-driven algorithms, like the algorithms described in Chapter 5, have shown potential to uncover similarly interpretable and useful models. Examples include related fluid systems [257, 256] as well as recent work that uncovers a Lorenz-like model of electroconvective chaos by Guan et al. [153]. Moreover, increasingly reduced order models are used to describe key mechanisms in plasma physics, including "predator-prey" dynamics in gyrokinetic simulations [225], direct data-driven discovery of

reduced MHD or kinetic equations from a plasma dataset [8], data-driven fluid models for the L-H mode transition in tokamaks [100], and the data-driven models for fully 3D MHD simulations of the HIT-SI experiment presented in Chapter 5. These models can be critical for providing insight into the physical system, including energy transfers and other nonlinear interactions.

Reduced-order models traditionally fall into two categories: projection-based model reduction and data-driven system identification. Projection-based model reduction is achieved by first computing the evolution of a governing PDE, often by spatially discretizing the domain, resulting in a high-dimensional system of ordinary differential equations (ODEs). Then a low-dimensional orthogonal basis is computed, often via the proper orthogonal decomposition (POD) [177, 414]. This process is also called the biorthogonal decomposition (BD or BOD) in the plasma physics community. Finally, the high-dimensional model is "Galerkin-projected" onto this basis [441, 34], resulting in an efficient reduced system that describes how the amplitudes of the POD modes evolve in time. However, this projection is intrusive since it requires knowledge of the governing physics and high-fidelity measurements or a high-fidelity numerical solver. The POD method will be reviewed and discussed along with POD-Galerkin methods for MHD models in Chapter 3.

In contrast, system identification techniques attempt to identify data-driven models directly from measurement data, often without knowledge of the governing equations. Increasingly, data-driven methods are producing effective bases beyond POD for different experimental or computational tasks [414]; modern methods include balanced POD [441, 372], spectral POD [425], dynamic mode decomposition (DMD) [393, 371, 427], the Koopman decomposition [229, 295, 62], resolvent analysis [284, 265], and neural-network-based autoencoders [266, 81, 243]. Data-driven techniques, including modern machine learning, are also being widely applied to discover dynamical systems models of complex physical systems [47, 394, 59, 360, 359, 450, 222, 438, 277, 114, 344, 325, 25, 226], with a particular emphasis on hybrid physics-inspired or physics-informed machine learning [359, 26, 257, 98, 298, 97]. In fluid mechanics, sparse model discovery has been used to develop interpretable nonlinear

models that enforce known physics by construction [59, 257]. This work provides similar models for plasmas described by MHD in Section 5.

1.6 Motivation for this work and broader perspective

There is an unprecedented opportunity to apply and formulate modern machine learning and reduced-order modeling techniques for understanding, forecasting, and controlling plasmas. Many reduced order modeling methods for fluids rely on the structure (particularly quadratic nonlinear structure) of the Navier-Stokes equations. There are many open questions about how to generalize these useful and powerful methods for plasmas that are well-described by various forms of MHD or Boltzmann-like equations.

Machine learning is now ubiquitous in plasma physics for a large number of tasks: interpreting diagnostic data [280], maximizing diagnostic information [308], prediction of disruption or other operational limits [364], and reduced order models for modeling phenomena as disparate as nonlinear opacities in ICF [224] to speeding up PIC simulations [233]. Almost all of this work is done with deep learning methods, although recent work that integrates Bayesian modeling is helping to fill in some important gaps in the literature and allowing for uncertainties on data-driven model predictions [345, 174, 346, 247, 396, 188]. Deep learning methods are useful and powerful but have some significant drawbacks: (1) they are often "physics-blind" and lack interpretability and (2) often require large troves of data to train successful models. Plasma devices such as tokamaks have fundamental limits on diagnostic access, tight operational margins in order to avoid disruption and damage, and important but intermittent or rare mode activity. For instance, in a DIII-D tokamak database built specifically for identifying Alfvén eigenmodes (described in Appendix C.2), more than 99% of the data still exhibits no activity whatsoever.

Despite these rare features, DIII-D has decades of diagnostic data obtained in vast ranges of operational parameter space, so control models from deep learning or other expensive methods can be trained offline before consolidation with a real-time control algorithm. These large and expansive databases are optimal targets for these deep data-driven approaches,

which excel at extracting patterns from high-dimensional spaces [144]. Given that a high-quality database is available, machine learning techniques have distinct advantages over other ROMs: (1) offline training allows for large nonlinear models, (2) multi-machine datasets facilitate finding universal plasma models across fusion devices [300], and (3) models can be generated for plasma instabilities and other plasma dynamics that are currently not well-understood or not amenable to any sort of linearization. Along these lines, Appendix C illustrates some collaborative work predicting Alfvén eigenmodes in the DIII-D tokamak.

However, there are fewer reduced order plasma modeling paradigms that allow for other attractive model features such as interpretability, sparse or low-dimensional models, physics-constrained models, models that can be trained online, and models that can be trained on minimal datasets. There are many benefits that can come from incorporating physical priors into system identification, including reducing the space of possible models and enforcing hard constraints such as $\nabla \cdot \boldsymbol{B} = 0$. Interestingly, the plasma physics community can follow the lead of the field of fluid mechanics, which has seen unprecedented progress in interpretable data-driven models that are connected in various ways to the underlying fluid physics. These methods are used along with more traditional deep learning, helping to fill in gaps of understanding and predictive capability. Because these models are often physics-informed, significant analytic work can be required to reformulate these methods for plasma physics. For instance, the incompressible Navier-Stokes equations has a single nonlinearity through the convective derivative term while much of the interesting dynamics in magnetohydrodynamic flows is driven by the Lorentz nonlinearity; while both nonlinearities are quadratic in the fluid fields, they lead to incredibly different dynamic behaviors, including significant differences in invariant quantities (e.g. helicity), turbulent cascades (e.g. inverse cascades), and equilibrium configurations (e.g. static pressure confinement). Data-driven plasma models should also be responsive to the fundamental anisotropy induced by magnetic fields in magnetohydrodynamic fluids. Thus, much of this work is focused on adapting data-driven models from fluid mechanics (convective nonlinearity) to magnetohydrodynamics (convective, Lorentz, and other nonlinearities). There is also exciting work to be done for

both fields in the realm of kinetics. Beyond interpretable reduced-order models, there are now modern algorithms for sparse sensor placement (for optimizing diagnostic information from a set of limited sensors), magnetic mode spectroscopy, and connections to be made with the large literature regarding plasma waves and instabilities. It is in this interpretable, physics-informed, and low-dimensional regime that this work paves a unique path in the intersection of plasma physics and machine learning.

The primary technical contributions of this work to the intersection of machine learning and plasma physics include: (Chapter 3 and Ref. [211]) formalization of a class of reduced-order models for a number of MHD systems, including deriving constraints on these low-dimensional models from global conservation laws, (Chapter 4 and Ref. [207]) exploration and development of a number of dynamic mode decomposition methods for magnetic spectroscopy and forecasting plasma measurements with linear models, (Chapter 5 and Ref. [211]) constrained data-driven identification of reduce-order models directly from plasma measurement data, (Sections 5.3−5.7 and Ref. [209]) building global stability theorems into system identification methods for producing provably stable data-driven models, and (Chapter 6 and Ref. [210]) incorporation of these innovations and other work in the field of system identification into the open-source `PySINDy` code.

1.7 Outline

The presentation of this work will proceed as follows: Chapter 2 will detail initial work with more traditional numerical simulations in plasma physics. A two-temperature Hall-MHD model is implemented in the 3D PSI-Tet code, and used to investigate how these changes alter the dynamics in the HIT-SI device. In chapter 3, projection-based ROMs for plasmas are introduced with a focus on classic Galerkin projection methods. Chapter 4 begins the exploration of data-driven models, detailing work exploring the dynamic mode decomposition for magnetic spectroscopy and forecasting. In chapter 5, sparse regression techniques are introduced for identifying dynamical and nonlinear models directly from data. Methods are demonstrated to constrain these techniques with global conservation laws in

fluids and plasmas, and analytic results from nonlinear stability theory. Chapter 6 concludes with a discussion and implementation of a number of robust variants of sparse regression, facilitating the analysis and modeling of a large number of nonlinear systems across many scientific disciplines.

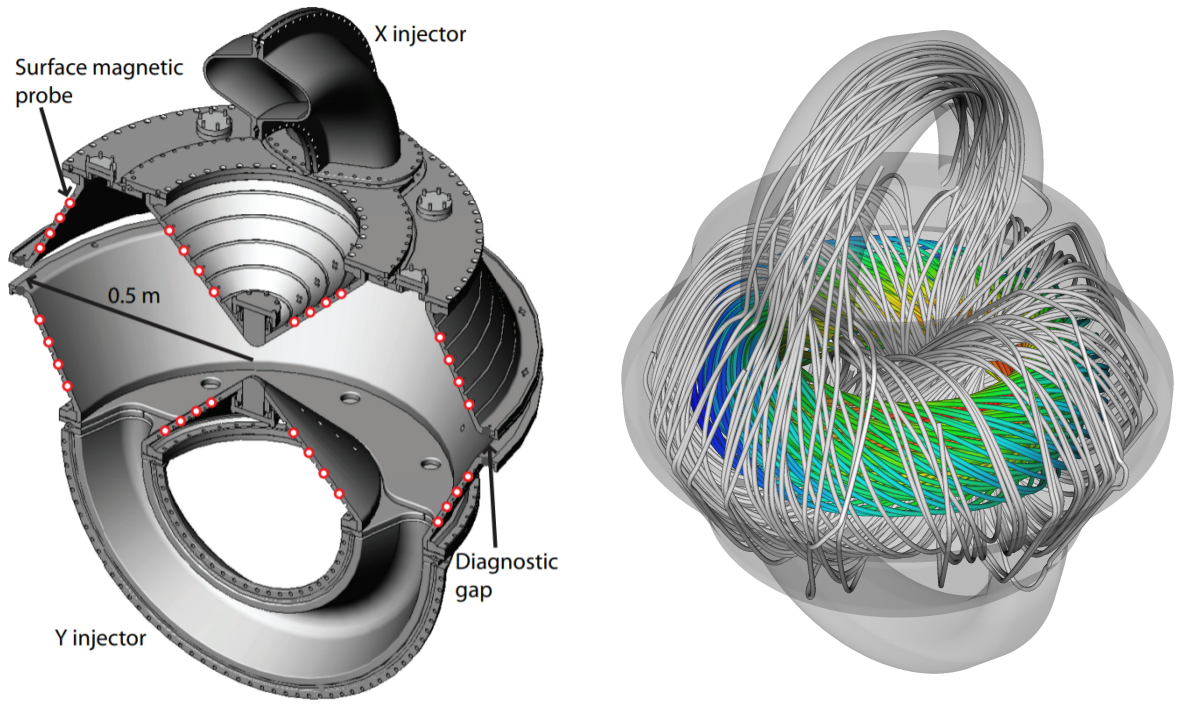
Chapter 2

TWO-TEMPERATURE HALL-MHD SIMULATIONS OF THE HIT-SI DEVICE

Before exploring the intersection of machine learning and plasma physics, a chapter is required to detail work done using more traditional numerical simulations. The study presented in this chapter has also been published as a paper in Kaptanoglu et al. [208] and begins with a description of the HIT-SI experiment in order to illustrate why a two-temperature Hall-MHD model is appropriate for this plasma device.

2.1 The HIT-SI experiment

HIT-SI was a laboratory plasma device at the University of Washington that formed and sustained spheromak plasmas for the study of plasma self-organization and steady inductive helicity injection [191]. It consisted of an axisymmetric flux conserver and two inductive injectors (actuators) mounted on each end as illustrated in the top left panel of Fig. 2.1a. A flux conserver is a volume bounded by a wall that, up to its resistive timescale, conserves the magnetic helicity defined in Eq. (1.37). The wall is usually made of a material with high conductivity. The HIT-SI flux conserver was constructed from 1/2" thick copper and the plasma-facing surface was coated with a thin insulating layer so that $\boldsymbol{J} \cdot \hat{\boldsymbol{n}} = 0$ and therefore the injector operation was purely inductive. Magnetic coils on each injector, generating helical fields linking through the flux conserver, were oscillated in phase at a frequency f_{inj} with values between $10 - 70$ kHz. The magnetic fields generated by the two injectors were spatially and temporally 90° out of phase, resulting in approximately constant power and magnetic helicity injection. The fields from these injectors provided the power and magnetic helicity to both form and sustain a spheromak during experimental discharges,



(a) (b)

Figure 2.1: (a) A cross section of the device shows the toroidal structure, the two magnetic helicity injectors, the surface probes, and the diagnostic gap. Reproduced from Wrobel et al. [447], with the permission of AIP Publishing. (b) Representative MHD equilibrium during sustainment showing an axisymmetric spheromak (rainbow) surrounded by field lines tied to the injector (gray).

with a quasi-steady-state period of roughly constant spheromak amplitude lasting $\lesssim 1$ ms. The sustainment period of each experimental discharge, indicated by the vertical black lines, is illustrated for a typical discharge in Fig. 2.2.

In HIT-SI, some of the characteristic time scales overlap, such as the toroidal Alfvén time τ_A and the injector frequency f_{inj}. The magnetic topology is fundamentally 3D and magnetic perturbation amplitudes are comparatively large ($|\delta \boldsymbol{B}|/|\boldsymbol{B}| \approx 10\%$ in HIT-SI and $|\delta \boldsymbol{B}|/|\boldsymbol{B}| \lesssim 0.1\%$ in typical tokamaks) [160]. Additional details of the experiment and its operation can be found in references [191, 448, 435].

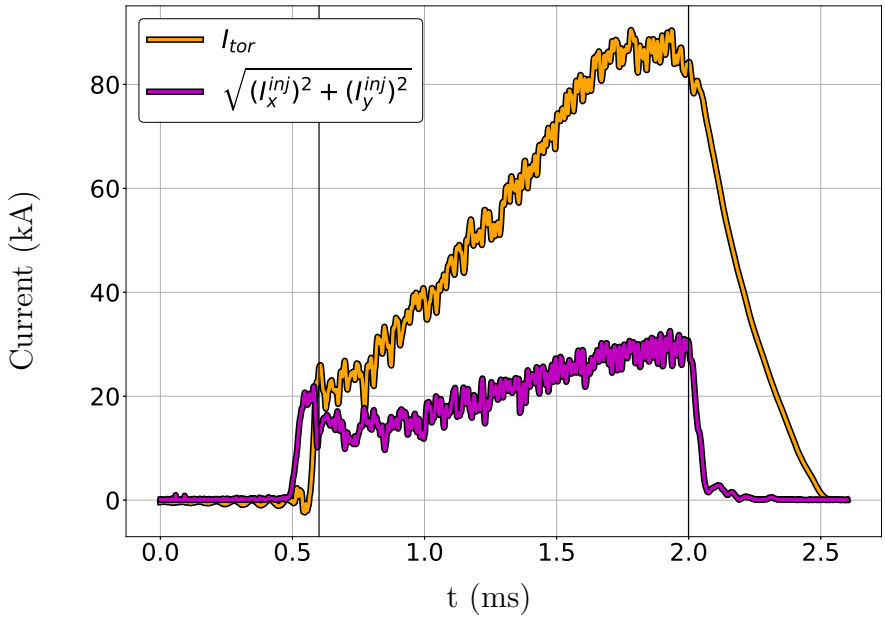


Figure 2.2: Toroidal current (yellow) and total injector current (purple) waveforms for the HIT-SI experiment, discharge 129499. The black vertical lines indicate the sustainment regime when the spheromak has formed and is being sustained by the injectors.

Previous work on HIT-SI simulations started with a resistive single fluid MHD model [186] followed by a model including Hall terms to capture some of the effects expected with the full two-fluid system of separate ions and electrons [5]. The importance of different physical terms in the MHD equations is interpreted from theoretical considerations [193] and the results of experimental validation [186, 5, 303]. Both experimentally and theoretically, two-fluid effects through the Hall terms have been shown to be important for spatial scales between the ion and electron inertial scales [127]. In HIT-SI, the ion inertial scale $d_i \approx 8$ cm is comparable with experimental length scales, such as the diameter of an injector mouth $d_{\text{inj}} \approx 14$ cm, and the characteristic magnetic scales of the spheromak $\lambda_{\text{sph}}^{-1} \approx 10$ cm and injectors $\lambda_{\text{inj}}^{-1} \approx 5$ cm [189, 31]. Numerical models with the Hall terms significantly improve agreement with HIT-SI measurements over the resistive MHD models [5], and validate well with many bulk measurements of the experiment [303, 32]. Modeling the full injector geometry, especially capturing the short-lived reconnection events in the injectors, led to improved agreement with

experimental measurements [163]. Additionally, from measurements of HIT-SI [181] (and a newer device called HIT-SI3 [120]) and theory [189, 191], ion temperatures are expected to be significantly higher than electron temperatures. Therefore, distinguishing between the ion and electron temperatures is a well-motivated change to the Hall-MHD equations, leading to new dynamics in HIT-SI simulations.

In general, investigating a collection of models and geometries of varying complexities facilitates new understanding of the plasma dynamics, as well as validation of the numerical models and codes. Although validation with experimental data is important for quantitative and predictive understanding of experimental results using simulations, it is not the primary goal of this chapter. Rather, Chapter 2 is part of a systematic and sustained effort to investigate the complexity of the physical models, geometry, and boundary conditions required to capture the important dynamical processes in this device. In particular, the aim of this chapter is to elucidate the underlying physics through identification of important terms in the dynamics.

As part of this work, the two-temperature Hall-MHD model in Eqs. (1.21) − (1.26) was implemented into the 3D PSI-Tet finite-element code [163] for simulations of the HIT-SI device. Subsequently, a number of parameter scans were performed, and a number of new dynamical phenomena were investigated.

2.2 HIT-SI simulations with the NIMROD and PSI-Tet codes

In order to perform simulations, the closures need to be clarified for a two-temperature model. For this study, an anisotropic Braginskii thermal conduction with temperature and magnetic field dependencies [52] $\chi_{\perp,s}$ and $\chi_{\parallel,s}$, Spitzer-like resistivity [406] $\eta = \eta_0/T_e^{\frac{3}{2}}$, and constant and isotropic viscosity $\nu = 550\ \mathrm{m}^2/\mathrm{s}$ are assumed. An approximate demagnetization effect, which enforces $\chi_{\perp,s} \leq \chi_{\parallel,s}$, is included in the thermal conduction coefficients because there are locations, like reconnection regions inside the injectors, where the magnetic field vanishes. The viscosity value used is approximately that obtained from Braginskii [52], assuming plasma parameters late in the discharge. This value matches that which was used

in prior studies. Previous HIT-SI simulations found that "while there are changes to the flow profile between isotropic and anisotropic viscosity models, they tend to have little influence on the bulk evolution of the plasma" [303]. The electrons are approximated in this model to receive the entirety of the Ohmic heating, while only ions receive viscous heating. These choices are well-motivated because the ratio of electron to ion Ohmic heat and the ratio of ion to electron viscous heat are both approximately $\sqrt{m_i/m_e}$. The heat exchange from ion-electron collisions, Q_{coll}, is an approximation obtained from the NRL plasma formulary [184].

The electron inertia term, the last term on the RHS in Eq. (1.29), incorporates two-fluid effects by introducing the electron inertial scale to the evolution equation and retaining the whistler-wave dispersion cutoff at the electron cyclotron frequency, $\omega = \omega_{ce}$. Despite the apparent physical motivation, this term should not strictly be included in the code at this level of approximation because there are other effective two-fluid terms of the same order that are omitted. The inclusion of this term is motivated instead because it helps to stabilize the magnetic field evolution; in both the PSI-Tet and NIMROD numerical codes, there is an enhancement factor for the electron inertia precisely for this reason. Enhanced electron inertia is a common numerical trick in MHD codes. Since the electron mass is artificially enhanced, high-frequency whistler waves are damped, reducing numerical stiffness in the magnetic field evolution. Moreover, to leading order $f_{m_e} m_e \dot{\boldsymbol{J}}/ne^2 \approx f_{m_e} m_e \omega_{\text{inj}} \boldsymbol{J}/ne^2 \propto \eta_f \boldsymbol{J}$, with η_f interpretable as additional resistivity for currents changing at a fixed frequency. In HIT-SI, for currents oscillating at the lowest frequency (14.5 kHz) used in the present work, η_f is smaller than η by a factor of 2-3. At higher frequencies, the terms are of similar magnitude. Fortunately, equilibrium and other slowly changing currents will see a smaller effective η_f. Typically $f_{m_e} = 36.72$ is used because this value was found in previous work to be well-converged for simulations similar to those presented here [5, 304]. A convergence study was also performed as part of this work, which showed no significant changes in the range $f_{m_e} = 18.36 - 73.44$. These and other relevant parameters are summarized in Table 2.1.

Parameter	**Value [Units]**
Injector Flux	0.5 [mWb]
Injector Current	8 [kA]
n_{wall}	0.75 [10^{19} m^{-3}]
T_{wall}	3 [eV]
ν	550 [m^2/s]
D	250 [m^2/s]
η_0	5.327×10^{-4} [Ω m eV$^{-\frac{3}{2}}$]
$m_i/(f_{m_e} m_e)$	100

Table 2.1: Fixed parameters for PSI-Tet and NIMROD frequency scan simulations.

2.2.1 PSI-Tet

PSI-Tet is a 3D high-order finite element code that supports multi-physics models on unstructured tetrahedral grids. The grids can be generated directly from Computer-Aided Design models and this discretization facilitates the accurate representation of complex 3D geometries like the HIT-SI device. The approximate experimental boundary conditions for HIT-SI are

$$\boldsymbol{B} \cdot \hat{\boldsymbol{n}} = 0, \qquad \boldsymbol{J} \cdot \hat{\boldsymbol{n}} = 0. \tag{2.1}$$

$\boldsymbol{B}$ is the magnetic field, $\boldsymbol{J}$ is the current density, and $\hat{\boldsymbol{n}}$ is a unit normal vector to the wall. The numerical implementation of these boundary conditions is enabled by the unique mixed element discretization used by PSI-Tet. In PSI-Tet, the injector flux and voltage circuits in HIT-SI are approximated by imposing a time-dependence for $\int \boldsymbol{B} \cdot \boldsymbol{dl}$ on independent sets of closed loops on each injector. This boundary condition fixes the current enclosed within these loops as a function of time and the values are chosen to reproduce experimental waveforms. The simulations initialize the magnetic field as a small Taylor state [31] of a few Amperes of toroidal current; this is not strictly necessary, nor does it determine the direction of the toroidal current after spheromak formation, but it tends to improve the initial numerics. More details can be found in the original works [163, 160].

For the remaining fields, PSI-Tet uses Dirichlet boundary conditions equal to the initial condition for velocity, temperature, and density, with the values

$$\boldsymbol{u}_{\text{wall}} = 0, \qquad T_{i,\text{wall}} = T_{e,\text{wall}} = 3 \text{ eV}, \qquad n_{\text{wall}} = 0.75 \times 10^{19} \text{ m}^{-3}, \tag{2.2}$$

These choices are typical for past HIT-SI simulations, and were chosen based on validation with experiment [186, 5, 304]. A more principled approach to these choices would require the evolution of a neutral fluid to model the atomic processes such as ionization and recombination, which are expected to be very important at the boundaries [370, 285]. This is a subject of future work [411].

An implicit Crank-Nicolson time advance is used with a maximum time step of 40 ns, determined to be suitable by convergence studies [160]. Typical time steps are 40 ns for low frequency simulations and 10 ns for high frequency simulations. The number of grid cells used in the PSI-Tet simulations is 103,887, corresponding to a uniform grid spacing of 2.8 cm with third order basis functions. This produces an approximate resolution of 9 mm, which resolves the electron inertial scale $d_e \approx 12$ mm, although most of the electron dynamics occurring at this scale are not included in Equations 1.21−1.26.

2.2.2 NIMROD

If the precise effects of the injector geometry are not important for a given study, considerable numerical efficiency can be gained by solving the evolution equations with algorithms optimized for axisymmetric geometries. NIMROD is a versatile extended MHD code used for simulating spheromaks [179, 5, 303, 304, 305] as well as many other axisymmetric or quasi-axisymmetric plasma systems [220, 217]. NIMROD discretizes equations in cylindrical coordinates (R, ϕ, Z) and the R-Z plane is composed of finite elements while the ϕ component is expanded in a finite Fourier series. Instead of implementing the boundary condition $\boldsymbol{J} \cdot \hat{\boldsymbol{n}}$, a thin layer of high resistivity with $\eta_{\text{wall}}/\eta_{\text{plasma}} \approx 10^5$ is used to impede current flow into the wall. To avoid numerical issues from a sharp jump in η, a matching boundary layer is

added to the mesh so that the variation occurs within a single cell. NIMROD is restricted to toroidally symmetric geometries, so the HIT-SI injectors are modeled as boundary conditions on the flux conserver.

The magnetic field boundary condition approximates the action of the magnetic helicity injectors through a combination of $\boldsymbol{B}\cdot\hat{\boldsymbol{n}}$ and $\boldsymbol{E}\times\hat{\boldsymbol{n}}$ conditions. The spatial profile of $\boldsymbol{B}\cdot\hat{\boldsymbol{n}}$ is generated through a Grad-Shafranov solution of the injector geometry and is detailed in reference [5]. The injector current, which is used to define $\boldsymbol{E}\times\hat{\boldsymbol{n}}$, has the same spatial profile of $\boldsymbol{B}\cdot\hat{\boldsymbol{n}}$ and differs by a scale factor and a phase delay in time. As in the experiment and PSI-Tet simulations, the two magnetic helicity injectors are set to a relative phasing of 90°.

The entire wall, including the injector "openings", uses a constant and uniform Dirichlet boundary condition for temperature and density with the values in Table 2.1. A minor complication is that the two-temperature simulation at $f_{\text{inj}} = 14.5$ kHz (in Sec. 2.3) was performed with $T_{\text{i,wall}} = T_{e,wall} = 1$ eV because of numerical issues; this minor change is justified further in Section 2.4.1 by showing that the dynamics are fairly insensitive to the wall temperatures. More importantly, the Dirichlet temperature and density boundary conditions for the injector may lead to significantly different dynamics than those observed in PSI-Tet. These boundary conditions enforce a cold, uniform plasma across the injector mouths. Parallel heat conduction then produces cold channels of plasma in the main volume which are linked to the injectors. This is in direct contrast to PSI-Tet simulations which observe highly dynamic plasma channels in the injectors which are much warmer than the wall temperature. These NIMROD simulations can be made more dynamically similar to the PSI-Tet simulations by adjusting the Dirichlet temperature and density boundary conditions on the injector mouths, but this tends to be a heuristic procedure.

The velocity is zero at the boundary except for a constant inward flow at each injector mouth to counter-act density holes that cause numerical issues. Without this inward flow, before the spheromak has formed, a density hole often forms in the core of the device from the establishment of increased temperature near the magnetic axis. This temperature pushes particles outwards due to a lack of pressure confinement, creating a deep density hole in the

core; during the formation event the density becomes so low that the Hall terms, proportional to n^{-1}, cause numerical instability. Regardless, inward flows are well-motivated by and comparable in magnitude to large flows seen in PSI-Tet simulations [163, 32]. The spatial profile of the normal velocity matches the absolute value of the normal magnetic field, and the peak velocity ≈ 19.5 km/s is in approximate agreement with the flow velocity observed on the experiment with ion doppler spectroscopy [181]. Like PSI-Tet, the NIMROD initial conditions are the same as the boundary conditions for the temperatures, density, and velocity (velocity is initialized to zero everywhere except on the injector mouths), with an initial Taylor state of a few Amperes.

The average cell size in the poloidal plane was 1.8 cm and used fourth order basis functions. Eleven Fourier modes in the toroidal direction were used, corresponding to a toroidal node spacing of approximately 8.3 cm at the mid-radius of the domain (27.5 cm at the outboard edge). Grid resolution studies for both PSI-Tet [163, 160] and NIMROD [5, 304] have shown convergence of results at these resolutions. Lastly, NIMROD solves the same system of equations that PSI-Tet does, with the exception of a divergence cleaning term added to the magnetic field evolution [405]. Before moving on to the results, terms in the power flows are identified in order to disambiguate various physical effects.

2.2.3 Power flows

Significant differences in ion and electron temperatures can often be explained by examining the heating and loss terms. The injector voltage circuit is responsible for most of the experimental power input, but in PSI-Tet only the current waveform of this circuit is known. The consequence is that the total injected power cannot be calculated directly in PSI-Tet. Instead, the total injector power is approximated from a power balance of the total time rate of change of the total energy with the ion and electron heat fluxes to the wall through

$$P_{\text{inj}} + \oint (\boldsymbol{h}_i + \boldsymbol{h}_e) \cdot \hat{\boldsymbol{n}} dS = \frac{d}{dt} \int \left[\frac{B^2}{2\mu_0} + \frac{n(m_i u^2 + m_e u_e^2)}{2} + \frac{n(T_i + T_e)}{\gamma - 1} \right] d^3\boldsymbol{x}. \tag{2.3}$$

The integration volume is the entire HIT-SI volume, and the integration surface is the entire HIT-SI boundary, including the injectors. Individual thermal energy flow terms are also tracked

$$\underbrace{\frac{\partial}{\partial t}\int \frac{n(T_i+T_e)}{\gamma-1}d^3\boldsymbol{x}}_{\text{thermal power}} = \underbrace{\oint (\boldsymbol{h}_e+\boldsymbol{h}_i)\cdot\hat{\boldsymbol{n}}dS}_{\text{heat flux to wall}} \tag{2.4}$$
$$-\underbrace{\int (\nabla\boldsymbol{u})^T:\boldsymbol{\Pi}d^3\boldsymbol{x}}_{\text{viscous heat}} + \underbrace{\int \eta J^2 d^3\boldsymbol{x}}_{\text{Ohmic heat}} + \underbrace{\frac{1}{\gamma-1}\int (T_i+T_e)D\nabla^2 n d^3\boldsymbol{x}}_{\text{diffusive heat}}$$
$$-\underbrace{\frac{1}{\gamma-1}\int\left[(T_i+T_e)(\boldsymbol{u}\cdot\nabla n+\gamma n\nabla\cdot\boldsymbol{u})+\frac{(\gamma-1)T_e}{ne}\boldsymbol{J}\cdot\nabla n\right]d^3\boldsymbol{x}}_{\text{density advection and compressive heat}}$$
$$-\underbrace{\frac{1}{\gamma-1}\int\left[n\boldsymbol{u}\cdot(\nabla T_i+\nabla T_e)-\frac{1}{e}\boldsymbol{J}\cdot\nabla T_e\right]d^3\boldsymbol{x}}_{\text{temperature advection}}.$$

Changes to the above relation from upwinding [54], applied to the density and temperature evolution equations, are included and contribute negligible heating power. Unphysical heat flow is introduced by the artificial diffusivity term, introduced for numerical stability, in Eq. 1.21. This artificial heat flow accounts for approximately 5% of the total power flowing at $f_{\text{inj}} = 14.5$ kHz and approximately 3% at $f_{\text{inj}} = 68.5$ kHz, a small but non-trivial contribution. Now that the power flows have been decomposed, the variations of these quantities can be used to explain dynamical changes in these new two-temperature simulations.

2.3 Injector frequency Scan

Driven nonlinear systems tend to exhibit resonances and reproduce harmonics or quasi-harmonics of the driving frequency [215]; experimental campaigns and simulations of HIT-SI with different injector frequencies provide understanding about these qualitative behavior changes. Experimental trends in injector frequency indicate that higher frequency operations tend to exhibit higher plasma impedance in the injectors, increased volume-averaged β,

reduced chord-averaged density fluctuations from interferometry, and, to a lesser extent, larger current gain $G = I_{\rm tor}/I_{\rm inj}$. Previous work [303, 304, 32] with the single-temperature models in NIMROD and PSI-Tet has indicated qualitative agreement with all of these trends when the injector frequency is increased from 14.5 kHz to 68.5 kHz; the two-temperature model facilitates further exploration and understanding of these qualitative trends. For the remainder of Chapter 2, all volume-averaged quantities will be denoted like $\langle\beta\rangle$, as will the chord-averaged density $\langle n\rangle$ which is used in place of the volume-averaged density. The volume averages of the various plasma quantities and power flows use the entire plasma volume, including the injectors. The toroidal current and current centroid are obtained from averages over a number of surface magnetic probes and will be denoted $I_{\rm tor}$ and $(\overline{R}, \overline{Z})$, respectively.

Most of this analysis will be concerned with the plasma dynamics observed in these simulations, but qualitative comparisons with experimental trends will also be made. It is important to note that the simulations presented in this Sec. 2.3 use typical parameters for a HIT-SI discharge at 68.5 kHz, i.e. at relatively low density and power. Furthermore, because the injectors are operated with a 90° phase shift with respect to one another, the total power injected in experimental HIT-SI approximately satisfies

$$P_{\rm inj} = V_{1,\rm inj} I_{1,\rm inj} + V_{2,\rm inj} I_{2,\rm inj} = V_{\rm inj} I_{\rm inj} [\cos^2(\phi) + \sin^2(\phi)] = V_{\rm inj} I_{\rm inj} = Z_{\rm inj} I_{\rm inj}^2 \propto f_{\rm inj} I_{\rm inj}^2, \quad (2.5)$$

where $Z_{\rm inj}$ is the plasma impedance in the injector and $Z_{\rm inj} \propto f_{\rm inj}$ is justified later in Eq. 2.9. These simulations keep the injector current and flux waveform amplitudes fixed, so that if the same scaling exists in the simulations, the power injected increases linearly with $f_{\rm inj}$. In fact, it is shown in Section 2.3.3 that the two-temperature PSI-Tet simulations produce injector power scaling with $f_{\rm inj}$ that is slightly super-linear. While the waveform amplitudes are constant, the relative phases between the waveforms are changed to match the average phases used for high performance HIT-SI experimental discharges at each frequency, and this may also affect the injector impedance scaling with injector frequency; more details on

the how the phases were chosen can be found in previous work [160].

2.3.1 Ion and electron temperatures

For low density or low temperature plasmas, the slow ion-electron collision rate τ_{ei} can result in separate temperatures for the two species on the timescales of interest. For the HIT-SI experiment, which sustains plasmas for 1-2 milliseconds, the approximate thermalization time $\tau_{ei} \approx 100 - 200\ \mu\text{s} > \tau_{\text{inj}}$ for all the injector frequencies investigated here and $\tau_{ei} \gg \tau_{\text{inj}}$ for high-frequency operation. This suggests that injector heating occurs faster than collisions can equilibrate the species, leading to potentially substantial differences in electron and ion temperatures. Therefore one expects that separate ion and electron temperature evolutions are important for understanding the specific temperature profiles (affecting η and β), as well as the flow of plasma energy and heat through the system.

The two-temperature model shows increasing differences between $\langle T_e \rangle$ and $\langle T_i \rangle$ as the frequency increases in Fig. 2.3. At 14.5 kHz $\langle T_e \rangle \approx \langle T_i \rangle$. However, at 68.5 kHz $\langle T_i \rangle$ and $\langle T_e \rangle$ match closely with the experimental measurements of $T_e \approx 5 - 10$ eV and $T_i \approx 20 - 30$ eV [181, 447]. Thomson scattering measurements on the newer HIT-SI3 device at 14.5 kHz also indicate electron temperatures of $5 - 10$ eV [120], in excellent agreement with the $\langle T_e \rangle$ calculated here. Volume-averaged electron temperature is fairly insensitive to injector frequency, while $\langle T_i \rangle$ shows a strong and approximately linear scaling with injector frequency. The mild increase for $\langle T_e \rangle$ can be partially explained through increases in the Ohmic heating and electron heat flux to the wall. Collisional heating with the ions is in the tens of kilowatts, compared to other heat flows in the megawatts, and therefore plays essentially no role in the electron heating. Ohmic heating and the total electron heat flux to the wall both increase by a factor of three or four from low to high frequency, restraining volume-averaged electron temperature to a modest increase at high frequency.

The larger volume-averaged ion temperatures at high frequency operation can be explained from the corresponding increased viscous heating through the following arguments. This anomalous ion heating is the conversion of magnetic energy to ion thermal energy

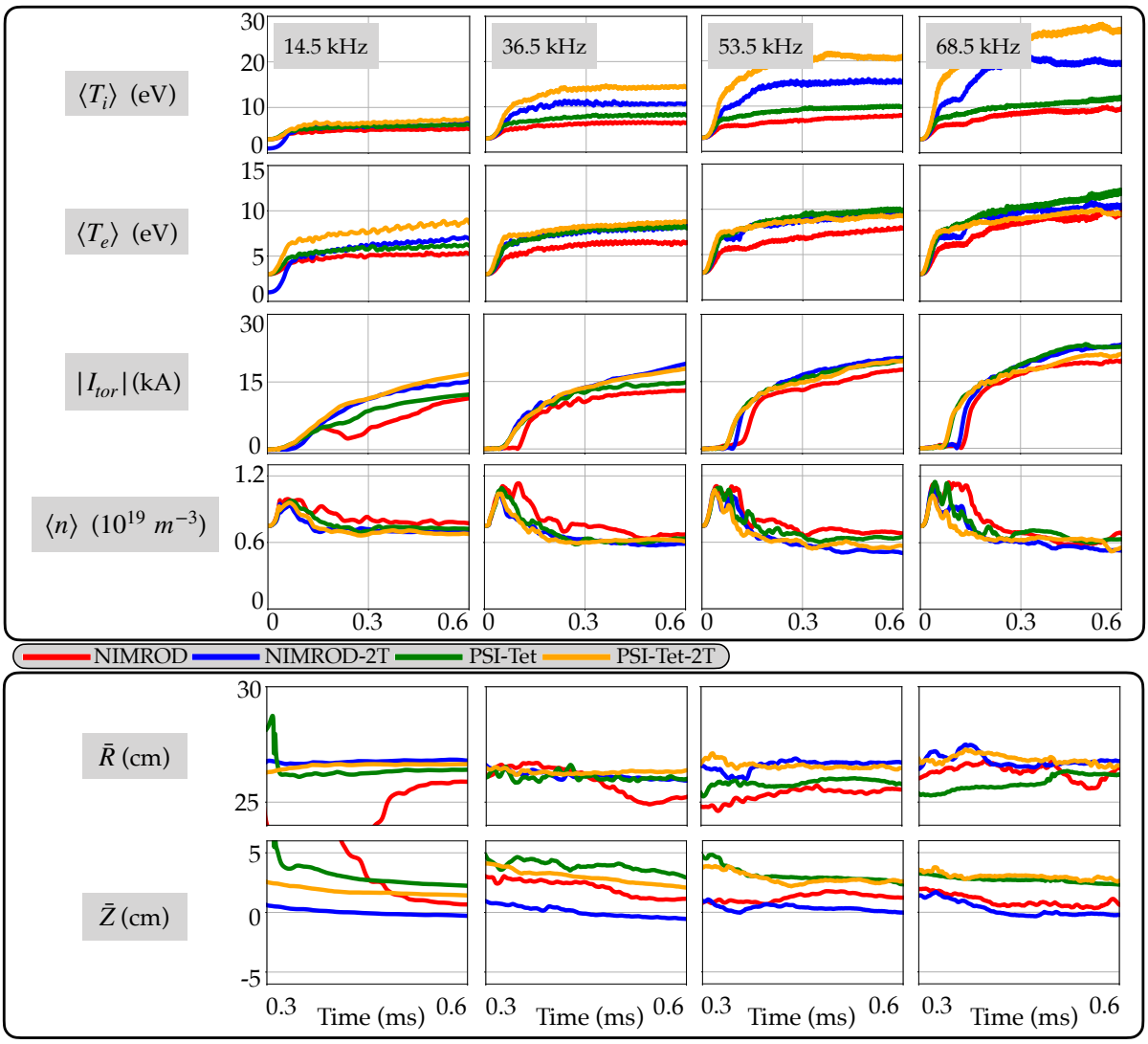


Figure 2.3: Summary of bulk metrics, comparing PSI-Tet and NIMROD single and two-temperature models. Density illustrations represent chord-averaged density $\langle n \rangle$ obtained from synthetic far-infrared interferometry. Magnitude of the toroidal current $|I_{\text{tor}}|$ and current centroid $(\overline{R}, \overline{Z})$ are calculated as averages of the four poloidal surface arrays.

through magnetic reconnection. Magnetic reconnection primarily converts the magnetic energy to ion thermal energy through shock or viscous heating of the reconnection outflow [332]. Work using a stochastic ion heating model for reconnection events indicates the fractional energy transfer is proportional to the species mass, providing a possible explanation for the hot ions and cold electrons [125]. Because most of this heating is lost through stochastic field lines in HIT-SI simulations, the thermal energies for both species builds up until the transport to the wall equals the rate of conversion from magnetic energy to thermal energies. The increased temperatures at higher frequency operation are likely a consequence of increased magnetic reconnection (and therefore increased viscous and Ohmic heating) at higher injector frequencies, since field reversals in the injectors are happening more frequently.

A likely source of additional heating is through the compressive heating $\propto n(T_e+T_i)\nabla\cdot\boldsymbol{u}$. Once the ions are warm, the compressive heat preferentially goes to the ions through the linear dependence on temperature, and Fig. 2.4 indicates that the total compressional heat $\propto n(T_e + T_i)\nabla \cdot \boldsymbol{u}$ is comparable to viscous and Ohmic heating. This is a significant new finding in these simulations, as viscous and Ohmic heating were originally assumed to be the dominant heating effects in the device. Compression also tends to be very high in the injectors during a field reversal. This is both because the plasma in the injectors tends to get relatively hot (see, for instance, Fig. 2.5) and because velocity gradients become very large during field reversals as plasma is accelerated from the walls to carry in oppositely directed flux. This creates an outer shell of plasma with flux in one direction, which compresses an inner channel of plasma flux in the opposite direction until it becomes unstable and global reversal occurs through reconnection [160]. The early and large (larger than Ohmic for the simulations with $f_{\text{inj}} \geq 36.5$ kHz) viscous heating of the ions facilitates the compressive heat to preferentially heat the ions even further. The temperature profiles are illustrated for the two-temperature PSI-Tet simulation at high frequency operation by the 2D contours in Fig. 2.5. The profiles show that the injectors contain hot channels of plasma and the spatial profiles of T_e and T_i are fairly similar.

To understand why the PSI-Tet two-temperature model exhibits larger $\langle T_e \rangle$ than the

single-temperature model at low frequency and smaller $\langle T_e \rangle$ at high frequency, note that, at low frequency, the Ohmic heating is greater than the viscous heating. The single-temperature model shares the Ohmic heating with the ions, reducing the amount of heat to the electrons. At high frequency, the viscous heating is shared between the electrons and ions, so that the single-temperature model now overheats the electrons. Although the NIMROD two-temperature model still produces slightly larger $\langle T_e \rangle$ than the single temperature model at high frequency, the trend with frequency is the same as the two-temperature PSI-Tet simulations (the gap between $\langle T_e \rangle$ in the single and two-temperature NIMROD simulations has shrunk as the frequency increases). It is unclear why the transition has not yet happened for the high-frequency NIMROD simulations. A possible explanation comes from the absence of the injector volumes, where significant ion heating due to compression and reconnection occurs in PSI-Tet.

2.3.2 Current centroid

The current centroid is an important quantity for modeling magnetic equilibria and understanding the magnetic topology in HIT-SI. The toroidal current I_{tor} and current centroid $(\overline{R}, \overline{Z})$ are calculated as averages of the results from each of the four poloidal surface probe arrays in order to be consistent with the experimental analysis. The experimentally measured current centroid is taken as the weighted average of the poloidal field measurements from each of the four poloidal surface magnetic probe arrays [448]

$$\overline{R} = \frac{1}{4} \sum_{k=1}^{4} \frac{\sum_{i=1}^{16} R_{ik} B_{\theta,ik}}{\sum_{i=1}^{16} B_{\theta,ik}}, \tag{2.6}$$

$$\overline{Z} = \frac{1}{4} \sum_{k=1}^{4} \frac{\sum_{i=1}^{16} Z_{ik} B_{\theta,ik}}{\sum_{i=1}^{16} B_{\theta,ik}}, \tag{2.7}$$

where the index k is summing over the four poloidal arrays and the index i is summing over the 16 magnetic probes in each array. Experimental measurements have indicated an

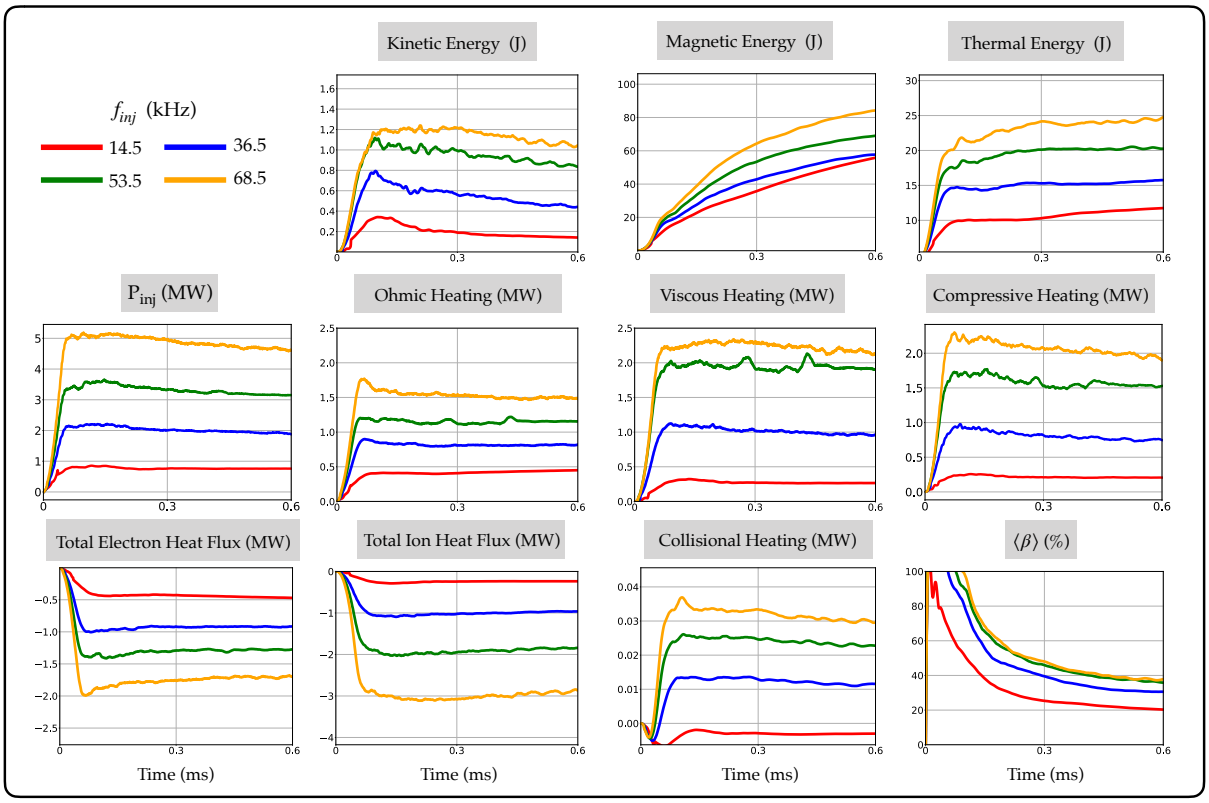


Figure 2.4: Summary of volume-integrated energies and powers for two-temperature PSI-Tet simulations.

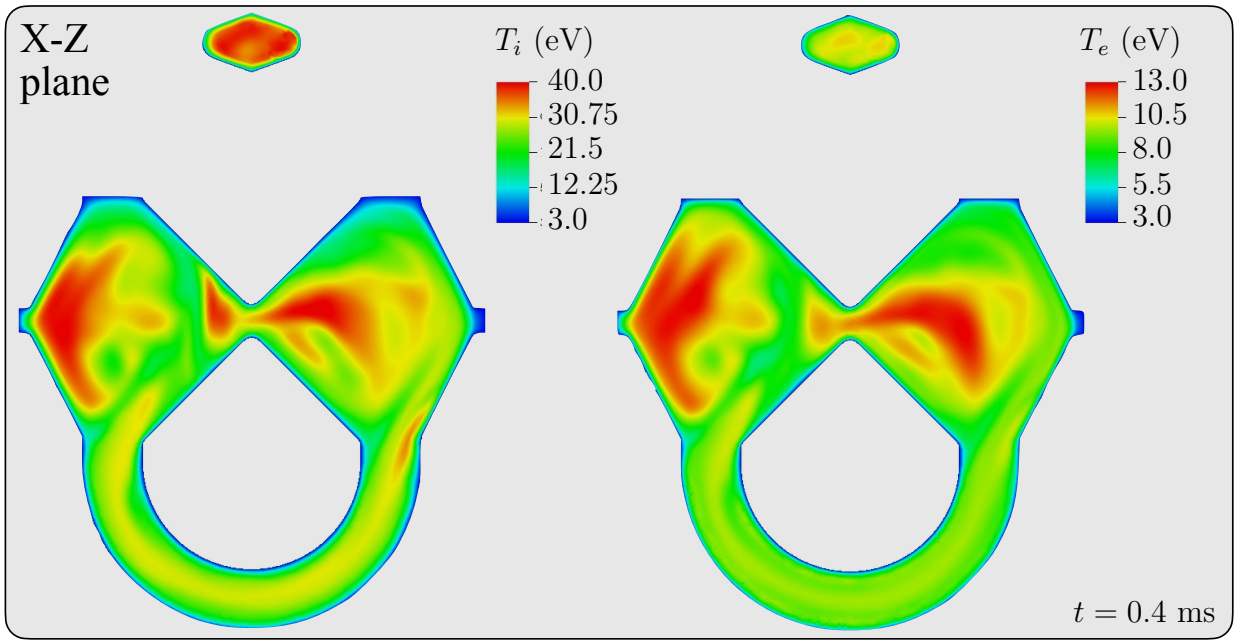


Figure 2.5: 2D contours of T_i and T_e in two-temperature PSI-Tet simulations with $f_{\text{inj}} = 68.5$ kHz indicate hot injectors and similar spatial distributions for T_i and T_e.

outward radial shift and vertical symmetrization of the current centroid at high frequency in HIT-SI [435]. Previous work comparing single-temperature and constant-temperature simulations did not find evidence of robust changes for different injector frequencies [32]. The current centroids computed using the two-temperature models are illustrated in Fig. 2.3. Compared to their single-temperature counterparts, they tend to exhibit a small outward radial shift. Both two-temperature models tend to produce vertical symmetrization of the current centroid compared to single temperature models, although high-frequency PSI-Tet simulations shift very little, or even reverse this trend. As before, there is no consistent evidence of symmetrization or outward radial shift as the injector frequency increases.

An interesting finding is that the PSI-Tet models tend to see the vertical component of the current centroid shifted upwards $2-4$ cm from the midplane, although the error bars are of a similar magnitude. The change in dynamics from the single to two-temperature PSI-Tet model appears to reduce the upward shift of $\overline{Z}$ at low injector frequency, but has negligi-

ble change at high frequency. The vertical shift observations in PSI-Tet appear to be the consequence of an asymmetry between the amount of power input by the two injectors in PSI-Tet [160]; injected power asymmetry is also observed experimentally. The current waveforms for the flux circuit tend to be larger for the injector on the bottom of the machine to produce the same flux in both circuits. This asymmetry depends in principle on the nonlinear plasma dynamics inside of each injector, and it is unknown how the stronger injector is determined. Experimental work suggests that the asymmetry is determined primarily by the toroidal current direction through the injector-spheromak coupling [190] but the two-temperature PSI-Tet simulations presented here with opposite toroidal current direction exhibit the same injector asymmetry. In PSI-Tet simulations the bottom injector will tend to input more power, and therefore perturb the system preferentially in the upwards direction. This is directly reflected in PSI-Tet by the vertical component of the current centroid. NIMROD only produces a slight current centroid asymmetry with the single-temperature model, suggesting that capturing the injector power asymmetry correctly may require modeling the dynamics inside the injectors.

The vertical shift of the current centroid is observed to a small extent (≈ 1 cm) experimentally [435]. The smaller experimental shift is likely because an increase in one of the injector voltage waveforms tends to occur with a corresponding decrease in the same injector current waveform, leading to approximately equal power from each injector despite the asymmetric waveforms. Dealing with this subtlety would require a more realistic model of the experimental circuit. Nonetheless, this effect provides an additional metric for experimental validation and an interesting direction for future experimental work, i.e. intentional asymmetric operation of the injectors to investigate the current centroid dependence.

The experimentally observed shift in the radial component of the current centroid has been postulated elsewhere as a Shafranov shift from confined plasma pressure in the interior of the device [435]. However, the simulations presented in this frequency scan do not exhibit evidence of closed flux surfaces or confined pressure. Thus, not surprisingly, the simulations indicate that the current centroid does not consistently shift outwards with higher frequency,

even with significantly larger $\langle\beta\rangle$. The volume-averaged plasma β is pictured in Fig. 2.4 and defined here through the plasma pressure $p = n(T_i + T_e)$ as

$$\langle\beta\rangle = \frac{\int (p - p_{\text{wall}}) d^3\boldsymbol{x}}{\int \frac{B^2}{2\mu_0} d^3\boldsymbol{x}}. \tag{2.8}$$

The lack of an outward shift of $\overline{R}$ as the frequency changes is consistent with single-temperature PSI-Tet observations in previous work [32]. The observed outward shift of $\overline{R}$ in the two-temperature PSI-Tet simulations is primarily due to changes in the spatial distribution of T_e, leading to changes in the parallel current distribution through η, and thus alter the force-free component of the equilibrium. Evidence for this claim is provided from similar simulations in Sec. 2.4.1.

The two-temperature PSI-Tet observations for $\langle\beta\rangle$ in Fig. 2.4 indicate a large increase from 14.5 kHz to 36.5 kHz (perhaps corresponding to a transition from lower to higher average $\langle\beta\rangle$ as has been demonstrated elsewhere [304]) and diminishing increases from 36.5 kHz to higher frequencies.

2.3.3 Injector impedance

Experimental data and imposed dynamo current drive [193, 192] predict that the injector impedance approximately satisfies

$$Z_{\text{inj}} = C_1 \mu_0 R_0 \left(\frac{1}{8\pi e r_0^3} \frac{\lambda_{\text{inj}}}{\lambda_{\text{sph}}} \frac{I_{\text{tor}}}{n} + 2\pi C_2 f_{\text{inj}} \right). \tag{2.9}$$

For NIMROD and PSI-Tet simulations the second term on the right-hand side is dominant [304], and so this can be further approximated to

$$Z_{\text{inj}} \approx 2\pi C_3 R_0 \mu_0 f_{\text{inj}}. \tag{2.10}$$

$R_0 = 0.5$ m and $r_0 = 0.25$ m are the major and minor radius of the HIT-SI device respectively and C_1, C_2, and $C_3 = C_1 C_2$ are fitting parameters. The Z_{inj} scaling with the frequency (Eq. 2.10) for two-temperature PSI-Tet simulations is indicated in Fig. 2.6 and indicates an average $C_3 \approx 0.23$, in rough agreement with the two-temperature NIMROD simulations with $C_3 \approx 0.125$. A lower impedance in NIMROD is reasonable, as the injector regions are not included. This means that the power used to generate magnetic fields in the injector volumes are not captured in the NIMROD simulations.

Fig. 2.6 indicates that this relationship is slightly super-linear in two-temperature PSI-Tet simulations but only linear in two-temperature NIMROD simulations (note that the 14 kHz NIMROD simulation is somewhat of an outlier, perhaps because this is the simulation with 1 eV walls, so it has been omitted in the average C_3). The physical cause of the additional scaling in PSI-Tet does not appear to correlate with average $|\boldsymbol{J}|/n$ as suggested by the theory of imposed dynamo current drive. Due to the continuous evolution of quantities and their profiles in these simulations as the injector frequency changes, an alternative physical correlation could not be found.

A natural extension of this analysis is an examination of the scaling of the current gain $G = I_{\text{tor}}/I_{\text{inj}}$. Helicity balance models [336, 447] indicate that for fixed injector waveforms and steady-state operation

$$G \propto \sqrt{\tau_{\text{L/R}} Z_{\text{inj}}} \propto \sqrt{T_e^{\frac{3}{2}} f_{\text{inj}}}, \tag{2.11}$$

where the second scaling follows from Eq. (2.10). However, these simulations are not at steady-state operation at $t = 0.6$ ms, as can be seen straightforwardly in the continued growth of the toroidal currents in Fig. 2.3. Further work should examine the scaling of gain and injector impedance at steady-state with the two-temperature models.

2.3.4 Spheromak formation

The nonlinear relaxation event, the interval during which the plasma self-organizes into a spheromak plasma, has been analyzed extensively. However, the exact timing of this event,

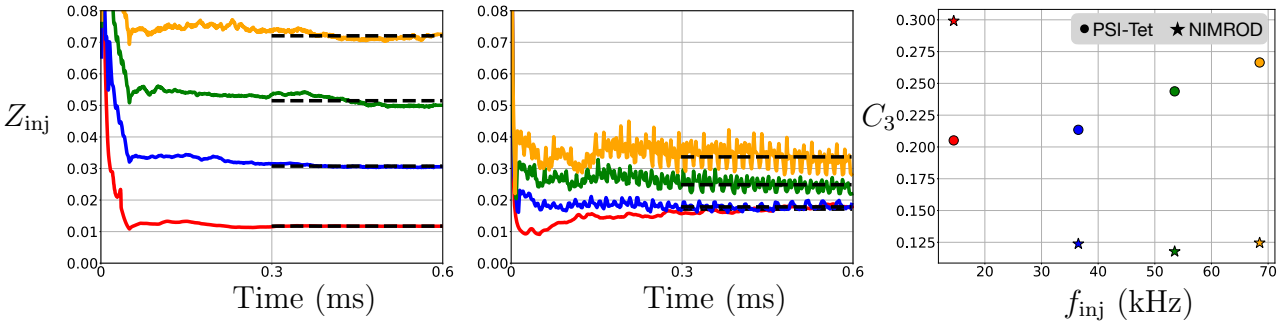


Figure 2.6: Left: Eq. 2.10 (dashed black lines) captures the first order time evolution of $Z_{\mathrm{inj}}(t)$ (color lines) for two-temperature PSI-Tet simulations. Middle: Equivalent to the left illustration for two-temperature NIMROD simulations. Bottom: Best fit values for C_3, corresponding to the black lines from above, indicate averaged $C_3 \approx 0.23$ and $C_3 \approx 0.125$ for two-temperature PSI-Tet and NIMROD simulations, respectively.

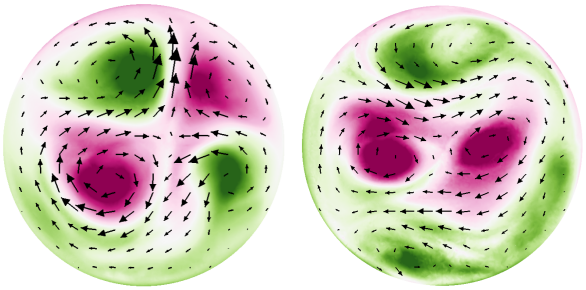
and the process by which the resulting toroidal current direction is determined, are not well-understood experimentally or theoretically.

Interestingly, at high injector frequency, NIMROD [304] and PSI-Tet single and two-temperature simulations all indicate the presence of a structure of two oppositely oriented flux tubes during spheromak formation. This structure also briefly appears during the formation event for low frequency HIT-SI simulations using the two-temperature PSI-Tet model. The structure is visualized using B_z at the $Z = 0$ midplane for both low and high frequency two-temperature PSI-Tet simulations in Fig. 2.7. A 3D isosurface of $\boldsymbol{J} \cdot \boldsymbol{B}/|\boldsymbol{B}|^2 = 13$ (an important quantity for spheromaks and Taylor states [31]) in Fig. 2.8 indicates that there are paths of constant $\boldsymbol{J} \cdot \boldsymbol{B}/|\boldsymbol{B}|^2$ which connect the two injectors, and suggest current loop formation [327] in the center of the device. These paths are reminiscent of the flux tubes, suggesting that the tubes may be generated by the magnetic field parallel to the current in these 3D pathways. Once the formation phase of the discharge gives way to the sustainment phase, these flux tubes are more difficult to identify, as the leading order magnetic field dependence now comes from the self-organized spheromak in the center of the device. However, work with DMD in Chapter 4 similarly identifies previously undiscovered

large-scale magnetic structures with $n_\phi = 2$ toroidal Fourier structure, oscillating at $2f_{\text{inj}}$ during sustainment [207].

This observation is notable because the HIT-SI experiment tends to produce both negative and positive toroidal discharges and this parity choice appears to depend on a number of factors, including the phases of the injector waveforms during the time of formation [304]. Previous studies have found that fixed parameter discharges operating a single injector always form and sustain spheromaks with toroidal current parity determined by the sign of the injected helicity [182, 118] but more detailed explanations for this behavior have been elusive. Moreover, HIT-SI simulations in PSI-Tet with these fixed parameters always produce negative toroidal current; equivalent simulations for the HIT-SI3 device, which has a different injector geometry, produce positive toroidal currents. This parity can often be switched in the simulations by changing the relative phases of the injector waveforms. Different parameter regimes indicate different parity preferences.

In the simulations presented here, at the time of spheromak formation, the two closest flux tubes merge in the center, while the other two spread out and merge along the edge. Which flux tubes merge in the center determines the direction of the toroidal current. All of these observations are consistent with the interpretation of each injector driving a flux tube pair, with parity fixed by the sign of helicity. However, this appears to be determined more by the relative phases of the injectors, rather than the number of injectors, since HIT-SI3 simulations also often exhibit this $n_\phi = 2$ activity [349]. During operations with both injectors, a quasi-random process then selects which flux tube pair merges and determines the sign of toroidal current. This quasi-random process likely depends on a number of nonlinearities, as different parameter spaces indicate different timing and sign of the toroidal current. Future work could investigate how these formation structures change with different injector phasing.



(a) $f_{\text{inj}} = 14.5$ kHz (b) $f_{\text{inj}} = 68.5$ kHz

Figure 2.7: Two-temperature PSI-Tet simulations: Contours of B_z with limits ± 100 Gauss and vector plots of $\boldsymbol{B}$ illustrated at time snapshots directly before spheromak formation, in the $Z = 0$ midplane. The flux tubes form a $n_\phi = 2$ toroidal Fourier structure, and oscillate at approximately $2f_{\text{inj}}$. During spheromak formation, the flux tubes merge to determine the direction of the toroidal current.

2.4 PSI-Tet Parameter scans

Investigating the parameter space of the magnetohydrodynamic models presented here is very important; fixed numerical terms require convergence studies to understand their impact on simulations and physical parameters should be scanned in order to understand the possible range of results from the experimental uncertainty in the measured values.

With the two-temperature PSI-Tet model, a number of parameter scans were performed to investigate the plasma dependence on these values. The wall temperature, wall density, and artificial diffusivity scans are performed only for $f_{\text{inj}} = 14.5$ kHz. With the exception of the wall density scan, which is performed at high injector power to investigate the possibility of closed flux formation, the simulations in this section use the parameters in Table 2.1 (although the parameter being scanned is changing from the table value). Previous NIMROD work with HIT-SI simulations has scanned the viscosity and indicated only small changes from the value of the viscosity and choice of isotropic or anisotropic viscosity [303]. However, scans performed in NIMROD may not see dependencies on the injector geometry and associated field reversals. Future work in PSI-Tet should explore this possibility.

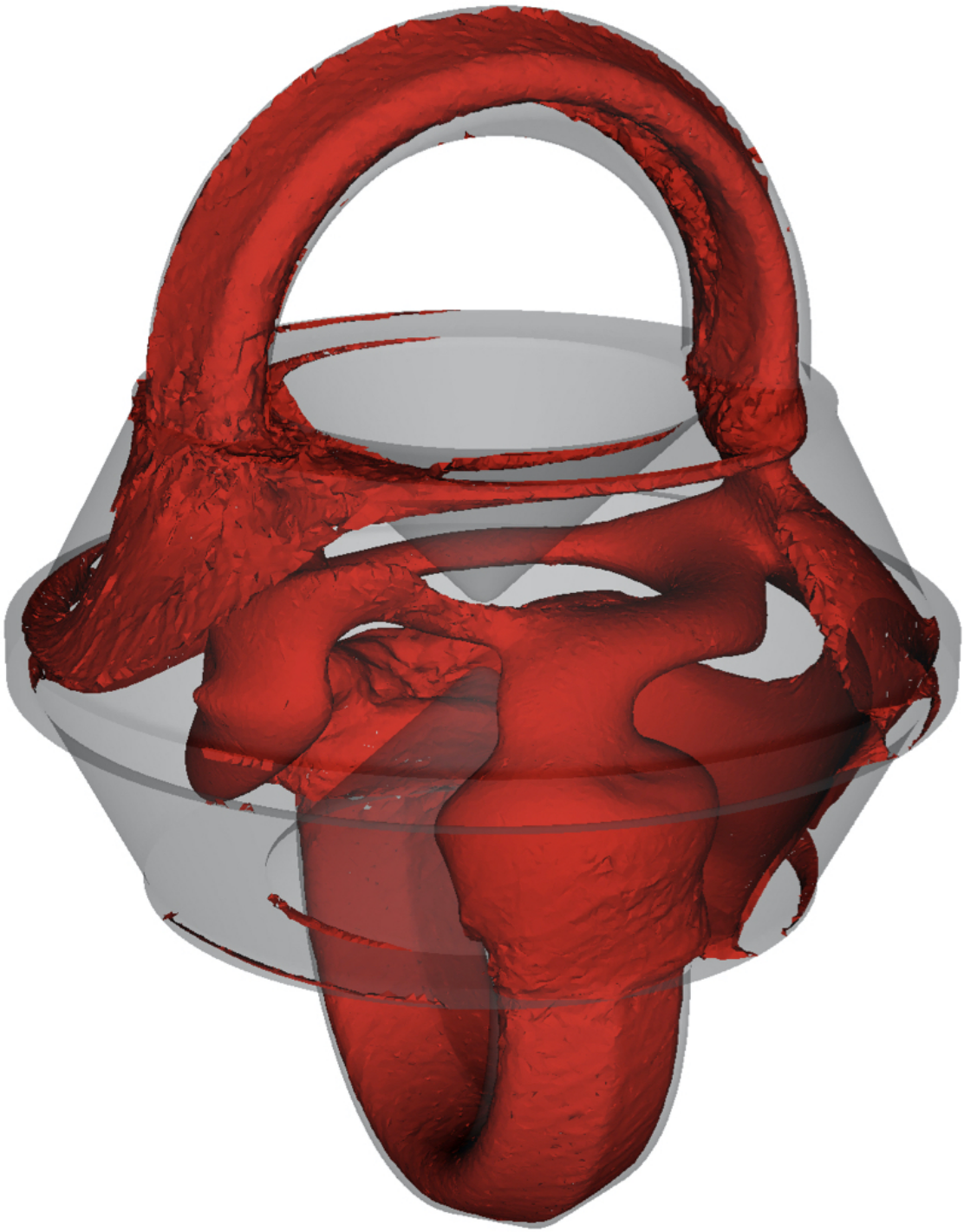


Figure 2.8: A 3D isosurface of $\boldsymbol{J} \cdot \boldsymbol{B}/|\boldsymbol{B}|^2 = 13$ at $f_{\text{inj}} = 14.5$ kHz, before spheromak formation ($t = 25$ μs), indicates paths connecting opposite injector mouths and suggests the formation of a current loop.

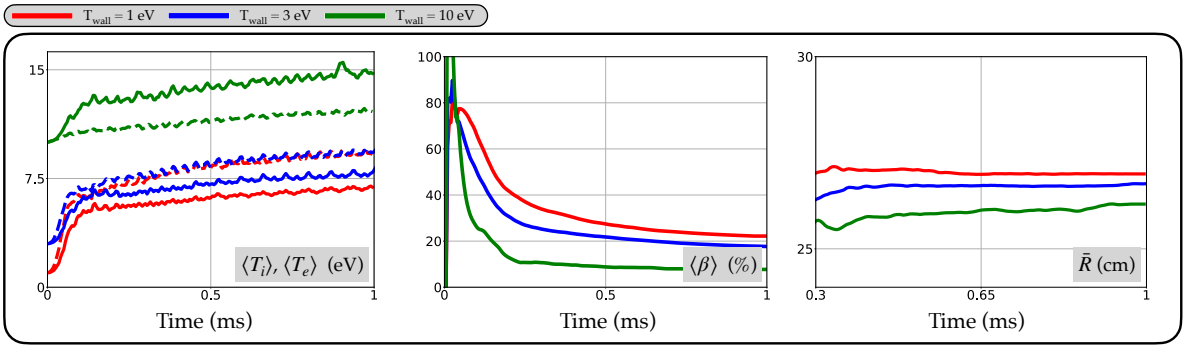


Figure 2.9: Time evolution of important quantities for the wall temperature scan of two-temperature HIT-SI simulations with $f_{\text{inj}} = 14.5$ kHz. $\langle T_i \rangle$ and $\langle T_e \rangle$ are in solid and dashed lines, respectively.

2.4.1 Wall temperature

The exact temperature boundary condition is unknown on the HIT-SI experiment. More sophisticated first-principles modeling of the temperature boundary condition would necessitate the evolution of a neutral fluid, as plasma-wall interactions involve recombination and other atomic processes which may strongly alter the density and temperature near the boundaries [411]. However, scanning the wall temperature provides understanding about how the internal plasma dynamics are affected and provides a sense of which value best validates with experimental observations.

Three different wall temperatures $T_{\text{i,wall}} = T_{\text{e,wall}} = 1$, 3, and 10 eV, were investigated. The only changes of any significance are shown in Fig. 2.9 and indicate that as the wall temperature is increased, a large inward shift of the current centroid occurs. The changes from 1 to 3 eV walls are particularly small, justifying the comparison between 1 eV and 3 eV wall temperature simulations in Sec. 2.3. In Fig. 2.10, time-averaged spatial profiles of the electron temperature and J_y are compared between the 1 eV and 10 eV cases. It appears that changes to the spatial distributions in T_e primarily lead to changes in the parallel current distribution through η. This likely produces a modified force-free part of the equilibrium. In contrast, $J_\perp$ (back to $\perp$ meaning perpendicular to the local magnetic field direction) is quite

similar as the wall temperature changes, as one might expect of a plasma simulation without significant indication of confined pressure. Therefore, the current distribution changes are posited to be responsible for the inward shift of the current centroid. The shift is inward because η decreases rapidly as $T_{e,wall}$ increases, facilitating stronger parallel current flows.

2.4.2 Wall density

Experimentally, HIT-SI discharges tend to reach higher current gains during low-density discharges [304, 181]. To investigate the density dependence in simulations, wall densities $n = 10^{19}$, 2.58×10^{19}, and 5.16×10^{19} m^{-3} are scanned at $f_{\text{inj}} = 14.5$ kHz. To better approximate experimental conditions, injector current and flux waveform amplitudes are increased by a factor of 2.6 as compared to the other 14.5 kHz simulations shown in previous sections, leading to higher input powers of $4 - 8$ MW. This is comparable with the low end of experimental discharges at low frequency, which typically input $5 - 15$ MW of injector power.

Fig. 2.11 indicates that the toroidal current is strongly dependent on the density. At low density, $\langle T_i \rangle$ exceeds 50 eV, and toroidal current is above 100 kA and growing. Despite large gains in the temperature, $\langle \beta \rangle$ decreases, indicating the magnetic field strength has increased significantly faster than the pressure. The lowest density simulation $\langle \beta \rangle$ falls to near the Mercier limit for the HIT-SI flux conserver [288, 282]. The large magnetic field strength, the spheromak gain of $G > 5$, and increased fluctuations on the average temperature evolution in Fig. 2.11 all indicate possible closed flux activity. Previous work found transient closed flux events from viscous effects but this occurred only for magnetic Prandtl number $Pr_m = \mu_0 \nu / \eta \gg 100$. In contrast, this work has Pr_m an order of magnitude smaller and exhibits significantly larger and more sustained closed flux periods. A Poincaré plot in Fig. 2.12 for $\boldsymbol{B}$ illustrates the presence of closed flux, and closed flux regions lasting $50 - 100$ μs, or $1 - 2$ injector periods can be seen in a video available at `doi.org/10.1063/5.0006311.1`. This video illuminates considerable variation in the closed flux surfaces; the flux surfaces vary from symmetric states to highly asymmetric states exhibiting complex magnetic island

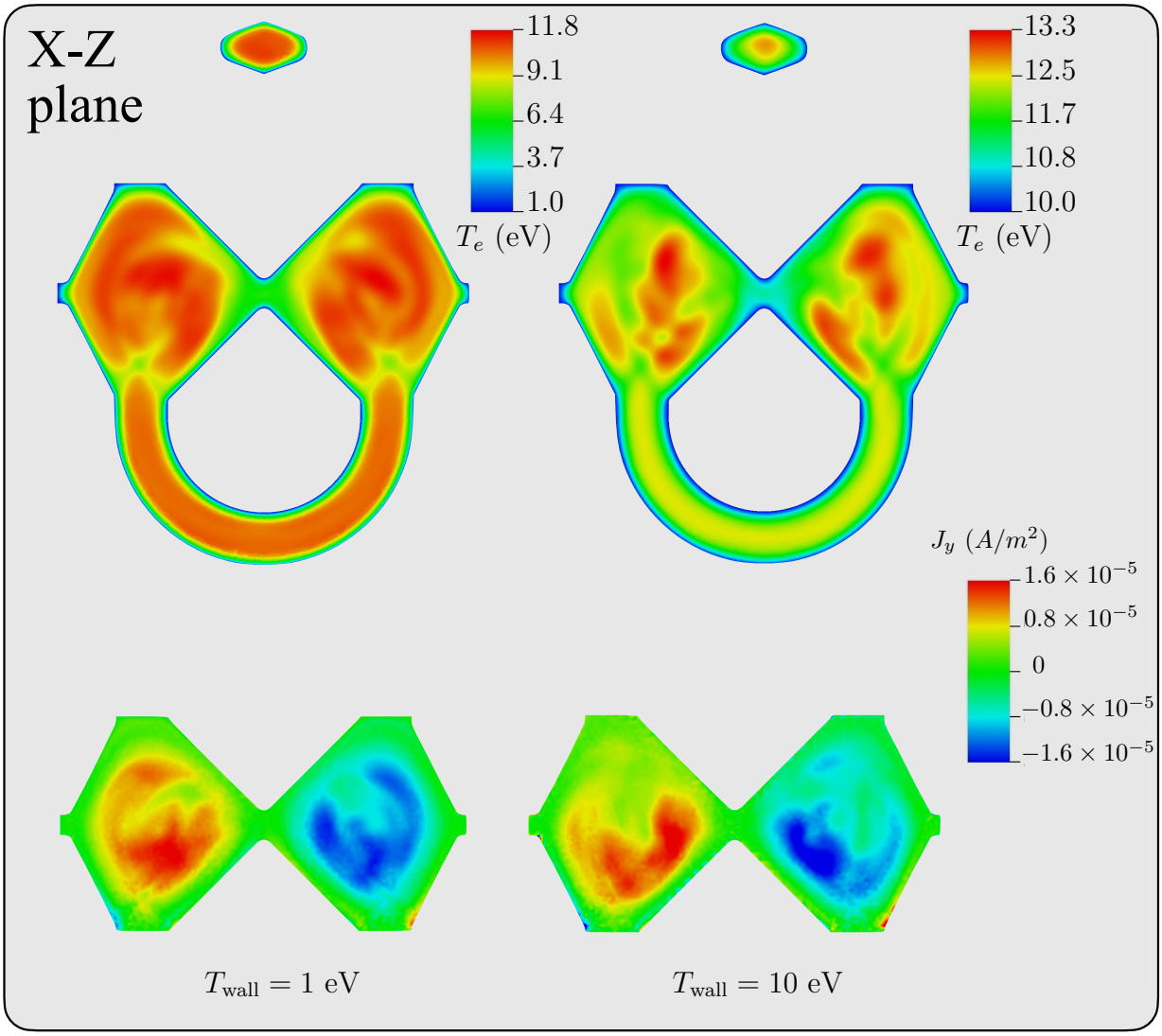


Figure 2.10: Time-averaged T_e and J_y on a toroidal slice for $T_{\text{wall}} = 1$ eV, 10 eV in two-temperature PSI-Tet simulations at $f_{\text{inj}} = 14.5$ kHz.

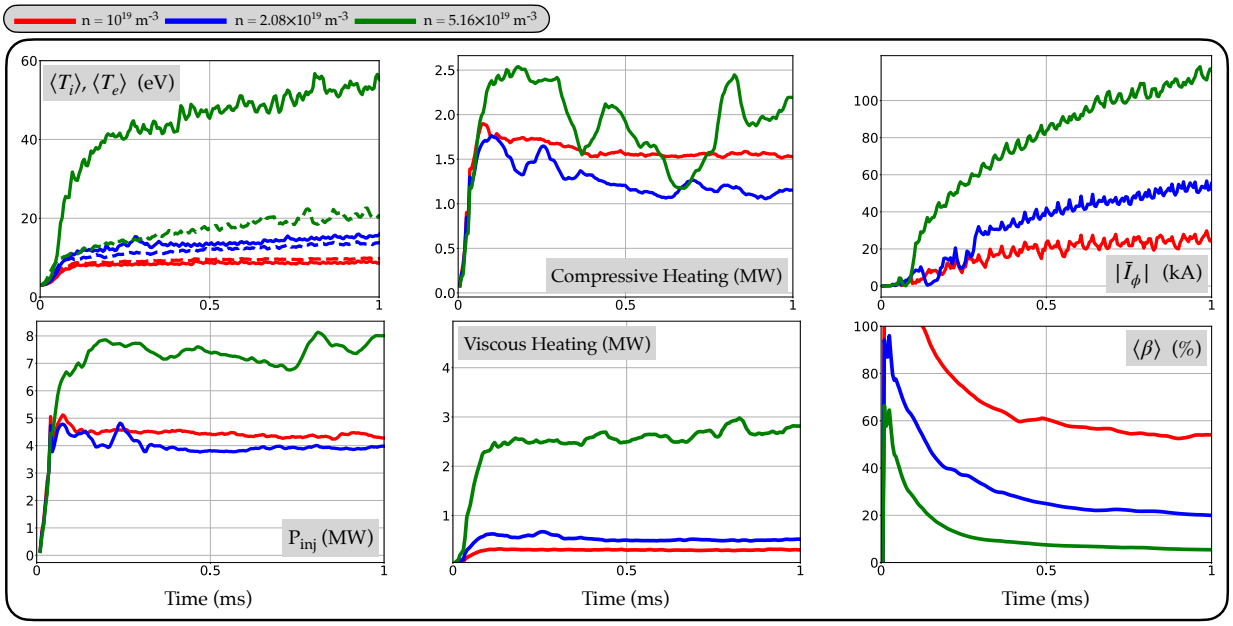


Figure 2.11: Time evolution of important quantities for the wall density scan of two-temperature PSI-Tet simulations with $f_{\text{inj}} = 14.5$ kHz. $\langle T_i \rangle$ and $\langle T_e \rangle$ are in solid and dashed lines, respectively.

structures.

The large increase in viscous heating (and therefore ion temperature) at low density can be mostly accounted for by the corresponding large increase in injector power. However, the simulation with $n = 2.58 \times 10^{19}$ m^{-3} actually produces a slight decrease in the injector power compared to when $n = 5.16 \times 10^{19}$ m^{-3}. A similar trend is seen in the compressive heating, although it exhibits a complicated temporal dependence not observed in any other of the simulations in Chapter 2. All of these observations suggest a threshold at a low enough density and high enough input power where the overall performance of the device increases sharply. An experimental study at 36.5 kHz operation did not find evidence of substantial toroidal current increases with low density deuterium [182], although the total input powers were less than 4 MW for all discharges examined. The record for experimental HIT-SI gain was $G = 3.9$, but this high frequency discharge also had only a few MW of input power [435],

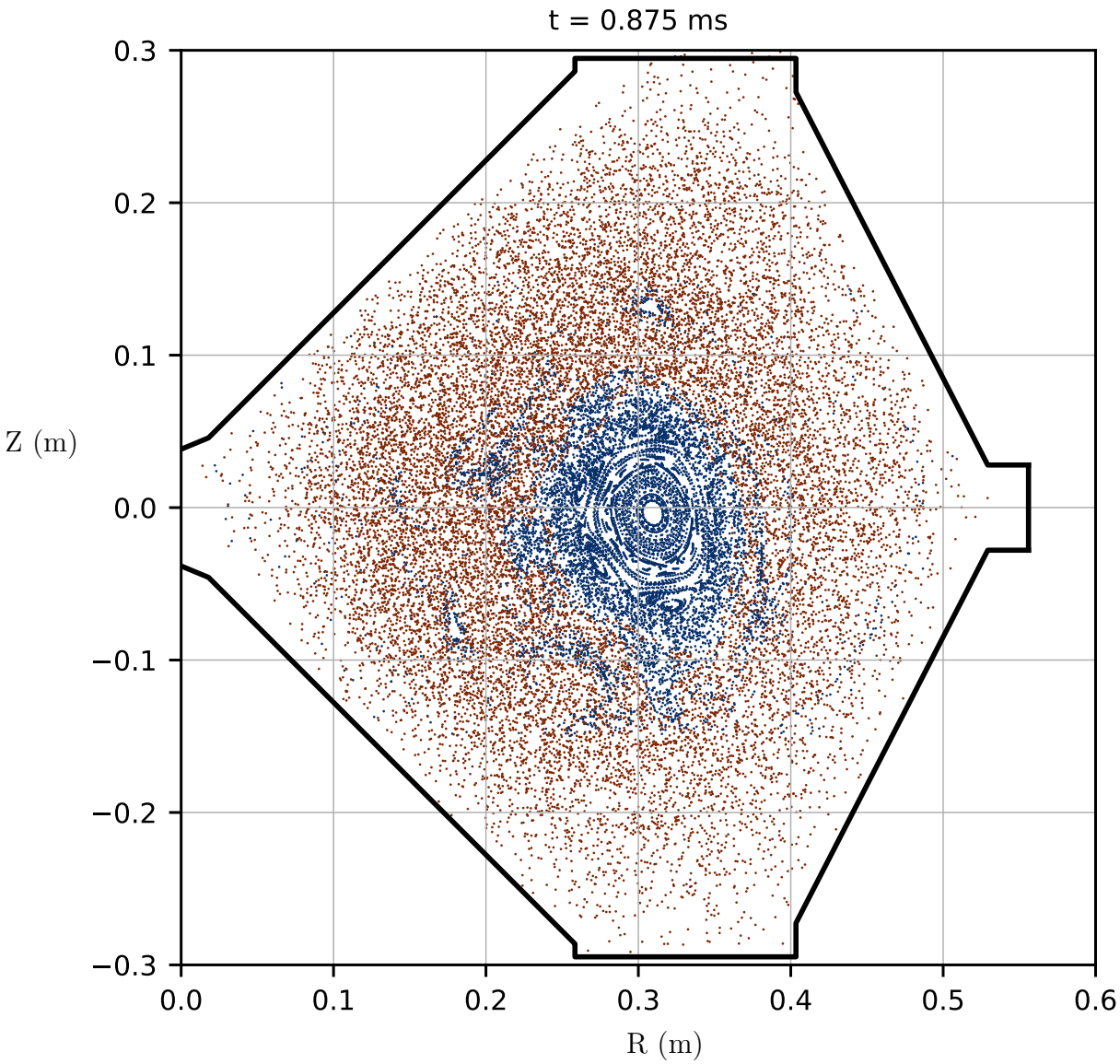


Figure 2.12: Poincaré plot of $\boldsymbol{B}$ illustrates closed flux surfaces with lifetimes $50 - 100$ μs in a low density, high power, two-temperature PSI-Tet simulation at $f_{\text{inj}} = 14.5$ kHz. Points in blue approximately indicate the closed field lines. Video available at `doi.org/10.1063/5.0006311.1`.

so the $G > 5$ regime has not been explored experimentally. Therefore, these observations at low density and high input power may indicate a route toward optimizing for higher performance experimental discharges. One complication in this route is that there are often large radiative losses in HIT-SI discharges, which are not modeled in the simulation. These losses may prevent the plasma from reaching the requisite viscous, Ohmic, and compressive heating necessary for this higher performance regime. These findings at low density, low frequency, and high power operation merit future work on the density boundary conditions and profiles used for HIT-SI simulations.

2.4.3 Particle diffusivity

From the previous analysis, the sensitivity of the spheromak performance to the density suggests that artificial particle diffusivity may change the dynamics considerably. Unfortunately, the artificial particle diffusivity in the continuity equation cannot be completed removed because it is found empirically to be necessary for the numerical stability of the Hall-MHD model; it is tolerable if it can be converged down to a small value where it has negligible effects on the simulation. However, previous work [304] only managed to reduce this value to approximately $D = 250$ m^2/s. Here successful runs are reported down to $D = 50$ m^2/s in PSI-Tet, which have been reproduced in NIMROD using additional hyper-diffusivity. These simulations are also compared with a run with $D = 1000$ m^2/s to understand the physical effects of this stabilizing term to regimes used in prior work [303]. This work was unable to reduce this value until the plasma dynamics are completely insensitive to it, but this comparison provides understanding about how diffusion affects the dynamics in these simulations. The quantities which indicated significant changes are summarized in Fig. 2.13.

In PSI-Tet and NIMROD this term is needed to avoid overshoot with grid-scale sharp features, generally near the wall or reconnecting regions. The two-temperature PSI-Tet simulations presented here have $\langle n \rangle$ which only mildly varies from $D = 1000$ $m^2/s \rightarrow 250$ m^2/s but is significantly lower at $D = 50$ m^2/s. This reduction in chord-averaged density produces an overall increase in performance (higher toroidal current and average

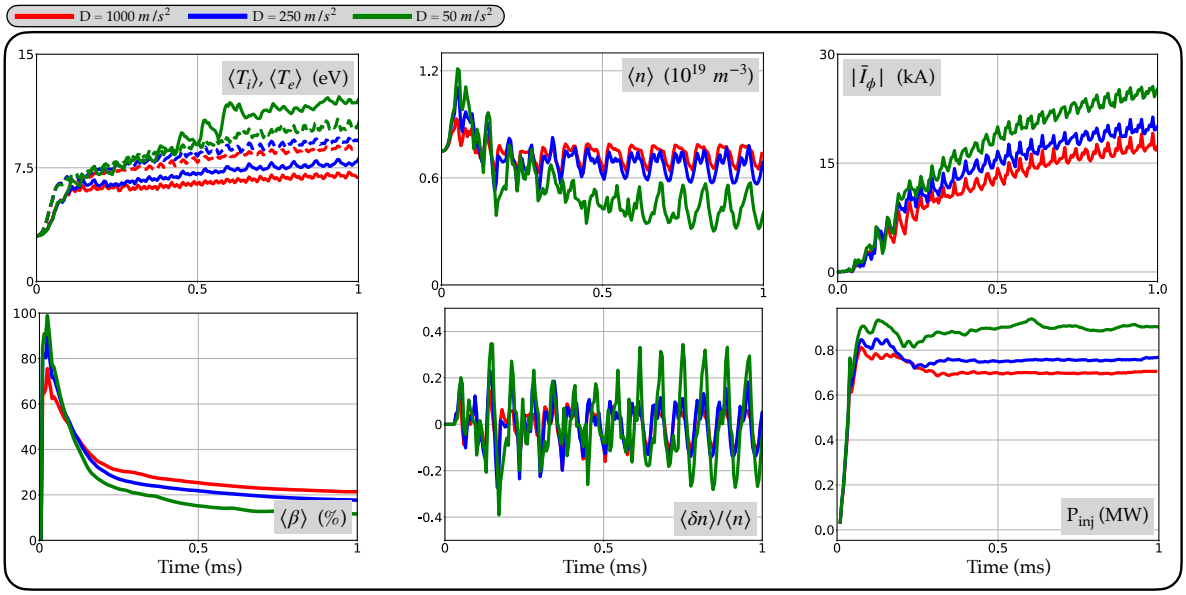


Figure 2.13: Time evolution of important quantities for the artificial diffusivity scan of two-temperature PSI-Tet simulations with $f_{\text{inj}} = 14.5$ kHz. $\langle T_i \rangle$ and $\langle T_e \rangle$ are in solid and dashed lines, respectively.

temperatures) consistent with the wall density scan. The volume-averaged temperatures increase significantly, and $\langle T_i \rangle > \langle T_e \rangle$ for $D = 50$ m^2/s. Relatively unchanged thermal pressure, with significantly larger magnetic pressure at low diffusivity, leads to reduced $\langle \beta \rangle \approx 10\%$.

Experimental interferometry exhibits large and rapid oscillations which are considerably larger than those observed in previous work or in this analysis. Artificial diffusivity in the simulations has been postulated as a possible explanation for this discrepancy. The density fluctuations $\langle \delta n \rangle$ shown on the raw density signal in Fig. 2.13 grow approximately proportionally as the diffusivity is decreased, and the relative density fluctuations $\langle \delta n \rangle / \langle n \rangle$ also increase significantly. Additionally, Fig. 2.14 indicates that in the low diffusivity case, there can be spatial variations in the density of almost two orders of magnitude. These large gradients in the density can lead to numerical instability. Since both absolute and relative density fluctuations increase as D decreases, it appears likely that further convergence of

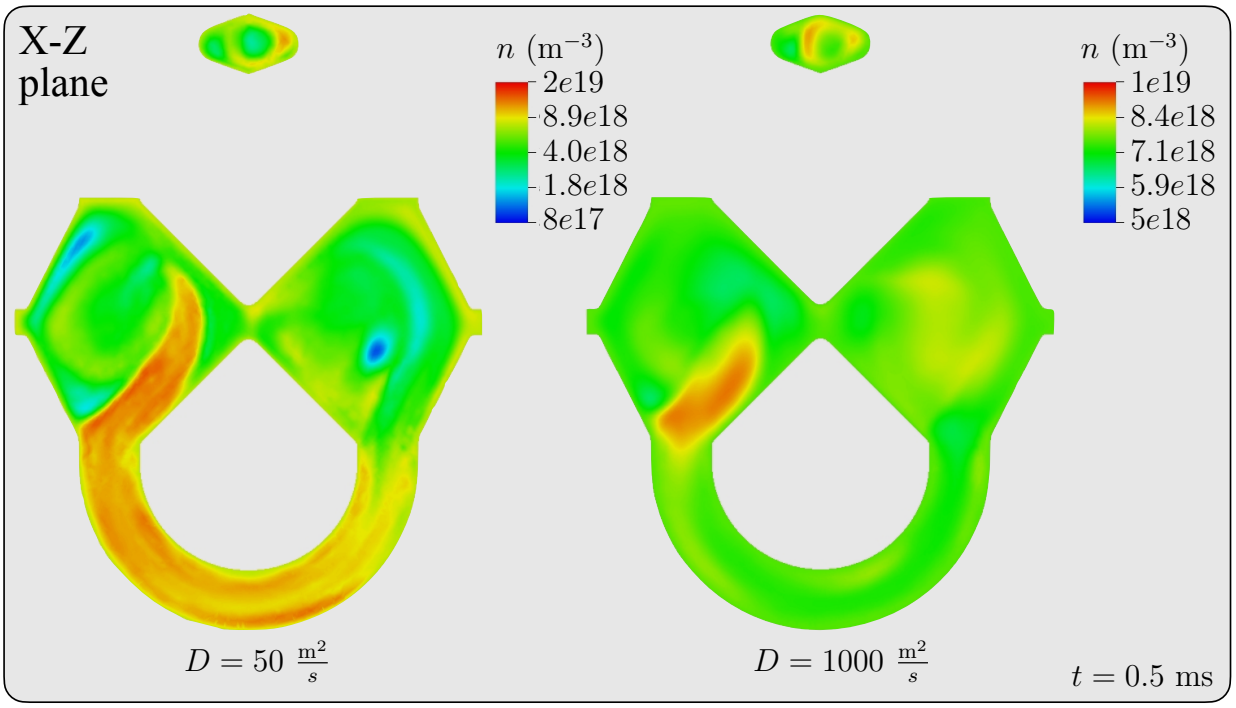


Figure 2.14: Comparison of the 2D spatial density profiles for $D = 50$ m$^2/s$ and $D = 1000$ m$^2/s$ in two-temperature PSI-Tet simulations at $f_{\text{inj}} = 14.5$ kHz. The contours indicate that pockets of low density form as D decreases. Contours also show significantly higher density due to compression in the injector regions.

$D \to 0$ will validate better with the large experimental oscillations. Another possibility is that matching the experimental density fluctuation size will require a density profile in the injectors that is a more faithful representation of the experimental fueling. Further progress will require careful reproduction of the experimental waveforms and a direct model of the experimental circuit.

2.5 Conclusions regarding two-temperature HIT-SI simulations

A two-temperature model was implemented in PSI-Tet and provides new insights into the plasma dynamics in HIT-SI simulation and experiment. Both the PSI-Tet and NIMROD two-temperature models differ from the single-temperature models through higher volume-averaged ion temperature, lower chord-averaged density, and, for the most part, axial sym-

metrization of the current centroid. Ion and electron temperatures are now in qualitative agreement with ion Doppler spectroscopy and initial Thomson scattering measurements. It was found that large compressive heating from the injector dynamics plays a significant role in the overall plasma evolution. This new model further indicates injector impedance and volume-averaged temperatures scale approximately linearly with injector frequency.

Parameter scans in the Dirichlet boundary condition for temperature and density, along with simulations exploring the artificial diffusivity, lead to new physical insights into the plasma dynamics in the HIT-SI device. At higher wall temperature, significant changes to the spatial distributions of T_e and $\boldsymbol{J}$ lead to an inward shift of the current centroid. Importantly, the plasma dynamics exhibit considerable dependence on the plasma density. Performance improves with higher volume-averaged temperatures, larger toroidal current, reduced oscillations of the current centroid position, and reduced heat flux to the wall. The low density, low frequency, high power simulation indicates a sharp rise in injector power, suggesting a sudden change in the dynamics towards a higher performance regime. These simulations motivate further experimental and numerical investigation of low density parameter regimes.

Reductions in the artificial diffusivity produce significant decreases in the chord-averaged density in the device, leading to higher temperatures and toroidal currents. The absolute and relative density fluctuations increase as D decreases, improving the qualitative agreement with experiment. Further reduction in the artificial diffusivity in PSI-Tet may potentially lead to numerical instability near these sharp density gradients, although these reductions have already been achieved by using a hyper-diffusivity term $D_h \nabla^4 n$. Hyper-diffusivity can provide a larger ratio between smoothing at the grid scale and the global scale and has been successfully applied to NIMROD simulations. The observations here strongly support the claim that variations in density produce a strong impact on the dynamics.

The dynamics of the HIT-SI device have been investigated with a two-temperature Hall-MHD model. A number of simulation parameters were varied, and qualitative experimental validation improved. Future work should perform detailed validation with the experiment

with the new two-temperature model, as has been done with previous models [161]. Towards this goal, a circuit model [179] of the injectors has recently been implemented into PSI-Tet so that injector drive can be captured more completely by the simulation [162]. This is expected to be important for experimental validation because the reconnection heating, propagated through the viscous heating, depends strongly on the phase between the injector waveforms when the injector reverses direction. The primary reversal phase was found in previous work [160] to last only a few μs, consistent with the Sweet-Parker [413, 343] reconnection timescale. Further physical understanding and improved validation with the experiment could also be obtained from an implementation of a more realistic closure for heat transport and anisotropic viscosity.

Finally, there is a concern that the results here are specific for the HIT-SI device, and do not generalize straightforwardly to the newer HIT-SI3 or HIT-SIU devices (these newer devices differ primarily in the number and operation of the injectors). Fortunately, recent simulations [349] of HIT-SI3 and HIT-SIU [33] with single and two-temperature Hall-MHD models in NIMROD and PSI-Tet appear fairly similar in scaling and dynamics to the HIT-SI simulations performed and illustrated here. Two-temperature simulations in NIMROD also significantly improve agreement with experimental data coming from both the HIT-SI and HIT-SI3 devices.

Chapter 3

GALERKIN MODELS FOR PLASMAS

To introduce the intersection of machine learning and plasma physics, it is natural to begin exploring analytic reduced-order models that can be derived from fluid or magnetohydrodynamic models. Many dynamical systems can be approximated by a system of coupled PDEs, or in the space-independent case, a system of ODEs. The scientific techniques to compute the numerical solutions, stability, and other analytic properties of these equations are extremely mature. Still, many of these dynamical systems exhibit behavior that is not fully understood, and the governing PDEs may be too complex to compute efficiently. For instance, realistic 3D gyrokinetic simulations in plasma physics can take days or months to compute [51]. In order to address such systems, scientists often use extra physical or heuristic assumptions to further reduce the complexity or dimensionality of the governing equations. The resulting ROMs can be used to efficiently compute approximations to the original dynamics, potentially facilitating parameter scans, physical understanding, and real-time control strategies.

Projection-based model reduction is a common approach for generating such models; a high-dimensional system, such as a spatially discretized set of PDEs, is projected onto a low-dimensional basis of modes [320, 321, 78, 79, 34, 374, 414, 415]. This "Galerkin projection" leads to a computationally efficient system of ODEs that describes how the mode amplitudes evolve in time [177]. However, these models can sometimes suffer from stability issues, causing solutions to diverge in finite-time.

A common class of projection-based ROMs are those derived from the proper orthogonal projection. POD-Galerkin ROMs have been increasingly used in the fluid dynamics field for the last thirty years and are well-understood, simple to compute, and often extend straight-

forwardly to a number of variations and improvements. POD-Galerkin models are far less common in plasma physics, and this may be because even a "simple" plasma model like single-fluid MHD is still significantly more complicated than the compressible Navier-Stokes equations. However, Galerkin and discontinuous Galerkin finite-element methods are often used to numerically solve the MHD or 5-moment model in plasma physics [263]. Although similar in flavor, POD-Galerkin models are typically low-dimensional models for the *global* solution of a fluid or plasma flow, while the discontinuous Galerkin finite-element method is a scheme for numerical discretization and solution of a system of PDEs in a traditional simulation code. While this global property of most POD-Galerkin models facilitates very large reductions in the system dimension (and therefore big computational efficiency gains), it is also this drastic reduction that casts doubt about the feasibility of POD-Galerkin models to adequately capture the fluid or plasma dynamics of experimental interest. This tradeoff between model efficiency and fidelity varies depending on the experimental goal and the system complexity, but despite these complications, POD-Galerkin models continue to be a valuable tool in the field of fluid mechanics. The work presented in this chapter has also been published as a paper in Kaptanoglu et al. [211].

3.1 Projection-based reduced-order models

Here, a brief description is provided to detail the projection-based reduction of a system of partial differential equations into a lower-dimensional ROM. In modern scientific computing, a set of governing partial differential equations is typically discretized into a high-dimensional system of coupled ordinary differential equations. Although not necessary for projection-based model reduction, this chapter explicitly considers dynamics with linear plus quadratic structure, as are found in many fluid and plasma systems:

$$\dot{\boldsymbol{q}} = \boldsymbol{C}^0 + \boldsymbol{L}^0\boldsymbol{q} + \boldsymbol{Q}^0(\boldsymbol{q}). \tag{3.1}$$

Here it is assumed that the PDE has already been discretized for numerical computation, resulting in a coupled system of N differential equations. The state of the system $\boldsymbol{q}(\boldsymbol{x}, t) \in \mathbb{R}^N$ is a high-dimensional vector that represents the fluid velocity or other set of spatiotemporal fields, for example sampled on a high-resolution spatial grid. Thus, $\boldsymbol{C}^0$, $\boldsymbol{L}^0$, and $\boldsymbol{Q}^0$ are typically high-dimensional operators used to perform a numerical simulation. The zero subscript distinguishes these operators from the Galerkin coefficients defined below in Eq. (3.12).

The goal of a projection-based ROM is to transform this high-dimensional system into a lower-dimensional system of size $r \ll N$ that captures the essential dynamics. One way to reduce the set of governing equations to a set of ODEs is by decomposition into a desired low-dimensional basis $\{\boldsymbol{\chi}_i(\boldsymbol{x})\}$ in a process commonly referred to as Galerkin expansion:

$$\boldsymbol{q}(\boldsymbol{x}, t) = \overline{\boldsymbol{q}}(\boldsymbol{x}) + \sum_{i=1}^{r} a_i(t)\boldsymbol{\chi}_i(\boldsymbol{x}). \tag{3.2}$$

Here, $\overline{\boldsymbol{q}}(\boldsymbol{x})$ is the mean field, $\boldsymbol{\chi}_i(\boldsymbol{x})$ are spatial modes, and $a_i(t)$ describe how the amplitude of these modes vary in time. The POD [177, 63] is frequently used to obtain the basis, since the modes $\boldsymbol{\chi}_i(\boldsymbol{x})$ are orthogonal and ordered by maximal energy content. The governing equations are then Galerkin projected onto the basis $\{\boldsymbol{\chi}_i(\boldsymbol{x})\}$ by substituting Eq. (3.2) into the PDE in Eq. (3.1) and using inner products to remove the spatial dependence. Orthogonal projection onto POD modes is the simplest and most common procedure, resulting in POD-Galerkin models, although Petrov-Galerkin projection [78, 79] improves model performance in some cases. Fourier-Galerkin models, also known in the MHD turbulence community as "shell models" [75, 244], are common and useful in the fluid and plasma physics fields but, as discussed further in Sec. 3.3.1, tend to be less practical as system ROMs. Now that the formation of projection-based ROMs has been discussed, the proper orthogonal decomposition will be reviewed.

3.2 Proper Orthogonal decomposition

The POD is already used extensively for interpreting plasma physics data across a range of parameter regimes [113, 246, 132, 432, 161], but some formalism is required to effectively use it for modeling and forecasting. Note that the POD is more commonly called the biorthogonal decomposition (abbreviated BD or BOD) in plasma physics. For POD, a set of point measurements at time t_k are arranged in a vector $\boldsymbol{q}_k \in \mathbb{R}^N$, called a snapshot, where the dimension N is now the product of the number of spatial locations and the number of variables measured at each point. For instance, magnetic field data could have been obtained from $N/3$ magnetic probes that measure the magnetic field components at a fixed location and sampling rate. Now assuming that the data is sampled at some times t_1, t_2, ..., t_M, arranged in a matrix $\boldsymbol{X} \in \mathbb{R}^{N\times M}$, and the average in time $\overline{\boldsymbol{q}}$ is subtracted off. The singular value decomposition (SVD) provides a low-rank approximation

$$\boldsymbol{X} = \overset{\text{time}\longrightarrow}{\begin{bmatrix} q_1(t_1) & q_1(t_2) & \cdots & q_1(t_M) \\ q_2(t_1) & q_2(t_2) & \cdots & q_2(t_M) \\ \vdots & \vdots & \ddots & \vdots \\ q_D(t_1) & q_D(t_2) & \cdots & q_N(t_M) \end{bmatrix}}\Bigg\downarrow \text{state} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\dagger, \tag{3.3}$$

where $\boldsymbol{U} \in \mathbb{R}^{N\times N}$ and $\boldsymbol{V} \in \mathbb{R}^{M\times M}$ are unitary matrices, and $\boldsymbol{\Sigma} \in \mathbb{R}^{N\times M}$ is a diagonal matrix containing non-negative and decreasing entries s_{jj} called the singular values of $\boldsymbol{X}$. $\boldsymbol{V}^\dagger$ denotes the complex-conjugate transpose of $\boldsymbol{V}$. The singular values indicate the relative importance of the corresponding columns of $\boldsymbol{U}$ and $\boldsymbol{V}$ for describing the spatiotemporal structure of $\boldsymbol{X}$.

It is often possible to discard small values of $\boldsymbol{\Sigma}$, resulting in a truncated matrix $\boldsymbol{\Sigma}_r \in \mathbb{R}^{r\times r}$.

With the first $r \ll \min(N, M)$ columns of $\boldsymbol{U}$ and $\boldsymbol{V}$, denoted $\boldsymbol{U}_r$ and $\boldsymbol{V}_r$,

$$\boldsymbol{X} \approx \boldsymbol{U}_r \boldsymbol{\Sigma}_r \boldsymbol{V}_r^\dagger. \tag{3.4}$$

The truncation rank r is typically chosen to balance accuracy and complexity [63]. The computational complexity of the SVD is $\mathcal{O}(NM^2 + M^3)$ [141], although there are randomized singular value decompositions [128, 252, 445] for very large problems that can be as fast as $\mathcal{O}(NM\log(r))$. Therefore, even for $r \gg 1$, the SVD typically produces significant computational speedup over codes which evolve the full spatiotemporal dynamics. The computational speed [141, 445] of the SVD also enables online computations to update a model for real-time control.

To proceed, a well-defined SVD requires that the measurements in $\boldsymbol{X}$ have the same physical dimensions. With a dimensionalized measurement vector $\boldsymbol{q}$, the matrix $\boldsymbol{X}^\dagger\boldsymbol{X}$ satisfies

$$\boldsymbol{X}^\dagger\boldsymbol{X} \approx \langle \boldsymbol{q}(t_k), \boldsymbol{q}(t_m) \rangle, \qquad k, m \in \{1, 2, ..., M\}. \tag{3.5}$$

The equality is not exact because the inner product (an integral) is approximated by the discrete sum from the matrix product of $\boldsymbol{X}^\dagger\boldsymbol{X}$. The temporal SVD modes, or chronos, $\boldsymbol{a}_j$ are the columns of $\boldsymbol{V}_r$. The spatial modes, or topos, $\boldsymbol{\chi}$ form the columns of $\boldsymbol{U}_r$. Finally,

$$\boldsymbol{q}(\boldsymbol{x}_i, t_k) \approx \overline{\boldsymbol{q}}(\boldsymbol{x}_i) + \sum_{j=1}^{r} a_j(t_k)\boldsymbol{\chi}_j(\boldsymbol{x}_i). \tag{3.6}$$

Any normalization of the $a_j(t_k)$ and the singular values has been absorbed into the definition of $\boldsymbol{\chi}_j(\boldsymbol{x}_i)$. By construction $\langle \boldsymbol{\chi}_i, \boldsymbol{\chi}_j \rangle \propto \delta_{ij}$. Note that, in principle, one can expand $\boldsymbol{q}$ in any set of modes, although orthonormal modes are preferred because this property facilitates the analysis below. Non-orthogonal modes are also suitable, but introduce a complication in the form of a mass matrix [367]. The advantage of the POD basis is that the modes

are ordered by energy content; a truncation of the system still captures a majority of the dynamics. With an expansion of the spatiotemporal state vector $\boldsymbol{q}$ in the POD basis, one can substitute this expansion into an analytic plasma model such as Hall-MHD and attempt to produce a POD-Galerkin model. To illustrate the POD expansion on a set of complex data, the POD modes from the velocity and magnetic field (coming from a 3D MHD simulation of the HIT-SI device) are shown in Fig. 3.1. The temporal modes consist primarily of harmonics of the injector frequency, and the spatial modes, visualized in 2D on the $Z = 0$ midplane of the device, exhibit interesting symmetries.

3.3 POD-Galerkin models

Traditional use of the POD on the MHD fields (velocity, magnetic, and temperature) would either require separate decompositions for $\boldsymbol{u}$, $\boldsymbol{B}$, and T, or an arbitrary choice of dimensionalization. This is not an ideal strategy because separate decompositions of the fields obfuscates the interpretability and increases the complexity of a low-dimensional model, and choosing the units of the combined matrix of measurement data can have a significant impact on the performance and energy spectrum of the resulting POD basis. Inspired by the inner product defined for compressible fluids [373], an inner product for MHD is defined through the following quantities

$$\boldsymbol{q}(\boldsymbol{x},t) = \begin{bmatrix} \boldsymbol{B}_u \\ \boldsymbol{B} \\ B_T \end{bmatrix}, \quad \boldsymbol{B}_u = \sqrt{\rho\mu_0}\boldsymbol{u}, \quad B_T = \sqrt{\frac{4\rho\mu_0 T}{m_i(\gamma-1)}}. \tag{3.7}$$

The total plasma energy is now related to the inner product $\langle \boldsymbol{q}, \boldsymbol{q} \rangle$ through

$$W = \frac{1}{2\mu_0}\langle \boldsymbol{q}, \boldsymbol{q} \rangle = \int \left(\frac{1}{2}\rho u^2 + \frac{B^2}{2\mu_0} + \frac{p}{\gamma - 1} \right) d^3\boldsymbol{x}. \tag{3.8}$$

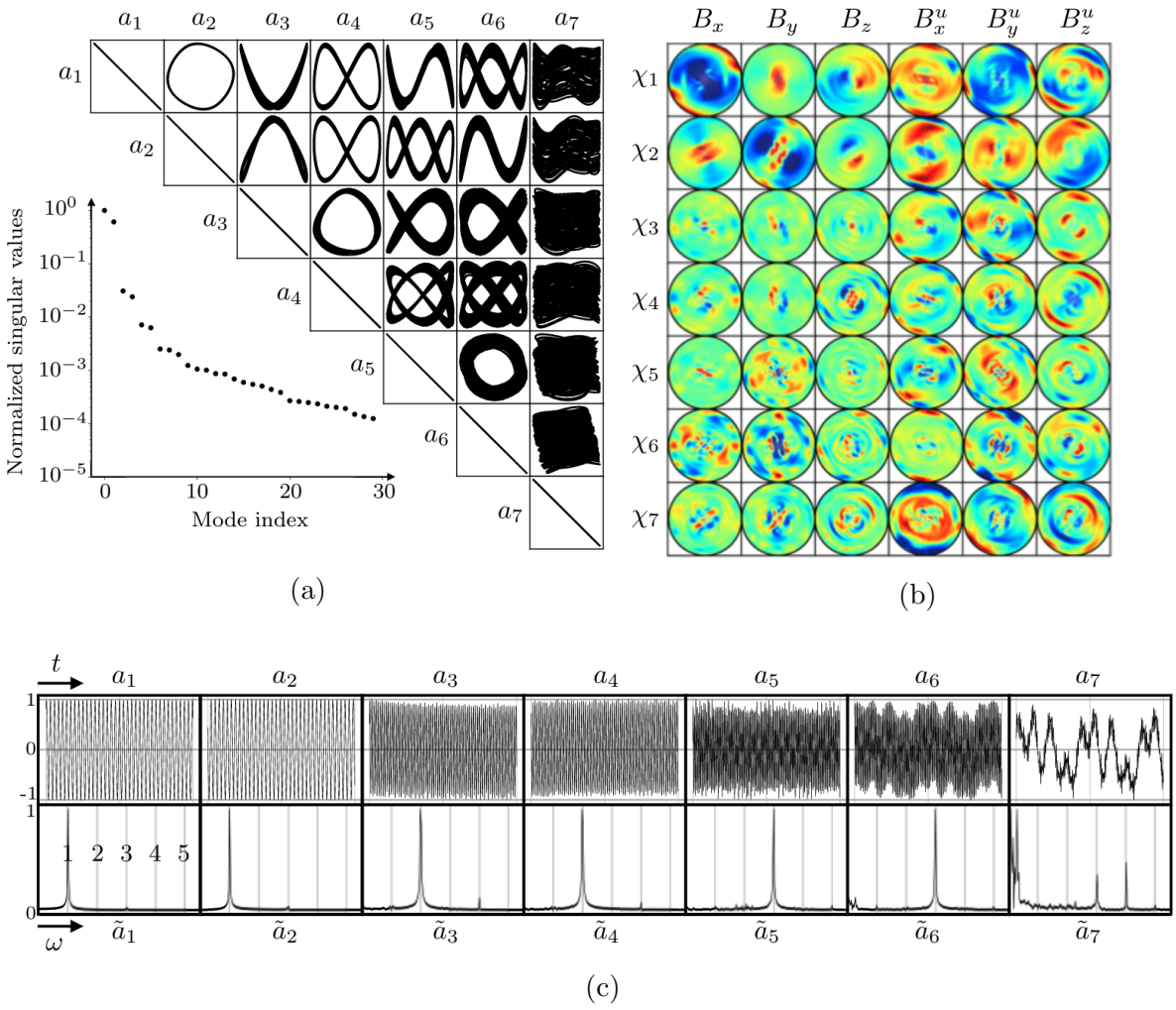


Figure 3.1: The first seven POD modes for a 3D isothermal Hall-MHD simulation of the HIT-SI device detailed in Sec. 2.1. The mean-flow-subtracted chronos indicate that the primary dynamics are forcing at the driving injector frequency and its harmonics; (a) Mode pair trajectories evolved in time and the corresponding singular values; (b) 3D spatial modes in the $Z = 0$ midplane illustrate a complicated mix of length scales; (c) Normalized temporal modes and corresponding Fourier transforms exhibit harmonics of the driving frequency, labeled 1-5.

Normalizing the MHD fields to magnetic field units produces a natural interpretation of inner products of the vector $\boldsymbol{q}$ as the total plasma energy. This formulation is also useful because reduced order models built for $\boldsymbol{q}$ can be constrained by conservation of energy via Eq. (3.8), as is illustrated in detail in Section 3.4. A separate POD of each of the MHD fields would lead to three sets of POD modes with independent time dynamics and mixed orthogonality properties. In contrast, the approach using Eq. (3.7) captures all the fields simultaneously, resulting in a single set of modes $a_i(t)$ in Eq. (3.6). Notably, the density is not included in $\boldsymbol{q}$ because it significantly complicates the definition of a suitable inner product and the Galerkin model derived below. This definition of $\boldsymbol{q}$ is still a suitable expansion for incompressible flows or flows for which the density varies slowly in time compared to the timescales associated with the Galerkin model.

With the expansion of the fields in a low-dimensional basis in Eq. (3.6), one can project the Hall-MHD equations onto these POD modes in order to obtain a POD-Galerkin model. Hall-MHD, assuming $T_e = T_i = T$ and the definitions in Eq. (3.7), can be written:

$$\dot{\rho}=-\nabla\cdot\left(\sqrt{\frac{\rho}{\mu_0}}\boldsymbol{B}_u\right), \tag{3.9}$$

$$\dot{\boldsymbol{B}}=\nabla\times\left[\frac{1}{\sqrt{\rho\mu_0}}\left(\boldsymbol{B}_u\times\boldsymbol{B}-d_i((\nabla\times\boldsymbol{B})\times\boldsymbol{B})\right)\right]+\eta_0\rho^{\frac{3}{2}}B_T^{-3}\nabla^2\boldsymbol{B}+\frac{d_i}{\sqrt{\rho\mu_0}}(\gamma-1)B_T\nabla B_T\times\frac{\nabla\rho}{2\rho},$$

$$\begin{aligned}\dot{\boldsymbol{B}}_u=&-\frac{1}{\sqrt{\rho\mu_0}}\left[\frac{1}{2}\boldsymbol{B}_u\nabla\cdot\boldsymbol{B}_u+\boldsymbol{B}_u\cdot\nabla\boldsymbol{B}_u-\frac{1}{4\rho}\boldsymbol{B}_u(\nabla\rho\cdot\boldsymbol{B}_u)-(\nabla\times\boldsymbol{B})\times\boldsymbol{B}+\frac{(\gamma-1)B_T^2}{2}\frac{\nabla\rho}{\rho}-(\gamma-1)B_T\nabla B_T\right]\\&+\nu\left[\nabla^2\boldsymbol{B}_u-\frac{\nabla^2\rho}{2\rho}\boldsymbol{B}_u+\frac{3\boldsymbol{B}_u}{4\rho^2}(\nabla\rho)^2+\frac{1}{\rho}(\nabla\rho\cdot\nabla)\boldsymbol{B}_u-\frac{1}{6\rho}\nabla(\nabla\rho\cdot\boldsymbol{B}_u)\right]\\&+\nu\left[\frac{1}{4\rho^2}(\nabla\rho\cdot\boldsymbol{B}_u)\nabla\rho+\frac{1}{3}\nabla(\nabla\cdot\boldsymbol{B}_u)-\frac{1}{6\rho}(\nabla\cdot\boldsymbol{B}_u)\nabla\rho\right],\end{aligned}$$

$$\dot{B}_T=-\frac{1}{\sqrt{\rho\mu_0}}\left[\boldsymbol{B}_u\cdot\nabla B_T-\gamma B_T\left(\nabla\cdot\boldsymbol{B}_u-\frac{\nabla\rho}{2\rho}\cdot\boldsymbol{B}_u\right)\right]-\frac{2}{B_T}(\nabla\cdot\boldsymbol{h}+Q_{\text{visc}})+4\eta_0\rho^{\frac{3}{2}}B_T^{-4}(\nabla\times\boldsymbol{B})^2,$$

where $\nabla\cdot\boldsymbol{B}=0$ is used and the modified definitions of the heat flux $\boldsymbol{h}$ and viscous heating

Q_{visc} are

$$\boldsymbol{h}=-\frac{(\gamma-1)B_T}{4}\left[\chi_{\parallel}\hat{\boldsymbol{b}}\hat{\boldsymbol{b}}+\chi_{\perp}\left(\boldsymbol{I}-\hat{\boldsymbol{b}}\hat{\boldsymbol{b}}\right)\right]\cdot\left(\nabla B_T-B_T\frac{\nabla\rho}{\rho}\right), \tag{3.10}$$

$$Q_{\text{visc}}=-\tilde{\nu}\left(\nabla\boldsymbol{B}_u-\boldsymbol{B}_u\frac{\nabla\rho}{2\rho}\right)^T:\left[\left(\nabla\boldsymbol{B}_u-\boldsymbol{B}_u\frac{\nabla\rho}{2\rho}\right)+\left(\nabla\boldsymbol{B}_u-\boldsymbol{B}_u\frac{\nabla\rho}{2\rho}\right)^T-\frac{2}{3}\boldsymbol{I}\left(\nabla\cdot\boldsymbol{B}_u-\boldsymbol{B}_u\cdot\frac{\nabla\rho}{2\rho}\right)\right].$$

See Appendix B for the derivation of these rather unintuitive equations and note that the incompressible, ideal limits reduce to a version of the usual Elsässer formulation [45]. Here $\tilde{\nu}=\rho\nu$ is the kinematic viscosity. Although many of the nonlinear terms are only quadratic in $\boldsymbol{q}$, here the isothermal limit and limit of time-independent density are considered to restrict models to the pure quadratic nonlinear case:

$$\dot{\boldsymbol{q}}=\boldsymbol{C}^0+\boldsymbol{L}^0(\boldsymbol{q})+\boldsymbol{Q}^0(\boldsymbol{q},\boldsymbol{q}), \tag{3.11}$$

$$\boldsymbol{C}^0=\begin{bmatrix}-\frac{(\gamma-1)B_T^2}{2}\sqrt{\frac{1}{\mu_0\rho}}\frac{\nabla\rho}{\rho}\\0\\0\end{bmatrix},$$

$$\boldsymbol{Q}^0(\boldsymbol{q},\boldsymbol{q})=\begin{bmatrix}-\frac{1}{\sqrt{\rho\mu_0}}\left(\boldsymbol{B}_u(\nabla\cdot\boldsymbol{B}_u)+\boldsymbol{B}_u\cdot\nabla\boldsymbol{B}_u-(\nabla\times\boldsymbol{B})\times\boldsymbol{B}\right)\\\nabla\times\left(\frac{1}{\sqrt{\rho\mu_0}}\left(\boldsymbol{B}_u\times\boldsymbol{B}-d_i(\nabla\times\boldsymbol{B})\times\boldsymbol{B}\right)\right)\\0\end{bmatrix},$$

$$\boldsymbol{L}^0(\boldsymbol{q})=\begin{bmatrix}\nu\left(\nabla^2\boldsymbol{B}_u-\cdots-\frac{1}{3\rho}(\nabla\cdot\boldsymbol{B}_u)\nabla\rho\right)\\\frac{\eta}{\mu_0}\nabla^2\boldsymbol{B}\\0\end{bmatrix}.$$

This is a slight abuse of notation, since these should be discretized operators. Other models, such as those assuming incompressibility and finite temperature evolution, can also be derived straightforwardly from the results here. Substituting Eq. (3.6) into Eq. (3.11) and utilizing

the orthonormality of the $\boldsymbol{\chi}_j$ produces:

$$\begin{aligned}
\dot{a}_i(t) &= C_i + \sum_{j=1}^{r} L_{ij} a_j + \sum_{j,k=1}^{r} Q_{ijk} a_j a_k, \qquad (3.12)\\
C_i &= \langle \boldsymbol{C}^0 + \boldsymbol{L}^0(\overline{\boldsymbol{q}}) + \boldsymbol{Q}^0(\overline{\boldsymbol{q}}, \overline{\boldsymbol{q}}), \boldsymbol{\chi}_i \rangle,\\
L_{ij} &= \langle \boldsymbol{L}^0(\boldsymbol{\chi}_j) + \boldsymbol{Q}^0(\overline{\boldsymbol{q}}, \boldsymbol{\chi}_j) + \boldsymbol{Q}^0(\boldsymbol{\chi}_j, \overline{\boldsymbol{q}}), \boldsymbol{\chi}_i \rangle,\\
Q_{ijk} &= \langle \boldsymbol{Q}^0(\boldsymbol{\chi}_j, \boldsymbol{\chi}_k), \boldsymbol{\chi}_i \rangle.
\end{aligned}$$

The model is quadratic in the temporal POD modes $a_i(t)$. In contrast to Eq. (3.12), a Galerkin model based on separate POD expansions for each field would involve significant mixing and a lack of orthonormality $\langle \boldsymbol{\chi}_i^{\boldsymbol{u}}, \boldsymbol{\chi}_j^{\boldsymbol{B}} \rangle \neq \delta_{ij}$ between the POD modes. Although Eq. (3.12) contains only quadratic nonlinearities, the influence of truncated low-energy modes can sometimes be modeled with cubic nonlinearities in the Galerkin model [320, 257]. This is discussed further below and in Sec. 5.4.2.

3.3.1 Relation to Fourier-Galerkin methods

Similar analytic Fourier-Galerkin models (also called MHD shell models) have been used for modeling incompressible MHD turbulence [351]. Shell models in MHD have primarily been used to describe the statistics of homogeneous and isotropic turbulence in spectral space, rather than as reduced order models [44]. The differences in application likely stem from shell models preserving the MHD invariants within each triad of wave vectors, but POD models providing a dataset-tailored and energy-optimal basis. However, in various homogeneous and symmetric limits, the POD reduces to the Fourier basis [96, 177]. In both Fourier-Galerkin and POD-Galerkin models, truncation of the model at some rank r can lead to under-resolving the dissipation rate or approximately breaking the global conservation laws, and a closure scheme may be required to re-introduce the full dissipation. Additionally, if energy is not conserved, as in some dissipative MHD models, the stability of the truncated system is no longer guaranteed.

3.4 Deriving constraints on projection-based models

The previous section has successfully obtained a POD-Galerkin model for the dynamic fields in Hall-MHD. However, there is substantial additional structure in the coefficients in Eq. (3.12) because some local properties and the global MHD conservation laws are in principle retained in this low-dimensional basis. With regards to local properties, vanishing $\nabla \cdot \boldsymbol{B}$ and the linear independence of the temporal POD modes produce

$$\nabla \cdot \boldsymbol{\chi}_i^B = 0, \qquad \forall i. \tag{3.13}$$

In other words, there is an analytic local divergence constraint for each of the $\boldsymbol{\chi}_i^B$, but this does not produce insight into the coefficients defined in Eq. (3.12), nor does it guarantee that Eq. (3.13) holds if $\boldsymbol{\chi}_i^B$ is obtained from noisy or low-resolution data. In contrast, global conservation laws produce substantial constraints on the structure of the Galerkin model coefficients.

3.4.1 Global conservation of energy

For an examination of the global conservation laws, consider isothermal Hall-MHD with a very slowly time-varying density. This model reduces to ideal MHD and incompressible, resistive, Hall MHD in the appropriate limits, and produces (Galtier [133] Eq. 3.22)

$$\begin{aligned} \frac{\partial W}{\partial t} = &- \int \left[\tilde{\nu} (\nabla \times \boldsymbol{u})^2 + \frac{\eta}{\mu_0} (\nabla \times \boldsymbol{B})^2 + \frac{4}{3} \tilde{\nu} (\nabla \cdot \boldsymbol{u})^2 \right] d^3\boldsymbol{x} \\ &- \oint \left[\left(\frac{1}{2} \rho u^2 + p \right) \boldsymbol{u} + \boldsymbol{P} - \frac{4}{3} \tilde{\nu} (\nabla \cdot \boldsymbol{u}) \boldsymbol{u} - \tilde{\nu} \boldsymbol{u} \times (\nabla \times \boldsymbol{u}) \right] \cdot \hat{\boldsymbol{n}} dS. \end{aligned} \tag{3.14}$$

The volume and surface is left intentionally vague but the rigid wall of a plasma device is a suitable boundary. Here $\hat{\boldsymbol{n}}$ is a unit normal vector to the boundary, and

$$\boldsymbol{P} = \frac{1}{\mu_0} \boldsymbol{E} \times \boldsymbol{B} = \frac{\boldsymbol{u}_e}{\mu_0} \cdot (B^2 \boldsymbol{I} - \boldsymbol{B}\boldsymbol{B}) - \frac{\eta}{\mu_0^2} (\nabla \times \boldsymbol{B}) \times \boldsymbol{B}, \tag{3.15}$$

is the Poynting vector ($\boldsymbol{E}$ is the electric field), which is often an imposed and experimentally-known function of space and time. Omission of the Hall term changes $\boldsymbol{u}_e$ to $\boldsymbol{u}$ in Eq. (3.15). Even with temperature evolution, the electron diamagnetic term in $\boldsymbol{P}$ does not alter the energy balance if Dirichlet conditions are used for ρ and T. To simplify, assume that $\boldsymbol{u}\cdot\hat{\boldsymbol{n}} = \boldsymbol{u}\times\hat{\boldsymbol{n}} = 0$, $\boldsymbol{J}\cdot\hat{\boldsymbol{n}} = 0$, and $\boldsymbol{B}\cdot\hat{\boldsymbol{n}} = 0$ at the wall, consistent with the HIT-SI experiment and HIT-SI simulations described in Sec. 2.1. and Section 5.2.1. Now assume steady-state, define $a_0(t) = 1$, and substitute Eq. (3.6) into Eq. (3.14),

$$
\begin{aligned}
0 \approx \frac{\partial W}{\partial t} &= \oint \frac{\eta}{\mu_0^2}((\nabla\times\boldsymbol{B})\times\boldsymbol{B})\cdot\hat{\boldsymbol{n}}dS \qquad (3.16)\\
&- \int\left[\frac{\nu}{\mu_0}\left(\nabla\times\boldsymbol{B}_u - \frac{\nabla\rho}{2\rho}\times\boldsymbol{B}_u\right)^2 + \frac{\eta}{\mu_0^2}(\nabla\times\boldsymbol{B})^2 + \frac{4}{3}\frac{\nu}{\mu_0}\left(\nabla\cdot\boldsymbol{B}_u - \frac{\nabla\rho}{2\rho}\cdot\boldsymbol{B}_u\right)^2\right]d^3\boldsymbol{x},\\
&= W^C + \sum_{i=1}^{r} W_i^L a_i + \sum_{i,j=1}^{r} W_{ij}^{\mathrm{Q}} a_i a_j = \sum_{i,j=0}^{r} W_{ij}^{Q} a_i a_j,
\end{aligned}
$$

The matrix in the last step is padded so that $W_{0i}^Q = 0$, $W_{i0}^Q = W_i^L$ for $i \in \{1, ..., r\}$, and $W_{00}^Q = W^C$. Eq. (3.16) is generally satisfied for anti-symmetric W_{ij}^Q, from which it follows

that

$$
\begin{aligned}
0 = W_{00}^{\mathrm{Q}} =& \frac{\eta}{\mu_0} \oint \left[(\nabla \times \overline{\boldsymbol{B}}) \times \overline{\boldsymbol{B}} \right] \cdot \hat{\boldsymbol{n}} dS \qquad (3.17)\\
&- \int \left[\nu \left(\nabla \times \overline{\boldsymbol{B}}_u - \frac{\nabla \rho}{2\rho} \times \overline{\boldsymbol{B}}_u \right)^2 + \frac{\eta}{\mu_0} (\nabla \times \overline{\boldsymbol{B}})^2 + \frac{4}{3} \nu \left(\nabla \cdot \overline{\boldsymbol{B}}_u - \frac{\nabla \rho}{2\rho} \cdot \overline{\boldsymbol{B}}_u \right)^2 \right] d^3\boldsymbol{x},\\
0 = W_{i0}^{\mathrm{Q}} =& \frac{\eta}{\mu_0} \oint \left[(\nabla \times \overline{\boldsymbol{B}}) \times \boldsymbol{\chi}_i^B + (\nabla \times \boldsymbol{\chi}_i^B) \times \overline{\boldsymbol{B}} \right] \cdot \hat{\boldsymbol{n}} dS\\
&- 2 \int \nu \left(\nabla \times \overline{\boldsymbol{B}}_u - \frac{\nabla \rho}{2\rho} \times \overline{\boldsymbol{B}}_u \right) \cdot \left(\nabla \times \boldsymbol{\chi}_i^{B_u} - \frac{\nabla \rho}{2\rho} \times \boldsymbol{\chi}_i^{B_u} \right) d^3\boldsymbol{x}\\
&+ \int \left[\frac{\eta}{\mu_0} (\nabla \times \overline{\boldsymbol{B}}) \cdot (\nabla \times \boldsymbol{\chi}_i^B) + \frac{4}{3} \nu \left(\nabla \cdot \overline{\boldsymbol{B}}_u - \frac{\nabla \rho}{2\rho} \cdot \overline{\boldsymbol{B}}_u \right) \cdot \left(\nabla \cdot \boldsymbol{\chi}_i^{B_u} - \frac{\nabla \rho}{2\rho} \cdot \boldsymbol{\chi}_i^{B_u} \right) \right] d^3\boldsymbol{x},\\
W_{ij}^{\mathrm{Q}} =& - W_{ji}^{\mathrm{Q}} = \frac{\eta}{\mu_0} \oint \left[(\nabla \times \boldsymbol{\chi}_i^B) \times \boldsymbol{\chi}_j^B \right] \cdot \hat{\boldsymbol{n}} dS\\
&- \int \left(\nabla \times \boldsymbol{\chi}_i^{B_u} - \frac{\nabla \rho}{2\rho} \times \boldsymbol{\chi}_i^{B_u} \right) \cdot \left(\nabla \times \boldsymbol{\chi}_j^{B_u} - \frac{\nabla \rho}{2\rho} \times \boldsymbol{\chi}_j^{B_u} \right) d^3\boldsymbol{x}\\
&+ \int \frac{\eta}{\mu_0} \left(\nabla \times \boldsymbol{\chi}_i^B \right) \cdot \left(\nabla \times \boldsymbol{\chi}_j^B \right) d^3\boldsymbol{x}\\
&+ \int \frac{4}{3} \nu \left(\nabla \cdot \boldsymbol{\chi}_i^{B_u} - \frac{\nabla \rho}{2\rho} \cdot \boldsymbol{\chi}_i^{B_u} \right) \cdot \left(\nabla \cdot \boldsymbol{\chi}_j^{B_u} - \frac{\nabla \rho}{2\rho} \cdot \boldsymbol{\chi}_j^{B_u} \right) d^3\boldsymbol{x}.
\end{aligned}
$$

Evaluating Eq. (3.17) and the Galerkin coefficients in Eq. (3.12) relies on the existence of all of the $\nabla \times \boldsymbol{\chi}_i$. These spatial POD modes are evaluated on a discrete set of spatial locations, but in practice one can always choose an interpolation such that the curl operator is well-defined. In such a case, $\nabla \times \boldsymbol{\chi}_i^B$ and $\nabla \times \boldsymbol{\chi}_i^{B_u}$ have natural interpretations as the spatial POD modes of the electromagnetic current and vorticity fields. However, here these computations only serve as formal manipulations, so it is not necessary to evaluate these curls. Moreover, the data-driven methods in Section 5 use sparse regression to determine these coefficients from data. Continuing on with the analysis, one can compute $a_i \dot{a}_i$ for

$i \in \{1, ..., r\}$,

$$a_i \dot{a}_i = \sum_{i,j=1}^{r} a_i \frac{\partial a_j}{\partial t} \int \boldsymbol{\chi}_i \cdot \boldsymbol{\chi}_j d^3\boldsymbol{x} = \int \frac{1}{2} \frac{\partial q^2}{\partial t} d^3\boldsymbol{x} = \frac{\partial W}{\partial t}, \tag{3.18}$$

$$a_i \dot{a}_i = a_i C_i + a_i L_{ij} a_j + a_i Q_{ijk} a_j a_k, \qquad i, j, k \in \{1, ..., r\}. \tag{3.19}$$

First, note that $W_{i0}^{\mathrm{Q}} = 0$ produces $C_i = 0$ for all $i \in \{1, ..., r\}$. There are no constant terms in the Galerkin model. This is a physical consequence of the assumption that $\overline{\boldsymbol{q}}$ is steady-state; nonzero constant terms would imply the possibility of unbounded growth in the energy norm. The anti-symmetry of W_{ij}^{Q} for $i, j \in \{1, ..., r\}$ constrains the quadratic structure of the energy $\boldsymbol{a}^T \boldsymbol{a}$,

$$\boldsymbol{a}^T \boldsymbol{L} \boldsymbol{a} \approx 0. \tag{3.20}$$

This physical interpretation is also clear. If the plasma is steady-state but has finite dissipation, the input power, here manifested through a purely quadratic Poynting flux $\boldsymbol{P} \propto \eta \boldsymbol{J} \times \boldsymbol{B}$, must be balancing these losses. Finally, there are no cubic terms in the time derivative of the energy in Eq. (3.16), implying

$$\boldsymbol{a}^T \boldsymbol{Q} \boldsymbol{a} \boldsymbol{a} = 0, \tag{3.21}$$

or equivalently,

$$Q_{ijk} + Q_{jik} + Q_{kij} = 0. \tag{3.22}$$

In other words, the quadratic nonlinearities in the Galerkin model of Eq. (3.12) are energy-preserving. This conclusion did not rely on any assumption of steady-state and energy-preserving structure in other quadratic nonlinearities is well-studied in fluid mechanics [390, 257, 209] and addressed more directly in Chapter 5. The lack of cubic, nonlinear energy

losses is a physical consequence coming from the boundary conditions $\boldsymbol{B}\cdot\hat{\boldsymbol{n}}=0$, $\boldsymbol{J}\cdot\hat{\boldsymbol{n}}=0$, $\boldsymbol{u}\cdot\hat{\boldsymbol{n}}=\boldsymbol{u}\times\hat{\boldsymbol{n}}=0$ (and constant temperature). Note that the orthogonality of the POD basis allowed us to conclude that each of the terms in Eq. 3.19 vanishes separately, rather than all together.

3.4.2 Global conservation of cross-helicity

An analogous derivation can be done to further constrain the model-building for systems which conserve cross-helicity, which includes incompressible ideal and resistive MHD turbulence, assuming suitable boundary conditions. Consider the local form of cross-helicity $H_c=\boldsymbol{u}\cdot\boldsymbol{B}$. Using Galtier [133] Eq. (3.36),

$$\begin{aligned}\frac{\partial H_c}{\partial t}=&-\nabla\cdot\left[\left(\frac{u^2}{2}+\frac{\gamma p}{(\gamma-1)\rho}\right)\boldsymbol{B}+\boldsymbol{u}\times(\boldsymbol{u}\times\boldsymbol{B})-\frac{d_i}{\sqrt{\rho\mu_0}}\boldsymbol{u}\times\big((\nabla\times\boldsymbol{B})\times\boldsymbol{B}\big)-\eta\boldsymbol{u}\times(\nabla\times\boldsymbol{B})\right] \\ &+\nu\nabla\cdot\left(\boldsymbol{B}\times(\nabla\times\boldsymbol{u})+\frac{4}{3}(\nabla\cdot\boldsymbol{u})\boldsymbol{B}\right)-\frac{d_i}{\sqrt{\rho\mu_0}}(\nabla\times\boldsymbol{u})\cdot\big((\nabla\times\boldsymbol{B})\times\boldsymbol{B}\big)-(\eta+\nu)(\nabla\times\boldsymbol{B})\cdot(\nabla\times\boldsymbol{u}).\end{aligned} \tag{3.23}$$

Consider again the simplifying case $\boldsymbol{J}\cdot\hat{\boldsymbol{n}}=0$, $\boldsymbol{B}\cdot\hat{\boldsymbol{n}}=0$, and $\boldsymbol{u}\cdot\hat{\boldsymbol{n}}=\boldsymbol{u}\times\hat{\boldsymbol{n}}=0$. If global cross-helicity is conserved, the integral form is

$$\begin{aligned}0\approx\int\frac{\partial H_c}{\partial t}d^3\boldsymbol{x}=&\int\nu\frac{\nabla\rho}{\rho}\cdot\left(\boldsymbol{B}\times(\nabla\times\boldsymbol{u})+\frac{4}{3}(\nabla\cdot\boldsymbol{u})\boldsymbol{B}\right)d^3\boldsymbol{x} \\ &-\int\left[\frac{d_i}{\sqrt{\rho\mu_0}}(\nabla\times\boldsymbol{u})\cdot\big((\nabla\times\boldsymbol{B})\times\boldsymbol{B}\big)+(\eta+\nu)(\nabla\times\boldsymbol{B})\cdot(\nabla\times\boldsymbol{u})\right]d^3\boldsymbol{x}.\end{aligned} \tag{3.24}$$

Substituting Eq. (3.6) into Eq. (3.24) produces terms up to cubic in the temporal POD modes,

$$0 \approx \int \frac{\partial H_c}{\partial t} d^3\boldsymbol{x} = \frac{\partial}{\partial t}(a_i a_j) \int \frac{1}{\sqrt{\rho\mu_0}} \boldsymbol{\chi}_i^{B_u} \cdot \boldsymbol{\chi}_j^{B} d^3\boldsymbol{x} \tag{3.25}$$

$$= A_{ij}^{H_c} \frac{\partial}{\partial t}(a_i a_j) \rightarrow \begin{bmatrix} A_{ij}^{H_c} C_j a_i \\ A_{ij}^{H_c} L_{jk} a_i a_k \\ A_{ij}^{H_c} Q_{jkl} a_i a_k a_l \end{bmatrix} \approx \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Note that if the system is energy-preserving, $C_j = 0$ for all j, so the first equality is already satisfied. The second equality determines that $A_{ij}^{H_c} L_{jk}$ is anti-symmetric under swapping i and k, and energy-preservation in Eq. (3.20) produces anti-symmetry under swapping j and k. The most straightforward solution is $L_{jk} = 0$ for all j,k; this solution is precisely the ideal limit corresponding to $\eta = \nu = 0$. Since $A_{ij}^{H_c}$ is not symmetric, this constraint can also apply to systems which conserve cross-helicity despite finite dissipation. In other words, cross-helicity can be injected into the device in a way that balances the dissipation, which is a form of "power balance" for the cross-helicity.

Lastly, $A_{ij}^{H_c} Q_{jkl}$, containing only the contribution from the Hall-term, exhibits the same structure as (and is compatible with) the constraint on the energy-preserving nonlinearities in Eq. (3.21). The simplest solution is $A_{ij}^{H_c} Q_{jkl} = 0$ for all i, k, l, since this corresponds to standard MHD without the Hall term. Like the analysis of the linear terms, this constraint indicates that it is possible that there are indices for which $A_{ij}^{H_c} Q_{jkl} \neq 0$ but overall satisfy $A_{ij}^{H_c} Q_{jkl} a_i a_k a_l = 0$, so that nonzero Hall contributions can still conserve cross-helicity. However, cross-helicity is not an ideal invariant of Hall-MHD since in the volume the Hall-term can act as both dissipation or a source and this would require the unlikely scenario that all of the volumetric contributions sum to approximately zero. Note that without the Hall term, the remaining terms in Eq. (3.24) are only up to quadratic in nonlinearity. Therefore, there are a class of common boundary conditions under which even *viscoresistive* MHD fluids exhibit quadratic nonlinearities that are cross-helicity preserving, in analogy

to energy-preserving, quadratic nonlinearities satisfying Eq. (3.21) that occur commonly in fluid mechanics [390]. Lastly, although inviscid Hall-MHD with these boundary conditions has other time-invariants through the mass, momentum, magnetic helicity, and generalized helicity, enforcing the helicity invariants may require alternative formulations to the one presented here, since derived fields like the vector potential are involved. Mass conservation cannot be straightforwardly related to the the POD-Galerkin model in Eq. (3.12) because the density is not included in the definition of $\boldsymbol{q} = [\boldsymbol{B}_u, \boldsymbol{B}, \boldsymbol{B}_T]$. However, in principle the momentum provides additional model constraints.

3.4.3 Global conservation of momentum

Similarly to the previous sections, the change in fluid momentum in time can be found in Galtier [133],

$$\frac{\partial(\rho \boldsymbol{u})}{\partial t} = -\nabla \cdot \left(\rho \boldsymbol{u}\boldsymbol{u} + \frac{B^2}{2\mu_0}\boldsymbol{I} - \frac{\boldsymbol{B}\boldsymbol{B}}{\mu_0} + p\boldsymbol{I} + \boldsymbol{\Pi} \right), \tag{3.26}$$

and the global change in fluid momentum is,

$$\begin{aligned} \frac{\partial}{\partial t} \int \rho \boldsymbol{u} d^3\boldsymbol{x} &= - \oint \left(\rho \boldsymbol{u}\boldsymbol{u} + \frac{B^2}{2\mu_0}\boldsymbol{I} - \frac{\boldsymbol{B}\boldsymbol{B}}{\mu_0} + p\boldsymbol{I} + \boldsymbol{\Pi} \right) \cdot \hat{\boldsymbol{n}} dS \\ &= \oint \left(\frac{B^2}{2\mu_0}\hat{\boldsymbol{n}} + \boldsymbol{\Pi} \cdot \hat{\boldsymbol{n}} \right) dS. \end{aligned} \tag{3.27}$$

Since time derivatives of the magnetic field are not involved, the Hall-term is irrelevant here, and the boundary conditions $\boldsymbol{B} \cdot \hat{\boldsymbol{n}} = 0$ and $\boldsymbol{u} \cdot \hat{\boldsymbol{n}} = 0$ have been used. If there is a balance between the viscous dissipation of momentum and momentum carried with the magnetic field (presumably this magnetic field is being externally sustained), then

$$0 \approx \frac{\partial}{\partial t} \int \rho \boldsymbol{u} d^3\boldsymbol{x} = \oint \left(\frac{B^2}{2\mu_0}\hat{\boldsymbol{n}} + \boldsymbol{\Pi} \cdot \hat{\boldsymbol{n}} \right) dS. \tag{3.28}$$

Moreover, from the POD-Galerkin model it follows that (neglecting the time dependence of ρ here)

$$0 \approx \int \frac{\partial(\rho \boldsymbol{u})}{\partial t} d^3\boldsymbol{x} = \int \frac{1}{\sqrt{\mu_0 \rho}} \dot{\boldsymbol{B}}_u d^3\boldsymbol{x} = \dot{a}_i \int \frac{1}{\sqrt{\mu_0 \rho}} \boldsymbol{\chi}_i^{B_u} d^3\boldsymbol{x} = C_i^M \dot{a}_i, \tag{3.29}$$
$$= C_i^M (C_i + L_{ij} a_j + Q_{ijk} a_j a_k).$$

Therefore, if one can reasonably approximate C_i^M from measurements of the density and the spatial POD modes of $\boldsymbol{B}_u$, then the terms in Eq. (3.29) should separately vanish. This is interesting but it is unclear what kind of plasma device conserves global momentum in this way and it is difficult in practice to approximate C_i^M. If plasma devices are balancing any quantity at all, it is usually the energy or magnetic helicity.

3.4.4 Conservation laws with velocity units

The previous sections have illustrated that the choice of magnetic field units in Eq. (3.7) allows one to relate global MHD conservation laws to the structure of the coefficients in the POD-Galerkin model. It is worth exploring any alterations in velocity units (in closer analogy to fluid dynamics) $\boldsymbol{q} = [\boldsymbol{u}, \boldsymbol{u}_A, u_s]$,

$$u_s^2 = \frac{4T}{m_i(\gamma - 1)}, \quad \boldsymbol{u}_A = \frac{\boldsymbol{B}}{\sqrt{\mu_0 \rho}}, \quad \frac{1}{2}\langle \boldsymbol{q}, \boldsymbol{q} \rangle = \frac{1}{2} \int \left(u^2 + u_A^2 + u_s^2 \right) d^3\boldsymbol{x}, \tag{3.30}$$

where u_s is defined as a scaled plasma sound speed. If ρ is uniform, $\rho \langle \boldsymbol{q}, \boldsymbol{q} \rangle / 2 = W$. The isothermal and time-independent density assumptions allow us to derive another quadratic model in $\boldsymbol{q}$, for which a POD-Galerkin model is readily available (the form is identical to Eq. (3.12) but the POD modes and coefficients have changed). Once again, assume $\boldsymbol{u} \cdot \hat{\boldsymbol{n}} = \boldsymbol{u} \times \hat{\boldsymbol{n}} = 0$, $\boldsymbol{J} \cdot \hat{\boldsymbol{n}} = 0$, and $\boldsymbol{B} \cdot \hat{\boldsymbol{n}} = 0$ on the boundary, so that

$$\int \frac{\rho}{2} \frac{dq^2}{dt} d^3\boldsymbol{x} = \frac{\partial W}{\partial t}. \tag{3.31}$$

This is equivalent to Eq. (3.18) in the particular case of time-independent density. Without this assumption, an extra term appears, proportional to $\int \boldsymbol{u} \cdot \nabla(u^2 + u_A^2) d^3\boldsymbol{x}$. Although from dimensional analysis this term is potentially very large, this may not be the case for many laboratory devices with strong anisotropy introduced by a large external magnetic field. For instance, steady-state toroidal plasmas with large closed flux surfaces would expect $\boldsymbol{u} \cdot \nabla u_A^2$ and $\boldsymbol{u} \cdot \nabla u^2$ to be small, as the fluid velocity is primarily along field lines and gradients in both the magnetic and velocity fields are primarily across field lines. For this reason, in certain devices the use of $\boldsymbol{q} = [\boldsymbol{u}, \boldsymbol{u}_A, u_s]$ could be a useful alternative to the formulation used in the main body of this work. It is possible that, in these units, the structure of the nonlinearities in the associated POD-Galerkin model may prove more amenable to analysis or computation.

3.4.5 Hyper-reduction techniques

Now that it has been illustrated how global conservation laws manifest as structure in Galerkin models, one could compute the coefficients in Eq. (3.12) and evolve the subsequent model. However, in order to calculate the model coefficients, spatial derivatives for ρ, $\boldsymbol{B}_u$, and $\boldsymbol{B}$ (and B_T if temperature is evolved) must be well-approximated in the region of experimental interest. In some cases, high-resolution diagnostics can resolve these quantities in a particular plasma region. Even if the high-quality data is available, for instance through simulations, computing these inner products and evaluating the nonlinear terms is expensive, because the fields have the original spatial dimension. This somewhat reduces the usefulness of projection-based model reduction. Fortunately, there are hyper-reduction techniques from fluid dynamics [34], such as the discrete empirical interpolation method (DEIM) [85], QDEIM [111], missing-point estimation [13] and gappy POD [440, 79], which can enable efficient computations. Instead of using hyper-reduction, this work turns to emerging and increasingly sophisticated machine learning methods in Section 5 to discover Galerkin models from data. There are three primary reasons to derive the POD-Galerkin model structure here anyways: 1) this analysis indicates that one can search plasma datasets for systems of ODEs

consisting only up to quadratic polynomials, as shown in Chapter 5, 2) it provides the theoretical basis for the incorporation of model constraints from global conservation laws, and 3) it provides a theoretical basis for projection-based model reduction and hyper-reduction techniques in future MHD work.

Chapter 4

DYNAMIC MODE DECOMPOSITION FOR PLASMAS

Although Chapter 3 focused on the POD basis, there are many different low-dimensional bases that can be used for different experimental or analytic tasks. POD modes tend to have a dominant frequency but they may also mix the frequency content, as the SVD optimization identifies orthogonal modes purely based on energy content. This frequency mixing in POD modes was one of the main motivations for the development of DMD in the fluids community. DMD is particularly attractive for physics research relevant to plasma waves and instabilities as it decomposes time-series signals into spatially correlated modes that are constrained to have periodic dynamics in time, possibly with a growth or decay rate. Thus, DMD isolates the data frequency content and results in a reduced set of spatial modes along with a linear model for how they evolve in time. Although DMD yields a linear model, the algorithm has strong connections to nonlinear dynamical systems via Koopman operator theory [371, 294, 295, 324, 326, 442, 417, 223]. There have also been several recent innovations and extensions to DMD that improve its ability to model complex systems, including for control [353], multi-resolution analysis [236], nonlinear observations [442], the incorporation of physical priors [16], and modal analysis from data that is undersampled in space [154, 60] and time [428]. Although DMD is known to be sensitive to noise [17], several algorithms exist to address this issue [171, 101, 12]. This collection of DMD algorithms may provide more efficient reduced-order models and deeper physical insight into the inner workings of plasmas. DMD has also been increasingly used in the plasma physics community, including for modal analysis [419, 207, 234, 312], identification of limit-cycle dynamics in 2D turbulent cylindrical plasma simulations [382], prediction and speedup of plasma kinetic simulations [314, 313], and other plasma simulation accelerations [103].

Interestingly, subspace system identification (SSI) methods have already been used across scientific domains for extracting single-frequency modes from data [355]. In plasma physics, SSI has been used for instability identification and magnetic spectroscopy in simulation and experimental tokamak data [329, 330, 437, 254]. Variants of SSI differ only with respect to the weighting scheme of the SVD [433] but the connections between SSI and the variants of DMD have been unclear despite the obvious similarities. Fortunately, a recent paper [400] indicates that, at least for autonomous systems, SSI and some forms of DMD [235, 165] are equivalent. This is good news for unifying SSI and the burgeoning DMD literature in plasma physics (which appear to be unaware of the other), as well as understanding when SSI or DMD may be more preferable for a given task. However, as detailed in the previous paragraph, there are many DMD variants and plenty of future work still to make additional connections with SSI. The study presented in this chapter has also been published as a paper in Kaptanoglu et al. [207]. This section begins with a description of the experimental and simulation datasets that were used so that the results generated with the DMD can be properly interpreted.

4.1 Data from the HIT-SI experiment and simulations

The HIT-SI experiment has an array of magnetic field probes that encircle four poloidal cross sections at toroidal angles $\phi = 0$, 45, 180, and 225 degrees, illustrated on the right in Fig. 4.1. On the left side in Fig. 4.1, the 18 surface probes are shown in one of the four identical poloidal cross sections. There are also additional probes, labeled $L05$ and $L06$, which are spaced out every 22.5^o, for a total of 96 probes. Each probe measures the components of the magnetic field that are locally tangential to the conducting wall. With the exception of the probes labeled L05 and L06, the tangential directions are the toroidal ϕ and poloidal θ directions, and $B_r \approx 0$. The experimental probes have a time resolution $\Delta t \approx 2\ \mu s$.

In addition to analyzing experimental data from these magnetic probes in Section 4.3, in Section 4.4 an investigation is performed to analyze coherent magnetic structures for the

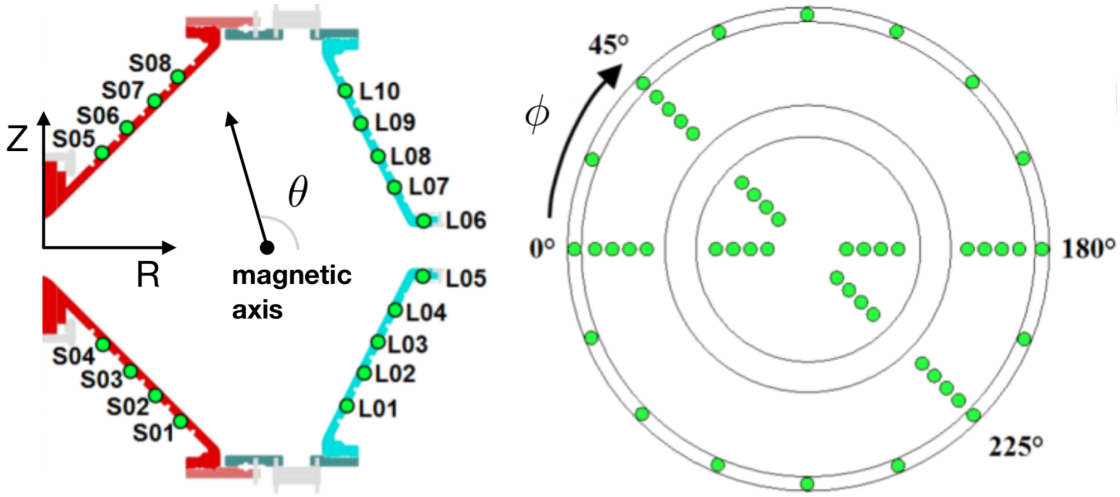


Figure 4.1: Surface probe locations in a HIT-SI cross section, and from a top view. There are 96 probes in total.

"BIG-HIT" extended MHD simulations using the NIMROD code [405]. BIG-HIT are mostly identical to a typical HIT-SI simulation, but the device has been enlarged by a factor of 2.5. Morgan et al. [305] provides more details on this simulation.

Simulation data is collected from the location of the experimental surface probes as well as a set of internal probes, all of which measure the magnetic field. In principle any set of normalized measurements may be used. For simplicity, all of the measurements and results in this Chapter 4 are reported in units of Gauss, and hereafter magnetic field units are omitted. In order to analyze the toroidal structure of the internal magnetic field in this simulation, 32 internal magnetic probe arrays are "placed" equally spaced toroidally at the $Z = 0$ midplane. Each synthetic array contains 160 measurement points at equally spaced radial locations between $0 \leq R \leq 1.34$ m. It will be shown that a sparse set of only 24 well-separated internal, synthetic, magnetic probes captures the mode structures that are observed with the full $160 \times 32 = 5120$ internal magnetic probes, providing evidence that this analysis is relevant to experimental devices with a small number of unevenly spaced measurements.

All experimental or simulation probe measurements at a fixed time t_k are arranged into

a snapshot column vector $\boldsymbol{q}_k \in \mathbb{R}^N$ with corresponding data matrix $\boldsymbol{X}$, as in Eq. (3.3). For both the experimental and simulation data without the internal probes, $N = 192$ and $M \approx 500 - 1000$, as typical discharges are $1 - 2$ ms with measurement resolution $\Delta t \approx 2$ μs. The small dataset used in Section 4.4 has 96 surface probes (192 measurements because each surface probe reports two components of the magnetic field) and 24 internal probes. The large simulation dataset has 96 surface probes and 5120 internal probes.

4.2 DMD algorithms

Proceeding with the exact DMD formulation and assuming that the state variables $\boldsymbol{q}_k$ are directly accessible for simplicity, (although adding observables is straightforward) the nonlinear evolution of the magnetic field may be approximated by a best-fit linear operator $\hat{\boldsymbol{A}}$ that evolves the state $\boldsymbol{q}_k$ forward in time:

$$\boldsymbol{q}_{k+1} \approx \hat{\boldsymbol{A}}\boldsymbol{q}_k. \tag{4.1}$$

The dynamic mode decomposition approximates the leading eigenvalues and eigenvectors of the linear operator $\hat{\boldsymbol{A}}$. To approximate $\hat{\boldsymbol{A}}$ from data, two matrices are constructed, $\boldsymbol{X}$ and $\boldsymbol{X}'$

$$\boldsymbol{X} = \begin{bmatrix} | & | & & | \\ \boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots & \boldsymbol{q}_{M-1} \\ | & | & & | \end{bmatrix}, \quad \boldsymbol{X}' = \begin{bmatrix} | & | & & | \\ \boldsymbol{q}_2 & \boldsymbol{q}_3 & \cdots & \boldsymbol{q}_{M} \\ | & | & & | \end{bmatrix},$$

which are related by =

$$\boldsymbol{X}' \approx \hat{\boldsymbol{A}}\boldsymbol{X}. \tag{4.2}$$

Observe that $\boldsymbol{X}$ and $\boldsymbol{X}'$ are identical to the original data matrix $\boldsymbol{X}$ defined in Eq. (3.3), except that either the first or last column is truncated off. The best-fit linear operator $\hat{\boldsymbol{A}}$

that satisfies Eq. (4.2) is the solution to the following least-squares optimization:

$$\hat{\boldsymbol{A}} = \underset{\hat{\boldsymbol{A}}}{\operatorname{argmin}} \|\boldsymbol{X}' - \hat{\boldsymbol{A}}\boldsymbol{X}\|_F = \boldsymbol{X}'\boldsymbol{X}^+ \approx \boldsymbol{X}'\boldsymbol{V}_r\boldsymbol{\Sigma}_r^{-1}\boldsymbol{U}_r^\dagger,$$

where $\boldsymbol{X}^+$ is the pseudoinverse of the matrix $\boldsymbol{X}$. However, when the measurement dimension N is large, then $\hat{\boldsymbol{A}}$ is too large to analyze directly, and instead $\hat{\boldsymbol{A}}$ is projected onto the first r singular vectors $\boldsymbol{U}_r$:

$$\tilde{\boldsymbol{A}} = \boldsymbol{U}_r^\dagger\hat{\boldsymbol{A}}\boldsymbol{U}_r = \boldsymbol{U}_r^\dagger\boldsymbol{X}'\boldsymbol{X}^+\boldsymbol{U}_r = \boldsymbol{U}_r^\dagger\boldsymbol{X}'\boldsymbol{V}_r\boldsymbol{\Sigma}_r^{-1} \tag{4.3}$$

Next, the eigendecomposition of $\tilde{\boldsymbol{A}}$ is computed:

$$\tilde{\boldsymbol{A}}\boldsymbol{W}_{\tilde{A}} = \boldsymbol{W}_{\tilde{A}}\boldsymbol{\Lambda}. \tag{4.4}$$

The diagonal matrix $\boldsymbol{\Lambda}$ contains the eigenvalues $\tilde{\lambda}_j$ of $\tilde{\boldsymbol{A}}$, which are also eigenvalues of $\hat{\boldsymbol{A}}$. The corresponding eigenvectors of $\tilde{\boldsymbol{A}}$ (and $\hat{\boldsymbol{A}}$) may be computed as

$$\boldsymbol{\Phi} = \boldsymbol{X}'\boldsymbol{V}_r\boldsymbol{\Sigma}_r^{-1}\boldsymbol{W}_{\tilde{A}}. \tag{4.5}$$

The columns $\boldsymbol{\varphi}_j$ of $\boldsymbol{\Phi}$ are DMD eigenvectors corresponding to DMD eigenvalues $\tilde{\lambda}_j$. It is then possible to reconstruct the state at time $k\Delta t$:

$$\boldsymbol{q}_k = \sum_{j=1}^{r} \boldsymbol{\varphi}_j\tilde{\lambda}_j^{k-1}b_j = \boldsymbol{\Phi}\boldsymbol{\Lambda}^{k-1}\boldsymbol{b}, \tag{4.6}$$

where $\boldsymbol{b}$ is a vector of DMD mode *amplitudes*. In the simplest case, it is possible to approximate $\boldsymbol{b} = \boldsymbol{\Phi}^+\boldsymbol{q}_1$, although the sparsity-promoting and optimized variants below will provide

more principled approaches to approximate $\boldsymbol{b}$. The data matrix $\boldsymbol{X}$ may then be written as

$$\boldsymbol{X} \approx \begin{bmatrix} | & & | \\ \boldsymbol{\varphi}_1 & \cdots & \boldsymbol{\varphi}_r \\ | & & | \end{bmatrix} \begin{bmatrix} b_1 & & \\ & \ddots & \\ & & b_r \end{bmatrix} \begin{bmatrix} 1 & \tilde{\lambda}_1 & \cdots & \tilde{\lambda}_1^{M-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \tilde{\lambda}_r & \cdots & \tilde{\lambda}_r^{M-1} \end{bmatrix}.$$

The eigenvalues $\tilde{\lambda}_j$ describe the discrete-time dynamical system in Eq. (4.1). It is often beneficial to analyze the corresponding continuous-time eigenvalues $\omega_j = \log(\lambda_j)/\Delta t$, with $\nu_j = \mathrm{Re}(\omega_j)/2\pi$, $f_j = \mathrm{Im}(\omega_j)/2\pi$. It is then possible to approximate the data matrix $\boldsymbol{X}$ as

$$\boldsymbol{X} \approx \underbrace{\begin{bmatrix} | & & | \\ \boldsymbol{\varphi}_1 & \cdots & \boldsymbol{\varphi}_r \\ | & & | \end{bmatrix}}_{\boldsymbol{\Phi}} \underbrace{\begin{bmatrix} b_1 & & \\ & \ddots & \\ & & b_r \end{bmatrix}}_{\mathrm{diag}(\boldsymbol{b})} \underbrace{\begin{bmatrix} e^{\omega_1 t_1} & \cdots & e^{\omega_1 t_{M-1}} \\ \vdots & \ddots & \vdots \\ e^{\omega_r t_1} & \cdots & e^{\omega_r t_{M-1}} \end{bmatrix}}_{\boldsymbol{T}(\boldsymbol{\omega})}$$

where $\mathrm{diag}(\boldsymbol{b})$ is a diagonal matrix of the mode amplitudes b_j and $\boldsymbol{T}(\boldsymbol{\omega})$ is a Vandermonde matrix. Note that the continuous-time eigenvalues are computed from the entire time series, and the index k in $\boldsymbol{\Lambda}^{k-1}$ serves only to indicate the amount of time elapsed (i.e. the eigenvalues are not recomputed every time step). The dynamics of each mode are separated, so that it is possible to isolate and examine a single spatiotemporal structure without the confounding effects of other modes. This will be particularly useful to characterize instability.

It is possible to obtain a better estimate of the mode amplitudes $\boldsymbol{b}$ with the following minimization problem:

$$\mathrm{argmin}_{\boldsymbol{b}} ||\boldsymbol{X} - \boldsymbol{\Phi}\,\mathrm{diag}(\boldsymbol{b})\,\boldsymbol{T}(\boldsymbol{\omega})||_F. \tag{4.7}$$

This formulation is the basis of the DMD extensions presented in the following sections.

4.2.1 Sparsity-promoting DMD

A central tension in reduced-order modeling is that including more modes often increases accuracy while reducing model interpretability. However, sparsity promotion through an addition L_1 penalty term has become a common technique for machine learning and data analysis [63] because it can produce sparse and interpretable models in terms of a few essential modes. Jovanovic et al. [199] introduced an L_1 penalty in the DMD optimization

$$\min_{\boldsymbol{b}} \left(||\boldsymbol{X} - \boldsymbol{\Phi}\, \mathrm{diag}(\boldsymbol{b})\, \boldsymbol{T}(\boldsymbol{\omega})||_F + \gamma_s ||\boldsymbol{b}||_1 \right) \tag{4.8}$$

to identify the key DMD modes. Here γ_s determines the level of sparsity. Following Jovanovic et al., this optimization problem was solved by writing it in a more convenient form which can be solved with the alternating direction method of multipliers [136]. This type of sparsity promotion is foundational to the methods in Chapter 5.

4.2.2 Optimized DMD

Depending on the scientific aims, the absence of a complete set of spatiotemporal DMD modes could be problematic. This lack of completeness implies that reconstructions of specific signals in the data matrix may be less accurate than the POD. This could be an issue if a very accurate fit of a subset of the data is desired, either for data-driven discovery or for control purposes. The optimized DMD of Askham and Kutz [12] addresses this issue by simultaneously considering the best-fit linear operator between all snapshots in time, as opposed to only considering sequential snapshots, as in the standard DMD. The optimized DMD results in excellent signal reconstructions, but at the cost of solving a potentially large, nonlinear optimization problem. The key to applying this method is a variable projection algorithm [12] that simplifies the nonlinear optimization. An additional benefit is that snapshots are no longer required to be evenly space in time. Defining $\boldsymbol{\Phi}_b = \boldsymbol{\Phi}\, \mathrm{diag}(\boldsymbol{b})$, the

nonlinear minimization problem is now

$$\min_{\boldsymbol{\omega},\boldsymbol{\Phi_b}} ||\boldsymbol{X} - \boldsymbol{\Phi_b}\boldsymbol{T}(\boldsymbol{\omega})||_F. \tag{4.9}$$

This problem is similar to the original DMD minimization, but the DMD fit is now optimized with respect to both $\boldsymbol{\omega}$ and $\boldsymbol{\Phi}_b$. This problem can be solved with a variable projection followed by the Levenberg-Marquardt algorithm [245, 278], which relies on a QR decomposition. For speed, the code implements a parallel QR decomposition called direct TSQR [37]. This algorithm is not guaranteed to find the global minima, but only a local one, so it benefits from an accurate initial guess. It is often useful to initialize using the results of the other DMD methods. Note that for real-valued data, the DMD methods give complex conjugate pairs. The implementation here breaks the complex conjugate symmetry, although it is possible to explicitly retain this symmetry [12]. This can be seen visually in Fig. 4.2a. However, reconstructions are built with the average of each mode and its complex conjugate, guaranteeing real-valued data.

There is often a trade-off between model interpretability and reconstruction accuracy. The sparsity-promoting DMD algorithm produces interpretable models, and the optimized DMD algorithm accurately reconstructs low-energy features and transient instabilities in the data. In this way, the strength of using a combination of DMD methods to understand a dynamic system is illustrated in the next sections. Throughout Chapter 4, blue, red, and green colors are used for the exact DMD, sparsity-promoting DMD, and optimized DMD, respectively.

4.3 Comparison of DMD algorithms on an experimental HIT-SI Discharge

The exact, sparsity-promoting, and optimized DMD variants are compared on real experimental HIT-SI data in order to understand their relative strengths and weaknesses in identifying interpretable and accurate reduced-order models. In the following, an analysis of diagnostic data from the high-performance HIT-SI discharge 129499 ($f_1^{\text{inj}} = 14.5$ kHz) is per-

formed. This discharge has been investigated extensively in previous studies [304, 435, 183].

The remaining analysis in this Chapter 4 will focus on the period of spheromak sustainment for the experimental and simulation data. Because DMD associates the spheromak and the injectors each with a single oscillation frequency, the modes are denoted f_0 ($f_0 \approx 0$) and f_1^{inj} throughout Chapter 4. The higher harmonics of the injector frequency are f_2^{inj}, f_3^{inj}, and so on.

4.3.1 DMD eigenvalues

The DMD eigenvalues determine the time evolution of the corresponding spatially coherent DMD modes. Figure 4.2a compares the eigenvalues for each of the three DMD methods; the eigenvalues are scaled by their amplitudes $|b_j|$. In each case, the SVD is truncated at $r = 20$ modes to avoid overfitting. The x-axis represents the imaginary component of the eigenvalue, and the y-axis represents the real component so that eigenvalues in the upper half plane are unstable and those in the lower half plane are stable.

The magnitude plot indicates that sparsity-promoting DMD is effective at isolating the three dominant modes, whereas exact and optimized DMD both result in spectra with many energetic modes. Although it appears that there are only two very energetic modes for optimized DMD, these modes decay extremely quickly, so that other modes become relatively more important. Thus, sparsity-promoting DMD is capable of extracting and isolating the leading large-scale magnetic structures in the experiment, providing enhanced interpretability. In contrast, the next section will show that optimized DMD is needed to extract and analyze small-scale transient modes for a more accurate fit. The large number of quickly decaying modes for all the methods indicates that despite the parsimonious $r = 20$ truncation, the dynamics can be well-fit with a lower-dimensional model. Another interpretation is that high-precision fits of the data require shorter time windows, suggesting a sliding-window DMD for forecasting and control.

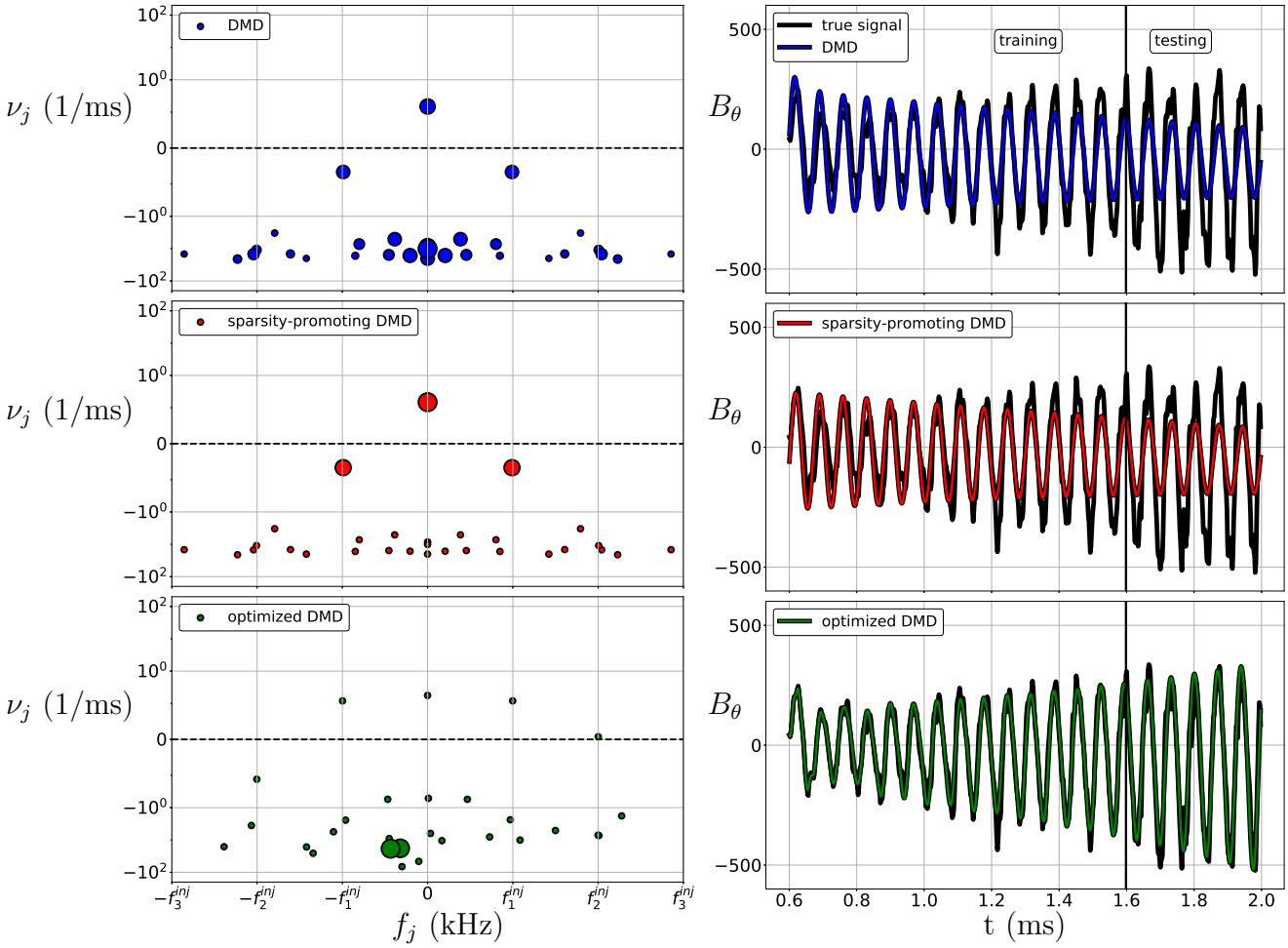
4.3.2 DMD reconstruction and forecasting

A common scientific aim is an accurate reconstruction of diagnostic signals using a subset of the modes from a reduced-order model. The advantage of the optimized DMD over the other algorithms is apparent from the reconstruction of a surface probe, as in Fig. 4.2b. The exact DMD and sparsity-promoting DMD capture the bulk evolution, but the optimized DMD also captures the deviations. The DMD methods are trained on a subset of the data and then evolved in time to forecast the remaining data. The optimized DMD provides the most accurate forecast. However, this model contains exponentially growing modes that will eventually diverge. Including more than $r = 20$ modes causes optimized DMD to overfit and results in more unstable modes. These observations are reaffirmed quantitatively on simulation data in the next section.

4.4 DMD analysis on BIG-HIT simulations

Physical understanding can be obtained from reduced-order models by extracting coherent structures and analyzing their spatiotemporal content. Here, we decompose the spatial dependence into Fourier modes because the spatial Fourier dependence is important in many experimental devices to determine MHD stability. For toroidal devices, such as HIT-SI, the toroidal and poloidal Fourier wavenumbers are denoted (n_ϕ, m_θ). In HIT-SI, when the safety factor q_s satisfies $q_s > 1$, it can be shown to be kink-unstable to the $(n_\phi, m_\theta) = (1, 1), (2, 2), (3, 3), ...$ modes [189]. Sawtooth oscillations from these resistive kink modes are also common in toroidal devices when $\min(q) < 1 < \max(q)$ [439, 116].

Here, the DMD methods described in Section III are quantitatively compared based upon their ability to characterize the simulated large-size version of HIT-SI, named BIG-HIT. These 3D simulations were performed using a Hall-MHD model, assuming constant and uniform temperature and density. Relevant constants include the plasma temperature $T_i = T_e = 71$ eV, density $n = 1.5 \times 10^{19}$ m^{-3}, resistivity $\eta = 8.9 \times 10^{-7}$ Ω-m, and injector frequency $f_1^{\text{inj}} = 14.5$ kHz. For more parameter details, see the original analysis and prior



(a) DMD eigenvalues weighted by amplitudes. (b) Reconstruction and forecasting on a signal.

Figure 4.2: Summary of the DMD analysis for the experimental discharge 129499. (a) The DMD eigenvalues plotted in the complex plane, $\nu_j = \mathrm{Re}(\omega_j)/2\pi$ and $f_j = \mathrm{Im}(\omega_j)/2\pi$, for the experimental shot at $f_1^{\mathrm{inj}} = 14.5$ kHz, weighted by $|b_j|$ until some minimum dot size; there are $r = 20$ modes. Modes above the dashed horizontal line are unstable. (b) The reconstruction and forecasting performance of each DMD method. The vertical black line indicates where forecasting begins. While optimized DMD provides the most accuracy, it also produces several growing modes that will eventually diverge.

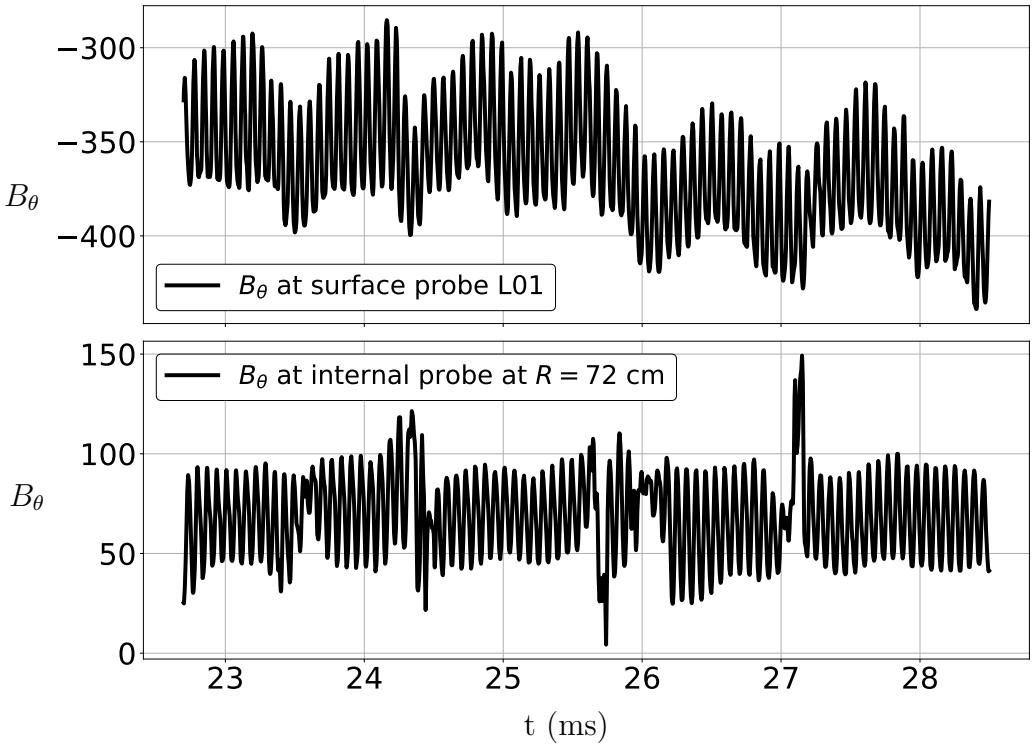


Figure 4.3: Raw BIG-HIT simulation data in the window 22.7 ms $\leq t \leq$ 28.5 ms indicates at least three large magnetic rearrangements captured poorly (well) by the synthetic surface (internal) probe illustrated here.

implementations of the model [5, 305].

To demonstrate the ability of these methods to work on small subsets of data, only simulation data in the range 22.7 ms $\leq t \leq$ 28.5 ms is used. Representative surface and internal probe B_θ time evolutions are depicted in Fig. 4.3 for this time range. The performance on a sparse and spatially well-separated dataset of 24 internal probes is compared with a large and uniformly spaced dataset of 5120 internal probes, in order to illustrate that the conclusions of this analysis on high-resolution data hold in the limit of low spatial resolution.

4.4.1 DMD reconstruction error

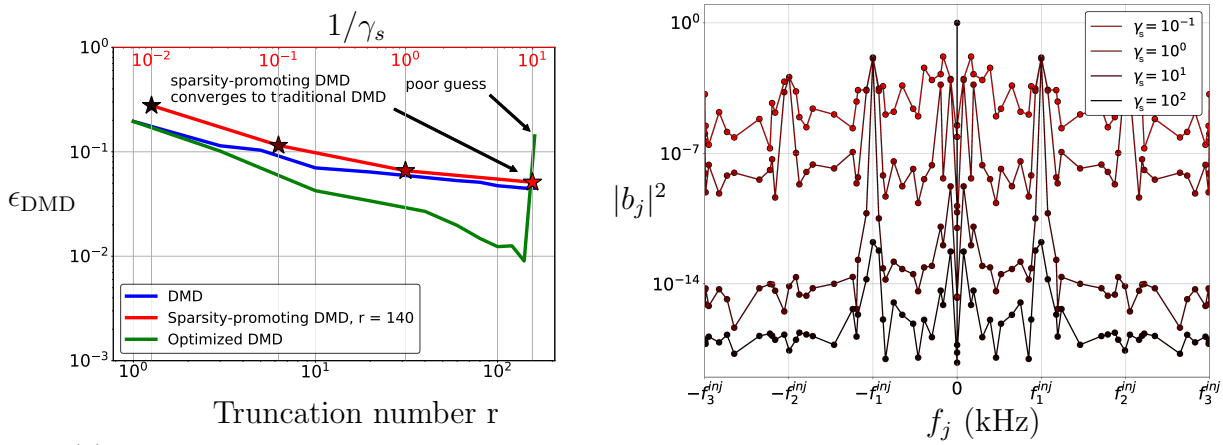
Each extension of the dynamic mode decomposition has its particular strengths and weaknesses. In the previous section, the sparsity-promoting DMD resulted in interpretable mod-

els, while the optimized DMD provided excellent reconstructions of the experimental data. Now using BIG-HIT simulation data in the shortened window 22.7 ms $\leq t \leq$ 23.5 ms, Figure 4.4 provides a quantitative comparison of the different DMD algorithms. For the exact and optimized methods, the relative reconstruction error is plotted as a function of the SVD truncation rank. The relative reconstruction error is defined as

$$\epsilon_{\text{DMD}} = \frac{||\boldsymbol{X} - \boldsymbol{\Phi}\,\text{diag}(\boldsymbol{b})\,\boldsymbol{T}(\boldsymbol{\omega})||_F}{||\boldsymbol{X}||_F}. \tag{4.10}$$

The sparsity-promoting scan is performed with fixed $r = 140$ while γ_s is varied. The factor $1/\gamma_s$ is labeled on the top axis in Fig. 4.4a (and colored red to indicate it applies only to the sparsity-promoting case) so that it can seen clearly that as $\gamma_s \to 0$, the sparsity-promoting results converge to exact DMD. The exact and optimized DMD methods use only the bottom axis, since the truncation number r is being varied. As γ_s increases, the reconstruction error of the sparsity-promoting DMD model also increases, as one would expect for a more parsimonious model.

At $r \approx 140$, the optimized DMD reconstruction error is an order of magnitude smaller than the exact DMD error ($\epsilon_{\text{DMD}} \approx 0.009$ against $\epsilon_{\text{DMD}} \approx 0.05$). In fact, optimized DMD with $r = 10$ obtains the same reconstruction error as exact DMD with $r = 140$. However, at $r = 160$, the optimized DMD error increases significantly. In this case, the initialization procedure chooses a poor first guess consisting of exponentially growing modes, which results in optimized DMD converging to a suboptimal minimum. Spurious unstable modes is a general issue with a number of DMD algorithms. These issues are often mitigated with a suitable rank truncation in the SVD (either manually or through sparsity promotion), a different window, or a more accurate initialization procedure. There are many DMD algorithm extensions that can provide additional improvements [235].



(a) Reconstruction errors for each method.

(b) Power spectra of sparsity-promoting DMD

Figure 4.4: Quantitative DMD analysis for BIG-HIT simulation data in the time window defined by 22.7 ms $\leq t \leq$ 23.5 ms. In (a) the optimized DMD obtains the most accurate reconstructions, and sparsity-promoting DMD is shown to converge to exact DMD as $\gamma_s \to 0$. The top (red) axis applies only to the sparsity-promoting DMD, while the bottom axis applies to the other methods. Stars correspond to the sparsity-promoting DMD power spectra in (b). For large r, a bad initial guess results in optimized DMD converging to local minimum with poor reconstruction. In (b) the normalized power spectrum of sparsity-promoting DMD illustrates overall suppression of DMD modes and fewer large peaks as γ_s increases.

4.4.2 DMD mode characterization

Physical insights into the dominant low-dimensional structures have the potential for improved understanding or control. In this section, sparsity-promoting DMD is used to characterize the dominant modes observed in BIG-HIT. Here, an analysis is presented of the B_θ measurements from the synthetic surface probes, as well as the B_z measurements from the synthetic internal probes at the $Z = 0$ midplane. Note that at $Z = 0$, B_z is either parallel or anti-parallel to B_θ and the sign choice does not affect the toroidal decomposition of this field (also, since all the internal probes are at $Z = 0$, the internal probes cannot be poloidally decomposed). The reconstructions at $Z = 0$ for the f_0–f_3^{inj} and f_{kink} (described later in Sec. 4.4.5) modes are illustrated in Fig. 4.5.

To analyze the spatial structure of each mode, reconstructions of the probe signals are created using only the relevant subset of DMD modes. These reconstructions map the probe locations to the proper location in (R, θ, ϕ) space. A Fourier decomposition in the toroidal direction is performed separately for each set of probes with the same radial location. The toroidal decomposition for the internal probes is

$$B_z(R, \phi, t) \approx \sum_{n_\phi=0}^{N_{\max}} \tilde{B}^z_{n_\phi}(R, t) \cos(n_\phi \phi - \zeta_{n_\phi}) \tag{4.11}$$

and the toroidal and poloidal decompositions for the surface probes are,

$$B_\theta(R, \theta, \phi, t) \approx \sum_{n_\phi=0}^{N_{\max}} \tilde{B}^\theta_{n_\phi}(R, \theta, t) \cos(n_\phi \phi - \zeta^{\text{tor}}_{n_\phi}), \tag{4.12a}$$

$$B_\theta(R, \theta, \phi, t) \approx \sum_{m_\theta=0}^{M_{max}} \tilde{B}^\theta_{m_\theta}(R, \phi, t) \cos(m_\theta \theta - \zeta^{\text{pol}}_{m_\theta}). \tag{4.12b}$$

To resolve the first $N_{\max}$ modes, $2N_{\max} + 1$ unique toroidal locations are required, and similarly for the poloidal direction. For evenly spaced measurements the coefficients can be found directly by orthogonality. A general method of obtaining the coefficients of these decompo-

sitions for irregularly spaced angular measurements can be found in the reference on HIT-SI surface probes [448]. Since the internal probes coefficients $\tilde{B}^z_{n_\phi}(R,t)$ are still too unwieldy to present clear results, especially for the set of 5120 internal probes, instead the absolute value of the $\tilde{B}^z_{n_\phi}(R,t)$ coefficients is used and averaged over the Fourier transforms obtained from different radial locations. This procedure results in

$$\langle \tilde{B}^z_{n_\phi}(t) \rangle = \frac{1}{N_{\text{rad}}} \sum_{i=1}^{N_{\text{rad}}} |\tilde{B}^z_{n_\phi}(R_i,t)| \tag{4.13}$$

where N_{rad} is the number of radial locations where a separate Fourier decomposition is performed. This quantity gives an average sense of the total toroidal dependence of the reconstructed B_z. For the poloidal Fourier decompositions, the four poloidal arrays of surface probes are separately decomposed and then similarly averaged (dropping the radial dependence since the probes in each array have identical radial locations)

$$\langle \tilde{B}^\theta_{m_\theta}(t) \rangle = \frac{1}{4} \sum_{i=1}^{4} |\tilde{B}^\theta_{m_\theta}(\phi_i,t)|. \tag{4.14}$$

Now that metrics are defined for quantifying the spatial Fourier dependence of the modes in the poloidal and toroidal directions, the spatial structures of the primary dynamics are investigated.

4.4.3 Sparsity-promoting DMD: First injector harmonic

The HIT-SI injectors drive large magnetic perturbations that sustain the spheromak. They are intentionally operated with an approximate $n_\phi = 1$ symmetry, but a full picture of the injector field structure is important for understanding the current drive and sustainment in this device. Reconstructions of the magnetic fields with only the injector mode reveal an overwhelming $n_\phi = 1$ dependence. There is also a phase shift of approximately 180^o between the core and edge region of the plasma, shown in Fig. 4.5, consistent with ion doppler spectroscopy measurements on the experiment [181]. A simulated version of this discharge

shows similar results [304]. This suggests a large-scale transition between the inner and outer regions of the plasma. This phase shift occurs at $R \approx 0.8 - 0.9$ m, close to the closed flux surfaces in the plasma.

4.4.4 Sparsity-promoting DMD: Second injector harmonics

Sub-harmonic, harmonic, or nearly-harmonic oscillations are a common feature observed in the nonlinear response to periodic inputs [215]. Modes oscillating at the harmonics of the injector frequency are often identified by the DMD algorithms. Surprisingly, the DMD mode corresponding to the second harmonic depends mostly on the even toroidal numbers with dominant $n_\phi = 2$. Moreover, Fig. 4.5 shows that there is a phase shift of approximately 180^o at $R \approx 1.05$ m and a smaller shift at $R \approx 0.3 - 0.4$ m. One possible interpretation is that the toroidally even part of the perturbation is filtered out where there is closed flux.

To investigate whether or not this mode corresponds to a physical structure, the BIG-HIT simulation $B^z_{n_\phi=2}$ is directly analyzed. A rich, previously unobserved, three-dimensional structure is discovered in the simulation, shown in Fig. 4.6. This structure wraps around the outside of the device by looping through the different injector mouths, and spirals down the core of the device, rotating at $f \approx f_2^{\text{inj}}$. A comparison of this structure at $Z = 0$ with the DMD reconstruction using only the f_2^{inj} mode shows surprising agreement. The reconstruction is able to capture much of the structure observed in the simulation, including the two phase shifts mentioned earlier.

This mode is also present in HIT-SI simulations that evolve the full Hall-MHD equations, as opposed to BIG-HIT, which does not evolve temperature and density. These simulations have carefully chosen parameters to match the experiment, and are thus expected to be the most representative of the HIT-SI experiment. Taken along with the results from Sec. 2.3.4, this is the first identification of a physical and coherent 3D structure in HIT-SI simulation or experiment beyond the dominant injector modes and the spheromak.

The f_3^{inj} mode exhibits mostly odd toroidal mode number dependence and a dominant $n_\phi = 3$ dependence at $R \approx 0.1$. However, this mode only accounts for approximately 1% of the

total B_z energy and has similar Fourier dependence as the f_1^{inj} mode, making it exceedingly difficult to verify if this mode is contained in $B^z_{n_\phi=3}$ in the simulation, as was done for the f_2^{inj} mode. This difficulty should come as no surprise because DMD decomposes the magnetic field into different oscillating and rotating structures that may have a complicated Fourier structure. Decomposing the magnetic field into Fourier components may obfuscate the coherent structures. This is in fact one of the primary motivations for reduced order models constructed from alternative bases.

The f_{kink} mode in Fig. 4.5 refers to the mode obtained by the optimized DMD, and is discussed in the next section. A summary of the toroidal and poloidal dependence for the f_0, f_1^{inj}, f_2^{inj}, and f_3^{inj} modes, as well as the f_{kink} mode analyzed in the next section, can be found in Fig. 4.8 for both datasets. All of the modes exhibit a broad poloidal spectrum that is expected for this device; the surface probes are arranged on the bowtie-shaped boundary of the HIT-SI device, and subsequently tend to observe a plasma with many poloidal wavenumbers.

4.4.5 Optimized DMD: Kink instability

Linear MHD stability is of considerable importance in the plasma physics community, especially for confinement devices. While the interpretable models of the previous section allowed for the identification of large-scale physical structures while avoiding overfitting, the optimized DMD is useful for accurate modeling of transient instabilities over smaller time windows.

In Fig. 4.7, Poincaré plots from BIG-HIT show closed flux surfaces that exhibit an $n_\phi = 1$ structure and quasi-periodic sawtooth activity from a $(1, 1)$ kink instability, which is consistent with the Kadomtsev or Wesson models [439]. The linear growth rate of the resistive $(n_\phi, m_\theta) = (1, 1)$ kink is [439]

$$\nu_{11} = \frac{1}{2\pi}\left(\frac{\eta q'(R_1)B_\theta(R_1)}{\mu_0^2 R_1^2 \rho}\right)^{\frac{1}{3}}. \tag{4.15}$$

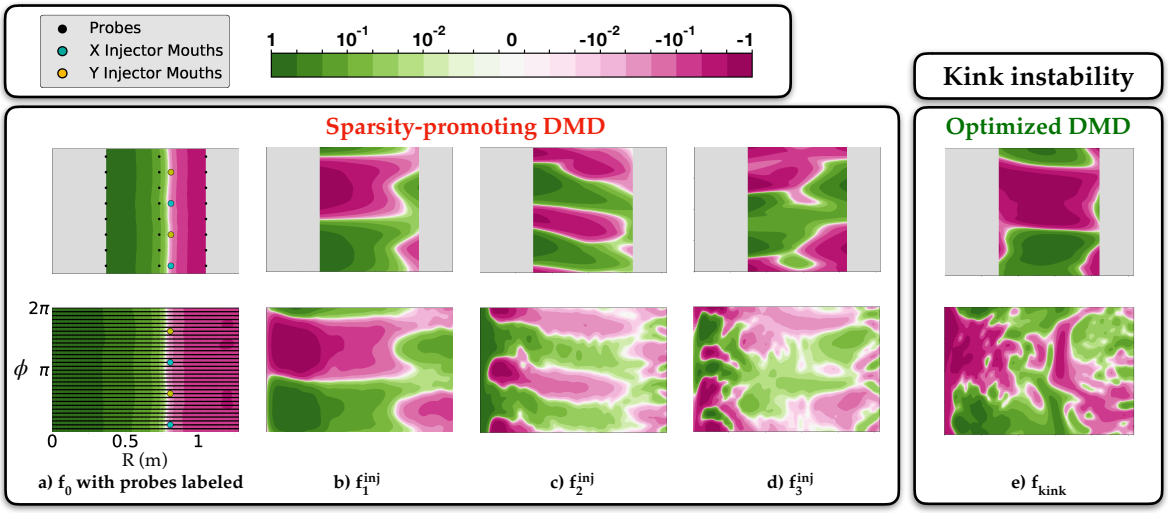


Figure 4.5: B_z at $Z = 0$ of the sparsity-promoting DMD modes $f_0, ..., f_3^{\text{inj}}$ and optimized DMD mode f_{kink}, with each mode separately normalized by its maximum absolute value. The small dataset illustrated in the top row has resolution $\Delta R \approx 37$ cm, $\Delta\phi = 45^o$. In the bottom row, $\Delta R \approx 0.8$ cm, $\Delta\phi = 11.25^o$. The sparsity-promoting method captures the vast majority of the spatial structure for each mode even with the small dataset. Fine-scale structure in the kink instability is not captured with the small dataset. The relative mode amplitudes, rather than the amplitudes normalized by their maxima, can be found in Fig. 4.8.

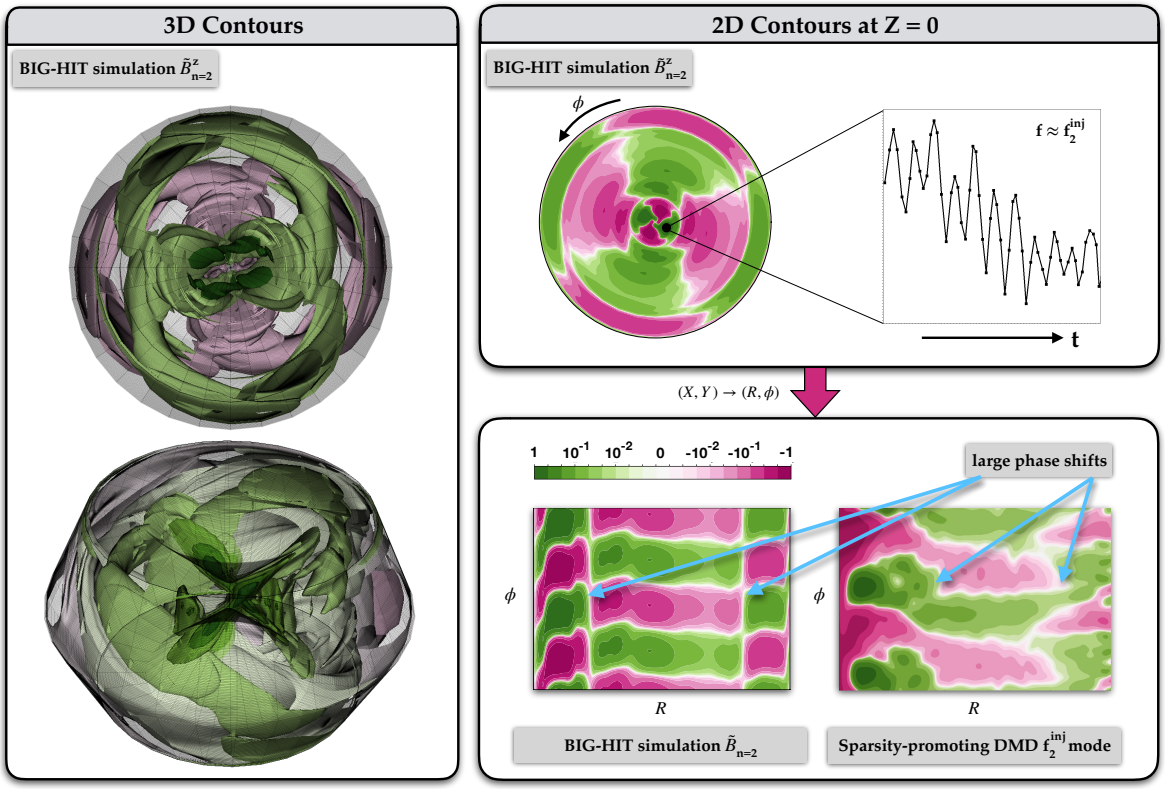


Figure 4.6: 3D snapshots of $\tilde{B}^z_{n_\phi=2}$ from BIG-HIT simulations illustrate a dynamic spiral structure that penetrates down the core and edge of the device, and connects through the injector mouths. Corresponding 2D contour plots at $Z = 0$ of the same simulation data indicate that much of this structure is oscillating roughly at f_2^{inj}, thereby connecting the f_2^{inj} mode found in the DMD analysis in Chapter 4 with a physical structure with the correct $n_\phi = 2$ dependence. Lastly, a comparison at $Z = 0$ between the simulation data $\tilde{B}^z_{n_\phi=2}$ and the sparsity-promoting DMD reconstruction using f_2^{inj} shows that much of the fine-scale structure can be captured correctly by sparsity-promoting DMD.

In the equation above, $R_1 \approx 0.9$ m is the radius which satisfies $q(R_1) \approx 1$, and rough estimates from the previous BIG-HIT analysis yield $q'(R_1) \approx \Delta q/\Delta R \approx 0.05/0.05 = 1$ and $B_\theta(R_1) \approx$ 100 G. Evaluating with these values gives $\nu_{11} \approx 1000\ s^{-1}$.

Optimized DMD captures an $(n_\phi, m_\theta) = (1, 1)$ instability and obtains its growth rate for both datasets in the window 26.8 ms $\leq t \leq$ 27.1 ms. This phenomenon is robust for a range of SVD truncation ranks from $10 \leq r \leq 50$; when $r < 10$, the fit does not capture the exponential growth, and when $r > 50$ the poor initial guess results in convergence to a sub-optimal minimum. For values of $10 \leq r \leq 50$, $\nu_{\text{kink}} \approx 600 - 1100\ s^{-1}$ using 5120 internal measurements, and $\nu_{\text{kink}} \approx 600 - 2000\ s^{-1}$ using 24 measurements, in excellent agreement with the estimate of $\nu_{11} \approx 1000\ s^{-1}$.

To account for some models resulting in many growing modes, only modes with $\nu_j > 100$ s^{-1} are retained; the major spheromak or injector modes are often below this threshold. Any modes oscillating within 1 kHz of the injector frequency are also rejected, in an attempt to control for modes directly driven by the injectors, which are known to have a $n_\phi = 1$ structure. The growth rate reported is the weighted average of the remaining growing modes. To validate this approach, the other two kink events observed in the full window 22.7 ms $\leq t \leq$ 28.5 ms are analyzed. Again, the results indicate growth rates $\nu_{\text{kink}} \approx 500 - 2000\ s^{-1}$ and similar spatial dependence.

The toroidal and poloidal Fourier decompositions of the modes, reported in Fig. 4.8, indicate an $(n_\phi, m_\theta) = (1, 1)$ structure, and the contour plots in Fig. 4.5 illustrate dominant $n_\phi = 1$ dependence in the closed flux region. The surface probes indicate an $m_\theta = 1$ structure of the instability in Fig. 4.8, despite the broad spectrum. With the low-resolution dataset, the surface probes exhibit dominant $n_\phi = 4$ dependence. Many of the surface probe signals have barely perceptible changes during the transient instability, and thus it is reasonable that the spatial dependence of the instability cannot be consistently identified with these probes. The surface probe decomposition exhibits a dominant $n_\phi = 1$ structure for the high-resolution dataset, which may be consistent with additional probes resulting in a better representation of the magnetic field dependence of the instability.

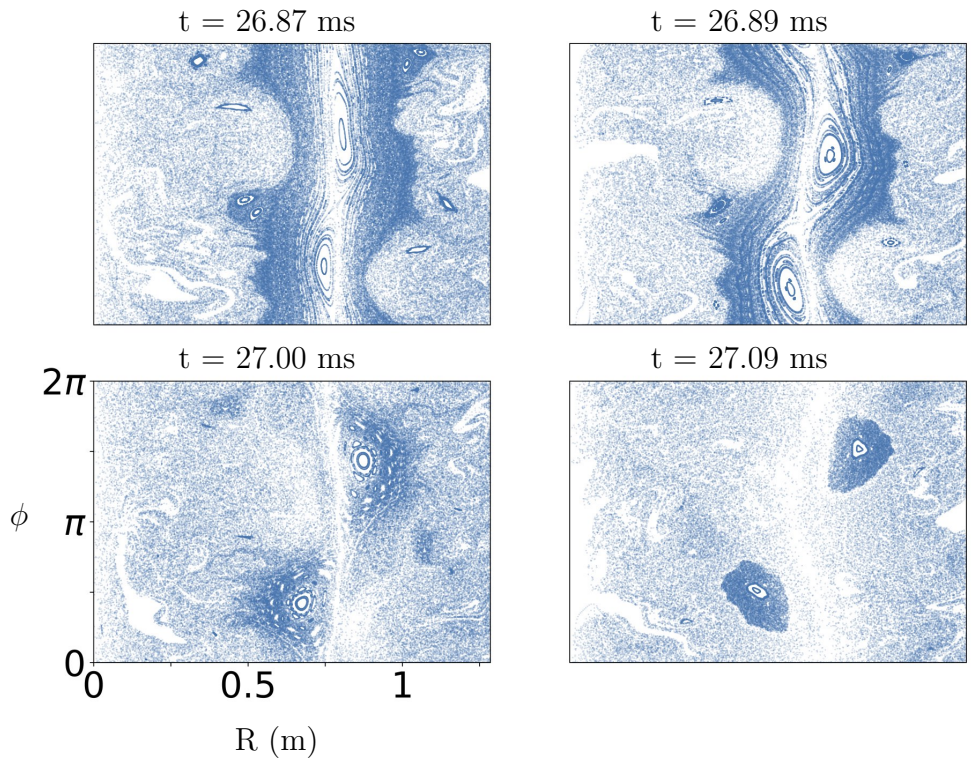


Figure 4.7: Poincaré plots, generated using the full simulation data on the midplane at four consecutive instances, illustrate the evolution of the resistive kink instability.

Originally, it appeared possible that the injectors directly drive the $n_\phi = 1$ kinking, but this analysis suggests that plasma generated activity is responsible. Another simulation using a set of four injectors (two each on top and bottom) driving a primarily $n_\phi = 2$ magnetic structure, indicates periods of closed flux followed by an opening of the flux surfaces by a $n_\phi = 1$ kink. This lends further evidence to identification of this kink mode independent of the primary injector magnetic configuration.

4.5 Conclusions from DMD work

The sparsity-promoting and optimized variants of the dynamic mode decomposition have been shown to enable the discovery of magnetic structures from a sparse set of measurements of a driven spheromak. Spatio-temporal modes corresponding to the injector harmonics were identified, along with the characterization of a resistive $(1, 1)$ kink instability. Further, the evolution of these modes is accurately captured by a low-rank, interpretable, and linear model, demonstrating the potential for forecasting and real-time control. Importantly, the effectiveness of DMD was demonstrated on data from both the HIT-SI experiment and accompanying BIG-HIT simulations.

The sparsity-promoting DMD is shown to provide an interpretable and physical model of the major magnetic modes, while avoiding overfitting. The f_1^{inj} structure corresponds to the dominant part of the driven injector fields. The f_2^{inj} mode on the midplane was used to uncover a previously unobserved 3D structure in the simulation that oscillates at f_2^{inj}, has $n_\phi = 2$ toroidal Fourier dependence, and spirals through the injectors near the boundary of the device. A very similar-looking mode was also observed during the nonlinear relaxation phase of the HIT-SI simulations in Sec. 2.3.4.

The optimized DMD demonstrates more accurate signal reconstruction that may be useful for forecasting and characterizing smaller-scale coherent structures. Unlike the other methods, the optimized DMD enables the full characterization of a resistive kink instability on a small time window, indicating the ability for robust data-driven identification of MHD instabilities, with implications for real-time control. If both the discovery of interpretable

dynamics and the accurate characterization of instabilities is desired, a joint use of both the sparsity-promoting and optimized DMD algorithms may be useful, as illustrated here.

While this chapter has focused on magnetic measurements from a number of simple probes, in principle, these methods only rely on a set of sparse experimental measurements of any relevant plasma quantity. Thus, they should be highly applicable to a wealth of different diagnostics and plasmas spanning much of the possible parameter space. Although all reduced-order methods discussed here result in global spatial modes, the analysis can be restricted to small-scale spatial structures by using a small number of nearby probes. Despite the localization of the resistive kink instability near the closed flux, it was successfully identified with these methods by radially averaging over the internal probe arrays, and visually confirming the $n_\phi = 1$ structure in the closed flux region. To capture transient modes such as the resistive kink identified in this paper, the methods can be applied on a small time window. This flexibility and generality make DMD an excellent choice for the discovery of coherent plasma structures and instabilities, and subsequent attempts to control them. For similar reasons, SSI is also increasingly used for modal analysis in the plasma physics community.

There are also a number of existing techniques that are specific to a certain class of dynamics arising in plasmas. For instance, bispectral analysis [218, 369] has traditionally been used for the identification and analysis of nonlinear wave interactions. For plasmas with important small-scale and transient turbulent structures, the (bi)orthogonal wavelet decomposition and reduced-order methods based on wavelets [121], such as multi-resolution DMD [236, 276] and multi-resolution biorthogonal decomposition [286], show significant promise. These methods would be ideal for the analysis of coherent structures with small spatial and temporal correlation lengths. The DMD framework has several other extensions that may improve the analysis of complex experimental plasmas, which often have limited measurements, complex multi-scale dynamics, and actuation. Other promising models for mapping nonlinear dynamics onto approximate linear representations are the many variants for approximating the Koopman operator [61, 62]. Discovering the underlying coherent structures also facilitates

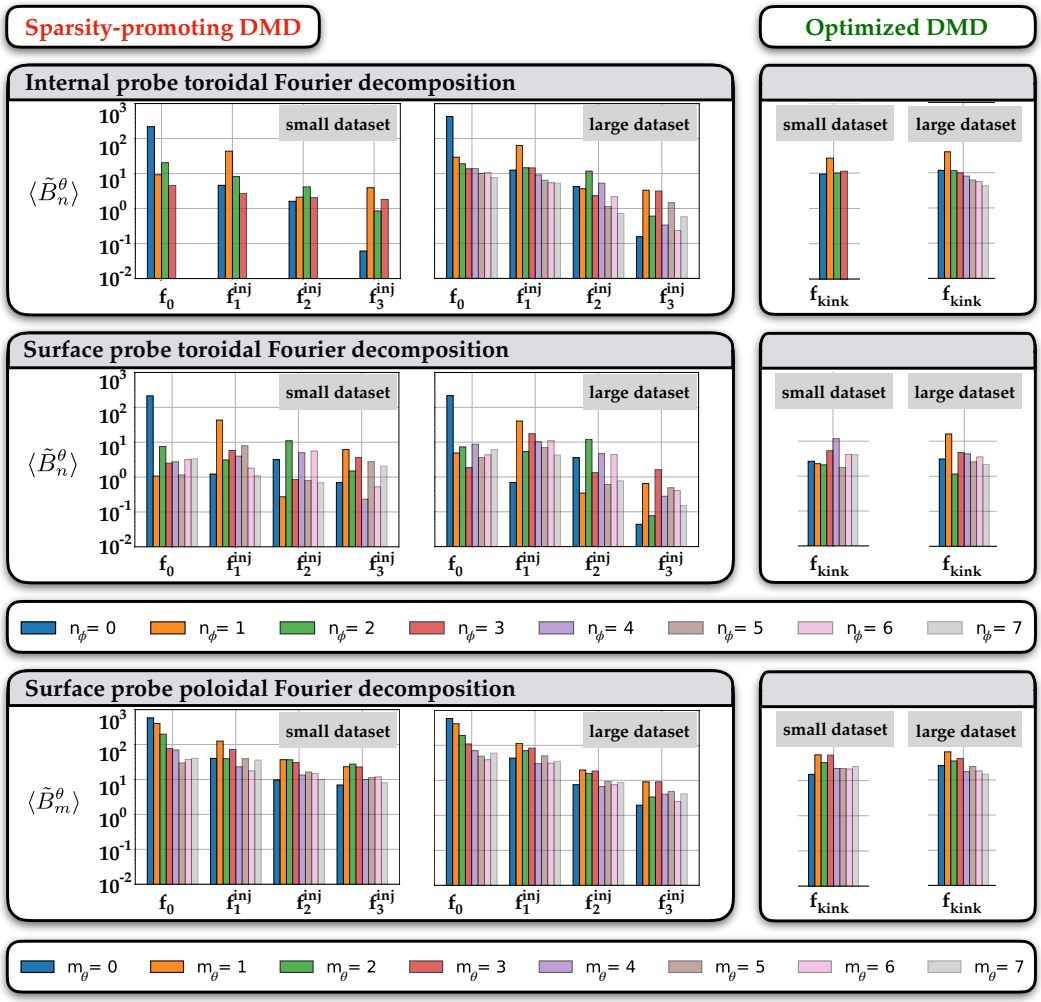


Figure 4.8: Averaged toroidal mode content $\langle \tilde{B}^{\theta}_{n_\phi} \rangle$ of the surface and internal probes for the f_0, f_1^{inj}, f_2^{inj}, f_3^{inj}, and f_{kink} DMD modes is shown for the two datasets (recall that both datasets have 192 surface probes but the small dataset has only 24 internal probes, while the large dataset has 5120 internal probes). The surface probe decomposition in the toroidal direction is in excellent agreement with the internal probe data except for the low-resolution kink mode. This illustrates the global structure of f_0, f_1^{inj}, f_2^{inj}, f_3^{inj} and the local spatial structure of the instability near the closed flux region. The f_0 mode is almost purely characterized by $n_\phi = 0$, reaffirming the physical interpretation of an axisymmetric spheromak. Modes f_1^{inj} and f_3^{inj} indicate dominant odd n_ϕ structure, which is expected because the injectors are driven to produce a mostly $n_\phi = 1$ magnetic structure. Interestingly, mode f_2^{inj} shows significant mode content in the even n_ϕ numbers. The instability f_{kink} is observed in both datasets to have a $(1, 1)$ structure. Magnitude disagreements between the datasets are a result of differences in the number and locations of the probes.

control [58]. The DMD algorithm has been previously extended to decompose signals while disambiguating the system dynamics from the effects of external forcing. This is ideal for discovering and then controlling plasma-generated dynamics. In fact, DMD with control (DMDc) has shown significant promise for externally forced systems [353, 310, 140]. For the purposes of this work, it was found that performance was similar when injector current waveforms were treated as actuators because the forcing is actually the primary dynamics in the system after the equilibrium is subtracted off. However, for other experimental devices, accounting for external actuation through DMDc or SSI may significantly improve discovery of plasma dynamics.

For reduced-order models, another clear direction for improvements over the traditional DMD is to extract *nonlinear* ROMs from data. In the next chapter, a popular system identification method is discussed that will allow for the identification of nonlinear ROMs with physical constraints, constraints from nonlinear stability theory, and other robustifying features.

Chapter 5

SPARSE SYSTEM IDENTIFICATION FOR PLASMAS AND FLUIDS

Increasingly, reduced-order models of complex systems, such as fluids and plasmas, are discovered from data with modern machine learning algorithms [393, 371, 295, 202, 59, 222, 376, 100, 257, 425, 108, 360, 344, 25, 114, 356, 8, 339, 354, 211, 251, 243, 172, 381, 226, 212], rather than classical projection-based methods that are intrusive and may require substantial knowledge of the governing equations. These data-driven approaches for modeling fluid dynamics [53, 65] range from generalized regression techniques [393, 59, 257] to deep learning [114, 243, 251, 381, 226, 361].

As illustrated in Chapter 3, one can project the high-dimensional MHD (or kinetic equations [318]) onto a low-dimensional basis in order to obtain an analytic ROM, although the high-dimensional simulations may still be mostly required for computing the coefficients appearing in the ROM. Chapter 4 showed that linear *data-driven* models can be extracted directly from data without assumption about the underlying governing equations. This can be quite useful when the underlying equations are unknown. The linear approximation makes this approach sub-optimal for much of nonlinear plasma dynamics, although the errors in this approximation can be somewhat addressed through local linearization, very high diagnostic sampling rates, sliding windows, and other more robust variants of DMD.

This chapter combines these two ideas, namely that: (1) analytic POD-Galerkin-type models can be computed from MHD and (2) it is often useful to find *nonlinear* models directly from data. First, the extraction of nonlinear governing equations from data is reviewed. Second, the focus is shifted from directly extracting the MHD or kinetic equations from the data (although this is an interesting line of work [8], especially for data-driven

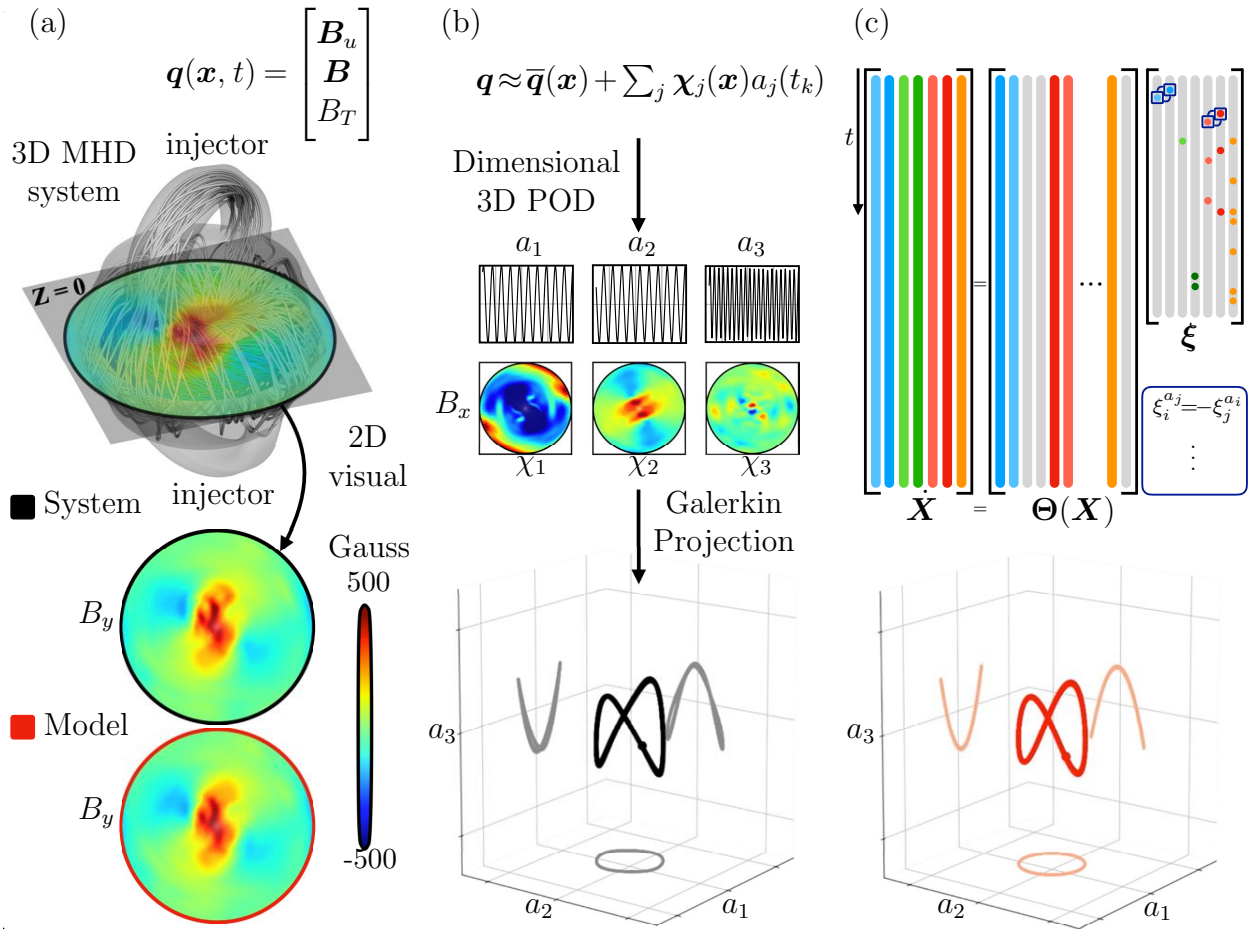


Figure 5.1: Proposed approach for filling in lower rungs of the plasma model hierarchy: (a) Collect data, (b) perform projection-based model reduction (c) discover data-driven models using physics-constrained system identification.

closure models), but rather extracting data-driven ROMs that approximate the MHD or kinetic equations. In other words, these nonlinear system identification methods will be used to directly discover POD-Galerkin models of the form in Eq. (3.12). The utility is that one can obtain very low-dimensional ROMs directly from the data and need not rely on high-resolution experimental or simulation datasets, which may not be available or may not exhibit the requisite resolution for calculating analytic ROMs.

5.0.1 Sparse Identification of nonlinear dynamics

The sparse identification of nonlinear dynamics (SINDy) method is a system identification technique for extracting governing equations from data. Empirically, dynamical systems are often well-described by ODEs or PDEs that are *sparse*, i.e. containing only a few active terms that dominate the dynamical evolution. The SINDy method discovers governing dynamical systems models of the form

$$\frac{d}{dt}\boldsymbol{q}(t) = \boldsymbol{f}(\boldsymbol{q}(t)). \tag{5.1}$$

Given data in the form of state measurements $\boldsymbol{q}(t) \in \mathbb{R}^N$, SINDy identifies a model for the dynamics, given by the function $\boldsymbol{f}$, which describes how the state of the system evolves in time. In particular, SINDy sparsely approximates the dynamics in a library of candidate basis functions $\boldsymbol{\theta}(\boldsymbol{q}) = [\theta_1(\boldsymbol{q}), \theta_2(\boldsymbol{q}), \ldots, \theta_{p_\Theta}(\boldsymbol{q})]$, so that

$$\boldsymbol{f}(\boldsymbol{q}) \approx \sum_{k=1}^{p_\Theta} \theta_k(\boldsymbol{q})\xi_k. \tag{5.2}$$

In order for this strategy to be successful, a reasonable approximation of $\boldsymbol{f}(\boldsymbol{q})$ must lie in the span of $\boldsymbol{\theta}$. Therefore, background scientific knowledge about expected terms in $\boldsymbol{f}(\boldsymbol{q})$ can be used to choose the library $\boldsymbol{\theta}$. To pose SINDy as a regression problem, time-series measurements of $\boldsymbol{q}$ and the corresponding time derivatives $\dot{\boldsymbol{q}}$ are arranged into matrices

$$\boldsymbol{X} = \begin{pmatrix} q_1(t_1) & q_2(t_1) & \cdots & q_N(t_1) \\ q_1(t_2) & q_2(t_2) & \cdots & q_N(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ q_1(t_M) & q_2(t_M) & \cdots & q_N(t_M) \end{pmatrix}, \quad \dot{\boldsymbol{X}} = \begin{pmatrix} \dot{q}_1(t_1) & \dot{q}_2(t_1) & \cdots & \dot{q}_N(t_1) \\ \dot{q}_1(t_2) & \dot{q}_2(t_2) & \cdots & \dot{q}_N(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ \dot{q}_1(t_M) & \dot{q}_2(t_M) & \cdots & \dot{q}_N(t_M) \end{pmatrix}.$$

Note that $\boldsymbol{X}$ has been redefined as the transpose of the original data matrix in Eq. (3.3). If $\boldsymbol{X}$ contains measurements with differing units, normalization or dimensionalization to energy units [373, 211] can improve the following sparse regression performance.

The derivatives $\dot{q}_i(t)$ can be approximated numerically or measured directly. The library functions are evaluated on the data, resulting in a candidate library of p_Θ terms, $\boldsymbol{\Theta}(\boldsymbol{X}) = [\theta_1(\boldsymbol{X}), \theta_2(\boldsymbol{X}), \ldots, \theta_{p_\Theta}(\boldsymbol{X})] \in \mathbb{R}^{M \times p_\Theta}$. Sparse regression is then performed to approximately solve the minimization problem

$$\underset{\boldsymbol{\Xi}}{\operatorname{argmin}} \left[\frac{1}{2} \|\dot{\boldsymbol{X}} - \boldsymbol{\Theta}(\boldsymbol{X})\boldsymbol{\Xi}\|^2 + \lambda R_s(\boldsymbol{\Xi}) \right], \tag{5.3}$$

where $\boldsymbol{\Xi} \in \mathbb{R}^{p_\Theta \times N}$ is a set of coefficients that determines the active terms in $\boldsymbol{f}$ and $R_s(\boldsymbol{\Xi})$ is a regularizer that biases the optimization towards sparse models. Although the traditional method uses $\boldsymbol{X}$ and $\dot{\boldsymbol{X}}$, in principle $\dot{\boldsymbol{X}}$ can be replaced by any matrix. For instance, for the wave equation one can use $\ddot{\boldsymbol{X}}$ and for steady-state problems one might use a polynomial term such as $\boldsymbol{X}^2$. The goal is to determine the coefficients that solve the optimization problem, $\boldsymbol{\Xi} = \left[\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \cdots, \boldsymbol{\xi}_{p_\Theta}\right]$, also written in vectorized form as $\boldsymbol{\Xi}[:] = \boldsymbol{\xi}$.

SINDy has been widely applied for model identification in applications such as chemical reaction dynamics [175], chemical networks [306], chemical processes [40, 387], biological transport [237], disease transmission [195], convective heat transfer [459], nonlinear optics [404], power systems [408], traveling waves [287], materials science [64], hydraulics [311], human behavior models [99], fluid dynamics [257, 259, 256, 117, 83, 108, 130], turbulence modeling [392, 29, 30], plasma physics [100, 211, 8], structural modeling [238], coupled aeronautic-structural systems [250], among others [311, 105, 337]. It has also been extended to handle more complex modeling scenarios such as PDEs [384, 376], stochastic differential equations [221, 55, 71], systems with inputs or control [204, 203, 122], systems with implicit dynamics [273, 201], hybrid systems [272, 422], to enforce physical constraints [257, 82, 211], to incorporate information theory [274] or global stability [209], to identify models from corrupt or limited data [426, 386, 107] and ensembles of initial conditions [449], to perform cross-validation with ensemble methods, and extending the formulation to include integral terms [385, 365, 291], tensor representations [135, 137], and stochastic forcing [48].

To use this method in conjunction with the methods discussed in Sec. 3, consider discov-

ering the POD-Galerkin models (systems of ODEs) for fluid and plasma models with this method. In that case, attention can be primarily focused on models that are quadratic in nonlinearity. As in Loiseau et al. [257], SINDy models for the dynamics of $\boldsymbol{q} = \boldsymbol{a}$ are developed, so that the dimension $N = r$ here, and $\boldsymbol{a}$ represents the coefficients or amplitudes of a modal Galerkin expansion in Eq. (3.2).

Throughout, this work deviates from the typical SINDy definitions by explicitly formulating the problem in terms of the vectorized $\boldsymbol{\xi} \in \mathbb{R}^{rp_\Theta}$, $\boldsymbol{\Theta}(\boldsymbol{X}) \in \mathbb{R}^{rM \times rp_\Theta}$, and $\dot{\boldsymbol{X}} \in \mathbb{R}^{rM}$,

$$\underset{\boldsymbol{\xi}}{\operatorname{argmin}} \left[\frac{1}{2} \|\boldsymbol{\Theta}\boldsymbol{\xi} - \dot{\boldsymbol{X}}\|_2^2 + \lambda R_s(\boldsymbol{\xi}) \right]. \tag{5.4}$$

The first term in the SINDy optimization problem in Eq. (5.4) fits a system of ODEs $\boldsymbol{\Theta}\boldsymbol{\xi}$ to the given data in $\dot{\boldsymbol{X}}$. The $R_s(\boldsymbol{\xi})$ term is typically chosen to be the l_0 norm, $\|\boldsymbol{\xi}\|_0$, which counts the number of nonzero elements of $\boldsymbol{\xi}$. The elements of $\boldsymbol{\xi}$ smaller than a threshold value, λ, are zeroed out. However, it is not technically a norm and leads to a non-convex optimization, so several convex relaxations have been proposed [59, 376, 458, 82, 457]. However, one can see that Eq. (5.4) does not rely on any specific physical principles beyond fitting the data and producing a model with a desired level of sparsity. In general, one would like to incorporate physical constraints directly into this optimization, so that models found with this technique *a priori* satisfy any known physical laws.

5.1 Constrained system identification

It is rare that a scientist knows *nothing* about a dynamical system of interest, and there a number of benefits that come from incorporating physical laws, priors, or constraints into system identification: (1) the identified models *a priori* satisfy the constraints, (2) extra constraints reduce the space of possible models, presumably restricting the models to a more physical subset, (3) there are a number of constraints that one would like to hold with high numerical precision, i.e. in magnetic field data it might be useful to prescribe $\nabla \cdot \boldsymbol{B} = 0$ or at least $\nabla \cdot \boldsymbol{B} \leq \epsilon_B$ for some small parameter ϵ_B, and (4) training data requirements

are typically reduced. Fortunately, there are a number of modern techniques to implement general constraints of the form

$$\mathcal{C}_i(\dot{\boldsymbol{X}}, \boldsymbol{X}, t, ...) \leq 0, \qquad i \in \{1, 2, ...\} \tag{5.5}$$

into both model reduction [242, 388] and system identification [257, 228, 82, 407].

Since the original SINDy algorithm, Loiseau et al. [257] introduced an extension to directly enforce constraints on the coefficients in $\boldsymbol{\xi}$. In particular, they enforced energy-preserving, skew-symmetry constraints on the quadratic terms for incompressible fluid flows, demonstrating improved model performance over standard Galerkin projection. For a SINDy library with constant, linear, and quadratic terms, there are a total of $p_\Theta = \frac{1}{2}(r^2 + 3r) + 1$ term. With the energy-preserving structure in the quadratic nonlinearities, as in Eq. (3.21), it can be shown that the number of constraints is $N_Q = r(r+1)(r+2)/6$ and therefore the number of free parameters is $rp_\Theta - N_Q = 2p_\Theta$. The constraint must also be converted into the SINDy formatting and this is done explicitly for arbitrary r in Appendix A. Recall that there is also an additional constraint on the linear parts of the model, Eq. (3.20), for fully energy-preserving plasma flows. All of the constraints are encoded together as $\boldsymbol{C}_s\boldsymbol{\xi} = \boldsymbol{d}$, $\boldsymbol{C}_s \in \mathbb{R}^{N_Q \times rp_\Theta}$, $\boldsymbol{d} \in \mathbb{R}^{N_Q}$, and the constrained SINDy algorithm solves the following minimization,

$$\underset{\boldsymbol{\xi}}{\operatorname{argmin}} \left[\frac{1}{2}\|\boldsymbol{\Theta}\boldsymbol{\xi} - \dot{\boldsymbol{X}}\|_2^2 + \lambda\|\boldsymbol{\xi}\|_1 + \delta_0(\boldsymbol{C}_s\boldsymbol{\xi} - \boldsymbol{d}) \right]. \tag{5.6}$$

In general, nonconvex regularizers can be used to promote sparsity in $\boldsymbol{\xi}$. However, modifications to Eq. (5.6) in Sec. 5.5 require a convex regularizer, so the remainder of this chapter uses the L^1 norm. The third term δ_0 is an indicator function that encodes hard constraints $\boldsymbol{C}_s\boldsymbol{\xi} = \boldsymbol{d}$. This term can be modified to use inequalities for imposing softer constraints. There are also variants of the constrained SINDy objective function in Eq. (5.6) that utilize sparse relaxed regularized regression (SR3) in order to improve performance [458, 82].

5.2 Initial constrained system identification results with 3D MHD simulations

The theoretical structure of this data-driven modeling framework is appealing, but its value to the community ultimately depends on the quality of the analysis when applied to plasma systems. Guided by the theory, a nonlinear, physics-constrained SINDy model is constructed for an isothermal Hall-MHD simulation of the HIT-SI experiment, described below.

5.2.1 HIT-SI simulations used for nonlinear system identification

Simulations of HIT-SI were performed using the Hall-MHD equations and solved by the NIMROD code [405]. Dirichlet boundary conditions for the variables include the plasma density $n_e = 2 \times 10^{19}$ m^{-3} and the temperatures $T_i = T_e = 14$ eV. Isotropic viscosity $\nu = 550$ m^2/s and Spitzer resistivity [406] is used. The remaining boundary conditions are $\boldsymbol{u} \times \hat{\boldsymbol{n}} = \boldsymbol{u} \cdot \hat{\boldsymbol{n}} = 0$, $\boldsymbol{J} \cdot \hat{\boldsymbol{n}} = 0$, and $\boldsymbol{B} \cdot \hat{\boldsymbol{n}} = 0$. For additional information on the numerical model see Sec. 2.2.2. The data for training and testing are obtained during the approximately steady-state phase of the simulation so that the energy constraints in Equations (3.20) and (3.21) are applicable.

The density, velocity, and magnetic field are sampled at a set of equally-spaced 3D points in the volume and sampling intervals $\Delta\phi = \pi/16$, $\Delta R \approx \Delta Z \approx 2$ cm. The result is that each component of $\boldsymbol{u}$ and $\boldsymbol{B}$ has 47712 samples. This high-resolution is ideal for visualization but substantial size reduction can be done with little or no change to the spatial or temporal POD modes. For instance, in Figures 3.1 and 5.2b, the $Z = 0$ visualizations of the 3D spatial POD modes are constructed from the 1440 sample locations at $Z = 0$; with a non-uniform set of 50 points in the midplane, the only change to the visualization is a smoothing out of the shortest wavelengths. The temporal resolution of the measurements is $\Delta t_k = 1$ μs. The analysis is essentially unchanged for time steps as large as 10 μs, but smaller time steps are required in HIT-SI to resolve harmonics of the injector frequency that appear in the temporal POD modes in Fig. 3.1. For instance, at $\Delta t_k = 10$ μs, the fourth injector harmonic

is sampled, on average, less than twice per period.

5.2.2 Results of constrained system identification

From these measurements of the density, velocity, and magnetic field, the topos and chronos are computed via the SVD in Eq. (3.4), obtaining a Galerkin expansion for the velocity and magnetic fields in magnetic field units, as in Eq. (3.6). Now a constrained SINDy model is identified for the first 16 chronos $a_i(t)$ and the forecasting is illustrated in Fig. 5.2a. The SINDy model accurately captures most of the $a_j(t)$ dynamics, with larger errors for the less energetic modes. Some of the low-frequency content in the $a_j(t)$ is not captured by this particular constrained SINDy model, but this is largely because the low frequencies are not well-resolved in the time range used for training. Despite this deficiency in the data, the SINDy model illustrates strong prediction performance in the $Z = 0$ midplane reconstructions of the simulation data in Fig. 5.2b and forecasts much of the time evolution for a high-dimensional simulation that used $589,824$ grid points, a tremendous efficiency gain of $\mathcal{O}(10^5)$. Furthermore, this model was obtained by training on a dataset representing a single discharge. Further improvements are likely accessible by training on a dataset of many discharges of varying trajectories.

A quality forecasting model has been found from the SINDy system identification method, but it is interesting to see how the model quality varies with the algorithm hyperparameters like the model sparsity λ and model rank r. Figure 5.3 illustrates how the normalized reconstruction errors of $\boldsymbol{X}$ and $\dot{\boldsymbol{X}}$ vary in the "Pareto-space" of (r, λ) for both the unconstrained and constrained SINDy algorithms, with the goal to explore the space of possible models obtained from this system identification technique. Although the exact reconstruction error values are unique to the simulation examined here, there are some interesting qualitative features that are expected to be quite general. The unconstrained SINDy algorithm indicates a significant region of (r, λ) where numerically unstable models are found. For $r \gtrapprox 10$, the models are typically either unstable or too sparse to be effective for forecasting. In contrast, by construction the constrained SINDy algorithm fully conserves the energy and

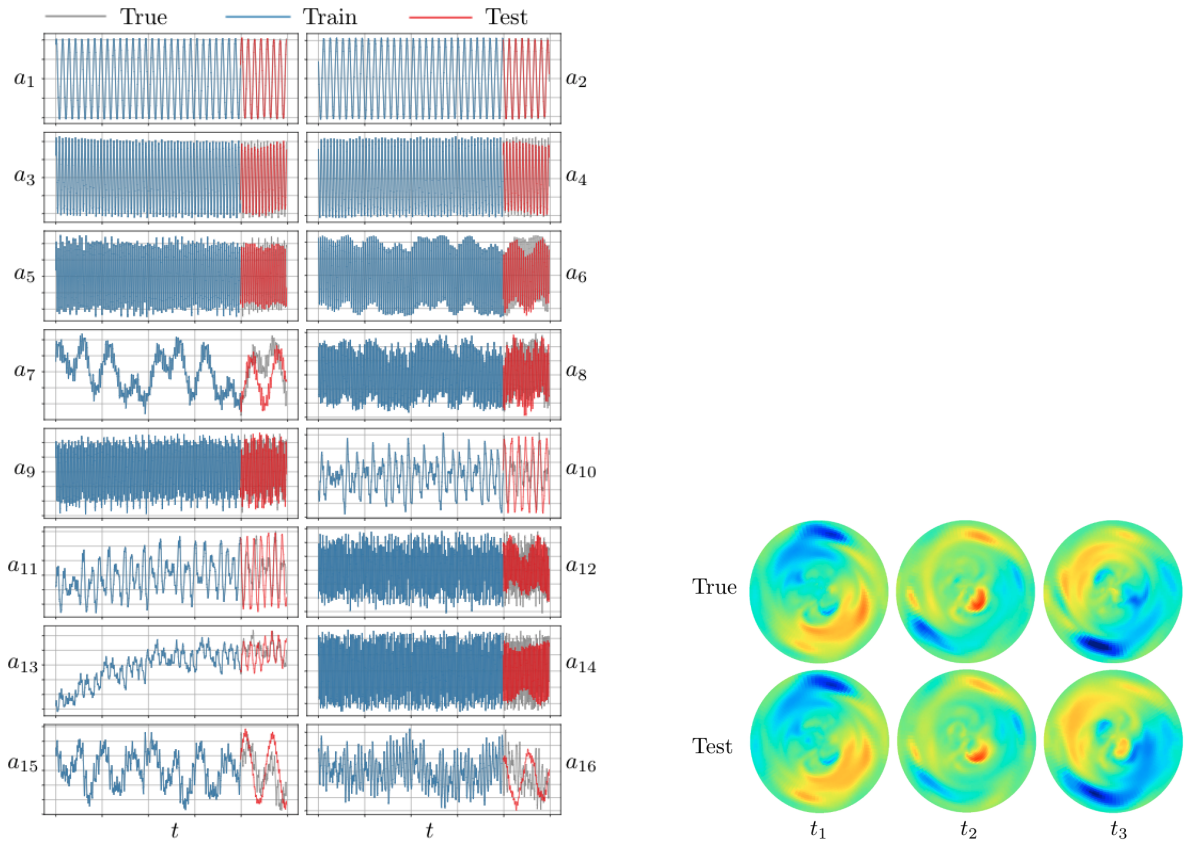


Figure 5.2: Summary of the constrained SINDy performance on a 3D Hall-MHD simulation of the HIT-SI device described in Sec. 2.1. (a) Constrained SINDy prediction of $a_1, ..., a_{16}$. The true evolution is in gray, the training data used for the model-finding is in blue, and the SINDy prediction is in red. (b) Constrained SINDy predictions of the z-component of $\boldsymbol{u}$ (Test) in the $Z = 0$ midplane are compared with the true evolution at three snapshots in time, indicating strong algorithm performance.

therefore exhibits no unstable models. This is promising for discovering models on historically challenging systems for machine learning methods − multi-scale or turbulent systems that require $r \gg 1$ to properly capture the dynamics. At first glance, it may appear that the constrained SINDy errors in $\dot{\boldsymbol{X}}$ are worse than the unconstrained errors, but the low-error values in the unconstrained case are precisely the unstable models. These nonsparse models are overfitting, leading to instability in the numerical integration. Finally, one can see that at $\lambda \approx 0.091$, all the SINDy models are rendered ineffective. This value is precisely at the driving frequency of the HIT-SI injectors in this simulation; if λ is larger than this frequency, SINDy thresholds off the primary dynamics in the system.

5.3 Stability-promoting system identification

It is often possible to improve the stability and performance of data-driven models by incorporating partially known physics, such as conservation laws and symmetries [257, 211, 212], or known physical structure [97]. Physics can be incorporated into machine learning algorithms through model structure, by augmenting training data with known symmetries, by adding constraints to the optimization as was done in the previous section, or by adding custom loss functions [65]. Thus, incorporating physics into machine learning and developing hybrid data-driven and operator-based approaches are rapidly growing fields of research [269, 24, 347, 257, 451, 359, 298, 325, 242, 97]. However, even physics-informed data-driven models often lack global stability guarantees, and the ability of these methods to find long-term bounded models typically depreciates as the state dimension increases. In this section, this issue is addressed by promoting stability guarantees during system identification. The work presented in Sections 5.3−5.7 has also been published as a paper in Kaptanoglu et al. [209].

5.4 The Schlegel and Noack trapping theorem

To address the issue of instability in analytic POD-Galerkin models, Schlegel and Noack [390] developed a "trapping theorem" with necessary and sufficient conditions for long-term model

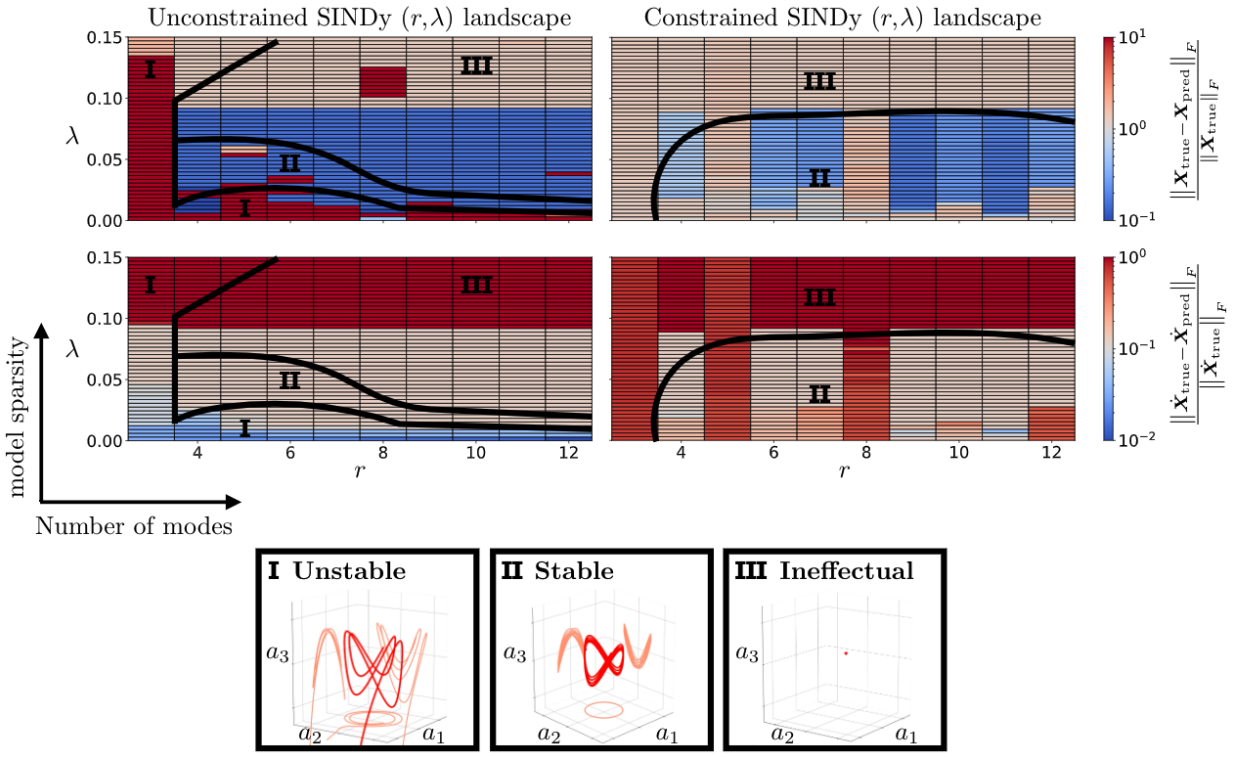


Figure 5.3: Summary of the (r, λ) space of unconstrained and constrained SINDy models from the HIT-SI simulation. The unconstrained models approximately separate into three distinct classes. Class I illustrates nonsparse and typically unstable models. Class II consists of sparse and accurate solutions. Class III denotes solutions which are too sparse to accurately capture the dynamics. Computed errors are for the testing part of the dataset; the colorbar (see online version for color) is limited to 10^1 as unstable model errors grow arbitrarily large. Constrained SINDy guarantees the energy norm is preserved and thus class I vanishes. Algorithmic advances of the type described in Chapter 6 can likely further expand the size of class II.

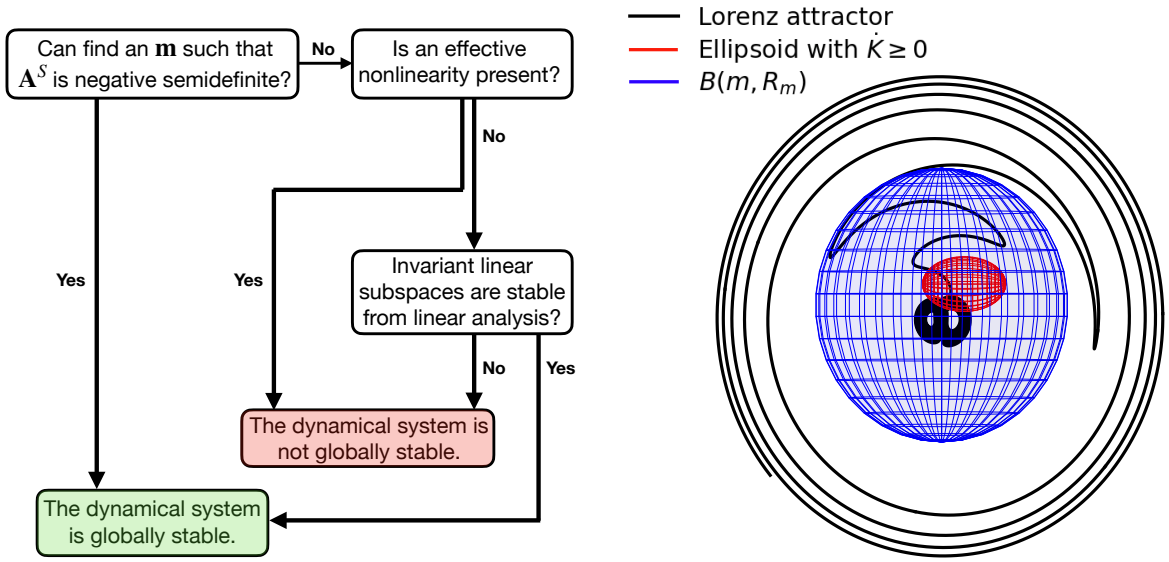


Figure 5.4: Left: Decision diagram to determine global stability, modified from Schlegel and Noack [390] and described in Section 5.4. Right: Illustration of a trapping region (blue sphere) for the Lorenz system; all outside trajectories monotonically approach this region, and after entering, remain inside. Trajectories inside the red ellipsoid experience positive energy growth, in this case precluding convergence to a fixed point.

stability for systems that exhibit quadratic, energy-preserving nonlinearities. This theorem can be used as an effective *post-hoc* diagnostic to examine whether a ROM is globally stable. The trapping theorem provides conditions for the existence of a global trapping region, towards which every system trajectory asymptotically and monotonically converges; once a trajectory enters this region, it remains inside for all time, guaranteeing that all trajectories are bounded. These types of guarantees are ideal for the application of real-time flow-control strategies. An example trapping region is illustrated by the blue sphere in Fig. 5.4 for the Lorenz system. For convenience, the terms "global stability", "long-term boundedness", and "monotonically trapping region" will be used interchangeably, although systems exhibiting trapping regions are a strict subset of globally stable systems (see Fig. 1 of Schlegel and Noack [390] for a useful organizational diagram of these various notions of stability).

In this section, the Schlegel and Noack [390] trapping theorem is used to diagnose and promote global stability of data-driven models with quadratic nonlinearities. Even though

their theorem was developed in the context of projection-based ROMs, it is important to emphasize that it can be applied directly to analyze data-driven model stability *post hoc*, and this work examines the conditions under which it holds. Next, this theorem is used to promote global stability in machine learned models by modifying the optimization loss function. This approach is illustrated with constrained SINDy, i.e. Eq. (5.6), by implementing an additional custom optimization loss term that promotes models that are globally stable by construction. Constrained SINDy generally produces more stable models than unconstrained SINDy, and reflects a broader trend that stability issues can often be improved by building physical constraints into system identification methods [257, 82].

The "trapping SINDy" algorithm generalizes previous stabilized or constrained reduced-order models for fluids by considering global rather than local stability, allowing for both transients and long-time attracting sets. Promoting global stability also improves robustness to noise over unconstrained or constrained SINDy. Recent works by Erichson et al. [119] and Sawant et al. [383] promote a more restrictive locally, asymptotically stable origin in fluid flows by adding similar loss terms to the optimization problem. Additionally, much of the literature has focused on the long-time energy properties of a dynamic attractor [22] by either prescribing that the system be *fully* energy-preserving (or Hamiltonian) [20, 80, 348, 3, 41, 90] or applying real-time control [241]. Mohebujjaman et al. [298] also used a simple version of the trapping theorem in order to constrain a hybrid projection-based and data-driven method; they promote global stability by requiring that the system be globally dissipative everywhere. This chapter builds on these studies, providing a framework for addressing the long-standing challenge of promoting global stability in data-driven models and summarizing the work in Kaptanoglu et al. [209].

This section is additionally motivated because there are many scenarios under which energy-preserving quadratic nonlinearities can arise. In fluid dynamics, the quadratic nonlinearity often represents the convective derivative $(\boldsymbol{u} \cdot \nabla)\boldsymbol{u}$ in the Navier-Stokes equations. This quadratic nonlinearity is energy-preserving for a large number of boundary conditions. Examples include no-slip conditions, periodic boundary conditions [283, 177], mixed no-slip

and periodic boundary conditions [378], and open flows in which the velocity magnitude decreases faster than the relevant surface integrals expand (e.g., two-dimensional rigid body wake flows and three-dimensional round jets) [391]. In magnetohydrodynamics, there are additional quadratic nonlinearities through $\nabla \times (\boldsymbol{u} \times \boldsymbol{B})$ and $\boldsymbol{J} \times \boldsymbol{B}$, which are also energy-preserving with common experimental boundary conditions such as a conducting wall [127], or a balance between dissipation and actuation in a steady-state plasma device, as illustrated in Eq. (3.19).

To begin, the Schlegel and Noack trapping theorem is reviewed in order to understand the circumstances under which it holds and how to incorporate it into system identification. This theorem provides necessary and sufficient conditions for energy-preserving, effectively nonlinear, quadratic systems to exhibit a trapping region $B(\boldsymbol{m}, R_m)$, a ball centered at $\boldsymbol{m} \in \mathbb{R}^r$ with radius R_m. Outside this region the rate of change of energy $\dot{K}$ is negative everywhere, so that the energy $\dot{K}$ is a Lyapunov function that renders this system globally stable. Recentering the origin by an arbitrary constant vector $\boldsymbol{m}$, the energy may be expressed in terms of the shifted state vector $\boldsymbol{y}(t) = \boldsymbol{a}(t) - \boldsymbol{m}$ as

$$K = \frac{1}{2}\boldsymbol{y}^T\boldsymbol{y}. \tag{5.7}$$

Taking a derivative and substituting in Eq. (3.12) produces

$$\frac{d}{dt}K = \boldsymbol{y}^T\boldsymbol{A}^S\boldsymbol{y} + \boldsymbol{d}_m^T\boldsymbol{y}, \tag{5.8}$$

$$\boldsymbol{A}^S = \boldsymbol{L}^S - \boldsymbol{m}^T\boldsymbol{Q}, \quad \boldsymbol{L}^S = \frac{1}{2}(\boldsymbol{L} + \boldsymbol{L}^T), \quad \text{and} \tag{5.9}$$

$$\boldsymbol{d}_m = \boldsymbol{C} + \boldsymbol{L}\boldsymbol{m} + \boldsymbol{Q}\boldsymbol{m}\boldsymbol{m}, \tag{5.10}$$

where Eq. (3.12) has been used, $\boldsymbol{m}^T\boldsymbol{Q}$ refers to $m_i Q_{ijk}$, and $\boldsymbol{Q}\boldsymbol{m}\boldsymbol{m}$ refers to $Q_{ijk}m_j m_k$. The trapping theorem may now be stated as:

Theorem 1. *There exists a monotonically trapping region at least as small as the ball* $B(\boldsymbol{m}, R_m)$ *if and only if the real, symmetric matrix* $\boldsymbol{A}^S$ *is negative definite with eigenvalues* $\lambda_r \leq \cdots \leq \lambda_1 < 0$*; the radius is then given by* $R_m = \|\boldsymbol{d}_m\|/|\lambda_1|$.[1]

In practice, the goal is then to find an origin $\boldsymbol{m}$ so that the matrix $\boldsymbol{A}^S$ is negative definite, guaranteeing a trapping region and global stability. Without effective nonlinearity, described at the beginning of Section 5.4.1, only the backwards direction holds; if one can find an $\boldsymbol{m}$ so that $\boldsymbol{A}^S$ is negative definite, the system exhibits a trapping region. However, such systems can be globally stable without admitting such an $\boldsymbol{m}$. Subsequently, the goal of Section 5.5 is to use this theorem to define a constrained machine learning optimization that identifies a reduced-order model with a guaranteed trapping region. Even when the requirements of the trapping theorem are not fully satisfied, the algorithm results in Section 5.6 indicate that this approach tends to produce models with improved stability properties.

To understand the R_m bound in Thm. 1, $\boldsymbol{y}$ is transformed into eigenvector coordinates $\boldsymbol{z} = \boldsymbol{T}_{\text{eig}}\boldsymbol{y}$, $\boldsymbol{h}^{\text{eig}} = \boldsymbol{d}_m\boldsymbol{T}_{\text{eig}}^T$, where $\boldsymbol{T}_{\text{eig}}$ are the eigenvectors of $\boldsymbol{A}^S$. Now Eq. (5.8) becomes

$$\frac{d}{dt}K = \sum_{i=1}^{r} h_i^{\text{eig}} z_i + \tilde{\lambda}_i z_i^2 = \sum_{i=1}^{r} \lambda_i \left(z_i + \frac{h_i^{\text{eig}}}{2\lambda_i} \right)^2 - \frac{(h_i^{\text{eig}})^2}{4\lambda_i}. \tag{5.11}$$

One can see that the trapping region will be determined by the equation of the ellipsoid where $\dot{K} = 0$,

$$1 = \sum_{i=1}^{r} \frac{1}{\alpha_i^2} \left(z_i + \frac{h_i^{\text{eig}}}{2\lambda_i} \right)^2, \tag{5.12}$$

$$\alpha_i \equiv \frac{1}{2}\sqrt{\frac{1}{\lambda_i}\sum_{j=1}^{r} \frac{(h_j^{\text{eig}})^2}{\lambda_j}} \leq \frac{1}{2|\lambda_1|}\|\boldsymbol{d}_m\|. \tag{5.13}$$

The origin at $\boldsymbol{y} = 0$ $(\boldsymbol{a} = \boldsymbol{m})$ lies on the ellipsoid, and in the worst case scenario lies at

[1] If a system is long-term bounded (not necessarily exhibiting a monotonically trapping region) and effectively nonlinear, only the existence of an $\boldsymbol{m}$ producing a negative *semi*definite $\boldsymbol{A}^S$ is guaranteed.

the tip of the major axis. Thus, to guarantee that a ball centered at this origin entirely contains this region, R_m is estimated as twice the size of the largest possible value of the half-axes α_i. Note that this definition of α_i differs from Schlegel and Noack [390] due to a minor typo in their Eq. 3.14. Fortunately, the only consequence is a change in the estimate of R_m. Lastly, recall that long-term bounded (not necessarily exhibiting a monotonically trapping region) and effectively nonlinear systems only guarantee an $\boldsymbol{m}$ exists such that $\boldsymbol{A}^S$ is negative semidefinite. In the case of mixed zero and nonzero eigenvalues, the ellipsoid becomes a paraboloid. The infinite extent of the paraboloid precludes a monotonic trapping region but not other forms of global stability. This edge case is not further discussed because in practice (numerically) there is no chance of arriving at an eigenvalue of exactly zero.

5.4.1 Interpretation of the trapping theorem

The Schlegel and Noack [390] theorem, summarized in Theorem 1, provides necessary and sufficient conditions for the projected ROM in Eq. (3.12) to be globally stable by admitting a trapping region. This theorem is necessary and sufficient for systems that exhibit effective nonlinearity, i.e., the system does not manifest invariant manifolds where there exists some i such that $Q_{ijk}a_j a_k = 0$ for all time, for which a linear stability analysis must be adopted. In other words, systems that start in purely linear model subspaces, and remain in those subspaces, do not exhibit effective nonlinearity − the quadratic part of the model never "turns on". Fortunately, realistic fluid flows exhibit effective nonlinearity, although there are some subtleties that are discussed in Section 5.4.2. In this case, one can always use the total fluid kinetic energy K as a Lyapunov function for the trapping region. This is ideal, as finding a suitable Lyapunov function is often the most difficult task in stability analysis. It is possible that other Lyapunov functions exist with tighter bounds on the size of a trapping region, but this section is primarily concerned with promoting models with long-term boundedness (i.e. models that exhibit a trapping region of any kind), rather than an algorithm for precisely capturing the shape and size of a trapping region. For a post-fit algorithm to find the optimal ellipsoidal estimate of the stability domain, see Kramer [231]

or the stability analysis work in Kalur [205, 206].

A generic nonlinear system may exhibit multiple fixed points, limit cycles, and other equilibrium point behavior. However, any physical system should produce bounded trajectories, and in principle the global stability property from the trapping theorem is agnostic to any *local* stability properties. Sections 5.4−5.7 solely considers systems that are globally stable, or equivalently, long-term (ultimately) bounded, by virtue of exhibiting globally trapping regions. Long-term boundedness means that there exists some T_0 and R_0 such that $\|\boldsymbol{a}(t)\| < R_0$ for all $t > T_0$. A trapping region encompasses an attractor or attracting set, which is typically defined as a set of the system phase space that many trajectories converge towards; this can be an equilibrium point, periodic trajectory, Lorenz's "strange attractor", or some other chaotic trajectory. Whenever it does not detract from the discussion, the qualifiers "globally", "monotonically" and "long-term" are omitted, as this is the only characterization of stability considered in Sections 5.3−5.7. Examples of physical systems that are globally stable but do not exhibit a trapping region include Hamiltonian systems and systems that do not fit into the trapping theorem assumptions (examined further in Section 5.4.2 and summarized in Fig. 5.4). For instance, fully energy-preserving systems satisfy $\dot{K} = 0$, so trajectories represent shells of constant distance from the origin. These trajectories are globally bounded but no trapping region exists.

5.4.2 Model truncation, effective nonlinearity, and closure

Before implementing the trapping theorem into system identification, the circumstances under which truncated projection-based models (such as those described in Chapter 3) will exhibit effective nonlinearity are investigated; the reader may skip this section if the subtleties of the trapping theorem are not of interest, although the discussion here is pertinent to Section 5.6. Effectively nonlinear dynamics are ideal because they can be decisively classified as globally stable or not, requiring no additional stability analysis. To clarify an important circumstance in which effective nonlinearity is not satisfied, consider a Fourier-

Galerkin model of the 1D Burgers' equation derived from the Fourier expansion $q(x,t) = \sum_k a_k(t)e^{ikx}$, and further examined in Section 5.6,

$$\dot{q} = -q\partial_x q + \nu\partial_{xx} q \quad \Longrightarrow \quad \dot{a}_k = -\nu k^2 a_k - \sum_{\ell=-\infty}^{\infty} i\ell a_\ell a_{k-\ell} \tag{5.14}$$

$$\Longrightarrow \quad \dot{K} = -\nu \sum_{k=-\infty}^{\infty} k^2 a_k^2 - \sum_{k,\ell=-\infty}^{\infty} i\ell a_\ell a_{k-\ell} a_k. \tag{5.15}$$

The particular "triadic" structure of the nonlinear term in the spectral domain, where the only nonzero terms acting on a_k are those whose wavenumbers sum to k, is identical to that arising in isotropic turbulence [420]. The triadic term in $\dot{K}$ transfers energy between length scales. Since the viscous term scales with k^2, energy is most effectively dissipated at the smallest scales; the combination of the two terms leads to the traditional energy cascade in which energy flows "downhill" from larger to smaller scales. This description implies that heavily truncating the Galerkin system leads to under-resolving the dissipation rate and a closure scheme may be required to re-introduce the full dissipation. Towards this goal, modern sparse regression and deep learning methods have been used to produce new closures for dynamical models [253, 380, 338, 114, 281, 29, 19]. While the traditional explanation for unstable Galerkin models derives from these truncated dissipative scales, increasingly there are alternate explanations including fundamental numerical issues with the Galerkin framework (potentially resolved in a Petrov-Galerkin framework) [149] and the Kolmogorov width issues of linear subspaces more generally [243]. If true, this is probably good news for (incompressible, dissipationless) Hall-MHD, in which additional invariants can lead to very complicated turbulent cascades. Interestingly, the notion of effective nonlinearity appears to be another approach from which one can attempt to resolve these disagreements about the sources of ROM instability.

To proceed with this theme, it is shown that the triadic structure of the model has repercussions for the presence of effective nonlinearity. Consider a truncation of the model

in Eq. (5.14),

$$\dot{a}_k = -\nu k^2 a_k - \sum_{\ell=-r}^{r} i\ell a_\ell a_{k-\ell}, \qquad k \in \{1, ..., r\}, \tag{5.16}$$

with the initial condition $a_j = 1$ for any $j \in \{\pm(\frac{r}{2}+1), \pm(\frac{r}{2}+2), ..., \pm r\}$, and $a_k = 0$, $k \neq j$. For simplicity it is assumed here that r is divisible by two. In this case the system has r invariant 1D subspaces for which

$$\dot{a}_j = -\nu j^2 a_j. \tag{5.17}$$

These invariant linear subspaces exist because higher wavenumber modes that *could* interact to transfer energy between coefficients have been discarded. In other words, *Fourier-Galerkin models with finite truncation do not exhibit effective nonlinearity.* In contrast, POD-Galerkin models weakly break the triadic structure of the nonlinearity [96], and therefore in general will weakly satisfy the trapping theorem criteria for effective nonlinearity, to the extent that they differ from the Fourier modes because of inhomogeneity in the system. There are also modern ROMs which attempt to retain the full dissipation by utilizing bases that intentionally mix length scales [21] – these types of models should be more likely to satisfy effective nonlinearity. Lastly, numerical errors appear to weakly restore effective nonlinearity, since errors break any triadic structure. Proceeding with this analysis is complicated because the numerical errors also weakly break the foundational assumption that Q_{ijk} is energy-preserving. Future investigations should be pursued to explore relationships between effective nonlinearity, the energy cascade, and closure models that reintroduce stabilizing dissipation to truncated models.

It is difficult to quantify "how close" a model is to exhibiting effective nonlinearity, since a lack of effective nonlinearity $Q_{ijk}a_j a_k = 0$ must hold for all time, for any i, and for any valid system trajectory. However, for an orthonormal set of temporal modes, and assuming there exists at least one index i such that $Q_{ijj} \neq 0$, the average strength of model effective

nonlinearity can be quantified through the metric

$$S_e = \frac{\min_i |Q_{ijk}\overline{a_j a_k}|}{\max_i |Q_{ijk}\overline{a_j a_k}|} = \frac{\min_i |Q_{ijj}|}{\max_i |Q_{ijj}|}. \tag{5.18}$$

The bar in $\overline{a_j a_k}$ denotes a temporal average. In Section 5.6 it will be illustrated that, for system identification, a lack of effective nonlinearity is not a terrible loss. The trapping SINDy algorithm in Section 5.5 minimizes $\dot{K}$ whether or not a negative definite $\boldsymbol{A}^S$ can be realized. However, without additional stability analysis, such models are no longer provably stable for any initial condition. Although Eq. (5.17) is a linearly stable system, this is not guaranteed for more general fluid models than the simple Burgers' equation considered here.

5.5 SINDy with stability guarantees (trapping SINDy)

Model constraints in system identification, such as global conservation laws or other physical considerations, often result in improved models, but do not generally guarantee global stability.

Enforcing stability in quadratic energy-preserving models is more complicated than Eq. (5.5). To see this, note that there a few different circumstances under which one might want to promote stability. If the true $\boldsymbol{A}^S$ *and* the optimal $\boldsymbol{m}$ are known, one can simply constrain the coefficients in Eq. (3.12) to produce this known negative definite $\boldsymbol{A}^S$. This would imply that one already knows the optimally-shifted eigenvalues of the system and an $\boldsymbol{m}$ that produces these negative eigenvalues; if this is the case, so much information about the system of ODEs is already known that machine learning methods are likely unnecessary.

But far more interesting are the cases in which 1) the underlying system is known to be globally stable and effectively nonlinear, so one wants to find the "correct" $\boldsymbol{m}$ and corresponding $\boldsymbol{A}^S$, or 2) it is not known if any $\boldsymbol{m}$ exists such that $\boldsymbol{A}^S$ is negative definite. In system identification, either of these cases can be addressed by searching for a model that both optimally fits the data and is globally stable. In this context, a mixed approach is adopted in the next section where the energy-preserving constraint is enforced and then the

optimization is separately biased towards models with a trapping region. This technique is a significant methodological extension because one can no longer rely on constraints of the form in Eq. (5.5).

Recall from Thm. 1 that $\boldsymbol{m}$ is an arbitrary, constant vector, of the same state size as $\boldsymbol{a}$, that specifies the center of a possible trapping region. Stability promotion is then achieved by jointly determining the sparse model coefficients $\boldsymbol{\xi}$ and state vector $\boldsymbol{m}$ such that $\boldsymbol{A}^S$ from Eq. (5.9) is negative definite.

To proceed with the trapping SINDy formulation, one must relate the model coefficients in $\boldsymbol{\xi}$ to the matrix $\boldsymbol{A}^S$ appearing in the trapping theorem. First, the projection operators are defined: $\boldsymbol{P}^L \in \mathbb{R}^{r\times r\times rM}$, $\boldsymbol{P}^Q \in \mathbb{R}^{r\times r\times r\times rM}$, and $\boldsymbol{P} \in \mathbb{R}^{r\times r\times rM}$. The operator $\boldsymbol{P}^L$ projects out the symmetric part of the linear coefficients through $\boldsymbol{L}^S = \boldsymbol{P}^L\boldsymbol{\xi}$. The same is true for the quadratic coefficients, $\boldsymbol{Q} = \boldsymbol{P}^Q\boldsymbol{\xi}$. The operator $\boldsymbol{P} = \boldsymbol{P}^L - \boldsymbol{m}^T\boldsymbol{P}^Q$ provides a concise representation of $\boldsymbol{A}^S$ through the following equation:

$$\boldsymbol{A}^S = \boldsymbol{L}^S - \boldsymbol{m}^T\boldsymbol{Q} = \boldsymbol{P}\boldsymbol{\xi} = (\boldsymbol{P}^L - \boldsymbol{m}^T\boldsymbol{P}^Q)\boldsymbol{\xi}. \tag{5.19}$$

A tentative version of the trapping SINDy optimization problem is now defined, in analogy to the constrained SINDy optimization in Eq. (5.6). It incorporates an additional loss term to reduce the maximal (most positive) eigenvalue λ_1 of the real, symmetric matrix $\boldsymbol{A}^S$:

$$\underset{\boldsymbol{\xi},\boldsymbol{m}}{\operatorname{argmin}} \left[\frac{1}{2}\|\boldsymbol{\Theta}\boldsymbol{\xi} - \dot{\boldsymbol{X}}\|_2^2 + \lambda\|\boldsymbol{\xi}\|_1 + \delta_0(\boldsymbol{C}_s\boldsymbol{\xi} - \boldsymbol{d}) + \frac{\lambda_1}{\zeta}\right]. \tag{5.20}$$

Note that a new hyperparameter ζ was introduced, which modulates the strength of the λ_1 loss term. Although λ_1 is a convex function of the matrix elements [335], $\boldsymbol{A}^S = (\boldsymbol{P}^L - \boldsymbol{m}^T\boldsymbol{P}^Q)\boldsymbol{\xi}$ is not affine in $\boldsymbol{\xi}' = [\boldsymbol{\xi}, \boldsymbol{m}]$. The result is that this new term is not convex, but *convex composite*. It is possible to approximately solve this problem with a variable projection technique, where one essentially treats $\boldsymbol{\xi}$ and $\boldsymbol{m}$ as independent, solves the convex problem in $\boldsymbol{\xi}$, and then substitutes $\boldsymbol{\xi}^*$, the solution at each iteration, into the optimization for $\boldsymbol{m}$. In

practice this algorithm performs fairly well, although the convergence properties are unclear. Eq. (5.20) is also amenable to other approaches, such as Gauss-Newton [68] or the prox-linear algorithm [112], because λ_1 is a convex function of the elements of $\boldsymbol{A}^S$ and $\boldsymbol{P\xi}$ is smooth in $\boldsymbol{m}$ and $\boldsymbol{\xi}$. Although a modified algorithm is instituted below, these convex-composite approaches are a promising future direction for effectively solving this nonconvex optimization problem.

In order to produce an algorithm with better performance and better understood convergence properties, a relax-and-split approach [457] is adopted, similar to the approach taken in Champion et al. [82]. An auxiliary variable $\boldsymbol{A}$ is introduced that represents the projection of $\boldsymbol{A}^S = \boldsymbol{P\xi}$ onto the space of negative definite matrices, and introduce two new terms in the optimization:

$$\underset{\boldsymbol{\xi},\boldsymbol{m},\boldsymbol{A}}{\operatorname{argmin}} \left[\frac{1}{2}\|\boldsymbol{\Theta\xi} - \dot{\boldsymbol{X}}\|_2^2 + \lambda\|\boldsymbol{\xi}\|_1 + \delta_0(\boldsymbol{C}_s\boldsymbol{\xi} - \boldsymbol{d}) + \frac{1}{2\zeta}\|\boldsymbol{P\xi} - \boldsymbol{A}\|_2^2 + \delta_{\mathcal{I}}(\boldsymbol{\Lambda}_A)\right]. \tag{5.21}$$

The new least-squares term enforces a "soft" constraint (or bias) towards $\boldsymbol{A}^S = \boldsymbol{P\xi}$ being negative definite by minimizing the difference between $\boldsymbol{P\xi}$ and its projection into the space of negative definite matrices. The auxiliary variable $\boldsymbol{A}$ is updated to approximate $\boldsymbol{A}^S = \boldsymbol{P\xi}$, and then, through the $\delta_{\mathcal{I}}$ term, enforced to be negative definite by requiring that the diagonalized matrix $\boldsymbol{\Lambda}_A = \boldsymbol{S}^{-1}\boldsymbol{AS}$ lies in $\mathcal{I} = (-\infty, -\gamma_A\,]$, $\gamma_A > 0$. Directly enforcing $\boldsymbol{P\xi}$ to be negative definite tends to badly distort the model fit to the data. Instead, the auxiliary variable $\boldsymbol{A}$ in Eq. (5.21) allows the algorithm to accurately fit the data with $\boldsymbol{\xi}$ and then relax the coefficients towards a negative definite $\boldsymbol{A}^S$ to promote global stability.

This flexible formulation also allows $\boldsymbol{A}$, the proxy for the projection of $\boldsymbol{P\xi}$ onto the space of negative definite matrices, to vary, and therefore fit the particular eigenvalues of the system in question. In other words, the proposed approach pushes $\boldsymbol{A}^S$ into the space of negative definite matrices in $\mathbb{R}^{r\times r}$ with minimal assumptions about the eigenvalues, only assuming that they are negative. Contrast this algorithm to a more restrictive approach that prescribes an $\boldsymbol{A}$, meaning a set of negative eigenvalues of $\boldsymbol{P\xi}$ is already known and

Trapping SINDy hyperparameters	
λ	Specifies the strength of sparsity-promotion through the regularizer $R_s(\boldsymbol{\xi})$. $\lambda = 0$ already works well for low-dimensional systems because the $\|\boldsymbol{P\xi} - \boldsymbol{A}\|_2^2$ term promotes stability.
ζ	Specifies how strongly to push the algorithm towards models with negative definite $\boldsymbol{A}^S$. If $\zeta \gg 1$, $\boldsymbol{\xi}^*$ is unaffected by the minimization over $\boldsymbol{m}$. If $\zeta \ll 1$, the problem is increasingly nonconvex.
γ_A	Determines how far to push the eigenvalues of $\boldsymbol{A}^S$ towards being negative definite. Typically $\gamma_A \lesssim 0.1$ works for a variety of problems regardless of the true eigenvalues of $\boldsymbol{A}^S$.

Table 5.1: Description of the trapping SINDy hyperparameters.

is compatible with the data. A description of each of the hyperparameters λ, ζ, and γ_A, is provided in Table 5.1. Note that Eq. (5.21) is not convex in $\boldsymbol{A}$, and this is the most challenging aspect of this formalism.

Now that the problem has been defined in Eq. (5.21), it needs to be solved. If the convex part of the optimization is denoted,

$$F(\boldsymbol{\xi}, \boldsymbol{m}, \boldsymbol{A}) = \|\boldsymbol{\Theta\xi} - \dot{\boldsymbol{X}}\|_2^2/2 + \lambda\|\boldsymbol{\xi}\|_1 + \delta_0(\boldsymbol{C}_s\boldsymbol{\xi} - \boldsymbol{d}) + \|\boldsymbol{P\xi} - \boldsymbol{A}\|_2^2/2\zeta, \tag{5.22}$$

and fix initial guesses for $\boldsymbol{m}$ and $\boldsymbol{A}$, then one can define the solution vector $\boldsymbol{\xi}^*$ through

$$\boldsymbol{\xi}^* = \underset{\boldsymbol{\xi}}{\operatorname{argmin}} \left[F(\boldsymbol{\xi}, \boldsymbol{m}, \boldsymbol{A})\right]. \tag{5.23}$$

If $\lambda = 0$, $\boldsymbol{\xi}^*$ is structurally identical to the $\boldsymbol{\xi}^*$ in Champion et al. [82]:

$$\boldsymbol{H} = (\boldsymbol{\Theta}^T\boldsymbol{\Theta} + \frac{1}{\zeta}\boldsymbol{P}^T\boldsymbol{P})^{-1}, \tag{5.24}$$

$$\boldsymbol{\xi}^* = \boldsymbol{H}\left[\boldsymbol{I} - \boldsymbol{C}_s^T(\boldsymbol{C}_s\boldsymbol{H}\boldsymbol{C}_s^T)^{-1}\boldsymbol{C}_s\boldsymbol{H}\right]\left[\boldsymbol{\Theta}^T\dot{\boldsymbol{X}} + \frac{1}{\zeta}\boldsymbol{P}^T\boldsymbol{A}\right] + \boldsymbol{H}\boldsymbol{C}_s^T(\boldsymbol{C}_s\boldsymbol{H}\boldsymbol{C}_s^T)^{-1}\boldsymbol{d}. \tag{5.25}$$

$\boldsymbol{H}$ is positive definite, $\boldsymbol{I}$ is the identity matrix, and $\boldsymbol{C}_s\boldsymbol{\xi}^* = \boldsymbol{d}$ can be verified using Eq. (5.25). The minimization over $\boldsymbol{\xi}$ with $\lambda \neq 0$ is still convex but not analytically tractable as in

Eq. (5.25). Since it is convex, it can be solved with standard convex optimization libraries such as CVXPY [109]. It can also be shown to reduce to a constrained quadratic program over the unit box with a positive semidefinite cost matrix. A barrier to this route is that typical numerical solvers either assume that the quadratic cost matrix is sparse or positive definite. Neither assumption is true here.

Now that the minimization over $\boldsymbol{\xi}$ has been solved, prox-gradient descent can be used on $(\boldsymbol{m}, \boldsymbol{A})$; each algorithm iteration, one alternates between solving for $\boldsymbol{\xi}^*$ and solving for $(\boldsymbol{m}^*, \boldsymbol{A}^*)$. Again, one can think about this problem as a variable projection [9, 454], which is a value function optimization over the remaining variables $(\boldsymbol{m}, \boldsymbol{A})$. To make this viewpoint more precise, define

$$\widetilde{F}(\boldsymbol{m}, \boldsymbol{A}) = F(\boldsymbol{\xi}^*, \boldsymbol{m}, \boldsymbol{A}), \tag{5.26}$$

The problem to solve is now written more simply as

$$\operatorname*{argmin}_{\boldsymbol{m}, \boldsymbol{A}} \left[\widetilde{F}(\boldsymbol{m}, \boldsymbol{A}) + \delta_{\mathcal{I}}(\boldsymbol{\Lambda}_A) \right].$$

Prox-gradient descent is applied to this nonconvex problem, so that

$$\boldsymbol{m}^* = \boldsymbol{m} - \alpha_m \nabla_{\boldsymbol{m}} \widetilde{F}(\boldsymbol{m}, \boldsymbol{A}), \tag{5.27}$$

$$\boldsymbol{A}^* = \operatorname{proj}_{\mathcal{I}} \left[\boldsymbol{A} - \alpha_A \nabla_{\boldsymbol{A}} \widetilde{F}(\boldsymbol{m}, \boldsymbol{A}) \right], \tag{5.28}$$

where α_m and α_A are step sizes. All that remains is to compute the gradients of the value function $\widetilde{F}$,

$$\nabla_{\boldsymbol{A}} \widetilde{F}(\boldsymbol{m}, \boldsymbol{A}) = \frac{1}{\zeta}(\boldsymbol{A} - \boldsymbol{P}\boldsymbol{\xi}^*), \tag{5.29}$$

$$\nabla_{\boldsymbol{m}} \widetilde{F}(\boldsymbol{m}, \boldsymbol{A}) = \frac{1}{\zeta} \boldsymbol{P}^Q \boldsymbol{\xi}^* (\boldsymbol{A} - \boldsymbol{P}\boldsymbol{\xi}^*). \tag{5.30}$$

These are Lipschitz continuous functions with Lipschitz constants L_A, L_m satisfying

$$\alpha_A \leq \frac{1}{L_A} \leq \zeta, \tag{5.31}$$

$$\alpha_m \leq \frac{1}{L_m} \leq \frac{\zeta}{\|(\boldsymbol{P}^Q\boldsymbol{\xi}^*)_{ijk}(\boldsymbol{P}^Q\boldsymbol{\xi}^*)_{ljk}\|_{\mathrm{F}}}, \tag{5.32}$$

in order for guaranteed convergence of fixed step-size, prox-gradient descent [14]. While the denominator in Eq. (5.32) varies with the update in $\boldsymbol{\xi}$, in practice, one can reduce α_m until convergence is found. The full trapping SINDy optimization is illustrated in Algorithm 1.

Algorithm 1 Trapping SINDy

Input: Numerical data $\dot{\boldsymbol{X}}$ and optional initial guesses for $\boldsymbol{m}$ and $\boldsymbol{A}$.

Output: Optimal model coefficients $\boldsymbol{\xi}^*$ and shift vector $\boldsymbol{m}^*$.

1: **procedure** SINDy($\dot{\boldsymbol{X}}$, λ, ζ, γ_A)

2: Construct $\boldsymbol{\Theta}(\boldsymbol{X})$, $\boldsymbol{P}$, $\boldsymbol{C}_s$, and $\boldsymbol{d}$.

3: **while** $|\boldsymbol{\xi}_k - \boldsymbol{\xi}_{k+1}| > \epsilon_{\text{tol}}^{\xi}$ and $|\boldsymbol{m}_k - \boldsymbol{m}_{k+1}| > \epsilon_{\text{tol}}^{m}$

4: $\boldsymbol{\xi}_{k+1} \Longleftarrow \text{argmin}_{\boldsymbol{\xi}_k} \left[F(\boldsymbol{\xi}_k, \boldsymbol{m}_k, \boldsymbol{A}_k)\right]$,

5: $\boldsymbol{S}_{k+1}(\boldsymbol{\Lambda}_A)_{k+1}(\boldsymbol{S}_{k+1})^{-1} \Longleftarrow \boldsymbol{A}_k - \alpha_A \nabla_{\boldsymbol{A}} \widetilde{F}(\boldsymbol{m}, \boldsymbol{A})|_{m_k, A_k}$,

6: $\boldsymbol{A}_{k+1} \Longleftarrow \boldsymbol{S}_{k+1}\text{proj}_{\mathcal{I}}\left[(\boldsymbol{\Lambda}_A)_{k+1}\right](\boldsymbol{S}_{k+1})^{-1}$,

7: $\boldsymbol{m}_{k+1} \Longleftarrow \boldsymbol{m}_k - \alpha_m \nabla_{\boldsymbol{m}} \widetilde{F}(\boldsymbol{m}, \boldsymbol{A})|_{m_k, A_k}$,

8: **end procedure**

In words, 2: initialize variables, 3: start iteration loop, 4: convex minimization for $\boldsymbol{\xi}_{k+1}$, 5: prox-gradient step for $\boldsymbol{A}_{k+1}$, 6: project $\boldsymbol{A}_{k+1}$ into $\mathcal{I}$, rotate into $\boldsymbol{P}\boldsymbol{\xi}_{k+1}$ basis, and 7: prox-gradient step for $\boldsymbol{m}_{k+1}$. Note that inequalities (5.31)–(5.32) should be satisfied, and there tends to be a sweet spot for ζ. It is often useful to start with $\zeta \gg 1$ and then reduce ζ until the model coefficients are significantly affected.

$\epsilon_{\text{tol}}^{\xi}$ and $\epsilon_{\text{tol}}^{m}$ are convergence tolerances. The $\boldsymbol{S}_{k+1}$ are the eigenvectors of $\boldsymbol{P}\boldsymbol{\xi}_{k+1}$ and are used to transform $\boldsymbol{A}_{k+1}$ into the same basis as $\boldsymbol{P}\boldsymbol{\xi}_{k+1}$. An example of the algorithm

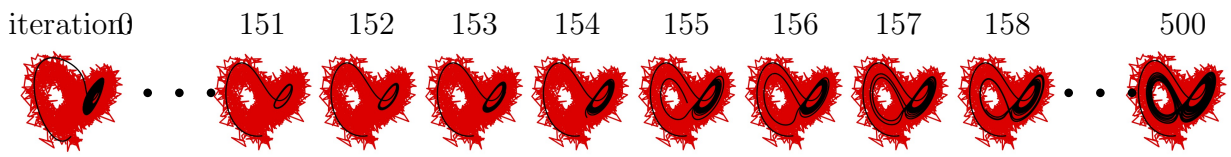


Figure 5.5: Illustration of trapping SINDy progress on noisy Lorenz data. The minimization results in the transition from a poor initial guess to identification of the correct attractor dynamics.

iterating on noisy data from the chaotic Lorenz system is shown in Fig. 5.5, demonstrating how the algorithm can transition from a poor initial guess that decays to a fixed point to a stable model converging to the correct attractor. An optional FISTA method [28, 316] is also implemented for reducing the convergence time in the $(\boldsymbol{m}, \boldsymbol{A})$ optimization. Algorithm 1 is computationally intensive compared to the traditional SINDy method, but it can be parallelized for speed in future work, following other SINDy variants [201]. Initial guesses are allowed for $\boldsymbol{m}$ and $\boldsymbol{A}$ in order to facilitate continuation of previous optimization runs. Along with these methods, the the λ_1 variant of the trapping algorithm in Eq. (5.20) is also implemented in the open-source PySINDy code [104, 210] described in Chapter 6.

A key insight to the trapping algorithm is that the energy-preserving constraint $\boldsymbol{C}_s\boldsymbol{\xi} = \boldsymbol{d}$ is non-negotiable. Although in practice small errors in $\boldsymbol{C}_s\boldsymbol{\xi} = \boldsymbol{d}$ do not significantly affect the optimization problem, the $\|\boldsymbol{P}\boldsymbol{\xi} - \boldsymbol{A}\|_2^2$ term in the optimization loses its physical interpretation if the coefficients are not exactly energy-preserving. Thus, the goal is to satisfy $\boldsymbol{C}_s\boldsymbol{\xi} = \boldsymbol{d}$ *exactly*, and then to push a potential model towards a more refined model that exhibits a trapping region, potentially at the expense of the fit to the data (this can also mitigate overfitting).

With regards to choosing hyperparameters, there is some work to do for each new problem. Fortunately, the results in Sec. 5.6 are fairly insensitive to the exact hyperparameter values. Rather, a common occurrence in sparse regression techniques is rediscovered – the existence of cutoffs in the hyperparameter space where the model quality sharply drops. These sharp boundaries are actually grounded in reality; if sparsity-promotion is increased

in the regression, it will eventually start to truncate out the smallest physical scales in the system, and at "large enough" values, it truncates the primary dynamics of interest. An example of this behavior was already illustrated in Sec. 5.2 and Fig. 5.3 when the threshold value is larger than the HIT-SI injector frequency. Another example of this behavior is shown in Sec. 6.2.1

More specifically for Algorithm 1, if the system has some small scales, poor choices of λ (sparsity-promotion) or γ_A (smallest eigenvalue of $\boldsymbol{A}^S$) can truncate these scales during the regression. A reasonable strategy, assuming no prior system knowledge, is to start with $\lambda = 0$, $\gamma_A \sim 0$, $\zeta \gg 1$, and then scan the values. For ζ (the strength of the long-term boundedness term), there tends to be a "sweet spot" regime. If $\zeta^{-1}\|\boldsymbol{P}\boldsymbol{\xi} - \boldsymbol{A}\|_2^2 \ll \|\boldsymbol{\Theta}\boldsymbol{\xi} - \dot{\boldsymbol{X}}\|_2^2$, then $\boldsymbol{\xi}^*$ is essentially unaffected by the minimizations over $\boldsymbol{m}$ and $\boldsymbol{A}$. In practice, this means that poor initial guesses for $\boldsymbol{\xi}^*$ do not improve as the full optimization problem is solved. In the opposite extreme, $\zeta^{-1}\|\boldsymbol{P}\boldsymbol{\xi} - \boldsymbol{A}\|_2^2 \gg \|\boldsymbol{\Theta}\boldsymbol{\xi} - \dot{\boldsymbol{X}}\|_2^2$, the optimization in Eq. (5.21) is increasingly nonconvex and potentially pulls $\boldsymbol{\xi}^*$ far away from the data. Finding the ζ regime where updating $\boldsymbol{m}$ perturbs $\boldsymbol{\xi}^*$, instead of leaving $\boldsymbol{\xi}^*$ unaffected or mangled, requires scanning ζ. Because each problem requires some data-craftsmanship, there are plans to parallelize the algorithm to efficiently scan large ranges in the hyperparameters. Finally, prior system knowledge can also constrain the hyperparameter space.

5.6 Trapping SINDy results

The utility of the trapping SINDy algorithm to identify stable, sparse, nonlinear models for a number of canonical fluid and plasma systems is now investigated. These examples illustrate that it is possible to both effectively discover stable models that exhibit trapping regions and improve the discovery of systems that do not satisfy Thm. 1 or the requirement of effective nonlinearity. For each system, SINDy is trained on a single trajectory with a random initial condition and evaluate the model on a different trajectory of the same temporal duration with a new random initial condition. It is difficult to quantity model performance for chaotic systems, such as the Lorenz system, where lobe-switching is extremely sensitive to initial

conditions and the coefficients of the identified model, and for systems with transients, for which the precise timing of instability must be matched to achieve the correct phase. Two reasonable definitions for the model quality are the relative Frobenius error in the model coefficients (for models with closed forms) and the time-averaged error in $\dot{\boldsymbol{X}}$,

$$\mathrm{E_m} = \frac{\|\boldsymbol{\Xi}_{\mathrm{True}} - \boldsymbol{\Xi}_{\mathrm{SINDy}}\|_{\mathrm{F}}}{\|\boldsymbol{\Xi}_{\mathrm{True}}\|_{\mathrm{F}}}, \tag{5.33}$$

$$\mathrm{E_f} = \frac{\overline{\|\dot{\boldsymbol{X}}_{\mathrm{True}} - \dot{\boldsymbol{X}}_{\mathrm{SINDy}}\|^2}}{\overline{\|\dot{\boldsymbol{X}}_{\mathrm{True}}\|^2}}. \tag{5.34}$$

It should be understood that the time-average in $\mathrm{E_f}$ is computed after dividing the numerator and denominator. When appropriate, the far more demanding relative prediction error is reported,

$$\mathrm{E_{pred}} = \frac{\|\boldsymbol{X}_{\mathrm{True}} - \boldsymbol{X}_{\mathrm{SINDy}}\|_F}{\|\boldsymbol{X}_{\mathrm{True}}\|_F}. \tag{5.35}$$

Table 5.2 summarizes the sampling, hyperparameters, and identified trapping regions for each example discussed in Sec. 5.6. Table 5.2 is intended to be instructive rather than exhaustive. For clarity, the training and testing trajectories used to generate this table do not have added noise, although Fourier modes from the Burgers' Equation and POD modes from the Von Kàrmàn street are obtained from direct numerical simulation (DNS), and subsequently contain minor numerical noise; the performance on noisy data will also be somewhat investigated in an example below. In the example below and additional examples in Chapter 6, noise is added to the training data to complicate some machine learning task. In all cases in this work, noise amplitudes are presented as percentages of the training data root-mean-square error (RMSE). In other words, zero-mean Gaussian noise is added to every training data point, with variance equal to a fraction of the training data RMSE. So added noise of $\mathcal{N}(0, 0.2 \times \mathrm{RMSE_{train}})$ is reported as 20% noise. When using experimental data, there are different sources of noise which may not be Gaussian.

To compare trapping region sizes R_m across different examples, $R_{\mathrm{eff}} = R_m/\sqrt{\sum_{i=1}^{r} \overline{y}_i^2}$

System	r	Δt	M	λ	ζ	γ_A	$\boldsymbol{m}^*$	R_m	R_{eff}	λ_1	E_m	E_f
Mean field	3	0.01	50K	0	10^{10}	1	[0,0,1.3]	1.3	218	-1	10^{-5}	10^{-12}
Atm. osc.	3	0.005	50K	0	10^8	0.1	[0,−0.9,0.4]	300	597	-0.01	10^{-4}	10^{-7}
Lorenz	3	0.005	50K	0	0.1	1	[−1.2,0.1,38]	106	4.4	-1	10^{-3}	10^{-5}
Triad MHD	6	0.001	50K	0	10^3	0.1	[0,...,0]	–	–	0	10^{-6}	10^{-10}
Burgers' Eq.	10	0.1	30K	0	500	0.1	[−0.2,0,...]	–	–	0.1	–	10^{-3}
Cyl. wake	5	0.1	30K	0.1	1	0.1	[−1.2,...,1.1]	29	17	-0.1	–	10^{-3}

Table 5.2: Description of the sampling, trapping SINDy hyperparameters, and identified trapping region for the dynamic systems examined in Section 5.6. Trajectory data does not include any added noise so $\lambda = 0$ works for most of the systems. The SINDy models are identified from a single trajectory. These parameters produce reasonable results for these systems, but a hyperparameter scan can lead to further improvements. The errors in the last two columns are approximate up to $\mathcal{O}(1)$ factors.

is reported, which is normalized to the approximate radius of the training data. The denominator denotes the root-mean-square of the temporal average of each component of the trajectory.

Mean field model

Often the trajectories of a nonlinear dynamical system, which has a linear part exhibiting some stable directions, will approach a slow manifold of reduced dimension with respect to the full state space. As an example of this behavior, consider the following linear-quadratic system originally proposed by Noack et al. [320] as a simplified model of the von Kàrmàn vortex shedding problem explored further in Sec. 5.6:

$$\frac{d}{dt}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \mu & -1 & 0 \\ 1 & \mu & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} -xz \\ -yz \\ x^2+y^2 \end{bmatrix}. \tag{5.36}$$

Systems of this form commonly arise in PDEs with a pair of unstable eigenmodes represented by x and y. The third variable z models the effects of mean-field deformations due to

nonlinear self-interactions of the instability modes. The system undergoes a supercritical Hopf bifurcation at $\mu = 0$; for $\mu \ll 1$ trajectories quickly approach the parabolic manifold defined by $z = x^2 + y^2$. All solutions asymptotically approach a stable limit cycle on which $z = x^2 + y^2 = \mu$. It is enough to notice that $\boldsymbol{m} = [0, 0, \mu + \epsilon]$, $\epsilon > 0$ produces

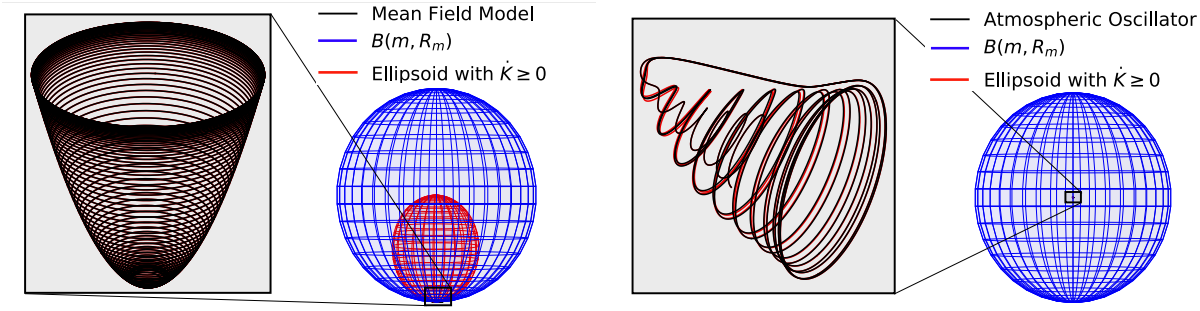
$$\boldsymbol{A}^S = \boldsymbol{L}^S - \boldsymbol{m}^T \boldsymbol{Q} = \begin{bmatrix} -\epsilon & 0 & 0 \\ 0 & -\epsilon & 0 \\ 0 & 0 & -1 \end{bmatrix}, \tag{5.37}$$

so this system exhibits a trapping region. A stable and accurate model is illustrated and identified by the trapping SINDy algorithm in Fig. 5.6a.

This system is of particular interest because it is a prototypical example of how quadratic interactions in a multi-scale system can give rise to effective higher-order nonlinearities. If the dynamics are restricted to the slow manifold, the system reduces to the cubic Hopf normal form [320, 176]

$$\frac{d}{dt} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \mu - (x^2 + y^2) & -1 \\ 1 & \mu - (x^2 + y^2) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \tag{5.38}$$

Systems of this type arise in weakly nonlinear pattern-forming systems and are often called Stuart-Landau equations. In this case, the nonlinear interactions are no longer energy-preserving, since the manifold restriction removes the fast, dissipative degree of freedom. One might intuitively expect that this type of manifold reduction would inherit the trapping properties of the underlying system, but it is unclear if a general theory of such situations has been worked out, even for the quadratic energy-preserving case.



(a) Trapping SINDy model (black) of a mean field system trajectory (red) with $\mu = 0.01$ and initial condition $[\mu, \mu, 0]$. The trajectory is shown within the estimated trapping region and ellipsoid where $\dot{K} \geq 0$. The prediction error is $\mathrm{E}_{\mathrm{pred}} \approx 0.6\%$.

(b) Same illustration for the atmospheric oscillator with random initial condition chosen from the unit ball. There is large scale separation in this system, so that $|\lambda_1| \ll |\lambda_2|, |\lambda_3|$. This leads to an overestimate of the trapping region size. The prediction error is $\mathrm{E}_{\mathrm{pred}} \approx 6\%$.

Figure 5.6: Identified models and trapping regions for the mean field and atmospheric oscillator systems.

Atmospheric oscillator model

A more complicated Lorenz-like system of coupled oscillators, motivated from atmospheric dynamics, is now examined:

$$\frac{d}{dt}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \mu_1 & 0 & 0 \\ 0 & \mu_2 & \omega \\ 0 & -\omega & \mu_2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} \sigma xy \\ \kappa yz + \beta z^2 - \sigma x^2 \\ -\kappa y^2 - \beta yz \end{bmatrix}. \tag{5.39}$$

For comparison, the parameters in Tuwankotta et al. [430], $\mu_1 = 0.05$, $\mu_2 = -0.01$, $\omega = 3$, $\sigma = 1.1$, $\kappa = -2$, and $\beta = -6$, are adopted, for which a limit cycle is known to exist. The trapping SINDy algorithm finds $\boldsymbol{m}$ such that $\boldsymbol{A}^S$ is negative definite for a wide range of parameter and hyperparameter choices, and accurate model results are illustrated in Fig. 5.6b alongside the mean-field model results.

So far, the trapping algorithm has successfully produced accurate and provably stable

models on simple systems that exhibit well-behaved attractors. In the next sections, noisier and higher-dimensional systems are investigated, which typically provide significant challenges for model discovery algorithms.

Noisy Lorenz attractor

The Lorenz 1963 system [261] is among the simplest systems exhibiting chaotic dynamics, developed to model thermal convection in the atmosphere based on computer simulations by Ellen Fetter and Margaret Hamilton:

$$\frac{d}{dt}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -\sigma & \sigma & 0 \\ \rho & -1 & 0 \\ 0 & 0 & -\beta \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} 0 \\ -xz \\ xy \end{bmatrix}. \tag{5.40}$$

For this system, it is possible to write $\boldsymbol{A}^S$ explicitly as

$$\boldsymbol{A}^S = \begin{bmatrix} -\sigma & \frac{1}{2}(\rho+\sigma-m_3) & \frac{1}{2}m_2 \\ \frac{1}{2}(\rho+\sigma-m_3) & -1 & 0 \\ \frac{1}{2}m_2 & 0 & -\beta \end{bmatrix}. \tag{5.41}$$

For Lorenz's choice of parameters, $\sigma = 10$, $\rho = 28$, $\beta = 8/3$, this system is known to exhibit a stable attractor. For $\boldsymbol{m} = [0, m_2, \rho+\sigma]$ (m_1 does not contribute to $\boldsymbol{A}^S$ so it is set to zero),

$$\boldsymbol{A}^S = \begin{bmatrix} -\sigma & 0 & \frac{1}{2}m_2 \\ 0 & -1 & 0 \\ \frac{1}{2}m_2 & 0 & -\beta \end{bmatrix}, \tag{5.42}$$

$$\lambda_1 = -1, \quad \lambda_\pm = -\frac{1}{2}\left[\beta+\sigma \mp \sqrt{m_2^2+(\beta-\sigma)^2}\right],$$

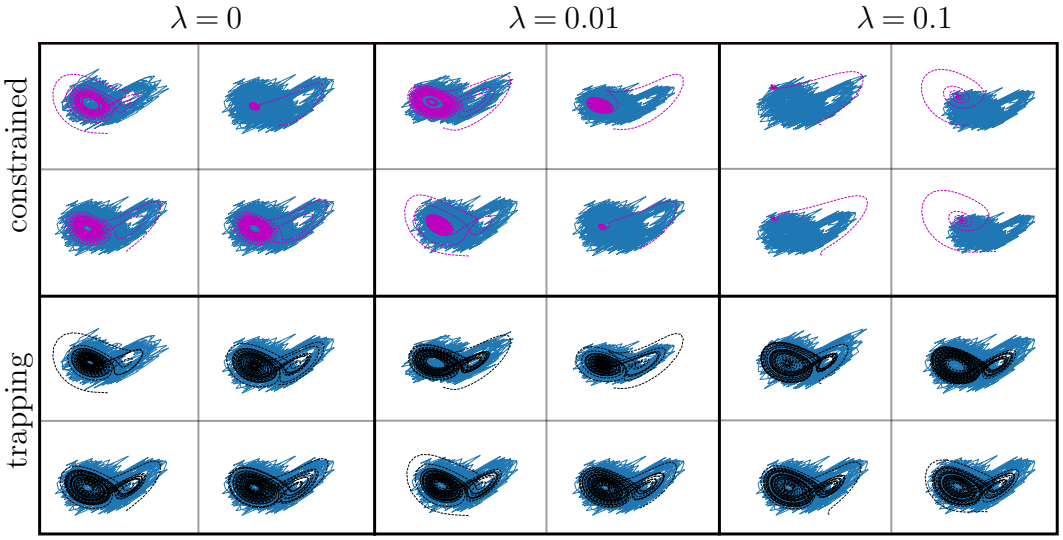


Figure 5.7: Comparison between the constrained SINDy (magenta) and trapping SINDy (black) results for the Lorenz system using three different values of the sparsity-promotion strength λ. Unconstrained SINDy results are not pictured because most of the models diverge. Each model is trained on a single Lorenz attractor with noise sampled from $\mathcal{N}(0, 4)$ ($\sim 20 - 30\%$ noise) and an initial condition of $[1, -1, 20]$ (blue). The illustrations depict the model performance on data evolved from four random initial conditions between $[-10, 10]$ (this testing data is not shown but the attracting set is unchanged). Trapping SINDy produces stable models that follow the underlying attractor for all values of λ.

so that if $\lambda_\pm < 0$, then $-2\sqrt{\sigma\beta} < m_2 < 2\sqrt{\sigma\beta}$. The algorithm successfully identifies the optimal $\boldsymbol{m}$, and identifies the inequality bounds on m_2 for stability. As this analysis is invariant to m_1, in principle the trapping region is given by a cylinder, extruded in the m_1 direction, rather than a sphere.

Further improvements in model quality can be obtained. Unconstrained, constrained, and trapping SINDy models are trained four times; the data for each is a single Lorenz attractor with four different noise instantiations. Then the performance of the resulting models is tested with a random initial condition in $[-10, 10] \times [-10, 10] \times [-10, 10]$. For direct comparison, the L^1 regularizer is used for each method. Fig. 5.7 illustrates the increased performance with the trapping SINDy algorithm over the constrained SINDy algorithm on noisy Lorenz data for varying threshold levels $\lambda = \{0, 0.01, 0.1\}$. The unconstrained method is not pictured because most of the identified models diverge at these high noise levels. At

all values of λ and most initial conditions, the unconstrained method overfits to the data and produces unstable and diverging models. The traditional constrained SINDy variant mostly manages to produce stable models but produces increasingly poor data fits as λ increases. In contrast, the trapping version continues to produce stable models that lie on the correct attractor. In this way, the additional optimization loss terms that promote stable models provide both a trapping region of known size and additional robustness to noise, even when the models appear otherwise stable, as with many of the constrained SINDy models that incorrectly decay to a fixed point.

Triadic MHD model

Magnetohydrodynamic systems exhibit quadratic nonlinearities that are often energy-preserving with typical boundary conditions. Consider a simple model of the nonlinearity in two-dimensional incompressible MHD, which can be obtained from Fourier-Galerkin projection of the governing equations onto a single triad of wave vectors. For the Fourier wave vectors $\boldsymbol{k}_1 = (1, 1)$, $\boldsymbol{k}_2 = (2, -1)$, and $\boldsymbol{k}_3 = (3, 0)$ and no background magnetic field, the Carbone and Veltri [75] system is

$$\begin{bmatrix} \dot{u}_1 \\ \dot{u}_2 \\ \dot{u}_3 \\ \dot{B}_1 \\ \dot{B}_2 \\ \dot{B}_3 \end{bmatrix} = \begin{bmatrix} -2\nu & 0 & 0 & 0 & 0 & 0 \\ 0 & -5\nu & 0 & 0 & 0 & 0 \\ 0 & 0 & -9\nu & 0 & 0 & 0 \\ 0 & 0 & 0 & -2\eta & 0 & 0 \\ 0 & 0 & 0 & 0 & -5\eta & 0 \\ 0 & 0 & 0 & 0 & 0 & -9\eta \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ B_1 \\ B_2 \\ B_3 \end{bmatrix} + \begin{bmatrix} 4(u_2u_3 - B_2B_3) \\ -7(u_1u_3 - B_1B_3) \\ 3(u_1u_2 - B_1B_2) \\ 2(B_3u_2 - u_3B_2) \\ 5(u_3B_1 - B_3u_1) \\ 9(u_1B_2 - B_1u_2) \end{bmatrix}, \tag{5.43}$$

where as usual ν is the viscosity and η is the resistivity. Without external forcing, this system is stable, dissipating to zero, so the inviscid limit $\nu = \eta = 0$ is considered. The system is now Hamiltonian and the algorithm correctly converges to $\boldsymbol{m} = 0$, $\boldsymbol{A}^S = 0$. The results in Fig. 5.8 provide a useful illustration that trapping SINDy converges to stable

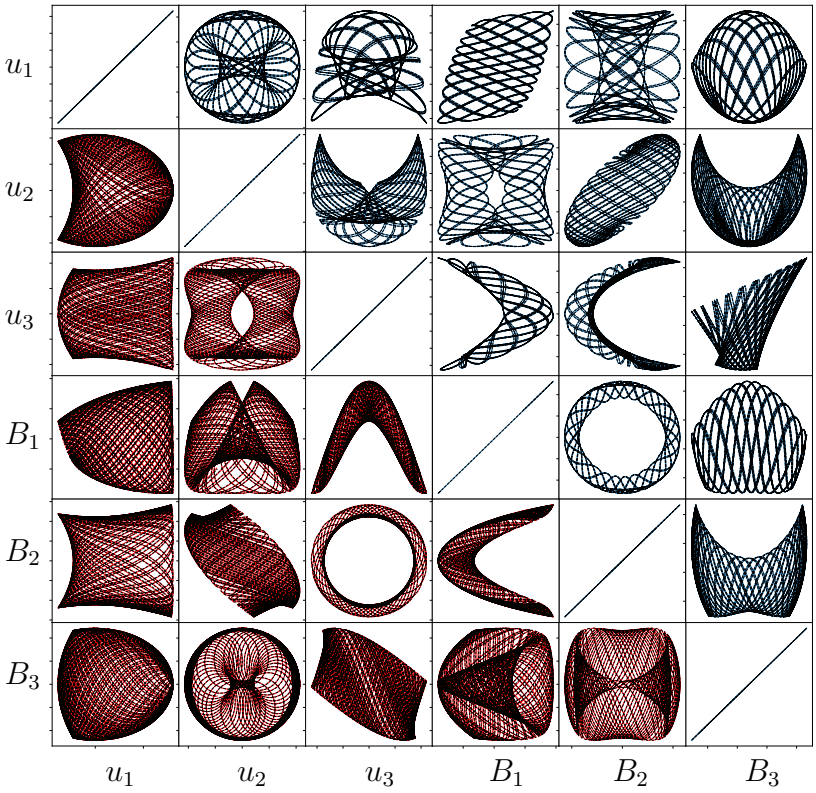


Figure 5.8: The triad model for 2D inviscid MHD training data (blue, upper triangle) and a trapping SINDy model (black) capturing Hamiltonian dynamics on testing data (red, lower triangle).

energy-preserving models even when the trapping theorem is not satisfied. These results also provide a reminder that there are a large number of dynamical systems beyond fluids, such as MHD, which may benefit from these types of techniques. The reason the algorithm converges to the correct behavior is because it is still minimizing $\dot{K}$; in this case trapping SINDy converges to $\dot{K} \approx 0$ and can make no further improvement.

Forced Burgers' equation

The viscous Burgers' equation has long served as a simplified one-dimensional analogue to the Navier-Stokes equations [67, 180]. The forced, viscous Burgers' equation on a periodic

domain $x \in [0, 2\pi)$ is:

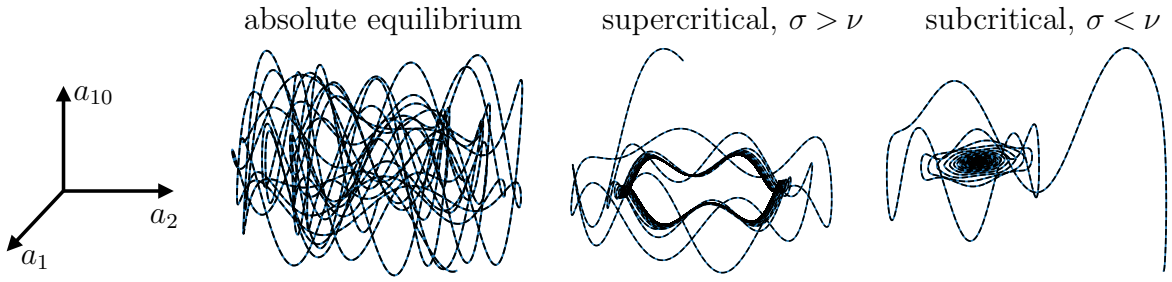
$$\frac{d}{dt}q = -(U_q + q)\partial_x q + \nu\partial_{xx}^2 q + g_q(x,t), \tag{5.44}$$

where the constant U_q models mean-flow advection. Now, this system is projected onto a Fourier basis and constant forcing is assumed to act on the largest scale, i.e., $g_q(x,t) = \sigma\left(a_1(t)e^{ix} + a_{-1}(t)e^{-ix}\right)$, as in Noack et al. [322]. After Fourier projection, the evolution of the coefficients $a_k(t)$ is given by the Galerkin dynamics

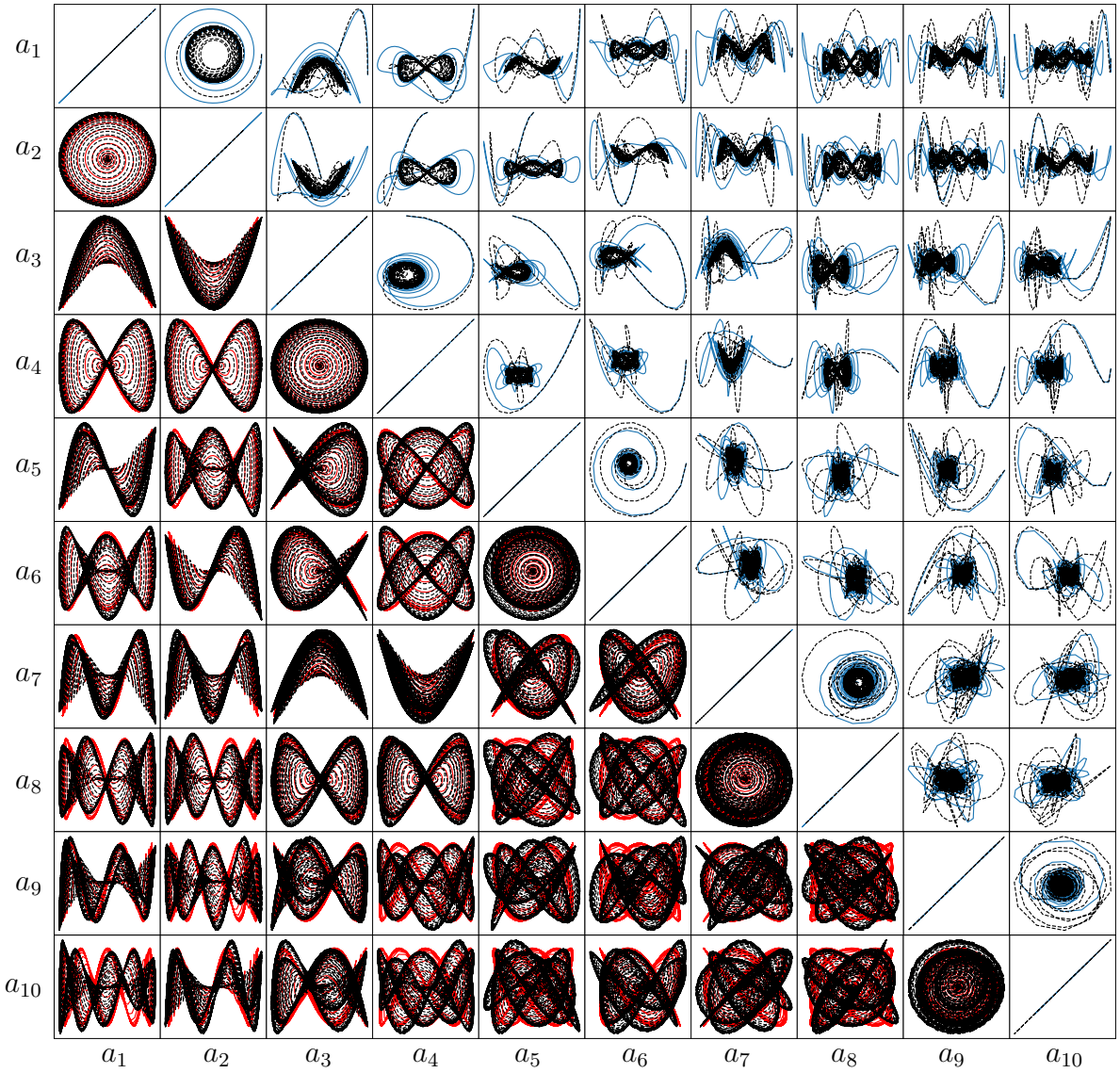
$$\dot{a}_k = \left(\delta_{|k|1}\sigma - \nu k^2 - ikU_q\right)a_k - \sum_{\ell=-r}^{r} i\ell a_\ell a_{k-\ell}. \tag{5.45}$$

In the subcritical case $\sigma < \nu$, the origin of this system is stable to all perturbations and all solutions decay for long times. However, in the supercritical case $\sigma > \nu$, the excess energy input from the forcing cascades to the smaller dissipative scales. The "absolute equilibrium" limit $\sigma = \nu = 0$ has a Hamiltonian structure; for long times the coefficients approach thermodynamic equilibrium and equipartition of energy [270]. This structure does not correspond to any physical behavior of the Navier-Stokes equations, although it does approximate some properties of the inviscid Euler equations [230]. Due to its rich dynamics, this modified Burgers' equation has also been investigated in the context of closure schemes for Galerkin models [322]. The PDE in Eq. (5.44) is simulated with a high-resolution Godunov-type finite volume method using a van Leer flux limiter, implemented in the open-source Clawpack solver [93].

The model performance is illustrated in Fig. 5.9a for the subcritical case with $\sigma = 0.01$ and $\nu = 0.025$, the supercritical case with $\sigma = 0.1$ and $\nu = 0.025$, and the absolute equilibrium. In all cases $U_q = 1$. For the subcritical condition, all the eigenvalues of $\boldsymbol{L}^S$ are negative, and thus the algorithm finds stable models. For the supercritical condition $\sigma > \nu$, there is some subtlety. The algorithm does not converge to a negative definite $\boldsymbol{A}^S$, although it finds a solution with $\dot{K} \leq 0$. As mentioned in Section 5.4.2, this system does not exhibit effective nonlinear-



(a) Trapping SINDy model (black) for the modified Burgers' equation in the three dynamic regimes. For improved illustration, the ground truth data (blue) is generated from the 10D Noack et al. [322] model rather than DNS.



(b) Temporal evolutions of each (a_i, a_j) pair for $i, j = 1, ..., 10$ obtained from DNS training data (blue, upper triangle), DNS testing data (red, lower triangle), and trapping SINDy prediction on both DNS datasets (black). The trapping algorithm struggles a bit with the transients, but obtains the correct attractor behavior.

ity. This lack of effective nonlinearity was also true for the MHD example in Section 5.6, since the initial condition with no magnetic field perturbation, $B_1(0) = B_2(0) = B_3(0) = 0$, remains on the purely hydrodynamic manifold. In the inviscid limit, one does not need to consider this subspace because the system already does not satisfy the trapping theorem by virtue of being Hamiltonian. Lastly, in the absolute equilibrium regime the trapping SINDy algorithm correctly identifies vanishing eigenvalues of $\boldsymbol{A}^S$. In practice, excellent models are found for all of the aforementioned systems and for all practical purposes these models are typically stable, regardless of effective nonlinearity or Hamiltonian dynamics, because the SINDy trapping algorithm is minimizing $\dot{K}$. However, without effective nonlinearity, there is no guarantee to produce a stable model for every possible initial condition.

An illustration is provided in Fig. 5.9b for the $r = 10$ model built from the DNS data in the supercritical regime with $\sigma = 0.1$, $\nu = 0.025$. It struggles a bit with the transient but otherwise the performance is accurate. Part of the reason for the poor fit to the transient is that $\lambda = 0$ is used here. The biasing towards stability appears to mitigate some of the need for sparsity-promotion; in other words, sparsity-promotion is not necessarily needed to produce a stable model, but may be needed for a more accurate or interpretable model, since the number of coefficients in Q_{ijk} is $\mathcal{O}(r^3)$ despite the constraints. Using finite λ may improve the model further, especially the transients. Next, the results section is concluded by addressing the challenging von Kàrmàn vortex shedding behind a circular cylinder.

Von Kàrmàn vortex street

The fluid wake behind a bluff body, characterized by a periodic vortex shedding phenomenon known as a von Kàrmàn street, is investigated. The two-dimensional incompressible flow past a cylinder is a stereotypical example of such behavior, and has been a benchmark problem for Galerkin models for decades [320]. The transition from a steady laminar solution to vortex shedding is given by a Hopf bifurcation, as a pair of eigenvalues of the linearized Navier-Stokes operator cross the real axis.

The transient energy growth and saturation amplitude of this instability mode is of

particular interest and has historically posed a significant modeling challenge. Early Galerkin models of vortex shedding, based on a POD expansion about the mean flow, captured the oscillatory behavior but were structurally unstable [106]. This was later resolved by Noack et al. [320], who recognized that the transient behavior could be explained by Stuart-Landau nonlinear stability theory, in which the unsteady symmetric flow is deformed to the neutrally stable mean flow via a nonlinear self-interaction of the instability mode. In that work, an 8-mode POD basis was augmented with a ninth "shift mode" parameterizing this mean flow deformation. This approach was later formalized with a perturbation analysis of the flow at the threshold of bifurcation [402].

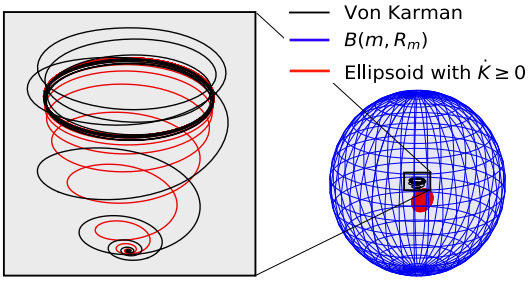
This modification encodes the intuition that the dynamics take place on the parabolic manifold associated with the Hopf bifurcation; without it, the energy quadratic models tends to overshoot and oscillate before approaching the post-transient limit cycle. Nevertheless, the 9-mode quadratic Galerkin model does resolve the transient dynamics, nonlinear stability mechanism, and post-transient oscillation, accurately reproducing all of the key physical features of the vortex street. Moreover, in Schlegel and Noack [390] stability of the quadratic model was proven with $m_9 = m_{\text{shift}} = \epsilon,\ \epsilon > 1$, and $m_i = 0$ for $i = \{1, ..., 8\}$. Recall from the discussion in Section 5.4.2 that POD-Galerkin models will generally weakly satisfy the effective nonlinearity criteria and it is unclear if the shift-mode complicates this picture.

Although the POD-Galerkin model is an accurate description of the flow past a cylinder, it is an intrusive model, in the sense that evaluating the projected dynamics requires evaluating individual terms in the governing equations, such as spatial gradients of the flow fields. POD-Galerkin models therefore tend to be highly sensitive to factors including mesh resolution, convergence of the POD modes, and treatment of the pressure and viscous terms. Recent work by Loiseau et al. [257, 259, 258] has bypassed the Galerkin projection step by using the SINDy algorithm to directly identify the reduced-order dynamics. This approach has been shown to yield compact, accurate models for low-dimensional systems ($r = 2$ or 3), but preserving accuracy and stability for higher-dimensional systems remains challenging. Higher-dimensional regression problems often become ill conditioned; for example, in the

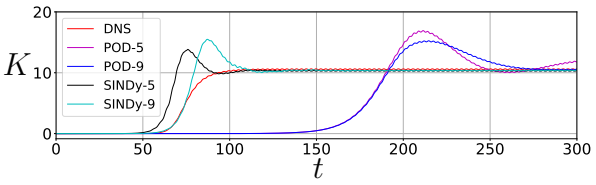
cylinder wake example, the higher modes 3-8 are essentially harmonics of the driving modes 1-2, and so it is difficult to distinguish between the various polynomials of these modes during regression. Because these higher harmonics are driven by modes 1-2, the 3D constrained quadratic SINDy model with modes 1-2 plus the shift mode from Loiseau et al. [257] already performs well enough to capture the energy evolution with minor overshoot and correct long-time behavior. Details of the DNS and the POD-Galerkin technique used to reproduce the 9D shift-mode model can be found in the original work [209].

With the trapping SINDy algorithm, new 5-dimensional and 9-dimensional models for the cylinder wake are obtained and compared against the same-size analytic POD-Galerkin models. The 5D trapping SINDy model is provably stable and the identified trapping region is illustrated in Fig. 5.10a. Additionally, the 5D SINDy and 9D POD-Galerkin models are compared in Fig. 5.10c. The 5D trapping SINDy model outperforms the 9D POD-Galerkin model by significantly improving the transient and improving the identification of the long-term attractor. For the 9D trapping SINDy model, the largest eigenvalue of $\boldsymbol{A}^S$ was reduced to $\mathcal{O}(10^{-2} - 10^{-4})$ but the algorithm was unable to produce accurate trapping SINDy models with fully negative definite $\boldsymbol{A}^S$. In practice, these models are functionally stable; a large set of random initial conditions were tested and no unbounded trajectories were discovered. Further searching in the hyperparameter space, or more algorithm iterations for better convergence, could potentially produce fully stable models.

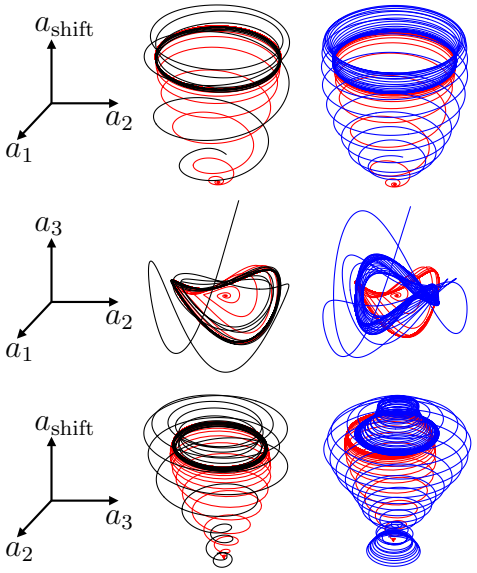
Despite this setback, the 9D trapping SINDy model performs quite well. The Galerkin model and the trapping SINDy model exhibit comparable performance and the SINDy model improves the transient prediction. The energies in Fig. 5.10b illustrate convergence to the true fluid flow energy for all the SINDy and POD-Galerkin models, with only the 9D trapping SINDy model capturing the precise timing of the transient. The flow reconstructions in Fig. 5.10d are quite accurate for both models. This is surprisingly strong performance with SINDy; recall that: 1) the Galerkin model is far more invasive a procedure than SINDy, requiring computation of spatial derivatives and inner products from the DNS, 2) the Galerkin model can still be quite sensitive to the DNS data, boundary conditions, and mesh size, and



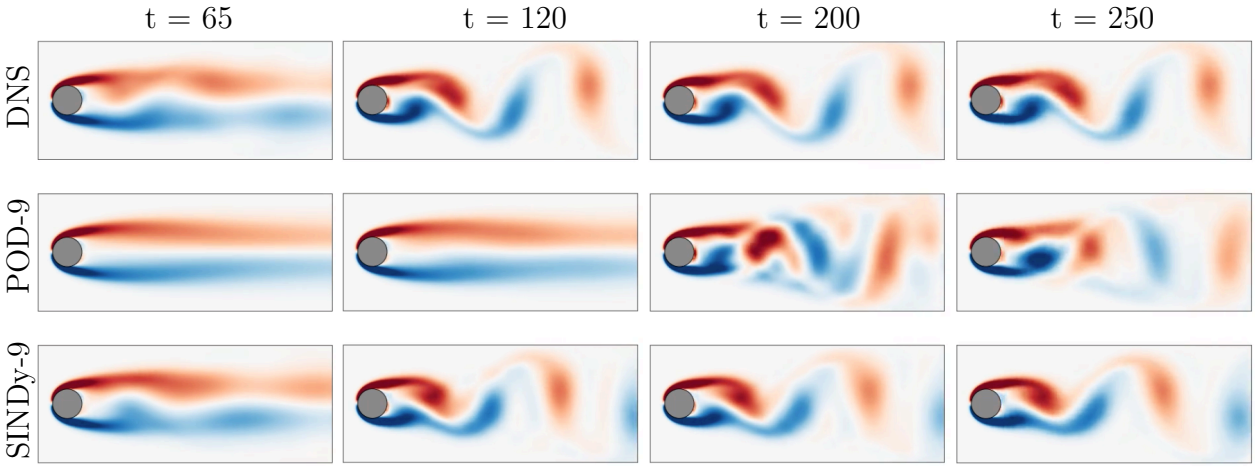
(a) Trapping SINDy 5D model (black) of a Von Kàrmàn trajectory (red). The trajectory is shown within the estimated trapping region and ellipsoid where $\dot{K} \geq 0$.



(b) Comparison of the energies for DNS and the 5 and 9 mode POD-Galerkin and trapping SINDy models.



(c) 5-mode trapping SINDy (black) and 9-mode POD-Galerkin (blue) models with a random initial condition, and the Von Kàrmàn trajectory used for training (red).



(d) Predictions of the vorticity field for the von Kàrmàn street at four snapshots in time, with a movie available [1]. The trapping SINDy model outperforms the 9D POD-Galerkin model, although an initial phase error in the trapping SINDy prediction (visible in the first snapshot) persists throughout the prediction.

Figure 5.10: Summary of the differences between DNS, POD-Galerkin models, and trapping SINDy models.

3) the 9D trapping SINDy model is far sparser and has far fewer "active" terms than the 9D POD-Galerkin model.

The difficulty in producing provably stable, 9D trapping SINDy models here appears to reveal an interesting optimization tradeoff. While sparsity-promotion tends to promote more accurate models and reduce the complexity of the nonconvex optimization problem (since there are fewer active terms to manage), it also deemphasizes the proposed metric for the strength of effective nonlinearity, S_e from Eq. (5.18), by reducing the values of unimportant model terms. For instance, the SINDy model here exhibits weak effective nonlinearity, $S_e \approx 10^{-5}$, compared with $S_e \approx 10^{-2}$ for the POD-Galerkin model. This small value of S_e may indicate increased difficulty in obtaining a fully negative definite $\boldsymbol{A}^S$. SINDy models with weaker sparsity-promotion exhibit larger S_e, but then it becomes exceedingly difficult to obtain accurate models in the nonconvex optimization problem. Without any sparsity-promotion this is an ill-conditioned, nonconvex optimization in a 330-dimensional space. In this way, there appears to be some tradeoff between sparsity-promotion and the strength of effective nonlinearity. Given these points, the sparse 5-mode and 9-mode SINDy models are promising first steps towards incorporating stability constraints into higher-dimensional data-driven models.

Before concluding, note that the eight-mode (no shift mode) POD-Galerkin model from Noack et al. [320], and all eight-mode models found by trapping SINDy, do not exhibit global stability. The problem fundamentally stems from the marginal stability of the mean flow and the very weak effective nonlinearity, both of which are somewhat addressed by the shift mode in the 9-mode model. This should be taken as a cautionary warning; success of these algorithms still relies on useful representations that capture the stability information of the underlying dynamics. This may require high-resolution data or alternative dynamic bases.

5.7 Trapping SINDy concluding remarks

This work developed physics-constrained system identification by biasing models towards fulfilling global stability criteria, and subsequently produces long-term bounded models with

no extra assumptions about the stability properties of equilibrium points and equilibrium trajectories. In order to produce globally stable models, a new trapping SINDy algorithm was implemented, based on the Schlegel-Noack trapping theorem [390]. Biasing models towards stability, and post-fit, proving that identified models are globally stable, will likely become increasingly important for both projection-based and data-driven models of fluids and plasmas. This approach, which relies on using the energy as a Lyapunov function for an entire class of models with fixed nonlinear structure, is challenging for application to higher-order nonlinearities where generic Lyapunov functions are often unknown. Fortunately, data-driven methods are now increasingly used to discover Lyapunov functions and barrier functions for nonlinear control [317, 216, 368, 228, 197, 416, 46, 84, 279]. These methods build a heuristic Lyapunov function for a given dataset, rendering the search for a Lyapunov function tractable but possibly at the cost of model generality.

The effectiveness of this optimization to identify stable models was demonstrated and additionally managed to improve the discovery of models that do not conform to the assumptions of the trapping theorem. The trapping SINDy algorithm resulted in more accurate and stable models for a range of systems, including simple benchmark problems, noisy data from chaotic systems, and DNS from full spatiotemporal PDEs. In these examples, it was found that this modified SINDy algorithm could effectively discover stable, accurate, and sparse models from significantly corrupted data. Even when an explicit stable trapping region was not found, improved stability was observed. Finally, relatively high-dimensional reduced-order models were explored, with $\mathcal{O}(10)$ degrees of freedom, which are typically challenging for unconstrained data-driven algorithms.

There is considerable future work for biasing machine learning methods to discover models that satisfy existence-style proofs of stability, especially those that require nonconvex optimization; the lack of convexity in the trapping SINDy algorithm seems to lead to deprecating algorithm speed and tractability as the size of the problem increases. There are many fluid flows which have known stable and unstable projection-based and data-driven reduced-order models, and which would benefit from a larger class of models with trapping

region guarantees. Future work should apply this methodology to heavily-researched systems such as the fluidic pinball [108, 358] and the lid-cavity flow [421, 262]. Other promising future work includes adapting this structure to many-body coupled Stuart-Landau equations for which stability theorems already exist [342]. However, the nonconvexity of this formulation may require adaptation to a deep learning approach for high-dimensional many-body problems that occur in fluids and modern neuronal models.

For all of the examples in this work, the trapping SINDy algorithm was trained on a single trajectory, although most data-driven methods can improve performance by processing data from multiple trajectories. Very large data can be effectively addressed with modern approaches, such as manifold Galerkin projection [258] and autoencoder [23, 296, 266, 81, 243] methods. These approaches may also address the significant Kolmogorov width limitations of linear transformations [350], and help ease the nonconvexity of the new optimization problem. There are also modern reduced-order modeling techniques, such as "lift & learn" [354], which produce quadratic ROMs regardless of the nonlinear structure of the underlying governing equations [147]. Similarly Koopman analysis aims to produce a map from the original state-space, where the dynamics are nonlinear, to a new coordinate system, typically infinite dimensional, where the dynamics become linear [229, 295, 222, 266, 249, 452, 417, 333].

However, adapting this methodology to alternative bases requires additional work to understand how the trapping theorem, or similar theorems, change under these (often nonlinear) coordinate transformations. For instance, Pan et al. [339] builds stable Koopman models by requiring that the real parts of the eigenvalues of the linear Koopman operator are non-positive, although the relationship between this linear stability and the trapping theorem is unclear. In related work, neural-network-based encoders are often used to reverse this mapping; encoders can input quadratically nonlinear fluid flow data and apply nonlinear transformations to find useful reduced-order models beyond what is capable with traditional projection-based methods [143]. A natural question that arises is: assuming the original energy-preserving, quadratically nonlinear fluid flow exhibits a trapping region, under what conditions does global stability hold in a new coordinate system given by $\boldsymbol{g}(\boldsymbol{y})$, for some

map $\boldsymbol{g} : \mathbb{R}^r \to \mathbb{R}^s$? The transformation could be an encoder, the reverse lifting map [354], or some other coordinate transform. Understanding how the stability properties manifest in the transformed system is a promising future direction for extending this stability theorem for ROMs with alternative dynamic bases.

Chapter 6

ROBUST SPARSE SYSTEM IDENTIFICATION WITH PYSINDY

Building physics priors into data-driven models with constraints has been illustrated in Sec. 5.1, and building stability priors into data-driven ROMs via new loss terms has been illustrated in Sec. 5.3. There are also many other ways to improve the robustness and quality of system-identified models: ensembling (sub-sampling) techniques [267, 366], Bayesian methods [456, 340, 319, 315, 173, 264, 299] and identification of weak form models [155, 156, 365, 366, 292, 291, 290]. Further, recent variants of the SINDy method are available that address systems with control inputs and model predictive control [204, 122], implicit ODEs [273, 201], and PDEs [376, 384]. This chapter details the incorporation of many of these advanced features into an easy-to-use and open-source code, and illustrates the use of these new features for advanced system identification across a number of scientific fields. A brief summary of the work presented in this chapter is also available as a pre-print [210].

6.1 Overview of the PySINDy Python code

A number of students and researchers at the University of Washington and elsewhere have collaboratively created the `PySINDy` code [104] as an open-source tool for applying the SINDy method. The original `PySINDy` package[1] was developed to identify a particular class of systems described by Eq. (5.1). In order to incorporate new developments with this technique, and accommodate the wide variety of possible dynamical systems, `PySINDy` has now been extended to a more general setting and added significant new functionality [210]. The code

[1] `https://github.com/dynamicslab/pysindy`

is thoroughly documented, contains extensive examples, and integrates a wide range of functionality, some of which may be found in a number of other local SINDy implementations[2]. In contrast to some of these existing implementations, `PySINDy` is completely open-source, adheres to software best-practices (for instance, providing unit tests and adhering to PEP8 stylistic standards), and minimally dependent on non-standard Python packages.

6.1.1 PySINDy features

The core object in the `PySINDy` package is the `SINDy` model class, which is implemented as a `scikit-learn` estimator. This design choice was made to ensure that the package is simple to use for a wide user base, as many potential users will be familiar with `scikit-learn`. It also expresses the `SINDy` model object at the appropriate level of abstraction so that users can embed it into more sophisticated pipelines in `scikit-learn`, such as for parameter tuning and model selection. The `PySINDy` implementation involves three major steps, resulting in three modeling decisions:

1. The *numerical differentiation* scheme used to approximate $\dot{\boldsymbol{X}}$, $\ddot{\boldsymbol{X}}$, ... from $\boldsymbol{X}$. For usage with PDEs, this also governs the approximation of spatial derivatives.

2. The candidate functions constituting the *feature library* $\boldsymbol{\Theta}$;

3. The *sparse regression* algorithm that is applied to solve (5.3) to find $\boldsymbol{\xi}$.

---

[2] `https://github.com/snagcliffs/PDE-FIND`,
`https://github.com/eurika-kaiser/SINDY-MPC`,
`https://github.com/dynamicslab/SINDy-PI`,
`https://github.com/SchatzLabGT/SymbolicRegression`,
`https://github.com/dynamicslab/databook_python`,
`https://github.com/sheadan/SINDy-BVP`,
`https://github.com/sethhirsh/BayesianSindy`,
`https://github.com/racdale/sindyr`,
`https://github.com/SciML/DataDrivenDiffEq.jl`,
`https://github.com/MathBioCU/WSINDy_PDE`,
`https://github.com/pakreinbold/PDE_Discovery_Weak_Formulation`,
`https://github.com/ZIB-IOL/CINDy`

The core `SINDy` object was designed to incorporate these three components in as modular a manner as possible. The SINDy object has one attribute corresponding to each component: `SINDy.differentiation_method` for numerical differentiation, `SINDy.feature_library` for the formation of the candidate function library, and `SINDy.optimizer` for the sparse regressor. `PySINDy` provides standard options and uses class inheritance for each step, making it easy to construct new sophisticated or "third-party" algorithms. In particular, at the time of writing, the methods in Fig. 6.1a have been implemented. In Fig. 6.1b the SINDy methods are organized by functionality.

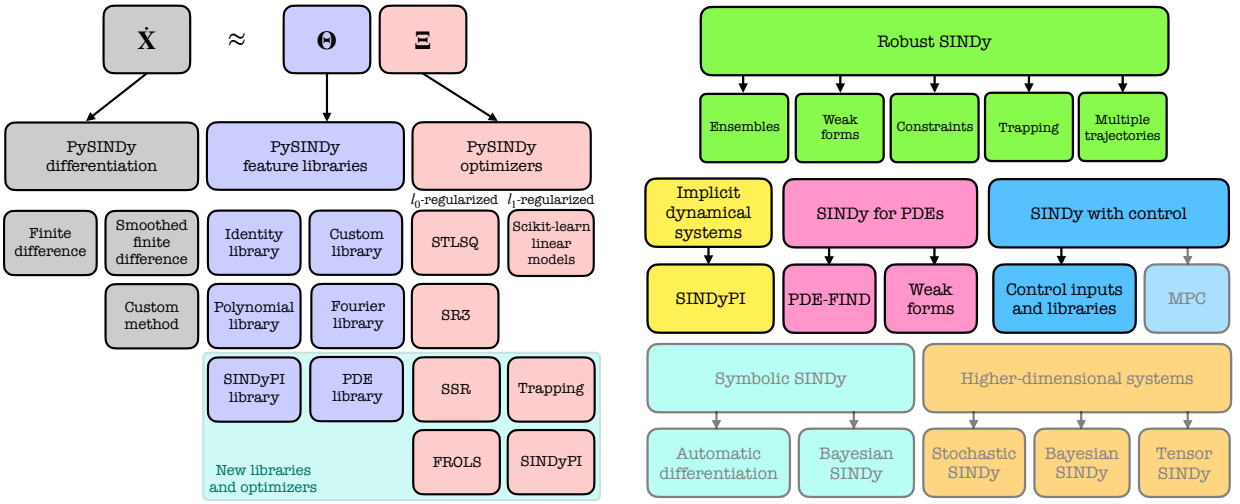
Given spatiotemporal data $\boldsymbol{q}(\boldsymbol{x}, t) \in \mathbb{R}^{M \times N}$, and optional control inputs $\boldsymbol{q}_u \in \mathbb{R}^{M \times N_u}$, `PySINDy` can now approximate algebraic systems of PDEs (and corresponding weak forms) in up to 3 spatial dimensions. The system is described by a function $\boldsymbol{g}$,

$$\boldsymbol{g}(\boldsymbol{q}, \dot{\boldsymbol{q}}, \boldsymbol{q}_x, \boldsymbol{q}_y, \boldsymbol{q}_{xx}, ..., \boldsymbol{q}_u) = \boldsymbol{0}. \tag{6.1}$$

ODEs, implicit ODEs, PDEs, and other dynamical systems are subsets of Eq. (6.1). Control terms and partial derivatives in the SINDy library can be accommodated by adding them as columns in $\boldsymbol{\Theta}(\boldsymbol{X})$, which becomes $\boldsymbol{\Theta}(\boldsymbol{X}, \dot{\boldsymbol{X}}, \boldsymbol{X}_x, ..., \boldsymbol{X}_u)$.

In addition, `PySINDy` has been extended to handle more complex modeling scenarios, including trapping SINDy for provably stable ODE models for fluids described in Sections 5.3−5.7, models trained using multiple dynamic trajectories, and the generation of many models with sub-sampling and ensembling methods for cross-validation and probabilistic system identification. In order to solve Eq. (6.1), `PySINDy` implements several different sparse regression algorithms. Greedy sparse regression algorithms, including step-wise sparse regression (SSR) [48] and forward regression orthogonal least squares (FROLS) [43], are now available. Figure 6.1 illustrates the `PySINDy` code structure, changes, and high-level goals for future work.

`PySINDy` includes extensive Jupyter notebook tutorials that demonstrate the usage of various features of the package and reproduce nearly the entirety of the examples from the



(a) (b)

Figure 6.1: Summary of SINDy features organized by (a) `PySINDy` structure and (b) functionality. (a) Hierarchy from the sparse regression problem solved by SINDy, to the submodules of `PySINDy`, to the individual optimizers, libraries, and differentiation methods implemented in the code. (b) Flow chart for organizing the SINDy variants and functionality in the literature. Bright color boxes indicate the features that have been implemented through this work, roughly organized by functionality. Semi-transparent boxes indicate features that have not yet been implemented.

original SINDy paper [59], trapping SINDy paper [209], and the PDE-FIND paper [376]. An extended example is included for the quasiperiodic shear-driven cavity flow [70].

The goal of the `PySINDy` package is to enable anyone with access to measurement data to engage in scientific model discovery. The package is designed to be accessible to inexperienced users, adhere to `scikit-learn` standards, include most of the existing SINDy variations in the literature, and provide a large variety of functionality for more advanced users. It is hoped that these changes will encourage researchers to use and contribute to the code in the future, pushing the boundaries of what is possible in system identification. For the remainder of this chapter, the various optimizers, candidate libraries, and other advanced functionality are explained with examples from the code. The chapter begins with a survey of the various

Optimizers	Constraints	PDEs	Ensemble	Control	Systems	Regularizers	Params
`sklearn`	×	✓	✓	✓	all	l_1, l_2	λ, (α)
`STLSQ`	×	✓	✓	✓	all	l_0, l_2	λ, (α)
`SR3`	✓	✓	✓	✓	all	l_0, l_1, l_2	λ, κ
SSR	×	✓	✓	✓	all	l_0, l_2	(κ), (α)
FROLS	×	✓	✓	✓	all	l_0, l_2	(κ), (α)
SINDyPI	×	✓	✓	✓	implicit	l_1, l_2	λ
Trapping	✓	✓	✓	✓	fluids	l_1, l_2	λ, ζ, γ_A

Table 6.1: Current capabilities of the SINDy optimizers implemented in `PySINDy`. Regularization with l_2 refers to the squared l_2 norm, i.e. ridge regression. The notation (κ) indicates the hyperparameter is optional. Bold font indicates a newly-implemented optimizer.

sparse regression optimizers.

6.2 A survey of sparse regression optimizers

The SINDy regression problem has been described in Eq. (5.3) and now a discussion is required to disambiguate between the wide range of sparse regression algorithms that can provide solutions to this regression problem. The original SINDy formulation solves Eq. (5.3) with the l_0 norm using a sequentially thresholded least squares (STLSQ) algorithm [59, 455]. Eq. (5.3) has also been solved with the l_0, l_1, and l_2 norms with many different sparse regression algorithms, including Lasso [423], sequentially thresholded ridge regression (STRidge) [376], sparse relaxed regularized regression (SR3) [458, 82], stepwise sparse regression (SSR) [48], Forward Regression Orthogonal Least Squares (FROLS) [43], blended conditional gradients (BCG) [76], or Bayesian methods [456, 340, 319, 315, 173, 264, 299]. In `PySINDy`, many of the optimizers allow for multiple regularizations, including l_0, l_1, l_2, as well as simple weighted variants that allow one to place stronger penalties on particular coefficients (more sophisticated weighted l_1 regularization can be found in Cortiella *et al.* [95]). In addition, users can use any of the objective formulations (`linear_model`) in `scikit-learn` (`Lasso` and `ElasticNet` use coordinate descent [446] and `Ridge` typically uses the conjugate gradient algorithm [399]).

The available optimizers are summarized in Table 6.1 and will be described in further detail below. The optimizers can be divided into those that were originally used to solve the nonconvex l_0-regularized problem, and those that were originally used to solve the l_1-regularized problem (the Lasso problem). For instance, although the SR3 algorithm can be used to solve the Lasso problem, SR3 fits more neatly into the former category because it was originally used to improve performance by relaxing the nonconvex setting. In order to illustrate the new optimizers, and before diving into the major changes required to accommodate PDEs in Sec. 6.3.4, the Lorenz 1963 equations (5.40) are used as the dynamical system.

6.2.1 l_0 based sparse regression algorithms

The l_0 norm is not strictly a norm; the l_0 regularizer simply returns the number of nonzero terms in the coefficients. Using the l_0 norm typically produces sparser solutions than using the l_1 norm. This tends to further lead to higher performance and more stable models, since there are no small-coefficient terms that can become active with new initial conditions or parameter regimes. The downside is that the l_0 norm transforms Eq. (5.3) into a nonconvex problem, for which only *local* convergence guarantees can be provided.

Sequentially-thresholded least-squares

The original SINDy paper [59] used the sequentially-thresholded least-squares algorithm, which has subsequentially been shown to converge rapidly to local minima [455]. This algorithm starts with a least-squares solution for $\boldsymbol{\xi}$ and then thresholds all coefficients that are smaller than the hyperparameter λ. The indices of the remaining non-zero coefficients are identified and then another least-squares solution for $\boldsymbol{\xi}$ is obtained on the remaining indices. The procedure is continued until the non-zero coefficients converge.

Sequentially-thresholded ridge regression

A common issue with solving Eq. (5.3) is that the candidate library $\boldsymbol{\Theta}$ is ill-conditioned. STRidge typically obtains better performance than STLSQ by improving the condition number of the linear system arising from the least squares problem. This is done by solving Eq. (5.3) with an extra Tikhonov regularization [142] term $\alpha\|\boldsymbol{\xi}\|_2^2$, with typical $\alpha \ll 1$. It was found empirically in Rudy *et al.* [376] to outperform STLSQ for PDE identification. STRidge is actually the default method that is called through the `STLSQ` optimizer in `PySINDy`, and only reduces to the true STLSQ method when $\alpha = 0$.

SR3

Sparse relaxed regularized regression, or SR3, is an attempt to provide a relaxation-type approach to the l_0 problem. The idea is to introduce a new auxiliary variable $\boldsymbol{W}$, and solve a relax-and-split [457] version of the problem,

$$\underset{\boldsymbol{\xi},\boldsymbol{W}}{\operatorname{argmin}} \|\dot{\boldsymbol{X}} - \boldsymbol{\Theta}(\boldsymbol{X})\boldsymbol{\xi}\|^2 + \kappa\|\boldsymbol{\xi} - \boldsymbol{W}\|_2^2 + \lambda\|\boldsymbol{W}\|_0. \tag{6.2}$$

In other words, perform the data fitting and the thresholding on two separate variables $\boldsymbol{\xi}$ and $\boldsymbol{W}$, and then use the strength of the κ hyperparameter to allow the coefficients to "relax" into agreement. This approach is additionally advantageous because other regularizers and model constraints can be added straightforwardly,

$$\underset{\boldsymbol{\xi},\boldsymbol{W}}{\operatorname{argmin}} \|\dot{\boldsymbol{X}} - \boldsymbol{\Theta}(\boldsymbol{X})\boldsymbol{\xi}\|^2 + \kappa\|\boldsymbol{\xi} - \boldsymbol{W}\|_2^2 + \lambda\|\boldsymbol{W}\|_0, \tag{6.3}$$
$$\boldsymbol{C}_s\boldsymbol{\xi} \leq \boldsymbol{d}$$

The original work [82] analytically solved the equality-constraints problem at each iteration. The implementation for both equality and inequality constraints is discussed further in Sec. 6.3.2. Although SR3 introduces an extra hyperparameter, the default $\kappa = 1$ often works

well. It also allows for a version of a "soft constraint", since the hard constraint is applied to $\boldsymbol{\xi}$ but the final sparse model coefficients are $\boldsymbol{W}$.

Step-wise sparse regression

Step-wise sparse regression (SSR) is a greedy method for sparsifying the solution to the least-squares SINDy optimization. Like `STLSQ`, the default is to use ridge regression but $\alpha = 0$ can be set to do pure least-squares. Instead of using a threshold parameter λ, SSR truncates (zeros out) the "least important" coefficient in the model at each iteration, until some criteria is met or no coefficients are left. Two SSR variants are implemented, where the least important coefficient at each iteration is 1) the nonzero coefficient with smallest absolute value [48], or 2) the coefficient which, when zeroed, results in a smaller model with the lowest residual error (similar to Gurevich *et al.* [155]). New criteria for coefficient sparsification can be easily added in the future. The advantage of SSR lies in its avoidance of hyperparameters, which often need to be tuned for particular dynamical systems.

FROLS

Forward regression orthogonal least squares (FROLS) is another greedy algorithm used to sparsify the solution to the SINDy ridge regression with the l_0 norm [43, 70], and therefore has similar drawbacks and benefits as SSR. FROLS iteratively selects the most correlated function in the library. At each step, the candidate functions are orthogonalized with respect to the already-selected functions. The selection criteria is the error reduction ratio, defined as the normalized increase in explained output variance due to the addition of a given function to the basis. In contrast to SSR, the algorithm starts with no terms and then adds terms to the regression until all terms are included in the model. The best model is chosen by computing all of the model mean-squared-errors, optionally weighted by an l_0 penalty via the parameter κ as in Rudy *et al.* [376]. This hyperparameter is an optional convenience; regardless, all the generated models are returned to the user and they may choose a best model according to any desired metric.

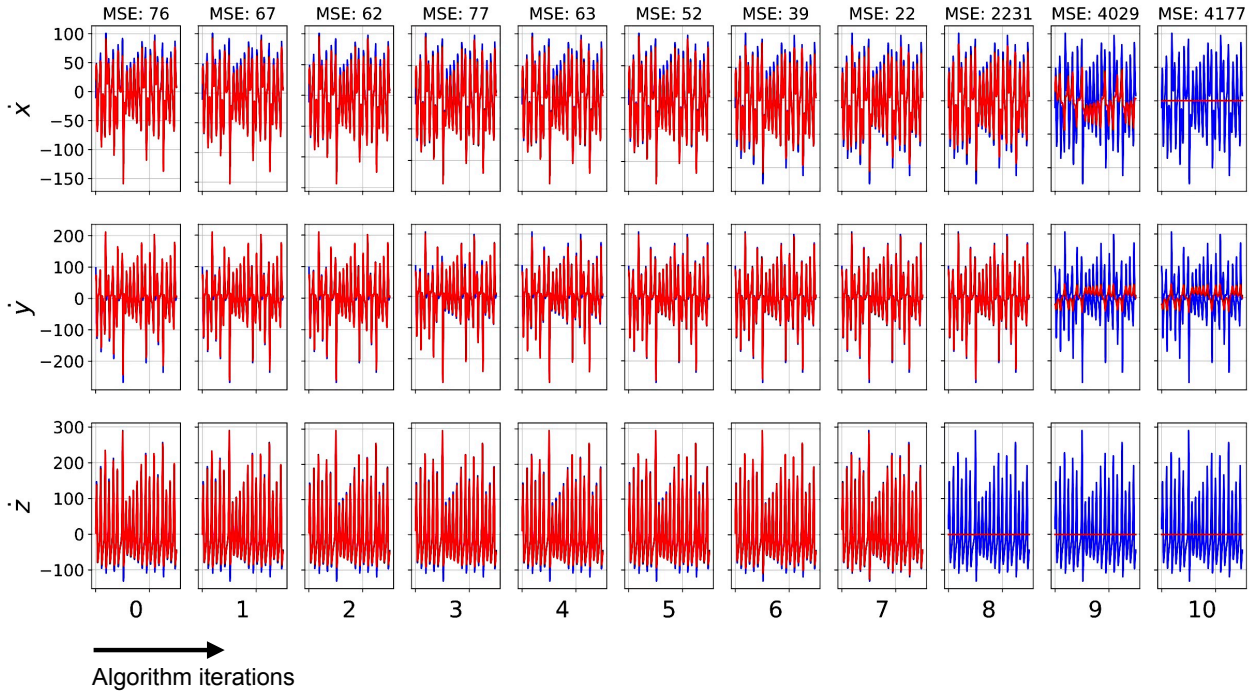


Figure 6.2: Illustration of the SSR model fit with training Lorenz data with 10% added noise and shown here against a noise-free testing Lorenz trajectory. The algorithm begins with all nonzero coefficients and at each new iteration the coefficient of the least important term (according to some criteria) is set to zero. On the top of each slice, the mean-squared error is tracked on a test trajectory as the algorithm progresses. Despite being quite sparse, the model at algorithm iteration 7 illustrates the strongest MSE performance on the testing trajectory. In the next iteration, the algorithm zeros out an essential dynamical term and the quality of the fit drops sharply.

6.2.2 l_1 based sparse regression algorithms

Using the l_1 norm has a great advantage − the optimization is convex. The downside is that sparsity is harder to enforce, and the l_1 norm can systematically bias the system towards certain solutions. Nonetheless, a number of algorithms have found high performance with versions of the Lasso formulation [76]. The default algorithm for `sklearn.linear_model.Lasso` is coordinate descent.

SINDy-PI

There are a class of ordinary differential equations that are not amenable to the traditional SINDy regression − implicit ODEs of the form

$$\dot{\boldsymbol{q}}(t) = \boldsymbol{f}(\boldsymbol{q}(t), \dot{\boldsymbol{q}}(t)). \tag{6.4}$$

Implicit SINDy [273], and the more advanced SINDy-PI [201], are useful for identifying implicit ODEs that depend on both $\boldsymbol{X}$ and $\dot{\boldsymbol{X}}$. The optimization strategy is fairly different than the typical approach, solving the minimization

$$\underset{\boldsymbol{\Xi}}{\operatorname{argmin}} \|\boldsymbol{\Theta}(\boldsymbol{X}, \dot{\boldsymbol{X}}) - \boldsymbol{\Theta}(\boldsymbol{X}, \dot{\boldsymbol{X}})\boldsymbol{\xi}\|^2 + \lambda\|\boldsymbol{\xi}\|_1, \quad s.t. \quad \operatorname{diag}(\boldsymbol{\xi}) = 0. \tag{6.5}$$

Solving the optimization problem in Eq. (6.5) generates p_Θ differential algebraic equations (DAEs), one equation for each of the candidate library terms. To illustrate the type of models that can be identified with SINDy-PI (and not other methods), consider the relatively simple Michaelis–Menten model for enzyme kinetics [198, 94]:

$$\dot{q} = 0.6 - 3q - \frac{10}{3} q\dot{q}, \tag{6.6}$$

In order to implement SINDy-PI, a library is defined that takes as input two separate libraries for $\boldsymbol{X}$ and $\dot{\boldsymbol{X}}$ and tensor products them together. With a fourth order polynomial library

in $\boldsymbol{X}$ (significant overkill) and a linear library in $\dot{\boldsymbol{X}}$, the algorithm approximately captures the correct model,

```
q' = 0.600  - 3.000 q - 3.300 qq'.
```

To integrate general DAEs, the native Python ODE solver can no longer be used. Instead, the system is converted into a symbolic set of equations and either (1) solved symbolically with SymPy [293] for $\dot{\boldsymbol{q}}$, after which the native ODE solver can be used, or (2) integrated with the `diffeqpy` Python package [357], which requires the `Julia` programming language [39]. This is not built into the main source code. Finally, note that solving for a DAE for every candidate library term can be computationally expensive, even though each optimization is separate; in the original work, the computation is sped up with parallel computing. Instead, an optional `SINDyPI` parameter `model_subset` is implemented, which is an array of indices that specifies a subset of the models to compute. Coupled with the new `PySINDy` capability to calculate high-order spatial derivatives, `SINDyPI` could be used for solving implicit boundary value problems as in Shea *et al.* [398] by changing all the temporal derivatives to spatial ones. Coupled with the PDE functionality described in Sec. 6.3.4, `SINDyPI` could also be used to identify very general partial differential algebraic equations (PDAEs).

6.3 Advanced functionality

Different options for sparse regression optimizers is helpful in some situations but does not directly address the most common issue with these techniques, the presence of noise. In the following sections, methods are illustrated that significantly increase the method robustness to noise and facilitate very advanced system identification attempts.

6.3.1 Ensembling

One way to improve system identification methods on noisy data is simply to generate a large number of models for different initial conditions, trajectories, noise levels, and more [123]. A final model can be chosen by the minimal error on the training or validation data, an

average over the models (bagging), a median over all the models (bragging), and so on. More formally, ensemble methods are the generation of n_{models} by sparse regression onto n_{subset} of the time samples in a signal. Both n_{models} and n_{subset} are user-defined. The current implementation chooses the subset of time points randomly; there is no issue that time points may not be consecutive because the optimization is point-wise. The default implementation is that $n_{\text{subset}} = M$, i.e. each sub-sampling uses the full temporal length but samples with replacement. Typically this selects approximately 60% of the data with duplicates.

The `PySINDy` implementation of ensemble methods is done entirely behind the scenes. The primary change with ensemble methods is that n_{models} models are generated and can be accessed via `model.coef_list`. The post-processing is left to the user; common choices to recombine the models are averaging (bagging), taking the median (bragging), or stability-metrics such as in Maddu *et al.* [267].

Library ensemble methods, i.e. the generation of n_{models} models by multiple regressions onto different subsets of the full candidate library $\boldsymbol{\Theta}$, are also implemented. Instead of specifying the subset of time slices to use, the number of terms to drop in $\boldsymbol{\Theta}$ is specified through `n_candidates_to_drop` (the default is `n_candidates_to_drop` = 1). The library ensemble method can be done separately or together with `ensemble`. Figure 6.3 illustrates a side-by-side comparison of the ensemble and library ensemble methods with different post-processing methods (bagging, bragging, etc.). For fixed parameters, library ensembling leads to predictably higher errors in the coefficients because key dynamic terms are missing in some of the regressions (this effect is amplified because a very small library is used). Figure 6.4 shows the distribution of coefficients generated with 100 ensemble models and 100 library ensemble models.

Finally, models can also be generated by ensembles of trajectories originating from disparate initial conditions. Fortunately, the `multiple_trajectories` parameter already allows training on time series data from many trajectories. Using `multiple_trajectories` with `ensemble` is then already similar in spirit to generating models from different trajectories.

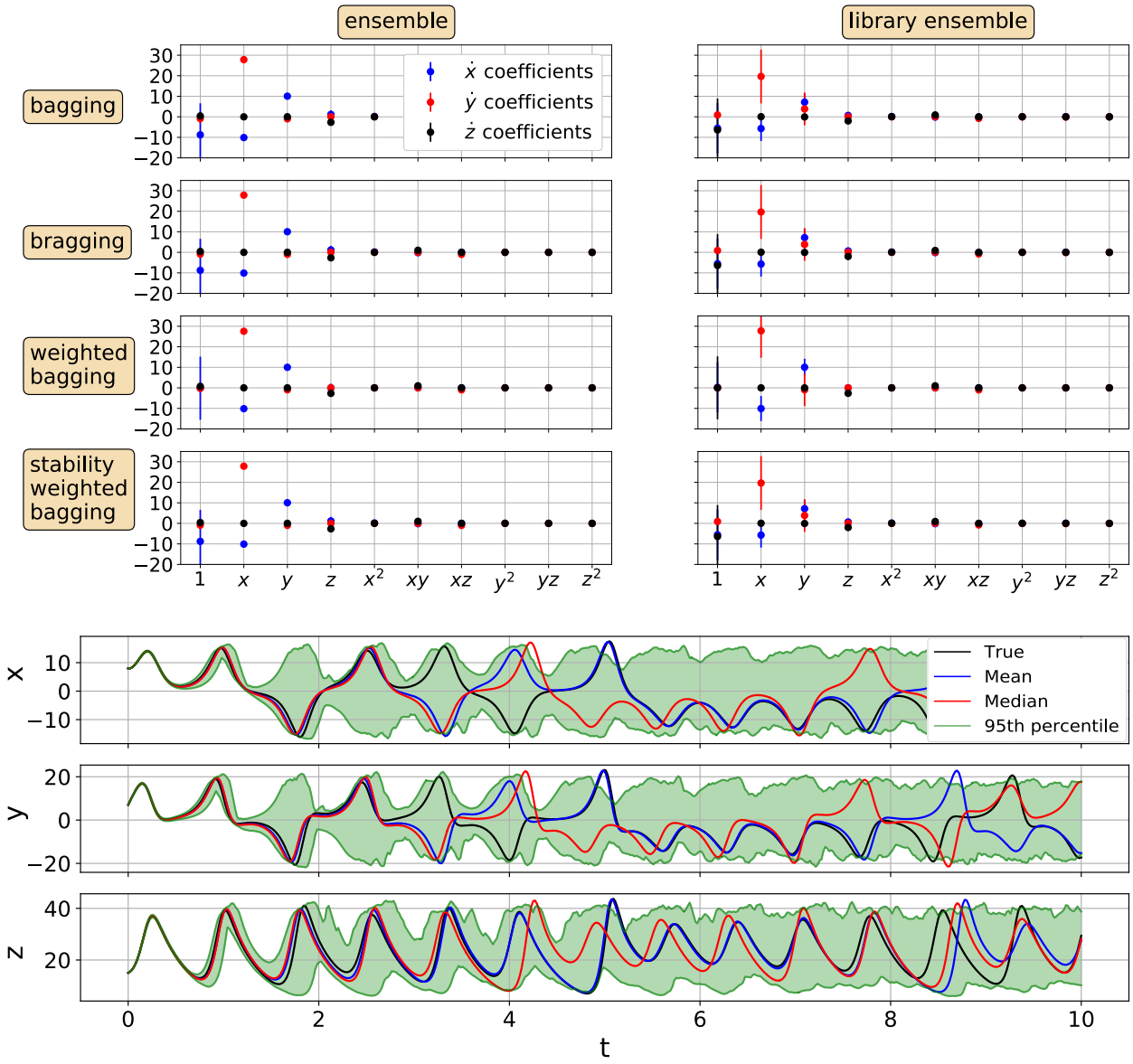


Figure 6.3: Top: Summary of the ensemble (left column) and library ensemble (right column) results using a quadratic polynomial library, the STLSQ optimizer, and the default ensemble functionality on the Lorenz system with 1% added noise. The summary compares different post-processing methods, including the average coefficient values (bagging), median coefficient values (bragging), weighted average by the mean-squared error of the fit, and weighted average where short-time unstable models are removed from the averaging. Bottom: With the same parameters, illustration of the Lorenz testing data predictions using the mean, median, and 95th percentile models.

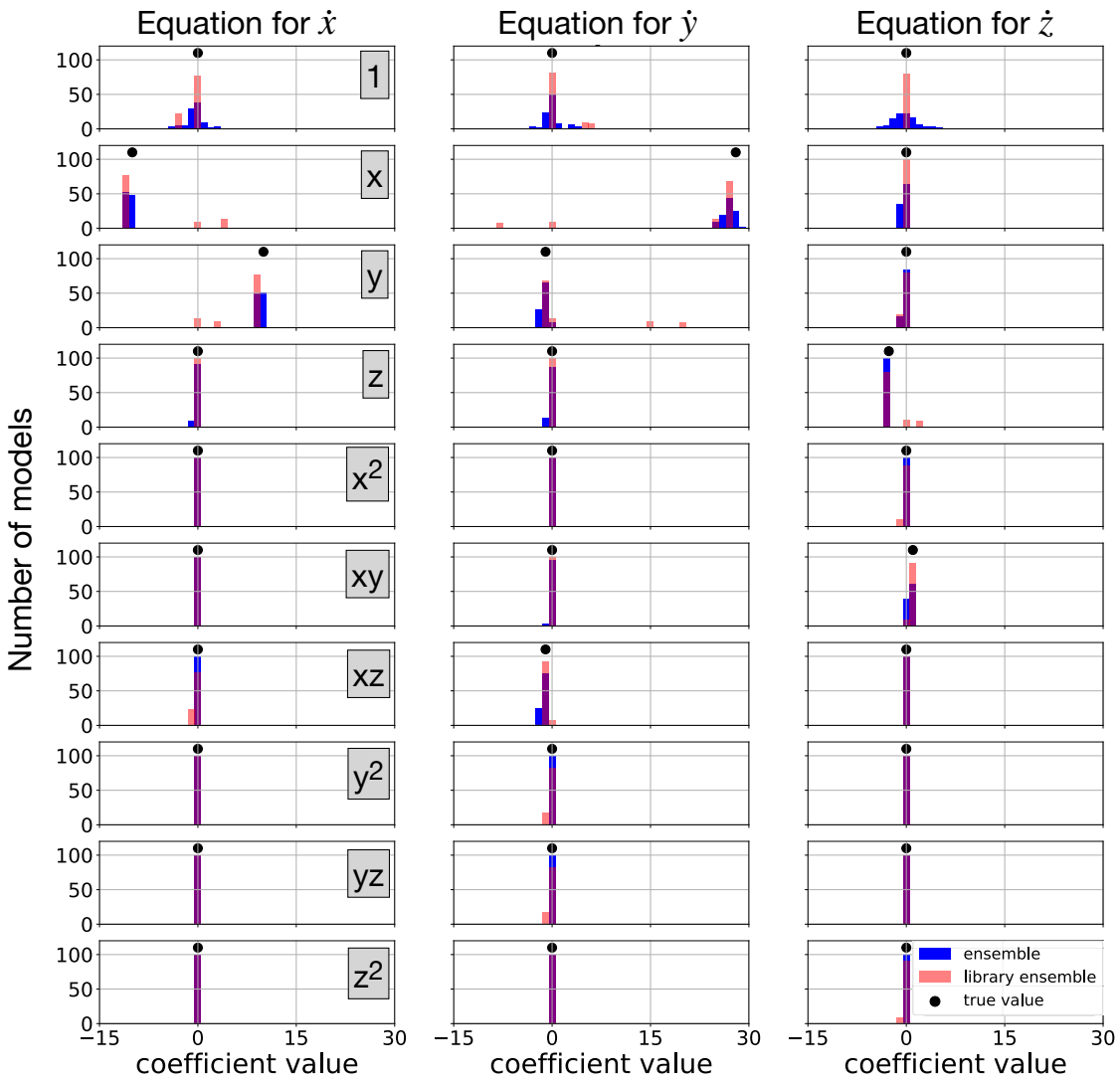


Figure 6.4: Ensemble (blue) and library ensemble (pink) coefficient distributions for 100 Lorenz models generated separately with each method, with the same data as in Figure 6.3.

6.3.2 Equality and inequality constraints

Often, some details are already known about the dynamical system, such as energy conservation or a region of attraction. Building physical constraints into system identification was already discussed in Sec. 5.1 and the formal optimization problem to solve is Eq. (5.6). The constrained nonconvex problem (when R_s is the l_0 norm) is effectively solved with a relax-and-split technique [457], and this is the approach for `SR3` and `ConstrainedSR3`. The primary work is defining $\boldsymbol{C}_s$ and $\boldsymbol{d}$, through `constraint_lhs` and `constraint_rhs` respectively. Below, the x coefficient in the $\dot{y}$ Lorenz equation is constrained to be exactly 28 and the coefficients of x and y are constrained to be equal and opposite in the $\dot{x}$ Lorenz equation, which correctly generates:

```
x' = -10.002 x + 10.002 y
y' = 28.000 x - 1.003 y - 1.000 xz
z' = -2.666 z + 0.999 xy
```

The fully constrained version of `SR3` may be too restrictive for many cases, particularly if the data is noisy. In the noisy case, exact satisfaction of a conservation law may not be possible or result in a good fit of the data, so it may be preferable to use inequality constraints. In `PySINDy`, using inequality constraints requires that the optimization remain convex, i.e. the l_1 or l_2 regularizers must be used. In this case, the problem is solved with the CVXPY [109] package. These details are hidden so that using inequality constraints are also straightforward. The previous code is repeated, but instead of exactly equal and opposite coefficients, the coefficients are required to be equal and opposite within a tolerance of 10^{-5}. The code correctly generates:

```
x' = -10.002 x + 10.002 y
y' = -0.015 1 + 27.991 x + -0.998 y + 0.002 z + -1.000 xz
z' = 0.008 1 + 0.006 x + -0.004 y + -2.666 z + 0.001 x^2 + 0.999 xy
```

```
x coefficient in x' equation is -10.001643
y coefficient in x' equation is 10.001653
```

Note that the error tolerance of the solver may need to be reduced in order to satisfy tight inequality constraints. These are rather contrived examples, but the Example 8 Jupyter notebook also contains code that computes the model constraint for energy-preserving quadratic nonlinearities in fluids and plasmas for arbitrary state size. Figuring out how to input some other constraint, or generalize it to arbitrary model dimension, will take some indexing work. Lastly, the `TrappingSR3` optimizer, based on the work in Sections 5.3−5.7, can also be used with inequality constraints, and in fact reduces to `ConstrainedSR3` in the limit that the stability-promoting hyperparameter satisfies $\zeta \gg 1$. In this sense, `TrappingSR3` can also be used for PDE identification and other tasks, but has no advantage over the other optimizers in these applications since the stabilizing part of the algorithm goes unused.

6.3.3 Control variables

It is straightforward to incorporate external inputs and control variables into the SINDy method; simply include terms depending on the control inputs $\boldsymbol{q}_u$ in the library of candidate terms [66], $\boldsymbol{\Theta}(\boldsymbol{q}) \to \boldsymbol{\Theta}(\boldsymbol{q}, \boldsymbol{q}_u)$. To illustrate the use of control variables in `PySINDy`, considering the following externally forced Lorenz system,

$$\dot{x} = -10x + 10y + 100u_0^2, \tag{6.7}$$

$$\dot{y} = x(28 - z) - y, \tag{6.8}$$

$$\dot{z} = xy - \frac{8}{3}z - u_1, \tag{6.9}$$

where $u_0(t) = \sin(2t)$ and $u_1(t) = t^2$. The code generates

```
x' = -9.999 x + 9.999 y + 99.990 u0^2
y' = 27.995 x - 1.000 y - 1.000 xz
```

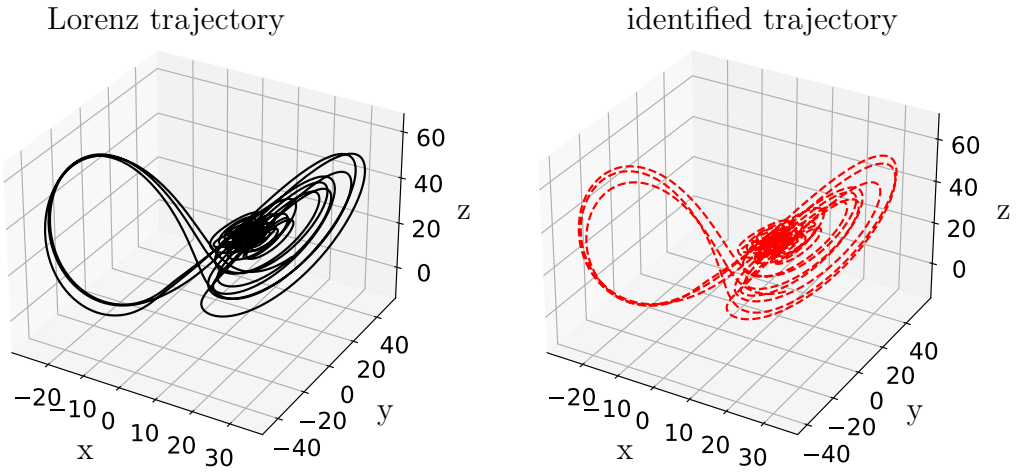


Figure 6.5: SINDy can accurately forecast the Lorenz system with external control inputs.

```
z' = -2.666 z - 1.000 u1 + 1.000 xy
```

Note that, by default, the library of terms $\boldsymbol{\Theta}$ treats the control inputs $\boldsymbol{q}_u$ identically to the state variables $\boldsymbol{q}$. For this example, the library is quadratic in polynomials of $\boldsymbol{q}$ and therefore all of the terms $[u_0, u_1, u_0x, u_0y, u_0z, u_1x, u_1y, u_1z, u_0u_1, u_0^2, u_1^2]$ are added to $\boldsymbol{\Theta}$. The downside of this behavior is that for higher-dimensional systems, or systems with large libraries, treating $\boldsymbol{q}_u$ like $\boldsymbol{q}$ can lead to many new terms. When different libraries for $\boldsymbol{q}_u$ and the inputs $\boldsymbol{q}$ are desired, one can use the `GeneralizedLibrary`, described below in Sec. 6.3.6.

To simulate forward a test trajectory with a new initial condition, there are two options, passing the control function directly to the integrator or passing a time series of control inputs to the integrator, which is subsequently interpolated onto the integration times. Clearly the former option is a better choice if the functional form of the control input is known. The simulation results for the controlled Lorenz system are illustrated in Fig. 6.5 and illustrate high performance.

6.3.4 PDE-FIND

So far, the focus has been on dynamical systems described by ordinary differential equations. While some PDEs can be converted to ODEs through clever transforms, it remains that most spatiotemporal systems are best described in a PDE framework. The original SINDy algorithm was explicitly designed for the identification of ODEs from data, but SINDy applied to PDEs is quite similar in flavor to the traditional method. No new optimizer is required, meaning that all of the existing optimizers, as well as the `ensemble` and `multiple_trajectories` functionality, can be used. In order to retain the structure of Eq. (5.3), the data must be flattened so that $\boldsymbol{q}(\boldsymbol{x}, t) \in \mathbb{R}^{M \times N}$, where N is the state dimension of $\boldsymbol{q}$ and M contains the rest of the dimensions. For instance, for a coupled PDE system of two variables, $N = 2$ and $M = N_x \times N_y \times N_z \times M_t$ for a spatiotemporal grid with N_x points in the x-direction, and so on. In addition, the derivative matrix $\dot{\boldsymbol{X}}$ (or $\ddot{\boldsymbol{X}}$, etc.) must be provided to `model.fit`, since the default code behavior is to take a temporal derivative throughout the entire flattened dimension M of $\boldsymbol{X}$.

To proceed, various spatial derivatives must be computed and incorporated into the library of terms $\boldsymbol{\Theta}$. Towards that goal, the user must specify the highest order of spatial derivative desired, `derivative_order` (defaults to 0 since weak-formulation ODEs can be found with this `PDELibrary` class as well), the spatial grid `spatial_grid`, and a library of terms $\boldsymbol{\Theta}(\boldsymbol{X})$ that will form the foundation of the PDE library. Up to fourth order spatial derivatives are implemented with centered finite differences. High order derivatives increasingly amplify noise and therefore provide less and less desirable candidates for regression. Using finite difference techniques on a grid with spacing $\mathcal{O}(\Delta x)$, and noise with amplitude $\mathcal{O}(\epsilon_n)$, produces i^{th} order derivatives with noise levels of approximately $\mathcal{O}(\epsilon_n(\Delta x)^{-i})$.

Spatial PDEs in 1D, 2D, and 3D space were implemented, although in three spatial dimensions the library of derivative terms becomes quite large and the computations become increasingly expensive. The number of partial derivatives for a i^{th} order derivative in j dimensions is $\mathcal{O}(i^j)$. All of this is complicated by the fact that derivative terms often appear

together in a PDE; for instance $\nabla \cdot \boldsymbol{u}$ contains j first derivative terms with the same coefficient. Rather than implement this separately, one can use model constraints with the `ConstrainedSR3` optimizer to require that the coefficients for each of the first derivatives are the same. In the same way, one can use constraints to build in local conservation laws such as $\nabla \cdot \boldsymbol{u} = 0$ or $\nabla \cdot \boldsymbol{B} = 0$. For additional performance, the columns of $\boldsymbol{\Theta}$ are normalized for the optimization with the optional boolean parameter `normalize_columns`; it was found empirically in Rudy *et al.* [376] that this usually improves the model selection by rescaling the columns of $\boldsymbol{\Theta}$ to be similar magnitudes. This option has been subsequently added to all `PySINDy` optimizers, although using `normalize_columns` is not currently compatible with using constraints.

All of the PDEs tested in the Example 10 Jupyter notebook are open-source datasets taken directly from Rudy *et al.* [376]. With some cross-validation, the PDE-relevant optimizers can successfully identify the 1D PDEs, including with some limited noise. For the Kuramoto-Sivashinsky (KS) equation,

$$q_t = -qq_x - q_{xx} - q_{xxxx}, \tag{6.10}$$

A model is trained on the first 60% of the data from Rudy et al. [376], which in total contains 1024 spatial grid points and 251 time steps. The KS model is identified correctly and the prediction for $\dot{\boldsymbol{q}}$ on the remaining testing data indicates strong performance in Fig. 6.6. To demonstrate a more advanced usage of this new functionality, consider the 2D reaction-diffusion system on a periodic domain:

$$\begin{aligned} u_t &= u + 0.1\nabla^2 u - (u^2 + v^2)(u - v), \\ v_t &= v + 0.1\nabla^2 v - (u^2 + v^2)(u + v). \end{aligned} \tag{6.11}$$

Identical parameters to Rudy et al. [376] are used ($x, y \in [-10, 10]$, $t \in [0, 10]$, 201 time points), except for a coarser grid of 128 spatial points. With ensembling, the mean and standard deviation of the model coefficients can be plotted as in Fig. 6.7. Using each of the

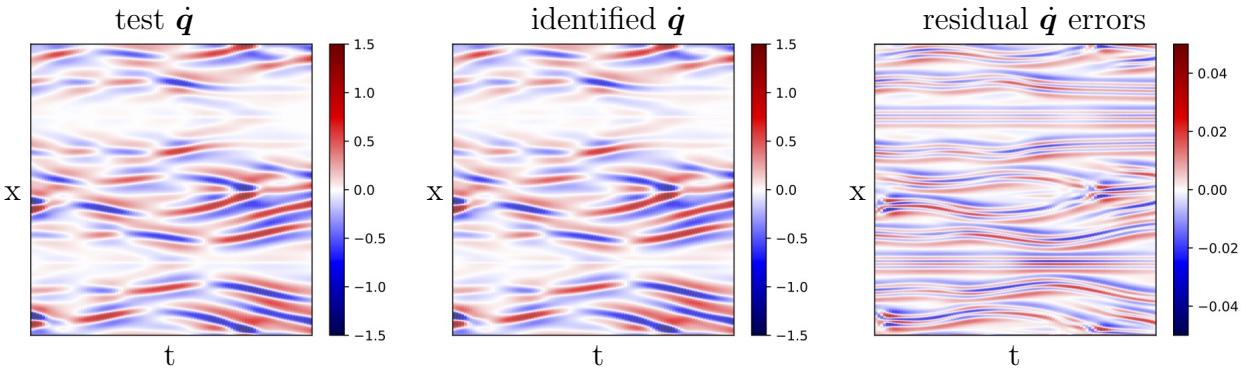


Figure 6.6: `PySINDy` can now be used for PDE identification. This new capability is illustrated by accurately capturing a set of testing data from the Kuramoto-Sivashinsky system, described by $q_t = -qq_x - q_{xx} - q_{xxxx}$. The identified model is $q_t = -0.98qq_x - 0.99q_{xx} - 1.0q_{xxxx}$.

different optimizers without ensembling generates:

```
STLSQ:
        u' = 1.013 u - 1.013 u^3 + 1.000 v^3 - 1.012 uv^2 + 0.999 u^2v
           + 0.101 u_xx + 0.102 u_yy
        v' = 1.013 v - 1.000 u^3 - 1.013 v^3 - 0.999 uv^2 - 1.013 u^2v
           + 0.102 v_xx + 0.101 v_yy


SR3 l0:
        u' = 1.013 u - 1.013 u^3 + 1.000 v^3 - 1.012 uv^2 + 0.999 u^2v
           + 0.101 u_xx + 0.102 u_yy
        v' = 1.013 v - 1.000 u^3 - 1.013 v^3 - 0.999 uv^2 - 1.013 u^2v
           + 0.102 v_xx + 0.101 v_yy


SR3 l1:
        u' = 1.013 u - 1.013 u^3 + 1.000 v^3 - 1.012 uv^2 + 0.999 u^2v
           + 0.101 u_xx + 0.102 u_yy
        v' = 1.013 v - 1.000 u^3 - 1.013 v^3 - 0.999 uv^2 - 1.013 u^2v
           + 0.102 v_xx + 0.101 v_yy


FROLS:
        u' = 0.994 u - 0.049 v - 0.992 u^3 + 1.048 v^3 - 0.991 uv^2 + 1.047 u^2v
```

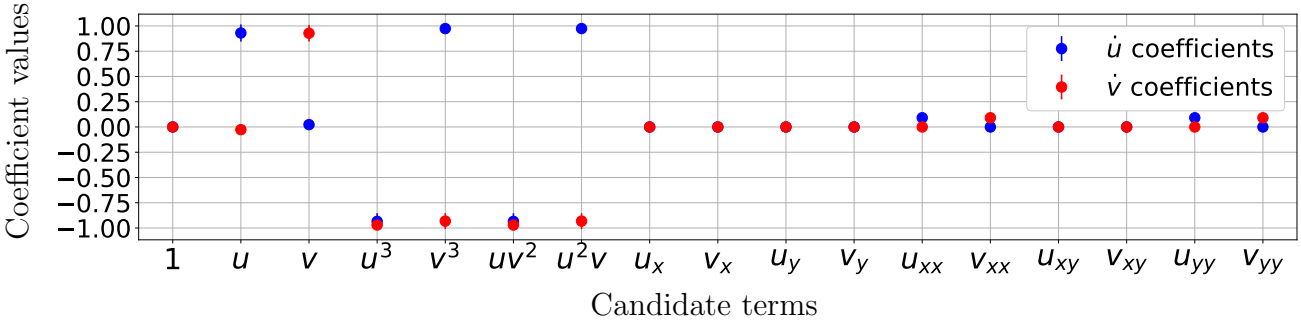


Figure 6.7: Mean and standard deviation computed from the ensembling SINDy models for the reaction-diffusion system.

```
    + 0.102 u_xx - 0.005 v_xx + 0.101 u_yy - 0.005 v_yy
v' = 0.048 u + 0.994 v - 1.048 u^3 - 0.992 v^3 - 1.047 uv^2 - 0.992 u^2v
    + 0.005 u_xx + 0.101 v_xx + 0.005 u_yy + 0.102 v_yy
```

```
ConstrainedSR3:
        u' = 0.995 u - 0.993 u^3 + 1.054 v^3 - 0.993 uv^2 + 1.054 u^2v
           + 0.101 u_xx + 0.101 u_yy
        v' = 0.995 v - 1.054 u^3 - 0.993 v^3 - 1.054 uv^2 - 0.993 u^2v
           + 0.101 v_xx + 0.101 v_yy
```

All of the optimizers essentially capture the true PDE model in Eq. (6.11). FROLS overfits slightly, but additional sparsity can be obtained by increasing the sparsity penalty κ or cross-validation by scanning over this parameter. Notice that constraints were used in `ConstrainedSR3` to guarantee that the coefficients for u_{xx} and u_{yy} (and similarly for v) are the same, giving us a pure Laplacian term. To the author's knowledge, this is the first use of `SR3` or `FROLS` for PDE system identification and the first use of SINDy constraints for PDE model identification.

6.3.5 Weak-form, integral SINDy

In the previous section, it was noted that the computation of high-order spatial derivatives is increasingly affected by noise in the data, leading to poor performance on noisy data. In

order to make PDE system identification more robust, Schaeffer *et al.* [385] introduced an integral form of SINDy that can identify PDE weak forms. If Eq. (5.3) is multiplied by a spatiotemporal weight $\boldsymbol{w}(\boldsymbol{x},t)$ and integrated over some spatiotemporal domain Ω_k, one obtains

$$\begin{aligned}\int_{\Omega_k} \boldsymbol{w}(\boldsymbol{x},t)\cdot\partial_t\boldsymbol{q}(\boldsymbol{x},t)dS &= -\int_{\Omega_k}\partial_t\boldsymbol{w}(\boldsymbol{x},t)\cdot\boldsymbol{q}(\boldsymbol{x},t)dS := \dot{Q}(t) \\ &\approx \left(\int_{\Omega_k}\boldsymbol{w}(\boldsymbol{x},t)\cdot\boldsymbol{\Theta}(\boldsymbol{q}(\boldsymbol{x},t))dS\right)\boldsymbol{\xi} := \Theta_Q\boldsymbol{\xi}.\end{aligned} \tag{6.12}$$

Critically, as was used in Eq. (6.12), the spatiotemporal weight $\boldsymbol{w}$ can be chosen to be smooth and vanish on Ω_k. Many of the derivative terms in $\boldsymbol{\Theta}(\boldsymbol{q}(\boldsymbol{x},t))$ can be integrated by parts so that the derivatives apply to $\boldsymbol{w}$, greatly reducing the noise amplification coming from derivatives of any noisy data. Now this process is repeated for K_d different choices of Ω_k, choosing each Ω_k to be a cube centered at (x_k, y_k, t_k) with width H_x, height H_y, and depth H_t (the precise geometry of Ω_k is not important so cubes are convenient). Following Reinbold *et al.* [365], a sufficient, smooth function that vanishes on the edges of Ω_k is $\boldsymbol{w}_k = (\tilde{x}^2-1)^P(\tilde{y}^2-1)^P(\tilde{t}^2-1)^P$ for some polynomial P, normalized $\tilde{x} = (x-x_k)/H_x$, normalized $\tilde{y} = (y-y_k)/H_y$, and normalized $\tilde{t} = (t-t_k)/H_t$. The generalizations to spatial dimensions other than 2D is straightforward. Arbitrary order polynomials are permitted although $P = 4$ is the default for compatibility with up to fourth order spatial derivatives. Using finite differences on $\boldsymbol{w}_k$ still amplifies the numerical noise from the grid. Fortunately, the simple form of $\boldsymbol{w}_k$ allows us to use the analytic form for derivatives of $\boldsymbol{w}_k$. Errors from the numerical approximations of the integrals can be reduced by using more points in the integration.

Stacking the results from $K_d \geq p_\Theta$ domains, the sparse regression problem becomes

$$\dot{\boldsymbol{Q}} \approx \boldsymbol{\Theta}_Q\boldsymbol{\xi}, \quad \dot{\boldsymbol{Q}} \in \mathbb{R}^{K_d\times N}, \quad \boldsymbol{\Theta}_Q \in \mathbb{R}^{K_d\times p_\Theta}, \quad \boldsymbol{\xi}\in\mathbb{R}^{p_\Theta\times N}. \tag{6.13}$$

The optional boolean parameter `weak_form` in the `PDELibrary` turns on this behavior. The

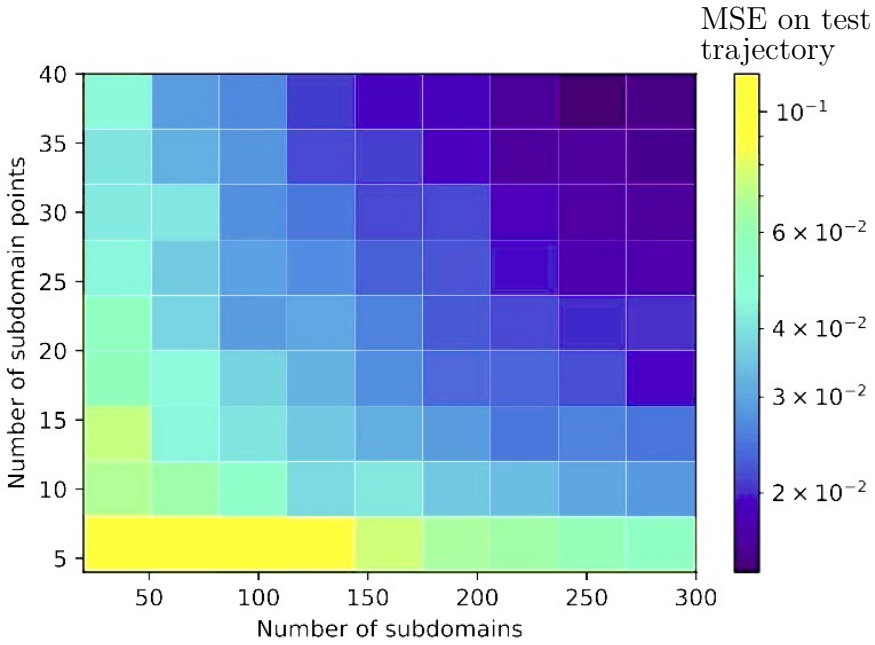


Figure 6.8: Illustration that as the number of subdomain points increases (the numerical approximation of the integrals in Eq. (6.13) is improving) and the number of subdomains K_d increases (the number of points to fit in the regression is increasing), the weak form implementation in PySINDy converges to better and better dynamical models. All models were generated with the same data – the Lorenz system with 20% added noise.

user can also specify K_d, the spatial grid, the temporal grid, H_x, H_y, H_t, and the number of integration points in each subdomain N_{sub}. If these parameters are not specified, the default weak form usage is that $K_d = 100$ subdomains are randomly selected, with $H_x = L_x/20$, $H_y = L_y/20$, $H_t = L_t/20$, $N_{\text{sub}} = 100$ (L_x and L_y being the lengths of the entire spatial domain, and L_t the temporal duration of the training data), such that no subdomain Ω_k overlaps with or passes through a boundary. If the KS equation in Eq. (6.10) is now revisited, one can illustrate how the coefficients of the model change as a function of added noise, as in Figure 6.9.

6.3.6 Generalized and tailored candidate libraries

A common issue is that the SINDy candidate library $\boldsymbol{\Theta}$ grows combinatorically with the addition of new terms. Moreover, the addition of control inputs in Sec. 6.3.3 and the addition

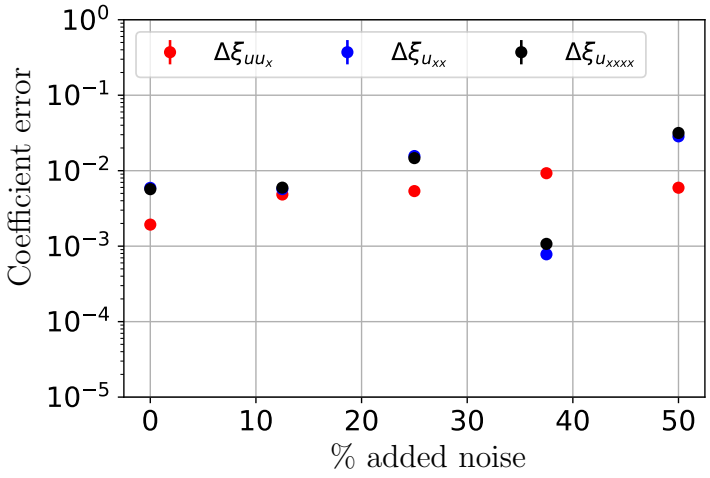


Figure 6.9: Error in the KS equation coefficients, defined through $\Delta\xi_{uu_x} = |\xi^{\text{true}}_{uu_x} - \xi^{\text{pred}}_{uu_x}|/\xi^{\text{true}}_{uu_x} = |1 - \xi^{\text{pred}}_{uu_x}|$. The models are quite robust even with very high (50%) added noise.

of 1, 2, or 3D spatial derivatives in Sec. 6.3.4 further exacerbates this issue. To see the issue here, consider a model that is up to quadratic in polynomials for a spatially-2D system of four PDEs. Suppose that up to second order derivatives are desired and candidate terms are postulated to include mixed-derivative terms and an external potential depending on the spatial coordinates $\boldsymbol{x}$. For this situation, the input to a `PolynomialLibrary` (ignore for the moment that the default `PySINDy` libraries cannot handle such an input) would be

$$\boldsymbol{q}_{\text{advanced}} = [\boldsymbol{q}, \boldsymbol{q}_x, \boldsymbol{q}_y, \boldsymbol{q}_{xx}, \boldsymbol{q}_{xy}, \boldsymbol{q}_{yy}, \boldsymbol{x}], \tag{6.14}$$

where each variable is actually in a 4D state space (except $\boldsymbol{x}$, which is 2D). The default is that `PolynomialLibrary` generates quadratic polynomials for each of these inputs, including all possible mixed terms. Therefore, $\boldsymbol{q}_{\text{advanced}} \in \mathbb{R}^{M\times 26}$, and $\boldsymbol{\Theta} \in \mathbb{R}^{M\times 378}$. Recall that in the PDE formulation $M = N_x \times N_y \times M_t$, so for a reasonable dataset, $M \sim 10^6$. One can see that this data is getting large and unwieldy. Furthermore, it would be useful if there was a systematic way to efficiently generate different libraries for different column inputs in Eq. (6.14); one

might expect Fourier dependence on $\boldsymbol{q}$ but polynomial dependence in $\boldsymbol{x}$.

Towards that end, a `GeneralizedLibrary` class was implemented, which relies on the `ConcatLibrary` and `TensoredLibrary` classes. A `ConcatLibrary` is generated from the concatenation of two existing `PySINDy` library objects, so that the candidate terms of both libraries are included in a single `ConcatLibrary`. In contrast, a `TensoredLibrary` is generated from the tensor product of the columns of two existing `PySINDy` library objects, so that every pair of candidate terms from the two libraries are multiplied together, and their product is added to the new `TensoredLibrary`. Fully utilizing the `GeneralizedLibrary` class requires: (1) a number of different libraries to stitch together, (2) a list of input variables that specifies which variables should be used to generate each library, (3) a list of which libraries to tensor together and add to the overall library. With this new functionality, one could define candidate libraries that are very tailored. For instance, one can fairly easily generate a very complex `GeneralizedLibrary` consisting of: a cubic polynomial library for the q_1, a custom library for q_2, a different custom library for the $\boldsymbol{q}_x$, a Fourier library for $\boldsymbol{q}_y$, a quadratic polynomial library for the spatial coordinates $\boldsymbol{x}$, a linear polynomial library for a set of control inputs $\boldsymbol{q}_u$, and any libraries obtained from tensoring together the previous libraries. This new functionality reduces the number of candidate terms and facilitates the use of very different candidate terms for each of the input features.

6.4 Closing remarks on robust system identification

All of this new and advanced functionality is useful for facilitating a broad range of system identification tasks across a wide range of scientific disciplines. However, for new users or scientists unfamiliar with many of these methods, it can be hard to decipher which methods are appropriate for their data. Towards that end, a user-friendly flow chart is illustrated in Fig. 6.10. If significant flexibility and tailoring is required in the candidate library, `GeneralizedLibrary` is recommended, despite it being harder to use than most of the other libraries.

There remain a number of SINDy variants and improvements that could be valuable to

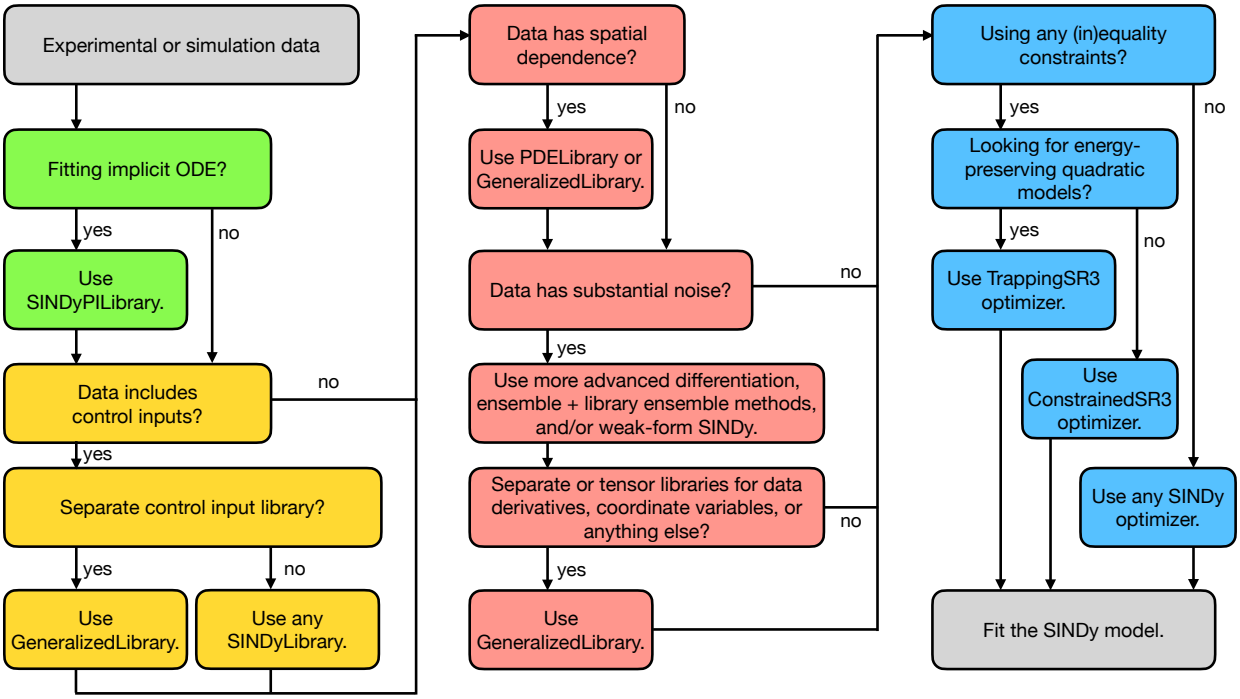


Figure 6.10: This flow chart summarizes how `PySINDy` users can start with a dataset and systematically choose the proper candidate library and sparse regression optimizer that are tailored for a specific scientific task.

incorporate into `PySINDy` in the future. Currently, high-order spatial derivatives can only be computed with finite differences in `PySINDy`, meaning that with any significant noise one must adopt the weak form PDE method, which can be computationally expensive. One solution to this problem is to learn both the unknown dynamics and the underlying noise distribution [377, 200]. A more common solution is to improve the quality of the high-order derivatives, reducing the noise amplification. This can be done to varying degrees by smoothing with a Gaussian kernel, total variation [375], Tikhonov differentiation [89], polynomial interpolation [56, 18], spline interpolation [410], or spectral methods [146] (the latter two methods are actually already available in `PySINDy` and can be extended to higher-order derivatives). Considerable progress could be made by adopting a fully symbolic problem formulation [173], and then using automatic differentiation [27] to achieve arbitrary precision derivatives. Promising research with high-dimensional systems includes stochastic methods [55, 71] and tensor libraries [135, 137]. Lastly, SINDy is increasingly incorporated with model-predictive control (MPC) [204, 387, 260, 122]. Incorporating basic MPC functionality into `PySINDy` would be a valuable addition to the code.

Chapter 7

SUMMARY AND CONCLUSIONS

This work has applied a broad set of machine learning methods for learning ROMs and performing prediction tasks in fluid and plasma systems. Linear data-driven models such as the dynamic mode decomposition appear useful for linear plasma models and magnetic spectroscopy, and future work should build additional connections with linear plasma stability and subspace system identification. It was further shown how to compute nonlinear, analytic, and data-driven fluid and plasma models that, by construction, satisfy global or local conservation laws. For instance, Sec. 5.2 illustrated globally energy-preserving models that accurately forecast fully 3D, isothermal, Hall-MHD simulations of the HIT-SI device. Moreover, a new system identification method was introduced that can produce nonlinear, data-driven models that are long-term bounded for any new initial condition; as far as the author is aware, this has not been done before in *any* nonlinear systems field. Furthermore, the theorem used for stability promotion relied only the identification of a suitable quadratic Lyapunov function, so work is in progress to extend this for flows that preserve a quantity other than the energy. For instance, there are nonlinear MHD flows that can preserve the magnetic helicity while not preserving the energy. Lastly, there are many other system identification methods that improve the robustness of extracting models from data. Many advanced routines have been incorporated into the open-source PySINDy code, which will continue to provide examples and advanced tools for anyone interested in applying system identification to a dataset. Some of this advanced functionality is brand new in the literature and can facilitate new system identification discoveries.

In total, this thesis investigated a number of machine learning methods for plasma physics applications, invented new types of reduced-order modeling, and illustrated how these mod-

ern techniques can be valuable in plasma physics and related scientific fields.

7.1 Future machine learning work for plasmas

There is still significant work to be done exploring the strengths and limitations of many of these methods for plasmas in different physical parameter regimes; ROMs for inertial confinement fusion will inevitably be different than those for magnetic confinement fusion, or those for low-temperature plasma etching. Since linear stability is so central to plasma physics, especially fusion-relevant plasmas, connecting these methods to the broad analytic literature in linear plasma stability and comparing forms of DMD and SSI would be very useful. Such a follow-up is now underway as part of a new NSF project based on the work presented in this thesis.

In particular, significant progress in understanding the strength and usefulness of these reduced-order models could be made by applying many of these techniques on a number of canonical plasma physics problems, including reconnection in driven [115, 160] and plasmoid generating [42, 331] current sheets and MHD vortex shedding simulations [110, 151, 379]. There remain open questions about if ROMs can correctly reproduce the qualitative features of different fluid and plasma models. For instance, future work should emphasize understanding how these models alter or retain the "direct energy cascade" coming from the interaction of terms in the Navier-Stokes equations, since often some of the dissipative scales of the system are smoothed, truncated, or otherwise modified. Preserving the features of the direct energy cascade in truncated Galerkin and data-driven models for incompressible fluid flows is a current field of research. Since even this "simple case" is unsettled, there is much research to be done in truncated Galerkin models for resistive or Hall-MHD. In resistive MHD the Kolmogorov direct cascade is modified either to a GS95 cascade [138, 139], or dynamical alignment [389]. The theory of dynamical alignment predicts a particularly complicated direct cascade that would be presumably very difficult to capture with data-driven reduced order models. In Hall-MHD there can additionally be inverse cascades and bidirectional cascades [352], further complicating the analysis of model stability and generalizability.

Long term goals of this avenue of research include: evaluating the effectiveness of advanced data decompositions for providing new physical insight into plasma dynamics, finding nonlinear models for plasma dynamics in the subsequent low-dimensional bases – purely from measurements, theoretical constraints, and physical symmetries, and using the same bases to find optimal sensor placements for experimental devices and sensor "paths" for spacecraft. Although the latter was not discussed in this work, there has been a modern revolution in sparse sensor placement algorithms [57, 275, 92, 334] that could be very useful for fusion-relevant plasmas where measurements and diagnostic access are limited, and optimizing diagnostic placement can lead to significant improvements in diagnostic capabilities, diagnostic resilience, and physical understanding [308].

BIBLIOGRAPHY

[1] See supplemental material at `http://link.aps.org/supplemental/10.1103/PhysRevFluids.6.094401` for a movie illustrating a methodological comparison of the vortex shedding behind a circular cylinder.

[2] Hamdi M Abdelhamid, Yohei Kawazura, and Zensho Yoshida. Hamiltonian formalism of extended magnetohydrodynamics. *Journal of Physics A: Mathematical and Theoretical*, 48(23):235502, 2015.

[3] Babak Maboudi Afkham and Jan S Hesthaven. Structure preserving model reduction of parametric Hamiltonian systems. *SIAM Journal on Scientific Computing*, 39(6):A2616–A2644, 2017.

[4] Eric Ahn, Azarakhsh Jalalvand, and Egemen Kolemen. Localizing Alfvén eigenmodes in plasma based on high resolution ECE spectrograms at DIII-D using autoencoders and image processing techniques. *Bulletin of the American Physical Society*, 2021.

[5] Cihan Akcay. *Extended magnetohydrodynamic simulations of the helicity injected torus (HIT-SI) spheromak experiment with the NIMROD code*. PhD thesis, University of Washington, Seattle, 2013.

[6] R Albanese, R Ambrosino, A Castaldo, G De Tommasi, ZP Luo, A Mele, A Pironti, BJ Xiao, and QP Yuan. ITER-like vertical stabilization system for the EAST tokamak. *Nuclear Fusion*, 57(8):086039, 2017.

[7] Frank Allgöwer, Thomas A Badgwell, Joe S Qin, James B Rawlings, and Steven J Wright. Nonlinear predictive control and moving horizon estimation—an introductory overview. In *Advances in control*, pages 391–449. Springer, 1999.

[8] E Paulo Alves and Frederico Fiuza. Data-driven discovery of reduced plasma physics models from fully-kinetic simulations. *arXiv preprint arXiv:2011.01927*, 2020.

[9] Aleksandr Aravkin, Dmitriy Drusvyatskiy, and Tristan van Leeuwen. Variable projection without smoothness. *arXiv preprint arXiv:1601.05011*, 2016.

[10] Marco Ariola, Giuseppe Ambrosino, Alfredo Pironti, Jonathan B Lister, and Parag Vyas. Design and experimental testing of a robust multivariable controller on a tokamak. *IEEE Transactions on Control Systems Technology*, 10(5):646–653, 2002.

[11] Marco Ariola and Alfredo Pironti. Plasma shape control for the JET tokamak: an optimal output regulation approach. *IEEE Control Systems Magazine*, 25(5):65–75, 2005.

[12] Travis Askham and J Nathan Kutz. Variable projection methods for an optimized dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems*, 17(1):380–416, 2018.

[13] Patricia Astrid, Siep Weiland, Karen Willcox, and Ton Backx. Missing point estimation in models described by proper orthogonal decomposition. *IEEE Transactions on Automatic Control*, 53(10):2237–2251, 2008.

[14] Hedy Attouch, Jérôme Bolte, and Benar Fux Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods. *Mathematical Programming*, 137(1):91–129, 2013.

[15] ME Austin and J Lohr. Electron cyclotron emission radiometer upgrade on the DIII-D tokamak. *Review of scientific instruments*, 74(3):1457–1459, 2003.

[16] Peter J Baddoo, Benjamin Herrmann, Beverley J McKeon, J Nathan Kutz, and Steven L Brunton. Physics-informed dynamic mode decomposition (piDMD). *arXiv preprint arXiv:2112.04307*, 2021.

[17] Shervin Bagheri. Effects of weak noise on oscillating flows: Linking quality factor, Floquet modes, and Koopman spectrum. *Physics of Fluids*, 26(9):094104, 2014.

[18] Zhe Bai, Thakshila Wimalajeewa, Zachary Berger, Guannan Wang, Mark Glauser, and Pramod K Varshney. Low-dimensional approach for reconstruction of airfoil data via compressive sensing. *AIAA journal*, 53(4):920–933, 2015.

[19] Joseph Bakarji and Daniel M Tartakovsky. Data-driven discovery of coarse-grained equations. *Journal of Computational Physics*, 434:110219, 2021.

[20] Maciej Balajewicz. Lyapunov stable Galerkin models of post-transient incompressible flows. *arXiv preprint arXiv:1312.0284*, 2013.

[21] Maciej Balajewicz and Earl H Dowell. Stabilization of projection-based reduced order models of the Navier–Stokes. *Nonlinear Dynamics*, 70(2):1619–1632, 2012.

[22] Maciej J Balajewicz, Earl H Dowell, and Bernd R Noack. Low-dimensional modelling of high-Reynolds-number shear flows incorporating constraints from the Navier-Stokes equation. *Journal of Fluid Mechanics*, 729:285, 2013.

[23] Pierre Baldi and Kurt Hornik. Neural networks and principal component analysis: Learning from examples without local minima. *Neural networks*, 2(1):53–58, 1989.

[24] Francesco Ballarin, Andrea Manzoni, Alfio Quarteroni, and Gianluigi Rozza. Supremizer stabilization of POD–Galerkin approximation of parametrized steady incompressible Navier–Stokes equations. *International Journal for Numerical Methods in Engineering*, 102(5):1136–1161, 2015.

[25] Yohai Bar-Sinai, Stephan Hoyer, Jason Hickey, and Michael P Brenner. Learning data-driven discretizations for partial differential equations. *Proceedings of the National Academy of Sciences*, 116(31):15344–15349, 2019.

[26] Peter W Battaglia, Jessica B Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, et al. Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261*, 2018.

[27] Atilim Gunes Baydin, Barak A Pearlmutter, Alexey Andreyevich Radul, and Jeffrey Mark Siskind. Automatic differentiation in machine learning: a survey. *Journal of machine learning research*, 18, 2018.

[28] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

[29] S Beetham and J Capecelatro. Formulating turbulence closures using sparse regression with embedded form invariance. *Physical Review Fluids*, 5(8):084611, 2020.

[30] Sarah Beetham, Rodney O Fox, and Jesse Capecelatro. Sparse identification of multiphase turbulence closures for coupled fluid–particle flows. *Journal of Fluid Mechanics*, 914, 2021.

[31] Paul M Bellan. *Spheromaks: a practical application of magnetohydrodynamic dynamos and plasma self-organization*. World Scientific, 2000.

[32] T. E. Benedett, C. J. Hansen, K. D. Morgan, and T. R. Jarboe. Effects of temperature and density evolution in MHD simulations of HIT-SI. *Physics of Plasmas*, 27(4):042508, 2020.

[33] Thomas Edwin Benedett. *Effects of nonaxisymmetry in equilibria and MHD evolution of spheromaks*. University of Washington, 2020.

[34] Peter Benner, Serkan Gugercin, and Karen Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM review*, 57(4):483–531, 2015.

[35] Peter Benner, Volker Mehrmann, and Danny C Sorensen. *Dimension reduction of large-scale systems*, volume 45. Springer, 2005.

[36] Peter Benner, Mario Ohlberger, Albert Cohen, and Karen Willcox. *Model reduction and approximation: theory and algorithms*. SIAM, 2017.

[37] Austin R Benson, David F Gleich, and James Demmel. Direct QR factorizations for tall-and-skinny matrices in MapReduce architectures. In *2013 IEEE international conference on big data*, pages 264–272. IEEE, 2013.

[38] Ricardo Betti and Jeffrey P Freidberg. Stability of Alfvén gap modes in burning plasmas. *Physics of Fluids B: Plasma Physics*, 4(6):1465–1474, 1992.

[39] Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B Shah. Julia: A fresh approach to numerical computing. *SIAM Review*, 59(1):65–98, 2017.

[40] Bhavana Bhadriraju, Mohammed Saad Faizan Bangi, Abhinav Narasingam, and Joseph Sang-Il Kwon. Operable adaptive sparse identification of systems: Application to chemical processes. *AIChE Journal*, 66(11):e16980, 2020.

[41] Harish S Bhat. Learning and interpreting potentials for classical Hamiltonian systems. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 217–228. Springer, 2019.

[42] A. Bhattacharjee, Yi-Min Huang, H. Yang, and B. Rogers. Fast reconnection in high-lundquist-number plasmas due to the plasmoid instability. *Physics of Plasmas*, 16(11):112102, 2009.

[43] Stephen A Billings. *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*. John Wiley & Sons, 2013.

[44] D Biskamp. Cascade models for magnetohydrodynamic turbulence. *Physical Review E*, 50(4):2702, 1994.

[45] Dieter Biskamp. *Magnetohydrodynamic turbulence*. Cambridge University Press, 2003.

[46] Nicholas M Boffi, Stephen Tu, Nikolai Matni, Jean-Jacques E Slotine, and Vikas Sindhwani. Learning stability certificates from data. *arXiv preprint arXiv:2008.05952*, 2020.

[47] Josh Bongard and Hod Lipson. Automated reverse engineering of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 104(24):9943–9948, 2007.

[48] Lorenzo Boninsegna, Feliks Nüske, and Cecilia Clementi. Sparse learning of stochastic dynamical equations. *The Journal of Chemical Physics*, 148(24):241723, 2018.

[49] M Bornatici, R Cano, O De Barbieri, and F Engelmann. Electron cyclotron emission and absorption in fusion plasmas. *Nuclear Fusion*, 23(9):1153, 1983.

[50] H-S Bosch, RC Wolf, T Andreeva, J Baldzuhn, D Birus, T Bluhm, T Bräuer, H Braune, V Bykov, A Cardella, et al. Technical challenges in the construction of the steady-state stellarator Wendelstein 7-X. *Nuclear Fusion*, 53(12):126001, 2013.

[51] C Bourdelle, J Citrin, B Baiocchi, A Casati, P Cottier, X Garbet, F Imbeaux, and JET Contributors. Core turbulent transport in tokamak plasmas: bridging theory and experiment with QuaLiKiz. *Plasma Physics and Controlled Fusion*, 58(1):014036, 2015.

[52] SI Braginskii and MA Leontovich. Reviews of plasma physics, 1965.

[53] MP Brenner, JD Eldredge, and JB Freund. Perspective on machine learning for advancing fluid mechanics. *Physical Review Fluids*, 4(10):100501, 2019.

[54] Alexander N. Brooks and Thomas J.R. Hughes. Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations. *Computer Methods in Applied Mechanics and Engineering*, 32(1):199 – 259, 1982.

[55] David B Brückner, Pierre Ronceray, and Chase P Broedersz. Inferring the dynamics of underdamped stochastic systems. *Physical review letters*, 125(5):058103, 2020.

[56] O Bruno and D Hoch. Numerical differentiation of approximated functions with limited order-of-accuracy deterioration. *SIAM Journal on Numerical Analysis*, 50(3):1581–1603, 2012.

[57] Bingni W Brunton, Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Sparse sensor placement optimization for classification. *SIAM Journal on Applied Mathematics*, 76(5):2099–2122, 2016.

[58] S. L. Brunton and B. R. Noack. Closed-loop turbulence control: Progress and challenges. *Applied Mechanics Reviews*, 67:050801–1–050801–48, 2015.

[59] S. L. Brunton, J. L. Proctor, and J. N. Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.

[60] S. L. Brunton, J. L. Proctor, J. H. Tu, and J. N. Kutz. Compressed sensing and dynamic mode decomposition. *Journal of Computational Dynamics*, 2(2):165–191, 2015.

[61] Steven L Brunton, Bingni W Brunton, Joshua L Proctor, Eurika Kaiser, and J Nathan Kutz. Chaos as an intermittently forced linear system. *Nature communications*, 8(1):19, 2017.

[62] Steven L Brunton, Marko Budišić, Eurika Kaiser, and J Nathan Kutz. Modern Koopman theory for dynamical systems. *arXiv preprint arXiv:2102.12086*, 2021.

[63] Steven L Brunton and J Nathan Kutz. *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press, 2019.

[64] Steven L Brunton and J Nathan Kutz. Methods for data-driven multiscale model discovery for materials. *Journal of Physics: Materials*, 2(4):044002, 2019.

[65] Steven L Brunton, Bernd R Noack, and Petros Koumoutsakos. Machine learning for fluid mechanics. *Annual Review of Fluid Mechanics*, 52:477–508, 2020.

[66] Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Sparse identification of nonlinear dynamics with control (SINDYc). *IFAC-PapersOnLine*, 49(18):710–715, 2016.

[67] J. M. Burgers. A mathematical model illustrating the theory of turbulence. *Advances in Applied Mechanics*, 1:171–199, 1948.

[68] James V Burke and Michael C Ferris. A Gauss—Newton method for convex composite optimization. *Mathematical Programming*, 71(2):179–194, 1995.

[69] Patrick James Byrne. *Study of External Kink Modes in Shaped HBT-EP Plasmas*. PhD thesis, Columbia University, 2017.

[70] Jared L Callaham, Steven L Brunton, and Jean-Christophe Loiseau. On the role of nonlinear correlations in reduced-order modeling. *arXiv preprint arXiv:2106.02409*, 2021.

[71] Jared L Callaham, J-C Loiseau, Georgios Rigas, and Steven L Brunton. Nonlinear stochastic modelling with Langevin regression. *Proceedings of the Royal Society A*, 477(2250):20210092, 2021.

[72] Debra Callahan. Achieving a burning plasma on the National Ignition Facility (NIF) laser. *Bulletin of the American Physical Society*, 2021.

[73] J. Candy and R. E. Waltz. Anomalous transport scaling in the DIII-D tokamak matched by supercomputer simulation. *Phys. Rev. Lett.*, 91:045001, Jul 2003.

[74] B Cannas, P C de Vries, A Fanni, A Murari, A Pau, and G Sias. Automatic disruption classification in JET with the ITER-like wall. *Plasma Physics and Controlled Fusion*, 57(12):125003, 2015.

[75] V Carbone and P Veltri. Relaxation processes in magnetohydrodynamics-a triad-interaction model. *Astronomy and Astrophysics*, 259:359–372, 1992.

[76] Alejandro Carderera, Sebastian Pokutta, Christof Schütte, and Martin Weiser. Cindy: Conditional gradient-based identification of non-linear dynamics – noise-robust recovery, 2021.

[77] Kevin Carlberg, Matthew Barone, and Harbir Antil. Galerkin v. least-squares Petrov–Galerkin projection in nonlinear model reduction. *Journal of Computational Physics*, 330:693–734, 2017.

[78] Kevin Carlberg, Charbel Bou-Mosleh, and Charbel Farhat. Efficient non-linear model reduction via a least-squares Petrov–Galerkin projection and compressive tensor approximations. *International Journal for numerical methods in engineering*, 86(2):155–181, 2011.

[79] Kevin Carlberg, Charbel Farhat, Julien Cortial, and David Amsallem. The GNAT method for nonlinear model reduction: effective implementation and application to computational fluid dynamics and turbulent flows. *Journal of Computational Physics*, 242:623–647, 2013.

[80] Kevin Carlberg, Ray Tuminaro, and Paul Boggs. Preserving Lagrangian structure in nonlinear model reduction with application to structural dynamics. *SIAM Journal on Scientific Computing*, 37(2):B153–B184, 2015.

[81] Kathleen Champion, Bethany Lusch, J Nathan Kutz, and Steven L Brunton. Data-driven discovery of coordinates and governing equations. *Proceedings of the National Academy of Sciences*, 116(45):22445–22451, 2019.

[82] Kathleen Champion, Peng Zheng, Aleksandr Y Aravkin, Steven L Brunton, and J Nathan Kutz. A unified sparse optimization framework to learn parsimonious physics-informed models from data. *IEEE Access*, 8:169259–169271, 2020.

[83] Haibin Chang and Dongxiao Zhang. Machine learning subsurface flow equations from data. *Computational Geosciences*, 23(5):895–910, 2019.

[84] Ya-Chien Chang, Nima Roohi, and Sicun Gao. Neural Lyapunov control. *arXiv preprint arXiv:2005.00611*, 2020.

[85] Saifon Chaturantabut and Danny C Sorensen. Discrete empirical interpolation for nonlinear model reduction. In *Proceedings of the 48th IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, pages 4316–4321. IEEE, 2009.

[86] Liu Chen and Fulvio Zonca. Physics of Alfvén waves and energetic particles in burning plasmas. *Reviews of Modern Physics*, 88(1):015008, 2016.

[87] CZ Cheng and MS Chance. Low-n shear Alfvén spectra in axisymmetric toroidal plasmas. *The Physics of fluids*, 29(11):3695–3701, 1986.

[88] CZ Cheng, Liu Chen, and MS Chance. High-n ideal and resistive shear Alfvén waves in tokamaks. *Annals of Physics*, 161(1):21–47, 1985.

[89] J Cheng, XZ Jia, and YB Wang. Numerical differentiation and its applications. *Inverse problems in Science and Engineering*, 15(4):339–357, 2007.

[90] Hoang K Chu and Mitsuhiro Hayashibe. Discovering interpretable dynamics by sparsity promotion on energy and the Lagrangian. *IEEE Robotics and Automation Letters*, 5(2):2154–2160, 2020.

[91] Jonathan Citrin, Sarah Breton, Federico Felici, Frederic Imbeaux, T Aniel, JF Artaud, B Baiocchi, C Bourdelle, Y Camenen, and J Garcia. Real-time capable first principle based modelling of tokamak turbulent transport. *Nuclear Fusion*, 55(9):092001, 2015.

[92] Emily Clark, Travis Askham, Steven L Brunton, and J Nathan Kutz. Greedy sensor placement with cost constraints. *IEEE Sensors Journal*, 19(7):2642–2656, 2018.

[93] Clawpack Development Team. Clawpack software, 2020. Version 5.7.1.

[94] Athel Cornish-Bowden. One hundred years of Michaelis–Menten kinetics. *Perspectives in Science*, 4:3–9, 2015.

[95] Alexandre Cortiella, Kwang-Chun Park, and Alireza Doostan. Sparse identification of nonlinear dynamical systems via reweighted l1-regularized least squares. *Computer Methods in Applied Mechanics and Engineering*, 376:113620, 2021.

[96] M Couplet, P Sagaut, and C Basdevant. Intermodal energy transfers in a proper orthogonal decomposition-Galerkin representation of a turbulent separated flow. *Journal of Fluid Mechanics*, 491:275, 2003.

[97] Miles Cranmer, Sam Greydanus, Stephan Hoyer, Peter Battaglia, David Spergel, and Shirley Ho. Lagrangian neural networks. *arXiv preprint arXiv:2003.04630*, 2020.

[98] Miles D Cranmer, Rui Xu, Peter Battaglia, and Shirley Ho. Learning symbolic physics with graph networks. *arXiv preprint arXiv:1909.05862*, 2019.

[99] Rick Dale and Harish S Bhat. Equations of mind: Data science for inferring nonlinear dynamics of socio-cognitive systems. *Cognitive Systems Research*, 52:275–290, 2018.

[100] Magnus Dam, Morten Brøns, Jens Juul Rasmussen, Volker Naulin, and Jan S Hesthaven. Sparse identification of a predator-prey system from simulation data of a convection model. *Physics of Plasmas*, 24(2):022310, 2017.

[101] Scott TM Dawson, Maziar S Hemati, Matthew O Williams, and Clarence W Rowley. Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition. *Experiments in Fluids*, 57(3):1–19, 2016.

[102] Andres De Bustos, E Ascasíbar, Alvaro Cappa, and Rafael Mayo-Garcia. Automatic identification of MHD modes in magnetic fluctuations spectrograms using deep learning techniques. *Plasma Physics and Controlled Fusion*, 2021.

[103] Sebastian De Pascuale, Kenneth Allen, David Green, and Jeremy Lore. Data-driven techniques for time domain decomposition of plasma physics simulation. *Bulletin of the American Physical Society*, 2021.

[104] Brian de Silva, Kathleen Champion, Markus Quade, Jean-Christophe Loiseau, J Nathan Kutz, and Steven Brunton. PySINDy: A Python package for the sparse identification of nonlinear dynamical systems from data. *Journal of Open Source Software*, 5(49):1–4, 2020.

[105] Brian M de Silva, David M Higdon, Steven L Brunton, and J Nathan Kutz. Discovery of physics from data: universal laws and discrepancies. *Frontiers in artificial intelligence*, 3:25, 2020.

[106] AE Deane, IG Kevrekidis, G Em Karniadakis, and SA Orszag. Low-dimensional models for complex geometry flows: application to grooved channels and circular cylinders. *Physics of Fluids A: Fluid Dynamics*, 3(10):2337–2354, 1991.

[107] Charles B Delahunt and J Nathan Kutz. A toolkit for data-driven discovery of governing equations in high-noise regimes. *arXiv preprint arXiv:2111.04870*, 2021.

[108] Nan Deng, Bernd R Noack, Marek Morzyński, and Luc R Pastur. Low-order model for successive bifurcations of the fluidic pinball. *Journal of Fluid Mechanics*, 884:A37, 2020.

[109] Steven Diamond and Stephen Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *The Journal of Machine Learning Research*, 17(1):2909–2913, 2016.

[110] Vincent Dousset and Alban Pothérat. Numerical simulations of a cylinder wake under a strong axial magnetic field. *Physics of Fluids*, 20(1):017104, 2008.

[111] Zlatko Drmac and Serkan Gugercin. A new selection operator for the discrete empirical interpolation method—improved a priori error bound and extensions. *SIAM Journal on Scientific Computing*, 38(2):A631–A648, 2016.

[112] Dmitriy Drusvyatskiy and Courtney Paquette. Efficiency of minimizing compositions of convex functions and smooth maps. *Mathematical Programming*, 178(1):503–558, 2019.

[113] T Dudok de Wit, A-L Pecquet, J-C Vallet, and R Lima. The biorthogonal decomposition as a tool for investigating fluctuations in plasmas. *Physics of Plasmas*, 1(10):3288–3300, 1994.

[114] Karthik Duraisamy, Gianluca Iaccarino, and Heng Xiao. Turbulence modeling in the age of data. *Annual Reviews of Fluid Mechanics*, 51:357–377, 2019.

[115] F. Ebrahimi, R. Raman, E. B. Hooper, C. R. Sovinec, and A. Bhattacharjee. Physics of forced magnetic reconnection in coaxial helicity injection experiments in national spherical torus experiment. *Physics of Plasmas*, 21(5):056109, 2014.

[116] A. W. Edwards, D. J. Campbell, W. W. Engelhardt, H. U. Fahrbach, R. D. Gill, R. S. Granetz, S. Tsuji, B. J. D. Tubbing, A. Weller, J. Wesson, and D. Zasche. Rapid collapse of a plasma sawtooth oscillation in the JET tokamak. *Phys. Rev. Lett.*, 57:210–213, July 1986.

[117] Yosef El Sayed M, Richard Semaan, and Rolf Radespiel. Sparse modeling of the lift gains of a high-lift configuration with periodic coanda blowing. In *2018 AIAA Aerospace Sciences Meeting*, page 1054, 2018.

[118] D.A. Ennis, B.S. Victor, J.S. Wrobel, C. Akcay, T.R. Jarboe, G.J. Marklin, B.A. Nelson, and R.J. Smith. New understandings and achievements from independent-injector drive experiments on HIT-SI. *Nuclear Fusion*, 50(7):072001, jun 2010.

[119] N Benjamin Erichson, Michael Muehlebach, and Michael W Mahoney. Physics-informed autoencoders for Lyapunov-stable fluid flow prediction. *arXiv preprint arXiv:1905.10866*, 2019.

[120] Christopher James Everson. *Experimental Assessment of Confinement and Heating on the HIT-SI3 Spheromak*. PhD thesis, University of Washington, 2020.

[121] Marie Farge and Kai Schneider. Wavelet transforms and their applications to MHD and plasma turbulence: a review. *Journal of Plasma Physics*, 81(6), 2015.

[122] Urban Fasel, Eurika Kaiser, J Nathan Kutz, Bingni W Brunton, and Steven L Brunton. SINDy with control: A tutorial. *arXiv preprint arXiv:2108.13404*, 2021.

[123] Urban Fasel, J Nathan Kutz, Bingni W Brunton, and Steven L Brunton. Ensemble-SINDy: Robust sparse model discovery in the low-data, high-noise limit, with active learning and control. *arXiv preprint arXiv:2111.10992*, 2021.

[124] A Fasoli, JB Lister, SE Sharapov, S Ali-Arshad, G Bosia, D Borba, DJ Campbell, N Deliyanakis, JA Dobbing, C Gormezano, et al. Overview of Alfven eigenmode experiments in JET. *Nuclear fusion*, 35(12):1485, 1995.

[125] G Fiksel, AF Almagri, BE Chapman, VV Mirnov, Y Ren, JS Sarff, and PW Terry. Mass-dependent ion heating during magnetic reconnection in a laboratory plasma. *Physical review letters*, 103(14):145002, 2009.

[126] Eleanor G Forbes, Uri Shumlak, Harry S McLean, Brian A Nelson, Elliot L Claveau, Raymond P Golingo, Drew P Higginson, James M Mitrani, Anton D Stepanov, Kurt K Tummel, et al. Progress toward a compact fusion reactor using the sheared-flow-stabilized Z-pinch. *Fusion Science and Technology*, 75(7):599–607, 2019.

[127] Jeffrey P Freidberg. *Ideal MHD*. Cambridge University Press, 2014.

[128] Alan Frieze, Ravi Kannan, and Santosh Vempala. Fast Monte-Carlo algorithms for finding low-rank approximations. *Journal of the ACM (JACM)*, 51(6):1025–1041, 2004.

[129] Yichen Fu, David Eldon, Keith Erickson, Kornee Kleijwegt, Leonard Lupin-Jimenez, Mark D Boyer, Nick Eidietis, Nathaniel Barbour, Olivier Izacard, and Egemen Kolemen. Machine learning control for disruption and tearing mode avoidance. *Physics of Plasmas*, 27(2):022501, 2020.

[130] Kai Fukami, Takaaki Murata, Kai Zhang, and Koji Fukagata. Sparse identification of nonlinear dynamics with low-dimensionalized flow representations. *Journal of Fluid Mechanics*, 926, 2021.

[131] C Galperti, S Coda, BP Duval, X Llobet, P Milne, O Sauter, JM Moret, and D Testa. Integration of a real-time node for magnetic perturbations signal analysis in the distributed digital control system of the TCV tokamak. *IEEE Transactions on Nuclear Science*, 64(6):1446–1454, 2017.

[132] C Galperti, C Marchetto, E Alessi, D Minelli, M Mosconi, F Belli, L Boncagni, A Botrugno, P Buratti, B Esposito, et al. Development of real-time MHD markers based on biorthogonal decomposition of signals from Mirnov coils. *Plasma Physics and Controlled Fusion*, 56(11):114012, 2014.

[133] Sébastien Galtier. *Introduction to modern magnetohydrodynamics*. Cambridge University Press, 2016.

[134] Manuel Garcia-Munoz, SE Sharapov, MA Van Zeeland, Enrique Ascasibar, A Cappa, L Chen, J Ferreira, Joaquin Galdon-Quiroga, Benedikt Geiger, J Gonzalez-Martin, et al. Active control of Alfvén eigenmodes in magnetically confined toroidal plasmas. *Plasma Physics and Controlled Fusion*, 61(5):054007, 2019.

[135] Patrick Gelß, Stefan Klus, Jens Eisert, and Christof Schütte. Multidimensional approximation of nonlinear dynamical systems. *Journal of Computational and Nonlinear Dynamics*, 14(6), 2019.

[136] Euhanna Ghadimi, André Teixeira, Iman Shames, and Mikael Johansson. Optimal parameter selection for the alternating direction method of multipliers (ADMM): quadratic problems. *IEEE Transactions on Automatic Control*, 60(3):644–658, 2014.

[137] A Goeßmann, M Götte, I Roth, R Sweke, G Kutyniok, and J Eisert. Tensor network approaches for learning non-linear dynamical laws. *arXiv preprint arXiv:2002.12388*, 2020.

[138] P Goldreich and S Sridhar. Toward a theory of interstellar turbulence. 2: Strong Alfvénic turbulence. *The Astrophysical Journal*, 438:763–775, 1995.

[139] P Goldreich and S Sridhar. Magnetohydrodynamic turbulence revisited. *The Astrophysical Journal*, 485(2):680, 1997.

[140] Andy Goldschmidt, Eurika Kaiser, Jonathan L Dubois, Steven L Brunton, and J Nathan Kutz. Bilinear dynamic mode decomposition for quantum control. *New Journal of Physics*, 23(3):033035, 2021.

[141] Gene H Golub et al. Matrix computations. *The Johns Hopkins*, 1996.

[142] Gene H Golub, Per Christian Hansen, and Dianne P O'Leary. Tikhonov regularization and total least squares. *SIAM journal on matrix analysis and applications*, 21(1):185–194, 1999.

[143] Francisco J Gonzalez and Maciej Balajewicz. Deep convolutional recurrent autoencoders for learning low-dimensional feature dynamics of fluid systems. *arXiv preprint arXiv:1808.01346*, 2018.

[144] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.

[145] TP Goodman, F Felici, O Sauter, JP Graves, TCV team, et al. Sawtooth pacing by real-time auxiliary power control in a tokamak plasma. *Physical review letters*, 106(24):245002, 2011.

[146] David Gottlieb and Steven A Orszag. *Numerical analysis of spectral methods: theory and applications*. SIAM, 1977.

[147] Pawan Goyal and Peter Benner. Learning low-dimensional quadratic-embeddings of high-fidelity nonlinear dynamics using deep learning. *arXiv preprint arXiv:2111.12995*, 2021.

[148] Harold Grad and Hanan Rubin. Hydromagnetic equilibria and force-free fields. *Journal of Nuclear Energy (1954)*, 7(3-4):284–285, 1958.

[149] Sebastian Grimberg, Charbel Farhat, and Noah Youkilis. On the stability of projection-based model order reduction for convection-dominated laminar and turbulent flows. *Journal of Computational Physics*, 419:109681, 2020.

[150] Daniel Grošelj, Alfred Mallet, Nuno F. Loureiro, and Frank Jenko. Fully kinetic simulation of 3D kinetic Alfvén turbulence. *Phys. Rev. Lett.*, 120:105101, Mar 2018.

[151] M. Gruszecki, V. M. Nakariakov, T. Van Doorsselaere, and T. D. Arber. Phenomenon of Alfvénic vortex shedding. *Phys. Rev. Lett.*, 105:055004, Jul 2010.

[152] S Gu, B Wan, Y Sun, N Chu, YQ Liu, T Shi, HH Wang, M Jia, and K He. A new criterion for controlling edge localized modes based on a multi-mode plasma response. *Nuclear Fusion*, 59(12):126042, 2019.

[153] Yifei Guan, Steven L Brunton, and Igor Novosselov. Sparse nonlinear models of chaotic electroconvection. *arXiv preprint arXiv:2009.11862*, 2020.

[154] F. Gueniat, L. Mathelin, and L. Pastur. A dynamic mode decomposition approach for large and arbitrarily sampled systems. *Physics of Fluids*, 27(2):025113, 2015.

[155] Daniel R Gurevich, Patrick AK Reinbold, and Roman O Grigoriev. Robust and optimal sparse regression for nonlinear PDE models. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(10):103113, 2019.

[156] Daniel R Gurevich, Patrick AK Reinbold, and Roman O Grigoriev. Learning fluid physics from highly turbulent data using sparse physics-informed discovery of empirical relations (SPIDER). *arXiv preprint arXiv:2105.00048*, 2021.

[157] Ammar H Hakim. Extended MHD modelling with the ten-moment equations. *Journal of Fusion Energy*, 27(1):36–43, 2008.

[158] Christopher Ham, Andrew Kirk, Stanislas Pamela, and Howard Wilson. Filamentary plasma eruptions and their control on the route to fusion energy. *Nature Reviews Physics*, 2(3):159–167, 2020.

[159] C. Hansen, D. P. Boyle, J. C. Schmitt, and R. Majeski. Equilibrium reconstruction with 3D eddy currents in the Lithium Tokamak eXperiment. *Physics of Plasmas*, 24(4):042513, 2017.

[160] C Hansen, G Marklin, B Victor, C Akcay, and T Jarboe. Simulation of injector dynamics during steady inductive helicity injection current drive in the HIT-SI experiment. *Physics of Plasmas*, 22(4):042505, 2015.

[161] C Hansen, B Victor, K Morgan, T Jarboe, A Hossack, G Marklin, BA Nelson, and D Sutherland. Numerical studies and metric development for validation of magnetohydrodynamic models on the HIT-SI experiment. *Physics of Plasmas*, 22(5):056105, 2015.

[162] Christopher Hansen, Aaron Hossack, Kyle Morgan, and Derek Sutherland. Development of a plasma-coupled circuit model for MHD simulations of inductive helicity injection. *Bulletin of the American Physical Society*, 2021.

[163] Christopher James Hansen. *MHD modeling in complex 3D geometries: Towards predictive simulation of SIHI current drive*. PhD thesis, University of Washington, 2014.

[164] Mahdi Hashemi. Enlarging smaller images before inputting into convolutional neural network: zero-padding vs. interpolation. *Journal of Big Data*, 6(1):1–13, 2019.

[165] Patrick Héas and Cédric Herzet. Optimal low-rank dynamic mode decomposition. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4456–4460. IEEE, 2017.

[166] William W Heidbrink, Michael A Van Zeeland, Max E Austin, Andreas Bierwage, Liu Chen, GJ Choi, Philipp Lauber, Zhihong Lin, George R McKee, and Donald A Spong. 'BAAE' instabilities observed without fast ion drive. *Nuclear Fusion*, 61(1):016029, 2020.

[167] WW Heidbrink. Basic physics of Alfvén instabilities driven by energetic particles in toroidally confined plasmas. *Physics of Plasmas*, 15(5):055501, 2008.

[168] WW Heidbrink, EJ Strait, MS Chu, and AD Turnbull. Observation of beta-induced Alfvén eigenmodes in the DIII-D tokamak. *Physical review letters*, 71(6):855, 1993.

[169] WW Heidbrink, EJ Strait, E Doyle, G Sager, and RT Snider. An investigation of beam driven Alfvén instabilities in the DIII-D tokamak. *Nuclear Fusion*, 31(9):1635, 1991.

[170] WW Heidbrink, MA Van Zeeland, ME Austin, NA Crocker, XD Du, GR McKee, and DA Spong. Stability of beta-induced Alfven eigenmodes (BAE) in DIII-D. *Nuclear Fusion*, 61(6):066031, 2021.

[171] Maziar S Hemati, Matthew O Williams, and Clarence W Rowley. Dynamic mode decomposition for large and streaming datasets. *Physics of Fluids (1994-present)*, 26(11):111701, 2014.

[172] Benjamin Herrmann, Peter J Baddoo, Richard Semaan, Steven L Brunton, and Beverley J McKeon. Data-driven resolvent analysis. *Journal of Fluid Mechanics*, 918, 2021.

[173] Seth M Hirsh, David A Barajas-Solano, and J Nathan Kutz. Sparsifying priors for Bayesian uncertainty quantification in model discovery. *arXiv preprint arXiv:2107.02107*, 2021.

[174] Udo Hoefel, Matthias Hirsch, Sehyun Kwak, Andrea Pavone, Jakob Svensson, Torsten Stange, Hans-Jürgen Hartfuß, Jonathan Schilling, Gavin Weir, Johan Willem Oosterbeek, et al. Bayesian modeling of microwave radiometer calibration on the example of the Wendelstein 7-X electron cyclotron emission diagnostic. *Review of Scientific Instruments*, 90(4):043502, 2019.

[175] Moritz Hoffmann, Christoph Fröhner, and Frank Noé. Reactive SINDy: Discovering governing reactions from concentration data. *The Journal of chemical physics*, 150(2):025101, 2019.

[176] Philip Holmes and John Guckenheimer. *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*, volume 42 of *Applied Mathematical Sciences*. Springer-Verlag, Berlin, Heidelberg, 1983.

[177] Philip Holmes, John L Lumley, Gahl Berkooz, and Clarence W Rowley. *Turbulence, coherent structures, dynamical systems and symmetry*. Cambridge university press, 2012.

[178] E B Hooper, T A Kopriva, B I Cohen, D N Hill, H S McLean, R D Wood, S Woodruff, and C R Sovinec. Magnetic reconnection during flux conversion in a driven spheromak. *Physics of Plasmas*, 12(9):092503, 2005.

[179] E Bickford Hooper, Bruce I Cohen, HS McLean, RD Wood, CA Romero-Talamas, and CR Sovinec. NIMROD resistive magnetohydrodynamic simulations of spheromak physics. *Physics of Plasmas*, 15(3):032502, 2008.

[180] E. Hopf. A mathematical example displaying features of turbulence. *Communications on Pure and Applied Mathematics*, 1(4):302–322, 1948.

[181] Aaron Hossack. *A study of plasma dynamics in HIT-SI using ion Doppler spectroscopy*. PhD thesis, University of Washington, Seattle, 2015.

[182] Aaron C Hossack, Taylor Firman, Thomas R Jarboe, James R Prager, Brian S Victor, Jonathan S Wrobel, and Timothy Ziemba. Reduction of plasma density in the helicity injected torus with steady inductance experiment by using a helicon pre-ionization source. *Review of Scientific Instruments*, 84(10):103506, 2013.

[183] AC Hossack, TR Jarboe, RN Chandra, KD Morgan, DA Sutherland, JM Penna, CJ Everson, and BA Nelson. Plasma response to sustainment with imposed-dynamo current drive in HIT-SI and HIT-SI3. *Nuclear Fusion*, 57(7):076026, 2017.

[184] Joseph D Huba. NRL plasma formulary. Technical report, Naval Research Lab Washington DC plasma physics division, 2006.

[185] Kelli D Humbird, Jayson Luc Peterson, BK Spears, and RG McClarren. Transfer learning to model inertial confinement fusion experiments. *IEEE Transactions on Plasma Science*, 48(1):61–70, 2019.

[186] VA Izzo and TR Jarboe. A numerical assessment of the Lundquist number requirement for relaxation current drive. *Physics of Plasmas*, 10(7):2903–2911, 2003.

[187] Azarakhsh Jalalvand, Alan Ali Kaptanoglu, Alvin Garcia, Andrew Oakleigh Nelson, Joseph Abbate, Max E Austin, Geert Verdoolaege, Steven Brunton, William W Heidbrink, and Egemen Kolemen. Alfvén eigenmode classification based on ECE diagnostics at DIII-D using deep recurrent neural networks. *Nuclear Fusion*, 2021.

[188] Sören Jalas, Manuel Kirchen, Philipp Messner, Paul Winkler, Lars Hübner, Julian Dirkwinkel, Matthias Schnepp, Remi Lehe, and Andreas R Maier. Bayesian optimization of a laser-plasma accelerator. *Physical review letters*, 126(10):104801, 2021.

[189] Thomas R Jarboe. Review of spheromak research. *Plasma Physics and Controlled Fusion*, 36(6):945, 1994.

[190] TR Jarboe, C Akcay, MA Chilenski, DA Ennis, CJ Hansen, NK Hicks, RZ Aboul Hosn, AC Hossack, GJ Marklin, BA Nelson, et al. Recent results from the HIT-SI experiment. *Nuclear Fusion*, 51(6):063029, 2011.

[191] TR Jarboe, WT Hamp, GJ Marklin, BA Nelson, RG O'Neill, AJ Redd, PE Sieck, RJ Smith, and JS Wrobel. Spheromak formation by steady inductive helicity injection. *Physical review letters*, 97(11):115003, 2006.

[192] TR Jarboe, CJ Hansen, AC Hossack, GJ Marklin, KD Morgan, BA Nelson, DA Sutherland, and BS Victor. A proof of principle of imposed dynamo current drive: Demonstration of sufficient confinement. *Fusion Science and Technology*, 66(3):369–384, 2014.

[193] TR Jarboe, BS Victor, BA Nelson, CJ Hansen, C Akcay, DA Ennis, NK Hicks, AC Hossack, GJ Marklin, and RJ Smith. Imposed-dynamo current drive. *Nuclear Fusion*, 52(8):083017, 2012.

[194] Stephen Jardin. *Computational methods in plasma physics*. CRC Press, 2010.

[195] Yu-Xin Jiang, Xiong Xiong, Shuo Zhang, Jia-Xiang Wang, Jia-Chun Li, and Lin Du. Modeling and prediction of the transmission dynamics of COVID-19 based on the SINDy-LM method. *Nonlinear Dynamics*, 105(3):2775–2794, 2021.

[196] R Jiménez-Gómez, E Ascasíbar, T Estrada, I García-Cortés, B Van Milligen, A López-Fraguas, I Pastor, and D López-Bruna. Analysis of magnetohydrodynamic instabilities in TJ-II plasmas. *Fusion science and technology*, 51(1):20–30, 2007.

[197] Wanxin Jin, Zhaoran Wang, Zhuoran Yang, and Shaoshuai Mou. Neural certificates for safe control policies. *arXiv preprint arXiv:2006.08465*, 2020.

[198] Kenneth A Johnson and Roger S Goody. The original Michaelis constant: translation of the 1913 Michaelis–Menten paper. *Biochemistry*, 50(39):8264–8269, 2011.

[199] Mihailo R Jovanović, Peter J Schmid, and Joseph W Nichols. Sparsity-promoting dynamic mode decomposition. *Physics of Fluids*, 26(2):024103, 2014.

[200] Kadierdan Kaheman, Steven L Brunton, and J Nathan Kutz. Automatic differentiation to simultaneously identify nonlinear dynamics and extract noise probability distributions from data. *arXiv preprint arXiv:2009.08810*, 2020.

[201] Kadierdan Kaheman, J Nathan Kutz, and Steven L Brunton. SINDy-PI: a robust algorithm for parallel implicit sparse identification of nonlinear dynamics. *Proceedings of the Royal Society A*, 476(2242):20200279, 2020.

[202] E. Kaiser, B. R. Noack, L. Cordier, A. Spohn, M. Segond, M. Abel, G. Daviller, J. Osth, S. Krajnovic, and R. K. Niven. Cluster-based reduced-order modelling of a mixing layer. *J. Fluid Mech.*, 754:365–414, 2014.

[203] Eurika Kaiser, J Nathan Kutz, and Steven L Brunton. Discovering conservation laws from data for control. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 6415–6421. IEEE, 2018.

[204] Eurika Kaiser, J Nathan Kutz, and Steven L Brunton. Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. *Proceedings of the Royal Society of London A*, 474(2219), 2018.

[205] Aniketh Kalur, Talha Mushtaq, Peter Seiler, and Maziar S Hemati. Estimating regions of attraction for transitional flows using quadratic constraints. *IEEE Control Systems Letters*, 2021.

[206] Aniketh Kalur, Peter Seiler, and Maziar S Hemati. Nonlinear stability analysis of transitional flows using quadratic constraints. *Physical Review Fluids*, 6(4):044401, 2021.

[207] A. A. Kaptanoglu, K. D. Morgan, C. J. Hansen, and S. L. Brunton. Characterizing magnetized plasmas with dynamic mode decomposition. *Physics of Plasmas*, 27(3):032108, 2020.

[208] Alan A Kaptanoglu, Tom E Benedett, Kyle D Morgan, Chris J Hansen, and Thomas R Jarboe. Two-temperature effects in Hall-MHD simulations of the HIT-SI experiment. *Physics of Plasmas*, 27(7):072505, 2020.

[209] Alan A Kaptanoglu, Jared L Callaham, Aleksandr Aravkin, Christopher J Hansen, and Steven L Brunton. Promoting global stability in data-driven models of quadratic nonlinear dynamics. *Physical Review Fluids*, 6(9):094401, 2021.

[210] Alan A. Kaptanoglu, Brian M. de Silva, Urban Fasel, Kadierdan Kaheman, Jared L. Callaham, Charles B. Delahunt, Kathleen Champion, Jean-Christophe Loiseau, J. Nathan Kutz, and Steven L. Brunton. PySINDy: A comprehensive Python package for robust sparse system identification. *arXiv preprint arXiv:2111.08481*, 2021.

[211] Alan A Kaptanoglu, Kyle D Morgan, Chris J Hansen, and Steven L Brunton. Physics-constrained, low-dimensional models for magnetohydrodynamics: First-principles and data-driven approaches. *Physical Review E*, 104(1):015206, 2021.

[212] Alan A Kaptanoglu, Kyle D Morgan, Christopher J Hansen, and Steven L Brunton. The structure of global conservation laws in Galerkin plasma models. *arXiv preprint arXiv:2101.03436*, 2021.

[213] Michael G Kapteyn, David J Knezevic, and Karen Willcox. Toward predictive digital twins via component-based reduced-order models and interpretable machine learning. In *AIAA Scitech 2020 Forum*, page 0418, 2020.

[214] Julian Kates-Harbeck, Alexey Svyatkovskiy, and William Tang. Predicting disruptive instabilities in controlled fusion plasmas through deep learning. *Nature*, 568(7753):526–531, 2019.

[215] Hassan K Khalil. Nonlinear systems. *Upper Saddle River*, 2002.

[216] S Mohammad Khansari-Zadeh and Aude Billard. Learning control Lyapunov function to ensure stability of dynamical system-based robot reaching motions. *Robotics and Autonomous Systems*, 62(6):752–765, 2014.

[217] Charlson C Kim, Yueqiang Liu, Paul B Parks, Lang L Lao, Michael Lehnen, and Alberto Loarte. Shattered pellet injection simulations with NIMROD. *Physics of Plasmas*, 26(4):042510, 2019.

[218] Young C Kim and Edward J Powers. Digital bispectral analysis and its applications to nonlinear wave interactions. *IEEE transactions on plasma science*, 7(2):120–131, 1979.

[219] H Kimura, Y Kusama, M Saigusa, GJ Kramer, K Tobita, M Nemoto, T Kondoh, T Nishitani, O Da Costa, T Ozeki, et al. Alfvén eigenmode and energetic particle research in JT-60U. *Nuclear Fusion*, 38(9):1303, 1998.

[220] Jacob R King, Keith H Burrell, Andrea M Garofalo, Richard J Groebner, SE Kruger, Alexei Y Pankin, and Philip B Snyder. NIMROD modeling of quiescent H-mode: reconstruction considerations and saturation mechanism. *Nuclear Fusion*, 57(2):022002, 2016.

[221] Anna Klimovskaia, Stefan Ganscha, and Manfred Claassen. Sparse regression based structure learning of stochastic reaction networks from single cell snapshot time series. *PLoS computational biology*, 12(12):e1005234, 2016.

[222] Stefan Klus, Feliks Nüske, Péter Koltai, Hao Wu, Ioannis Kevrekidis, Christof Schütte, and Frank Noé. Data-driven model reduction and transfer operator approximation. *Journal of Nonlinear Science*, 28(3):985–1010, 2018.

[223] Stefan Klus, Feliks Nüske, Péter Koltai, Hao Wu, Ioannis Kevrekidis, Christof Schütte, and Frank Noé. Data-driven model reduction and transfer operator approximation. *Journal of Nonlinear Science*, 2018.

[224] G Kluth, KD Humbird, BK Spears, JL Peterson, HA Scott, MV Patel, J Koning, M Marinak, L Divol, and CV Young. Deep learning for NLTE spectral opacities. *Physics of Plasmas*, 27(5):052707, 2020.

[225] Sumire Kobayashi, Özgür D Gürcan, and Patrick H Diamond. Direct identification of predator-prey dynamics in gyrokinetic simulations. *Physics of Plasmas*, 22(9):090702, 2015.

[226] Dmitrii Kochkov, Jamie A Smith, Ayya Alieva, Qing Wang, Michael P Brenner, and Stephan Hoyer. Machine learning–accelerated computational fluid dynamics. *Proceedings of the National Academy of Sciences*, 118(21), 2021.

[227] E Kolemen, SL Allen, BD Bray, ME Fenstermacher, DA Humphreys, AW Hyatt, CJ Lasnier, AW Leonard, MA Makowski, AG McLean, et al. Heat flux management via advanced magnetic divertor configurations and divertor detachment. *Journal of Nuclear Materials*, 463:1186–1190, 2015.

[228] J Zico Kolter and Gaurav Manek. Learning stable deep dynamics models. In *Advances in Neural Information Processing Systems*, volume 32, pages 11128–11136, 2019.

[229] B. O. Koopman. Hamiltonian systems and transformation in Hilbert space. *Proceedings of the National Academy of Sciences*, 17(5):315–318, 1931.

[230] R. H. Kraichnan and S. Chen. Is there a statistical mechanics of turbulence? *Physica D: Nonlinear Phenomena*, 37:160–172, 1989.

[231] Boris Kramer. Stability domains for quadratic-bilinear reduced-order models. *SIAM Journal on Applied Dynamical Systems*, 20(2):981–996, 2021.

[232] Richard Michael Jack Kramer, Eric C Cyr, Sean Miller, Edward Geoffrey Phillips, Gregg Arthur Radtke, Allen C Robinson, and John N Shadid. A plasma modeling hierarchy and verification approach. 2020.

[233] R Kube, RM Churchill, and B Sturdevant. Machine learning accelerated particle-in-cell plasma simulations. *arXiv preprint arXiv:2110.12444*, 2021.

[234] Akira Kusaba, Tetsuji Kuboyama, and Shigeru Inagaki. Sparsity-promoting dynamic mode decomposition of plasma turbulence. *Plasma and Fusion Research*, 15:1301001–1301001, 2020.

[235] J Nathan Kutz, Steven L Brunton, Bingni W Brunton, and Joshua L Proctor. *Dynamic mode decomposition: data-driven modeling of complex systems*, volume 149. SIAM, 2016.

[236] J Nathan Kutz, Xing Fu, and Steven L Brunton. Multiresolution dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems*, 15(2):713–735, 2016.

[237] John H Lagergren, John T Nardini, G Michael Lavigne, Erica M Rutter, and Kevin B Flores. Learning partial differential equations for biological transport models from noisy spatio-temporal data. *Proceedings of the Royal Society A*, 476(2234):20190800, 2020.

[238] Zhilu Lai and Satish Nagarajaiah. Sparse structural system identification method for nonlinear dynamic systems with hysteresis/inelastic behavior. *Mechanical Systems and Signal Processing*, 117:813–842, 2019.

[239] PT Lang, AW Degeling, JB Lister, YR Martin, PJ Mc Carthy, ACC Sips, W Suttrop, GD Conway, L Fattorini, O Gruber, et al. Frequency control of type-I ELMs by magnetic triggering in ASDEX upgrade. *Plasma physics and controlled fusion*, 46(11):L31, 2004.

[240] LL Lao, H St John, RD Stambaugh, AG Kellman, and W Pfeiffer. Reconstruction of current profile parameters and plasma shapes in tokamaks. *Nuclear fusion*, 25(11):1611, 1985.

[241] Davide Lasagna, Deqing Huang, Owen R Tutty, and Sergei Chernyshenko. Sum-of-squares approach to feedback control of laminar wake flows. *Journal of Fluid Mechanics*, 809:628–663, 2016.

[242] Kookjin Lee and Kevin Carlberg. Deep conservation: A latent-dynamics model for exact satisfaction of physical conservation laws. *arXiv preprint arXiv:1909.09754*, 2019.

[243] Kookjin Lee and Kevin T Carlberg. Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders. *Journal of Computational Physics*, 404:108973, 2020.

[244] Thomas Lessinnes, Franck Plunian, and Daniele Carati. Helical shell models for MHD. *Theoretical and Computational Fluid Dynamics*, 23(6):439–450, 2009.

[245] Kenneth Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of applied mathematics*, 2(2):164–168, 1944.

[246] JP Levesque, N Rath, D Shiraki, S Angelini, J Bialek, PJ Byrne, BA DeBono, PE Hughes, ME Mauel, GA Navratil, et al. Multimode observations and 3D magnetic control of the boundary of a tokamak plasma. *Nuclear Fusion*, 53(7):073037, 2013.

[247] William E Lewis, Patrick F Knapp, Stephen A Slutz, Paul F Schmit, Gordon A Chandler, Matthew R Gomez, Adam J Harvey-Thompson, Michael A Mangan, David J Ampleford, and Kristian Beckwith. Deep-learning-enabled Bayesian inference of fuel magnetization in magnetized liner inertial fusion. *Physics of Plasmas*, 28(9):092701, 2021.

[248] Pan Li, Yadong Li, Jiangang Li, Guojiang Wu, Wei Chen, Jingsen Geng, Fei Chen, Yuhao Wang, Bin Zhang, Liqing Xu, et al. Dynamics between toroidal Alfvén eigenmode evolution and turbulence suppression under RMP on EAST. *Nuclear Fusion*, 2021.

[249] Qianxiao Li, Felix Dietrich, Erik M. Bollt, and Ioannis G. Kevrekidis. Extended dynamic mode decomposition with dictionary learning: A data-driven adaptive spectral decomposition of the Koopman operator. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(10):103111, 2017.

[250] Shanwu Li, Eurika Kaiser, Shujin Laima, Hui Li, Steven L Brunton, and J Nathan Kutz. Discovering time-varying aerodynamics of a prototype bridge by sparse identification of nonlinear dynamical systems. *Physical Review E*, 100(2):022220, 2019.

[251] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

[252] Edo Liberty, Franco Woolfe, Per-Gunnar Martinsson, Vladimir Rokhlin, and Mark Tygert. Randomized algorithms for the low-rank approximation of matrices. *Proceedings of the National Academy of Sciences*, 104(51):20167–20172, 2007.

[253] Julia Ling, Andrew Kurzawski, and Jeremy Templeton. Reynolds averaged turbulence modelling using deep neural networks with embedded invariance. *Journal of Fluid Mechanics*, 807:155–166, 2016.

[254] Tong Liu, ZR Wang, Mark D Boyer, Stefano Munaretto, Zheng-Xiong Wang, B-H Park, NC Logan, SeongMoo Yang, and J-K Park. Identification of multiple eigenmode growth rates towards real time detection in DIII-D and KSTAR tokamak plasmas. *Nuclear Fusion*, 61(5):056009, 2021.

[255] A Loarte, B Lipschultz, AS Kukushkin, GF Matthews, PC Stangeby, N Asakura, GF Counsell, G Federici, A Kallenbach, K Krieger, et al. Power and particle control. *Nuclear Fusion*, 47(6):S203, 2007.

[256] Jean-Christophe Loiseau. Data-driven modeling of the chaotic thermal convection in an annular thermosyphon. *Theoretical and Computational Fluid Dynamics*, 34(4):339–365, 2020.

[257] Jean-Christophe Loiseau and Steven L Brunton. Constrained sparse Galerkin regression. *Journal of Fluid Mechanics*, 838:42–67, 2018.

[258] Jean-Christophe Loiseau, Steven L Brunton, and Bernd R Noack. From the POD-Galerkin method to sparse manifold models. *Handbook of Model-Order Reduction*, 2:1–47, 2019.

[259] Jean-Christophe Loiseau, Bernd R Noack, and Steven L Brunton. Sparse reduced-order modeling: sensor-based dynamics to full-state estimation. *Journal of Fluid Mechanics*, 844:459–490, 2018.

[260] Jeremy Lore, Sebastian De Pascuale, Paul Laiu, Birdy Phathanapirom, Steven Brunton, John Canik, Sacit Cetiner, Nathan Kutz, and Peter Stangeby. Model predictive

control of boundary plasmas using reduced models derived from SOLPS-ITER. *Bulletin of the American Physical Society*, 2021.

[261] Edward N Lorenz. Deterministic nonperiodic flow. *Journal of atmospheric sciences*, 20(2):130–141, 1963.

[262] Stefano Lorenzi, Antonio Cammi, Lelio Luzzi, and Gianluigi Rozza. POD-Galerkin method for finite volume approximation of Navier–Stokes and RANS equations. *Computer Methods in Applied Mechanics and Engineering*, 311:151–179, 2016.

[263] John Loverich, Ammar Hakim, and Uri Shumlak. A discontinuous Galerkin method for ideal two-fluid plasma equations. *Communications in Computational Physics*, 9(2):240–268, 2011.

[264] Peter Lu and Pierre FJ Lermusiaux. Bayesian learning of stochastic dynamical models. *Physica D: Nonlinear Phenomena*, page 133003, 2021.

[265] Mitul Luhar, Ati S Sharma, and Beverley J McKeon. Opposition control within the resolvent analysis framework. *Journal of Fluid Mechanics*, 749:597–626, 2014.

[266] Bethany Lusch, J Nathan Kutz, and Steven L Brunton. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature communications*, 9(1):1–10, 2018.

[267] Suryanarayana Maddu, Bevan L Cheeseman, Ivo F Sbalzarini, and Christian L Müller. Stability selection enables robust learning of partial differential equations from limited noisy data. *arXiv preprint arXiv:1907.07810*, 2019.

[268] B Madsen, M Salewski, WW Heidbrink, L Stagner, M Podestà, D Lin, AV Garcia, Per Christian Hansen, J Huang, et al. Tomography of the positive-pitch fast-ion velocity distribution in DIII-D plasmas with Alfvén eigenmodes and neoclassical tearing modes. *Nuclear Fusion*, 60(6):066024, 2020.

[269] Andrew J Majda and John Harlim. Physics constrained nonlinear regression models for time series. *Nonlinearity*, 26(1):201, 2012.

[270] Andrew J Majda and Ilya Timofeyev. Remarkable statistical behavior for truncated Burgers–Hopf dynamics. *Proceedings of the National Academy of Sciences*, 97(23):12413–12417, 2000.

[271] E Maljaars, F Felici, TC Blanken, Christian Galperti, O Sauter, MR De Baar, F Carpanese, TP Goodman, D Kim, SH Kim, et al. Profile control simulations and experiments on TCV: a controller test environment and results using a model-based predictive controller. *Nuclear Fusion*, 57(12):126063, 2017.

[272] Niall M Mangan, Travis Askham, Steven L Brunton, J Nathan Kutz, and Joshua L Proctor. Model selection for hybrid dynamical systems via sparse regression. *Proceedings of the Royal Society A*, 475(2223):20180534, 2019.

[273] Niall M Mangan, Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Inferring biological networks by sparse identification of nonlinear dynamics. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, 2(1):52–63, 2016.

[274] Niall M Mangan, J Nathan Kutz, Steven L Brunton, and Joshua L Proctor. Model selection for dynamical systems via sparse regression and information criteria. *Proceedings of the Royal Society A*, 473(2204):1–16, 2017.

[275] Krithika Manohar, Bingni W Brunton, J Nathan Kutz, and Steven L Brunton. Data-driven sparse sensor placement for reconstruction: Demonstrating the benefits of exploiting known patterns. *IEEE Control Systems Magazine*, 38(3):63–86, 2018.

[276] Krithika Manohar, Eurika Kaiser, Steven L Brunton, and J Nathan Kutz. Optimized sampling for multiscale dynamics. *Multiscale Modeling & Simulation*, 17(1):117–136, 2019.

[277] Andreas Mardt, Luca Pasquali, Hao Wu, and Frank Noé. VAMPnets: Deep learning of molecular kinetics. *Nature Communications*, 9(5), 2018.

[278] Donald W Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.

[279] Stefano Massaroli, Michael Poli, Michelangelo Bin, Jinkyoo Park, Atsushi Yamashita, and Hajime Asama. Stable neural flows. *arXiv preprint arXiv:2003.08063*, 2020.

[280] Francisco A Matos, Diogo R Ferreira, Pedro J Carvalho, and JET Contributors. Deep learning for plasma tomography using the bolometer system at JET. *Fusion engineering and design*, 114:18–25, 2017.

[281] Romit Maulik, Omer San, Adil Rasheed, and Prakash Vedula. Subgrid modelling for two-dimensional turbulence using neural networks. *Journal of Fluid Mechanics*, 858:122–144, 2019.

[282] RM Mayo and GJ Marklin. Numerical calculation of Mercier beta limits in spheromaks. *The Physics of fluids*, 31(6):1812–1815, 1988.

[283] William D McComb. The physics of fluid turbulence. *Oxford*, 1990.

[284] Beverley J McKeon and Ati S Sharma. A critical-layer framework for turbulent pipe flow. *Journal of Fluid Mechanics*, 658:336, 2010.

[285] E T Meier, R J Goldston, E G Kaveeva, M A Makowski, S Mordijck, V A Rozhansky, I Yu Senichenkov, and S P Voskoboynikov. Analysis of drift effects on the tokamak power scrape-off width using SOLPS-ITER. *Plasma Physics and Controlled Fusion*, 58(12):125012, 2016.

[286] MA Mendez, M Balabane, and J-M Buchlin. Multi-scale proper orthogonal decomposition of complex fluid flows. *Journal of Fluid Mechanics*, 870:988–1036, 2019.

[287] Ariana Mendible, Steven L Brunton, Aleksandr Y Aravkin, Wes Lowrie, and J Nathan Kutz. Dimensionality reduction and reduced-order modeling for traveling wave physics. *Theoretical and Computational Fluid Dynamics*, 34(4):385–400, 2020.

[288] Claude Mercier. On a representation of toroidal surfaces. applications to magnetohydrodynamic equilibria. *Nuclear Fusion*, 3(2):89, 1963.

[289] Claude Mercier. Equilibrium and stability of a toroidal magnetohydrodynamic system in the neighbourhood of a magnetic axis. *Nuclear Fusion*, 4(3):213, 1964.

[290] Daniel A Messenger and David M Bortz. Learning mean-field equations from particle data using WSINDy. *arXiv preprint arXiv:2110.07756*, 2021.

[291] Daniel A Messenger and David M Bortz. Weak SINDy for partial differential equations. *Journal of Computational Physics*, page 110525, 2021.

[292] Daniel A Messenger and David M Bortz. Weak SINDy: Galerkin-based data-driven model selection. *Multiscale Modeling & Simulation*, 19(3):1474–1497, 2021.

[293] Aaron Meurer, Christopher P Smith, Mateusz Paprocki, Ondřej Čertík, Sergey B Kirpichev, Matthew Rocklin, AMiT Kumar, Sergiu Ivanov, Jason K Moore, Sartaj Singh, et al. SymPy: symbolic computing in Python. *PeerJ Computer Science*, 3:e103, 2017.

[294] Igor Mezić. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, 41(1-3):309–325, 2005.

[295] Igor Mezic. Analysis of fluid flows via spectral properties of the Koopman operator. *Annual Review of Fluid Mechanics*, 45(1):357–378, 2013.

[296] Michele Milano and Petros Koumoutsakos. Neural network modeling for near wall turbulent flow. *Journal of Computational Physics*, 182(1):1–26, 2002.

[297] Sean T Miller and Uri Shumlak. A multi-species 13-moment model for moderately collisional plasmas. *Physics of Plasmas*, 23(8):082303, 2016.

[298] Muhammad Mohebujjaman, Leo G Rebholz, and Traian Iliescu. Physically constrained data-driven correction for reduced-order modeling of fluid flows. *International Journal for Numerical Methods in Fluids*, 89(3):103–122, 2019.

[299] Rambod Mojgani, Ashesh Chattopadhyay, and Pedram Hassanzadeh. Closed-form discovery of structural errors in models of chaotic systems by integrating Bayesian sparse regression and data assimilation, 2021.

[300] K. J. Montes, C. Rea, R. S. Granetz, R. A. Tinguely, N. Eidietis, O. M. Meneghini, D. L. Chen, B. Shen, B. J. Xiao, K. Erickson, and M. D. Boyer. Machine learning for disruption warnings on Alcator C-Mod, DIII-D, and EAST. *Nuclear Fusion*, 59(9):096015, July 2019.

[301] D Moreau, F Crisanti, X Litaudon, D Mazon, P De Vries, R Felton, E Joffrin, L Laborde, M Lennholm, A Murari, et al. Real-time control of the q-profile in JET for steady state advanced tokamak operation. *Nuclear fusion*, 43(9):870, 2003.

[302] KD Morgan, AC Hossack, CJ Hansen, BA Nelson, and DA Sutherland. High-speed feedback control of an oscillating magnetic helicity injector using a graphics processing unit. *Review of Scientific Instruments*, 92(5):053530, 2021.

[303] KD Morgan, TR Jarboe, AC Hossack, RN Chandra, and CJ Everson. Validation of extended magnetohydrodynamic simulations of the HIT-SI3 experiment using the NIMROD code. *Physics of Plasmas*, 24(12):122510, 2017.

[304] Kyle Morgan. *Finite-beta simulations of HIT-SI and HIT-SI3 using the NIMROD code*. PhD thesis, University of Washington, Seattle, 2018.

[305] Kyle Morgan, Thomas Jarboe, and Cihan Akcay. Formation of closed flux surfaces in spheromaks sustained by steady inductive helicity injection. *Nuclear Fusion*, 59(6):066037, 2019.

[306] Megan Morrison and J Nathan Kutz. Nonlinear control of networked dynamical systems. *IEEE Transactions on Network Science and Engineering*, 8(1):174–189, 2020.

[307] Philip J Morrison and John M Greene. Noncanonical Hamiltonian density formulation of hydrodynamics and ideal magnetohydrodynamics. *Physical Review Letters*, 45(10):790, 1980.

[308] S Munaretto, EJ Strait, and NC Logan. Optimizing the differential connection schemes for detecting 3D magnetic perturbations in DIII-D. *Review of Scientific Instruments*, 92(7):073504, 2021.

[309] A. Murari, M. Lungaroni, E. Peluso, P. Gaudio, J. Vega, S. Dormido-Canto, M. Baruzzo, and M. Gelfusa and. Adaptive predictors based on probabilistic SVM for real time disruption mitigation on JET. *Nuclear Fusion*, 58(5):056002, 2018.

[310] Abhinav Narasingam and Joseph Sang-Il Kwon. Development of local dynamic mode decomposition with control: Application to model predictive control of hydraulic fracturing. *Computers & Chemical Engineering*, 106:501–511, 2017.

[311] Abhinav Narasingam and Joseph Sang-Il Kwon. Data-driven identification of interpretable reduced-order models using sparse regression. *Computers & Chemical Engineering*, 119:101–111, 2018.

[312] H Natsume, H Tanaka, S Kajita, and N Ohno. Application of dynamic mode decomposition to rotating structures in detached linear plasmas. *Physics of Plasmas*, 27(4):042301, 2020.

[313] Indranil Nayak, Mrinal Kumar, and Fernando L Teixeira. Detection and prediction of equilibrium states in kinetic plasma simulations via mode tracking using reduced-order dynamic mode decomposition. *Journal of Computational Physics*, 447:110671, 2021.

[314] Indranil Nayak and Fernando L Teixeira. Dynamic mode decomposition for prediction of kinetic plasma behavior. In *2020 International Applied Computational Electromagnetics Society Symposium (ACES)*, pages 1–2. IEEE, 2020.

[315] Rajdip Nayek, Ramon Fuentes, Keith Worden, and Elizabeth J Cross. On spike-and-slab priors for Bayesian equation discovery of nonlinear dynamical systems via sparse linear regression. *Mechanical Systems and Signal Processing*, 161:107986, 2021.

[316] Yu Nesterov. Gradient methods for minimizing composite functions. *Mathematical Programming*, 140(1):125–161, 2013.

[317] Klaus Neumann, Andre Lemme, and Jochen J Steil. Neural learning of stable dynamical systems based on data-driven Lyapunov candidates. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1216–1222. IEEE, 2013.

[318] Julio L Nicolini, Dong-Yeop Na, and Fernando L Teixeira. Model order reduction of electromagnetic particle-in-cell kinetic plasma simulations via proper orthogonal decomposition. *IEEE Transactions on Plasma Science*, 47(12):5239–5250, 2019.

[319] Robert K Niven, Ali Mohammad-Djafari, Laurent Cordier, Markus Abel, and Markus Quade. Bayesian identification of dynamical systems. *Multidisciplinary Digital Publishing Institute Proceedings*, 33(1):33, 2020.

[320] Bernd R Noack, Konstantin Afanasiev, Marek Morzyński, Gilead Tadmor, and Frank Thiele. A hierarchy of low-dimensional models for the transient and post-transient cylinder wake. *Journal of Fluid Mechanics*, 497:335–363, 2003.

[321] Bernd R Noack, Marek Morzynski, and Gilead Tadmor. *Reduced-order modelling for flow control*, volume 528. Springer Science & Business Media, 2011.

[322] Bernd R Noack, Michael Schlegel, Boye Ahlborn, Gerd Mutschke, Marek Morzyński, Pierre Comte, and Gilead Tadmor. A finite-time thermodynamics of unsteady fluid flows. *Journal of Non-Equilibrium Thermodynamics*, 33(2):103–148, 2008.

[323] Bernd R Noack, Witold Stankiewicz, Marek Morzynski, and Peter J Schmid. Recursive dynamic mode decomposition of a transient cylinder wake. *Journal of Fluid Mechanics*, 809:843–872, 2016.

[324] Frank Noé and Feliks Nuske. A variational approach to modeling slow processes in stochastic dynamical systems. *Multiscale Modeling Simulation*, 11(2):635–655, 2013.

[325] Frank Noé, Simon Olsson, Jonas Köhler, and Hao Wu. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science*, 365(6457):eaaw1147, 2019.

[326] Feliks Nüske, Bettina G Keller, Guillermo Pérez-Hernández, Antonia SJS Mey, and Frank Noé. Variational approach to molecular kinetics. *Journal of chemical theory and computation*, 10(4):1739–1752, 2014.

[327] J B O'Bryan and C R Sovinec. Simulated flux-rope evolution during non-inductive startup in Pegasus. *Plasma Physics and Controlled Fusion*, 56(6):064005, 2014.

[328] O. Ohia, J. Egedal, V. S. Lukin, W. Daughton, and A. Le. Demonstration of anisotropic fluid closure capturing the kinetic structure of magnetic reconnection. *Phys. Rev. Lett.*, 109:115004, Sep 2012.

[329] K Erik J Olofsson, Per R Brunsell, and James R Drake. Experimental modal analysis of resistive wall toroidal pinch plasma dynamics. *Nuclear Fusion*, 53(7):072003, 2013.

[330] K Erik J Olofsson, Jeremy M Hanson, Daisuke Shiraki, Francesco A Volpe, David A Humphreys, Robert J La Haye, Matthew J Lanctot, Edward J Strait, Anders S Welander, Egemen Kolemen, et al. Array magnetics modal analysis for the DIII-D tokamak based on localized time-series modelling. *Plasma Physics and Controlled Fusion*, 56(9):095012, 2014.

[331] J. Olson, J. Egedal, S. Greess, R. Myers, M. Clark, D. Endrizzi, K. Flanagan, J. Milhone, E. Peterson, J. Wallace, D. Weisberg, and C. B. Forest. Experimental demonstration of the collisionless plasmoid instability below the ion kinetic scale during magnetic reconnection. *Phys. Rev. Lett.*, 116:255001, Jun 2016.

[332] Y. Ono, H. Tanabe, Y. Hayashi, T. Ii, Y. Narushima, T. Yamada, M. Inomoto, and C. Z. Cheng. Ion and electron heating characteristics of magnetic reconnection in a two flux loop merging experiment. *Phys. Rev. Lett.*, 107:185001, Oct 2011.

[333] Samuel E Otto and Clarence W Rowley. Linearly recurrent autoencoder networks for learning dynamics. *SIAM Journal on Applied Dynamical Systems*, 18(1):558–593, 2019.

[334] Samuel E Otto and Clarence W Rowley. Inadequacy of linear methods for minimal sensor placement and feature selection in nonlinear systems; a new approach using secants. *arXiv preprint arXiv:2101.11162*, 2021.

[335] Michael L Overton. On minimizing the maximum eigenvalue of a symmetric matrix. *SIAM Journal on Matrix Analysis and Applications*, 9(2):256–268, 1988.

[336] RG O'Neill, GJ Marklin, TR Jarboe, C Akcay, WT Hamp, BA Nelson, AJ Redd, RJ Smith, BT Stewart, JS Wrobel, et al. A fully relaxed helicity balance model for an inductively driven spheromak. *Physics of Plasmas*, 14(11):112304, 2007.

[337] Shaowu Pan, Nicholas Arnold-Medabalimi, and Karthik Duraisamy. Sparsity-promoting algorithms for the discovery of informative Koopman-invariant subspaces. *Journal of Fluid Mechanics*, 917, 2021.

[338] Shaowu Pan and Karthik Duraisamy. Data-driven discovery of closure models. *SIAM Journal on Applied Dynamical Systems*, 17(4):2381–2413, 2018.

[339] Shaowu Pan and Karthik Duraisamy. Physics-informed probabilistic learning of linear embeddings of nonlinear dynamics with guaranteed stability. *SIAM Journal on Applied Dynamical Systems*, 19(1):480–509, 2020.

[340] W. Pan, Y. Yuan, J. Gonçalves, and G. Stan. A sparse Bayesian approach to the identification of nonlinear state-space systems. *IEEE Transactions on Automatic Control*, 61(1):182–187, January 2016.

[341] Mihir Pandya. *Low edge safety factor disruptions in the Compact Toroidal Hybrid: Operation in the low-q regime, passive disruption avoidance and the nature of MHD precursors*. PhD thesis, Auburn University, 2016.

[342] Elena Panteley, Antonio Loria, and Ali El-Ati. Practical dynamic consensus of Stuart–Landau oscillators over heterogeneous networks. *International Journal of Control*, 93(2):261–273, 2020.

[343] Eugene N Parker. Sweet's mechanism for merging magnetic fields in conducting fluids. *Journal of Geophysical Research*, 62(4):509–520, 1957.

[344] Jaideep Pathak, Brian Hunt, Michelle Girvan, Zhixin Lu, and Edward Ott. Model-free prediction of large spatiotemporally chaotic systems from data: a reservoir computing approach. *Physical review letters*, 120(2):024102, 2018.

[345] A Pavone, U Hergenhahn, Maciej Krychowiak, U Hoefel, S Kwak, J Svensson, P Kornejew, V Winters, R Koenig, M Hirsch, et al. Measurements of visible bremsstrahlung and automatic Bayesian inference of the effective plasma charge Zeff at W7-X. *Journal of Instrumentation*, 14(10):C10003, 2019.

[346] A Pavone, J Svensson, S Kwak, MRCW Brix, RC Wolf, and JET Contributors. Neural network approximated Bayesian inference of edge electron density profiles at JET. *Plasma Physics and Controlled Fusion*, 62(4):045019, 2020.

[347] Benjamin Peherstorfer and Karen Willcox. Data-driven operator inference for nonintrusive projection-based model reduction. *Computer Methods in Applied Mechanics and Engineering*, 306:196–215, 2016.

[348] Liqian Peng and Kamran Mohseni. Symplectic model reduction of Hamiltonian systems. *SIAM Journal on Scientific Computing*, 38(1):A1–A27, 2016.

[349] James Michael Penna. *Characterization of the Plasma Response to Injector Frequency in the HIT-SI3 Spheromak Experiment*. PhD thesis, University of Washington, 2021.

[350] Allan Pinkus. *N-widths in Approximation Theory*, volume 7. Springer Science & Business Media, 2012.

[351] Franck Plunian, Rodion Stepanov, and Peter Frick. Shell models of magnetohydrodynamic turbulence. *Physics Reports*, 523(1):1–60, 2013.

[352] Annick Pouquet, Duane Rosenberg, Julia E Stawarz, and Raffaele Marino. Helicity dynamics, inverse, and bidirectional cascades in fluid and magnetohydrodynamic turbulence: a brief review. *Earth and Space Science*, 6(3):351–369, 2019.

[353] Joshua L Proctor, Steven L Brunton, and J Nathan Kutz. Dynamic mode decomposition with control. *SIAM Journal on Applied Dynamical Systems*, 15(1):142–161, 2016.

[354] Elizabeth Qian, Boris Kramer, Benjamin Peherstorfer, and Karen Willcox. Lift & Learn: Physics-informed machine learning for large-scale nonlinear dynamical systems. *Physica D: Nonlinear Phenomena*, 406:132401, 2020.

[355] S Joe Qin. An overview of subspace identification. *Computers & chemical engineering*, 30(10-12):1502–1513, 2006.

[356] Christopher Rackauckas, Yingbo Ma, Julius Martensen, Collin Warner, Kirill Zubov, Rohit Supekar, Dominic Skinner, Ali Ramadhan, and Alan Edelman. Universal differential equations for scientific machine learning. *arXiv preprint arXiv:2001.04385*, 2020.

[357] Christopher Rackauckas and Qing Nie. Differentialequations.jl – a performant and feature-rich ecosystem for solving differential equations in Julia. *The Journal of Open Research Software*, 5(1), 2017.

[358] Cedric Raibaudo, Peng Zhong, Bernd R Noack, and Robert John Martinuzzi. Machine learning strategies applied to the control of a fluidic pinball. *Physics of Fluids*, 32(1):015108, 2020.

[359] M Raissi, P Perdikaris, and GE Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

[360] Maziar Raissi and George Em Karniadakis. Hidden physics models: Machine learning of nonlinear partial differential equations. *Journal of Computational Physics*, 357:125–141, 2018.

[361] Maziar Raissi, Alireza Yazdani, and George Em Karniadakis. Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations. *Science*, 367(6481):1026–1030, 2020.

[362] SS Ravindran. Real-time computational algorithm for optimal control of an MHD flow system. *SIAM Journal on Scientific Computing*, 26(4):1369–1388, 2005.

[363] C. Rea and R. S. Granetz. Exploratory machine learning studies for disruption prediction using large databases on DIII-D. *Fusion Science and Technology*, 74(1-2):89–100, 2018.

[364] C. Rea, R. S. Granetz, K. Montes, R. A. Tinguely, N. Eidietis, J. M. Hanson, and B. Sammuli. Disruption prediction investigations using machine learning tools on DIII-D and Alcator C-Mod. *Plasma Physics and Controlled Fusion*, 60(8):084004, June 2018.

[365] Patrick AK Reinbold, Daniel R Gurevich, and Roman O Grigoriev. Using noisy or incomplete data to discover models of spatiotemporal dynamics. *Physical Review E*, 101(1):010203, 2020.

[366] Patrick AK Reinbold, Logan M Kageorge, Michael F Schatz, and Roman O Grigoriev. Robust learning from noisy, incomplete, high-dimensional experimental data via physically constrained symbolic regression. *Nature communications*, 12(1):1–8, 2021.

[367] Dietmar Rempfer and Hermann F Fasel. Dynamics of three-dimensional coherent structures in a flat-plate boundary layer. *Journal of Fluid Mechanics*, 275:257–283, 1994.

[368] Spencer M Richards, Felix Berkenkamp, and Andreas Krause. The Lyapunov neural network: Adaptive stability certification for safe learning of dynamical systems. In *Proceedings of the 2nd Conference on Robot Learning (CoRL 2018)*, volume 87, pages 466–476. PMLR, 2018.

[369] Ch P Ritz, EJ Powers, and Roger D Bengtson. Experimental measurement of three-wave coupling and energy cascading. *Physics of Fluids B: Plasma Physics*, 1(1):153–163, 1989.

[370] T D Rognlien and M E Rensink. Edge-plasma models and characteristics for magnetic fusion energy devices. *Fusion Engineering and Design*, 60(4):497–514, 2002.

[371] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D.S. Henningson. Spectral analysis of nonlinear flows. *Journal of fluid mechanics*, 641(1):115–127, 2009.

[372] Clarence W Rowley. Model reduction for fluids, using balanced proper orthogonal decomposition. *International Journal of Bifurcation and Chaos*, 15(03):997–1013, 2005.

[373] Clarence W Rowley, Tim Colonius, and Richard M Murray. Model reduction for compressible flows using POD and Galerkin projection. *Physica D: Nonlinear Phenomena*, 189(1-2):115–129, 2004.

[374] Clarence W Rowley and Scott TM Dawson. Model reduction for flow analysis and control. *Annual Review of Fluid Mechanics*, 49:387–417, 2017.

[375] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

[376] Samuel H Rudy, Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Data-driven discovery of partial differential equations. *Science Advances*, 3(4):e1602614, 2017.

[377] Samuel H Rudy, J Nathan Kutz, and Steven L Brunton. Deep learning of dynamics and signal-noise decomposition with time-stepping constraints. *Journal of Computational Physics*, 396:483–506, 2019.

[378] Bernd Rummler and Andreas Noske. Direct Galerkin approximation of plane-parallel-Couette and channel flows by Stokes eigenfunctions. *Notes on numerical fluid mechanics*, 64:3–19, 1998.

[379] Tanmoy Samanta, Hui Tian, and Valery M. Nakariakov. Evidence for vortex shedding in the sun's hot corona. *Phys. Rev. Lett.*, 123:035102, Jul 2019.

[380] Omer San and Romit Maulik. Neural network closures for nonlinear model order reduction. *Advances in Computational Mathematics*, 44(6):1717–1750, 2018.

[381] Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter Battaglia. Learning to simulate complex physics with graph networks. In *International Conference on Machine Learning*, pages 8459–8468. PMLR, 2020.

[382] M Sasaki, Y Kawachi, R O Dendy, H Arakawa, N Kasuya, F Kin, K Yamasaki, and S Inagaki. Using dynamical mode decomposition to extract the limit cycle dynamics of modulated turbulence in a plasma simulation. *Plasma Physics and Controlled Fusion*, 61(11):112001, Oct. 2019.

[383] Nihar Sawant, Boris Kramer, and Benjamin Peherstorfer. Physics-informed regularization and structure preservation for learning stable reduced models from data with operator inference. *arXiv preprint arXiv:2107.02597*, 2021.

[384] Hayden Schaeffer. Learning partial differential equations via data discovery and sparse optimization. In *Proc. R. Soc. A*, volume 473, page 20160446. The Royal Society, 2017.

[385] Hayden Schaeffer and Scott G McCalla. Sparse model selection via integral terms. *Physical Review E*, 96(2):023302, 2017.

[386] Hayden Schaeffer, Giang Tran, and Rachel Ward. Extracting sparse high-dimensional dynamics from limited data. *SIAM Journal on Applied Mathematics*, 78(6):3279–3295, 2018.

[387] Lukas Scheffold, Tiago Finkler, and Uwe Piechottka. Gray-box system modeling using symbolic regression and nonlinear model predictive control of a semibatch polymerization. *Computers & Chemical Engineering*, 146:107204, 2021.

[388] Alexander Schein, Kevin T Carlberg, and Matthew J Zahr. Preserving general physical properties in model reduction of dynamical systems via constrained-optimization projection. *International Journal for Numerical Methods in Engineering*, accepted 2021.

[389] Alexander A Schekochihin. MHD turbulence: A biased review. *arXiv preprint arXiv:2010.00699*, 2020.

[390] Michael Schlegel and Bernd R Noack. On long-term boundedness of Galerkin models. *Journal of Fluid Mechanics*, 765:325–352, 2015.

[391] Hermann Schlichting and Klaus Gersten. *Boundary-layer theory*. Springer, 2016.

[392] Martin Schmelzer, Richard P Dwight, and Paola Cinnella. Discovery of algebraic Reynolds-stress models using sparse symbolic regression. *Flow, Turbulence and Combustion*, 104(2):579–603, 2020.

[393] Peter J. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, 2010.

[394] Michael Schmidt and Hod Lipson. Distilling free-form natural laws from experimental data. *Science*, 324(5923):81–85, 2009.

[395] VD Shafranov. Equilibrium of a plasma toroid in a magnetic field. *Sov. Phys. JETP*, 37(10):775, 1960.

[396] RJ Shalloo, SJD Dann, J-N Gruse, CID Underwood, AF Antoine, Christopher Arran, Michael Backhouse, CD Baird, MD Balcazar, Nicholas Bourgeois, et al. Automation and control of laser wakefield accelerators using Bayesian optimization. *Nature communications*, 11(1):1–8, 2020.

[397] SE Sharapov, B Alper, HL Berk, DN Borba, BN Breizman, CD Challis, A Fasoli, NC Hawkes, TC Hender, J Mailloux, et al. Alfvén wave cascades in a tokamak. *Physics of Plasmas*, 9(5):2027–2036, 2002.

[398] Daniel E Shea, Steven L Brunton, and J Nathan Kutz. SINDy-BVP: Sparse identification of nonlinear dynamics for boundary value problems. *Physical Review Research*, 3(2):023255, 2021.

[399] Jonathan Richard Shewchuk et al. An introduction to the conjugate gradient method without the agonizing pain, 1994.

[400] Sungho Shin, Qiugang Lu, and Victor M Zavala. Unifying theorems for subspace identification and dynamic mode decomposition. *arXiv preprint arXiv:2003.07410*, 2020.

[401] Uri Shumlak, Robert Lilly, Noah Reddell, E Sousa, and Bhuvana Srinivasan. Advanced physics calculations using a multi-fluid plasma model. *Computer Physics Communications*, 182(9):1767–1770, 2011.

[402] D. Sipp and A. Lebedev. Global stability of base and mean flows: a general approach and its applications to cylinder and open cavity flows. *Journal of Fluid Mechanics*, 593:333–358, 2007.

[403] Vít Škvára, Václav Šmídl, Tomáš Pevnỳ, Jakub Seidl, Aleš Havránek, and David Tskhakaya. Detection of Alfvén eigenmodes on COMPASS with generative neural networks. *Fusion Science and Technology*, 76(8):962–971, 2020.

[404] Mariia Sorokina, Stylianos Sygletos, and Sergei Turitsyn. Sparse identification for nonlinear optical communication systems: SINO method. *Optics express*, 24(26):30433–30443, 2016.

[405] CR Sovinec, AH Glasser, TA Gianakon, DC Barnes, RA Nebel, SE Kruger, DD Schnack, SJ Plimpton, A Tarditi, Ming-Sheng Chu, et al. Nonlinear magnetohydrodynamics simulation using high-order finite elements. *Journal of Computational Physics*, 195(1):355–386, 2004.

[406] Lyman Spitzer. *Physics of fully ionized gases*. Courier Corporation, 2006.

[407] Vishal Srivastava and Karthik Duraisamy. Generalizable physics-constrained modeling using learning and inference assisted by feature space engineering. *arXiv preprint arXiv:2103.16042*, 2021.

[408] Alex M Stanković, Aleksandar A Sarić, Andrija T Sarić, and Mark K Transtrum. Data-driven symbolic regression for identification of nonlinear dynamics in power systems. In *2020 IEEE Power & Energy Society General Meeting (PESGM)*, pages 1–5. IEEE, 2020.

[409] EJ Strait, JD King, JM Hanson, and NC Logan. Spatial and temporal analysis of DIII-D 3D magnetic diagnostic data. *Review of Scientific Instruments*, 87(11):11D423, 2016.

[410] Fangzheng Sun, Yang Liu, and Hao Sun. Physics-informed spline learning for nonlinear dynamics discovery. *arXiv preprint arXiv:2105.02368*, 2021.

[411] D.A. Sutherland. *Measurements of neutral particles and simulations of plasma-neutral dynamics in the HIT-SI3 experiment*. PhD thesis, University of Washington, Seattle, 2019.

[412] Donald Gary Swanson. *Plasma waves*. CRC Press, 2003.

[413] PA Sweet. Electromagnetic phenomena in cosmical physics (IAU Symp. 6), ed, 1958.

[414] Kunihiko Taira, Steven L Brunton, Scott Dawson, Clarence W Rowley, Tim Colonius, Beverley J McKeon, Oliver T Schmidt, Stanislav Gordeyev, Vassilios Theofilis, and Lawrence S Ukeiley. Modal analysis of fluid flows: An overview. *AIAA Journal*, 55(12):4013–4041, 2017.

[415] Kunihiko Taira, Maziar S Hemati, Steven L Brunton, Yiyang Sun, Karthik Duraisamy, Shervin Bagheri, Scott TM Dawson, and Chi-An Yeh. Modal analysis of fluid flows: Applications and outlook. *AIAA journal*, 58(3):998–1022, 2020.

[416] Naoya Takeishi and Yoshinobu Kawahara. Learning dynamics models with stable invariant sets. *arXiv preprint arXiv:2006.08935*, 2020.

[417] Naoya Takeishi, Yoshinobu Kawahara, and Takehisa Yairi. Learning Koopman invariant subspaces for dynamic mode decomposition. In *Advances in Neural Information Processing Systems*, pages 1130–1140, 2017.

[418] SX Tang, TA Carter, NA Crocker, WW Heidbrink, JB Lestz, RI Pinsker, KE Thome, MA Van Zeeland, and EV Belova. Stabilization of Alfvén eigenmodes in DIII-D via controlled energetic ion density ramp and validation of theory and simulations. *Physical Review Letters*, 126(15):155001, 2021.

[419] Roy Taylor, J Nathan Kutz, Kyle Morgan, and Brian A Nelson. Dynamic mode decomposition for plasma diagnostics and validation. *Review of Scientific Instruments*, 89(5):053501, 2018.

[420] H. Tennekes and J. L. Lumley. *A First Course in Turbulence*. MIT Press, 1972.

[421] Filippo Terragni, Eusebio Valero, and José M Vega. Local POD plus Galerkin projection in the unsteady lid-driven cavity problem. *SIAM Journal on Scientific Computing*, 33(6):3538–3561, 2011.

[422] Gregor Thiele, Arne Fey, David Sommer, and Jörg Krüger. System identification of a hysteresis-controlled pump system using SINDy. In *2020 24th International Conference on System Theory, Control and Computing (ICSTCC)*, pages 457–464. IEEE, 2020.

[423] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.

[424] Y Todo. Introduction to the interaction between energetic particles and Alfvén eigenmodes in toroidal plasmas. *Reviews of Modern Plasma Physics*, 3(1):1–33, 2019.

[425] Aaron Towne, Oliver T Schmidt, and Tim Colonius. Spectral proper orthogonal decomposition and its relationship to dynamic mode decomposition and resolvent analysis. *Journal of Fluid Mechanics*, 847:821–867, 2018.

[426] Giang Tran and Rachel Ward. Exact recovery of chaotic systems from highly corrupted data. *Multiscale Modeling & Simulation*, 15(3):1108–1129, 2017.

[427] J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz. On dynamic mode decomposition: theory and applications. *Journal of Computational Dynamics*, 1(2):391–421, 2014.

[428] Jonathan H Tu, Clarence W Rowley, J Nathan Kutz, and Jessica K Shang. Spectral analysis of fluid flows using sub-Nyquist-rate PIV data. *Experiments in Fluids*, 55(9):1–13, 2014.

[429] AD Turnbull, EJ Strait, WW Heidbrink, MS Chu, HH Duong, JM Greene, LL Lao, TS Taylor, and SJ Thompson. Global Alfvén modes: Theory and experiment. *Physics of Fluids B: Plasma Physics*, 5(7):2546–2553, 1993.

[430] JM Tuwankotta. Chaos in a coupled oscillators system with widely spaced frequencies and energy-preserving non-linearity. *International Journal of Non-Linear Mechanics*, 41(2):180–191, 2006.

[431] Karel Lucas van de Plassche, Jonathan Citrin, Clarisse Bourdelle, Yann Camenen, Francis J Casson, Victor I Dagnelie, Federico Felici, Aaron Ho, Simon Van Mulders, and JET Contributors. Fast modeling of turbulent transport in fusion plasmas using neural networks. *Physics of Plasmas*, 27(2):022310, 2020.

[432] B Ph Van Milligen, E Sánchez, A Alonso, MA Pedrosa, C Hidalgo, A Martín De Aguilera, and A López Fraguas. The use of the biorthogonal decomposition for the identification of zonal flows at TJ-II. *Plasma Physics and Controlled Fusion*, 57(2):025005, 2014.

[433] Peter Van Overschee and Bart De Moor. A unifying theorem for three subspace system identification algorithms. *Automatica*, 31(12):1853–1864, 1995.

[434] BS Victor, C Akcay, CJ Hansen, TR Jarboe, BA Nelson, and KD Morgan. Development of validation metrics using biorthogonal decomposition for the comparison of magnetic field measurements. *Plasma Physics and Controlled Fusion*, 57(4):045010, 2015.

[435] BS Victor, TR Jarboe, CJ Hansen, C Akcay, KD Morgan, AC Hossack, and BA Nelson. Sustained spheromaks with ideal n=1 kink stability and pressure confinement. *Physics of Plasmas*, 21(8):082504, 2014.

[436] Libo Wang, XQ Xu, Ben Zhu, Chenhao Ma, and Yi-an Lei. Deep learning surrogate model for kinetic Landau-fluid closure with collision. *AIP Advances*, 10(7):075108, 2020.

[437] ZR Wang, NC Logan, S Munaretto, YQ Liu, YW Sun, S Gu, J-K Park, JM Hanson, QM Hu, T Strait, et al. Identification of multiple eigenmode growth rates in DIII-D and EAST tokamak plasmas. *Nuclear Fusion*, 59(2):024001, 2019.

[438] Christoph Wehmeyer and Frank Noé. Time-lagged autoencoders: Deep learning of slow collective variables for molecular kinetics. *The Journal of Chemical Physics*, 148(241703):1–9, 2018.

[439] John Wesson and David J Campbell. *Tokamaks*, volume 149. Oxford university press, 2011.

[440] Karen Willcox. Unsteady flow sensing and estimation via the gappy proper orthogonal decomposition. *Computers & Fluids*, 35(2):208–226, 2006.

[441] Karen Willcox and Jaime Peraire. Balanced model reduction via the proper orthogonal decomposition. *AIAA journal*, 40(11):2323–2330, 2002.

[442] Matthew O Williams, Ioannis G Kevrekidis, and Clarence W Rowley. A data–driven approximation of the Koopman operator: Extending dynamic mode decomposition. *Journal of Nonlinear Science*, 25(6):1307–1346, 2015.

[443] KL Wong, RJ Fonck, SF Paul, DR Roberts, ED Fredrickson, R Nazikian, HK Park, M Bell, NL Bretz, R Budny, et al. Excitation of toroidal Alfvén eigenmodes in TFTR. *Physical review letters*, 66(14):1874, 1991.

[444] Benjamin JQ Woods, Vinícius N Duarte, Eric D Fredrickson, Nikolai N Gorelenkov, Mario Podestà, and Roddy GL Vann. Machine learning characterization of Alfvénic and sub-Alfvénic chirping and correlation with fast-ion loss at NSTX. *IEEE Transactions on Plasma Science*, 48(1):71–81, 2020.

[445] Franco Woolfe, Edo Liberty, Vladimir Rokhlin, and Mark Tygert. A fast randomized algorithm for the approximation of matrices. *Applied and Computational Harmonic Analysis*, 25(3):335–366, 2008.

[446] Stephen J Wright. Coordinate descent algorithms. *Mathematical Programming*, 151(1):3–34, 2015.

[447] J. S. Wrobel, C. J. Hansen, T. R. Jarboe, R. J. Smith, A. C. Hossack, B. A. Nelson, G. J. Marklin, D. A. Ennis, C. Akcay, and B. S. Victor. Relaxation-time measurement via a time-dependent helicity balance model. *Physics of Plasmas*, 20(1):012503, 2013.

[448] Jonathan Scott Wrobel. *A study of HIT-SI plasma dynamics using surface magnetic field measurements*. University of Washington, 2011.

[449] Kailiang Wu and Dongbin Xiu. Numerical aspects for approximating governing equations using data. *Journal of Computational Physics*, 384:200–221, 2019.

[450] Or Yair, Ronen Talmon, Ronald R Coifman, and Ioannis G Kevrekidis. Reconstruction of normal forms by learning informed observation geometries from data. *Proceedings of the National Academy of Sciences*, 114(38):E7865–E7874, 2017.

[451] Liu Yang, Dongkun Zhang, and George Em Karniadakis. Physics-informed generative adversarial networks for stochastic differential equations. *SIAM Journal on Scientific Computing*, 42(1):A292–A317, 2020.

[452] Enoch Yeung, Soumya Kundu, and Nathan Hodas. Learning deep neural network representations for Koopman operators of nonlinear dynamical systems. In *2019 American Control Conference (ACC)*, pages 4832–4839. IEEE, 2019.

[453] Z Yoshida and E Hameiri. Canonical Hamiltonian mechanics of Hall magnetohydrodynamics and its limit to ideal magnetohydrodynamics. *Journal of Physics A: Mathematical and Theoretical*, 46(33):335502, 2013.

[454] Jize Zhang, Andrew M Pace, Samuel A Burden, and Aleksandr Aravkin. Offline state estimation for hybrid systems via nonsmooth variable projection. *Automatica*, 115:108871, 2020.

[455] Linan Zhang and Hayden Schaeffer. On the convergence of the SINDy algorithm. *Multiscale Modeling & Simulation*, 17(3):948–972, 2019.

[456] Sheng Zhang and Guang Lin. Robust data-driven discovery of governing physical laws with error bars. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 474(2217):20180305, 2018.

[457] Peng Zheng and Aleksandr Aravkin. Relax-and-split method for nonconvex inverse problems. *Inverse Problems*, 36(9):095013, 2020.

[458] Peng Zheng, Travis Askham, Steven L Brunton, J Nathan Kutz, and Aleksandr Y Aravkin. A unified framework for sparse relaxed regularized regression: SR3. *IEEE Access*, 7:1404–1423, 2019.

[459] Victor Zucatti, Hugo FS Lui, Diogo B Pitz, and William R Wolf. Assessment of reduced-order modeling strategies for convective heat transfer. *Numerical Heat Transfer, Part A: Applications*, 77(7):702–729, 2020.

Appendix A

DERIVATION OF SINDY CONSTRAINTS

In Sec. 3.4, constraints were derived for the POD-Galerkin model coefficients from global conservation laws; the goal here is to rewrite these constraints to be compatible with the formulation of the SINDy system identification method. The conclusions for the global conservation of energy were: 1) no constant terms, 2) an anti-symmetry constraint on the linear part of the coefficient matrix $\boldsymbol{\xi}$ in Eq. (3.20), and 3) a more complicated energy-preserving structure in the quadratic coefficients in Eq. (3.21). Consider a quadratic library in a set of r modes, ordered as $\boldsymbol{\Theta}(\boldsymbol{a}) = [a_1, ..., a_r, a_1a_2, ..., a_{r-1}a_r, a_1^2, ..., a_r^2] \in \mathbb{R}^{p_\Theta}$. Note that this arrangement of the polynomials in $\boldsymbol{\Theta}$ differs from Loiseau et al. [257], so the indexing and subscripts are also different here. First, the constraint on the linear part of the Galerkin model in Eq. (3.12), $\boldsymbol{a}^T\boldsymbol{L}\boldsymbol{a} \approx 0$, will be considered. It can be rewritten in the SINDy notation as

$$0 = \begin{bmatrix} a_1 & \cdots & a_r \end{bmatrix} \begin{bmatrix} \xi_1^{a_1} & \cdots & \xi_r^{a_1} \\ \vdots & \ddots & \vdots \\ \xi_r^{a_r} & \cdots & \xi_r^{a_r} \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_r \end{bmatrix}. \tag{A.1}$$

Therefore, it can be concluded that $\xi_i^{a_j} = -\xi_j^{a_i}$ for $i, j \in \{1, ..., r\}$ and $\xi_i^{a_j}$ can be identified by accessing the $(i-1)r + j$ index in the vector of model coefficients $\boldsymbol{\xi}$. Note that only the first r^2 elements of $\boldsymbol{\xi}$ are accessed. For models of linear and quadratic polynomials, $p_\Theta = (r^2 + 3r)/2$ and the number of constraints from anti-symmetry of the linear coefficients is $N_L = (r^2 + r)/2$. Thus there are now only $rp_\Theta - N_L = r(r^2 + 2r - 1)/2$ free parameters. Since the constrained SINDy algorithm solves linear equality constraints of the form $\boldsymbol{C}_s\boldsymbol{\xi} = \boldsymbol{d}$,

one can write this out explicitly for $r = 3$,

$$\begin{bmatrix} \mathbf{1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & \mathbf{1} & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{1} & 0 & \cdots \\ 0 & \mathbf{1} & 0 & \mathbf{1} & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & \mathbf{1} & 0 & 0 & 0 & \mathbf{1} & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & \mathbf{1} & 0 & \mathbf{1} & 0 & 0 & \cdots \end{bmatrix} \begin{bmatrix} \xi_1^{a_1} \\ \xi_1^{a_2} \\ \xi_1^{a_3} \\ \xi_2^{a_1} \\ \xi_2^{a_2} \\ \xi_2^{a_3} \\ \xi_3^{a_1} \\ \xi_3^{a_2} \\ \xi_3^{a_3} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{A.2}$$

The boundary conditions $\boldsymbol{u} \cdot \hat{\boldsymbol{n}} = 0$, $\boldsymbol{J} \cdot \hat{\boldsymbol{n}} = 0$, $\boldsymbol{B} \cdot \hat{\boldsymbol{n}} = 0$ guaranteed that the quadratic nonlinearities were energy-preserving, and thus that cubic terms in Eq. (3.14) vanish,

$$\sum_{i,j,k=0}^{r} Q_{ijk} a_i a_j a_k \approx 0. \tag{A.3}$$

This constraint is significantly more involved to reformat. Written in SINDy notation, this is equivalent to

$$0 = \begin{bmatrix} a_1 & \cdots & a_r \end{bmatrix} \begin{bmatrix} \xi_{r+1}^{a_1} & \xi_{r+2}^{a_1} & \cdots & \xi_{p_\Theta}^{a_1} \\ \vdots & \vdots & \vdots & \vdots \\ \xi_{r+1}^{a_r} & \xi_{r+2}^{a_r} & \cdots & \xi_{p_\Theta}^{a_r} \end{bmatrix} \begin{bmatrix} a_1 a_2 \\ \vdots \\ a_{r-1} a_r \\ a_1^2 \\ \vdots \\ a_r^2 \end{bmatrix}. \tag{A.4}$$

Expand this all out and group the like terms, i.e. terms which look like a_i^3, $a_i a_j^2$ or $a_i a_j a_k$, $i, j, k \in \{1, ..., r\}$, $i \neq j \neq k$. All of the like terms can be straightforwardly shown to be

linearly independent, so one can consider three constraints separately for the three types of terms. The number of each of these respective terms is $\binom{r}{1} = r$, $2\binom{r}{2} = r(r-1)$, and $\binom{r}{3} = r(r-1)(r-2)/6$, for a total of $r(r+1)(r+2)/6 = N_Q$ constraints. With both constraints, there are $rp_\Theta - N_L - N_Q = r(r-1)(2r+5)/6$ free parameters, and $N_c = N_L + N_Q$ constraints. Further considering the quadratic case, one finds that coefficients which adorn a_i^3 must vanish, $\xi^{a_i}_{p_\Theta - r + i} = 0$. Now define

$$\tilde{\xi}_{ijk} = \xi^{a_i}_{r + \frac{j}{2}(2r-j-3)+k-1}. \tag{A.5}$$

The second type of constraint, with $i \neq j$, produces

$$\xi^{a_i}_{p_\Theta - r + j} = \begin{cases} \tilde{\xi}_{jij} & i < j \\ \tilde{\xi}_{jji} & i > j, \end{cases} \tag{A.6}$$

while the third type of constraint produces

$$\tilde{\xi}_{ijk} + \tilde{\xi}_{jik} + \tilde{\xi}_{kij} = 0. \tag{A.7}$$

This relation is equivalent to the energy-preserving condition in Eq. (3.21) and is an arbitrary r generalization to the $r = 3$ constraint used in Loiseau et al. [257]. For the specific case where the plasma system is Hamiltonian (for instance in ideal [307], Hall [453], and extended [2] MHD without dissipation) and the measurements are assumed to be sufficient to represent the Hamiltonian, one could alternatively use formulations of SINDy to directly discover the Hamiltonian [90] and subsequently derive the equations of motion. Lastly, if the global energy conservation constraint on the quadratic terms in the SINDy coefficient matrix $\boldsymbol{\Xi}$ is written $\boldsymbol{C}_s\boldsymbol{\xi} = \boldsymbol{0}$, then the quadratic cross-helicity constraints in Eq. (3.25) can be written $\boldsymbol{A}^{H_c}\boldsymbol{C}_s\boldsymbol{\xi} = A^{H_c}_{jk} C_{kl} \xi_l = 0_j$.

Appendix B

CONVERTING THE HALL-MHD EQUATIONS INTO MAGNETIC FIELD UNITS

Compressible Hall-MHD in the limit of constant and uniform temperature is

$$\dot{\rho} = -\nabla \cdot (\rho \boldsymbol{u}), \tag{B.1}$$

$$\dot{\boldsymbol{u}} = -\boldsymbol{u} \cdot \nabla \boldsymbol{u} + \frac{1}{\rho}\left[\frac{1}{\mu_0}(\nabla \times \boldsymbol{B}) \times \boldsymbol{B} - \frac{2T}{m_i}\nabla\rho + \tilde{\nu}\nabla^2\boldsymbol{u} + \frac{\tilde{\nu}}{3}\nabla(\nabla \cdot \boldsymbol{u})\right], \tag{B.2}$$

$$\dot{\boldsymbol{B}} = -\nabla \times \left[-\boldsymbol{u} \times \boldsymbol{B} + \frac{\eta}{\mu_0}\nabla \times \boldsymbol{B} + \frac{m_i}{e\rho\mu_0}(\nabla \times \boldsymbol{B}) \times \boldsymbol{B} - \frac{2T}{e}\frac{\nabla\rho}{\rho}\right]. \tag{B.3}$$

The goal is to convert these equations into another system of equations that explicitly evolves the configuration vector $\boldsymbol{q} = [\boldsymbol{B}_u, \boldsymbol{B}, B_T]$ originally defined in Eq. (3.7) to facilitate a single POD-Galerkin model for all the MHD fields. Since T is constant and uniform here, B_T is omitted from the configuration vector. So $\boldsymbol{q} = [\boldsymbol{B}_u, \boldsymbol{B}]$ and the continuity equation is

$$\dot{\rho} = -\nabla \cdot \left(\sqrt{\frac{\rho}{\mu_0}}\boldsymbol{B}_u\right) \tag{B.4}$$

Now the new magnetic field evolution is a simple change since $\boldsymbol{B}_u$ is substituted at a single location and $\nabla \times (\nabla\rho/\rho) = 0$,

$$\dot{\boldsymbol{B}} = \nabla \times \left[\frac{1}{\sqrt{\rho\mu_0}}\boldsymbol{B}_u \times \boldsymbol{B} - \frac{m_i}{e\rho\mu_0}(\nabla \times \boldsymbol{B}) \times \boldsymbol{B}\right] + \frac{\eta}{\mu_0}\nabla^2\boldsymbol{B}. \tag{B.5}$$

However the velocity equation becomes very complicated

$$\begin{aligned}\dot{\boldsymbol{B}}_u &= \sqrt{\mu_0}\partial_t(\sqrt{\rho}\boldsymbol{u}) = \sqrt{\frac{\mu_0}{4\rho}}(\dot{\rho}\boldsymbol{u}+2\rho\dot{\boldsymbol{u}}) \\ &= \sqrt{\frac{\mu_0}{4\rho}}\left[-\boldsymbol{u}\nabla\cdot(\rho\boldsymbol{u})+2\rho(-\boldsymbol{u}\cdot\nabla\boldsymbol{u}+\frac{1}{\rho}[\frac{1}{\mu_0}(\nabla\times\boldsymbol{B})\times\boldsymbol{B}-\frac{2T}{m_i}\nabla\rho+\tilde{\nu}\nabla^2\boldsymbol{u}+\frac{\tilde{\nu}}{3}\nabla(\nabla\cdot\boldsymbol{u})])\right] \\ &= \sqrt{\frac{\mu_0}{4\rho}}\left[-\boldsymbol{u}(\nabla\rho\cdot\boldsymbol{u}+\rho\nabla\cdot\boldsymbol{u})-2\rho\boldsymbol{u}\cdot\nabla\boldsymbol{u}+\frac{2}{\mu_0}(\nabla\times\boldsymbol{B})\times\boldsymbol{B}-\frac{4T}{m_i}\nabla\rho+2\tilde{\nu}\nabla^2\boldsymbol{u}+\frac{2\tilde{\nu}}{3}\nabla(\nabla\cdot\boldsymbol{u})\right].\end{aligned} \tag{B.6}$$

Let's evaluate it term by term:

$$\begin{aligned}&-\boldsymbol{u}(\nabla\rho\cdot\boldsymbol{u}+\rho\nabla\cdot\boldsymbol{u})-2\rho\boldsymbol{u}\cdot\nabla\boldsymbol{u} \\ &= -\frac{1}{\mu_0\sqrt{\rho}}\boldsymbol{B}_u(\frac{1}{\sqrt{\rho}}\nabla\rho\cdot\boldsymbol{B}_u+\rho\nabla\cdot\frac{\boldsymbol{B}_u}{\sqrt{\rho}})-2\frac{\sqrt{\rho}}{\mu_0}\boldsymbol{B}_u\cdot\nabla\frac{\boldsymbol{B}_u}{\sqrt{\rho}} \\ &= -\frac{1}{\mu_0\sqrt{\rho}}\boldsymbol{B}_u(\frac{1}{\sqrt{\rho}}\nabla\rho\cdot\boldsymbol{B}_u+\sqrt{\rho}\nabla\cdot\boldsymbol{B}_u-\frac{\nabla\rho}{2\sqrt{\rho}}\cdot\boldsymbol{B}_u)-2\frac{\sqrt{\rho}}{\mu_0}\boldsymbol{B}_u\cdot(-\frac{\nabla\rho}{2\rho^{\frac{3}{2}}}\boldsymbol{B}_u+\frac{\nabla\boldsymbol{B}_u}{\sqrt{\rho}}) \\ &= -\frac{1}{\mu_0\rho}\boldsymbol{B}_u(\nabla\rho\cdot\boldsymbol{B}_u)-\frac{1}{\mu_0}\boldsymbol{B}_u\nabla\cdot\boldsymbol{B}_u+\frac{3}{2\mu_0\rho}\boldsymbol{B}_u(\nabla\rho\cdot\boldsymbol{B}_u)-\frac{2}{\mu_0}\boldsymbol{B}_u\cdot\nabla\boldsymbol{B}_u \\ &= -\frac{1}{\mu_0}\left[\boldsymbol{B}_u\nabla\cdot\boldsymbol{B}_u+2\boldsymbol{B}_u\cdot\nabla\boldsymbol{B}_u-\frac{1}{2\rho}\boldsymbol{B}_u(\nabla\rho\cdot\boldsymbol{B}_u)\right]\end{aligned} \tag{B.7}$$

To address the first viscous term, the following simplification is required,

$$\nabla^2\frac{1}{\sqrt{\rho}} = \nabla\cdot(-\frac{1}{2\sqrt{\rho}\rho}\nabla\rho) = -\frac{1}{2\sqrt{\rho}\rho}\nabla^2\rho+\frac{3}{4\sqrt{\rho}\rho^2}\nabla\rho\cdot\nabla\rho. \tag{B.8}$$

Now $\nabla^2\boldsymbol{u}$ can be evaluated in full,

$$\nabla^2\boldsymbol{u} = \frac{1}{\sqrt{\mu_0}}\nabla^2\frac{\boldsymbol{B}_u}{\sqrt{\rho}} = -\frac{\nabla^2\rho}{2\sqrt{\mu_0\rho}\rho}\boldsymbol{B}_u+\frac{3\boldsymbol{B}_u}{4\sqrt{\mu_0\rho}\rho^2}\nabla\rho\cdot\nabla\rho-\frac{1}{\sqrt{\mu_0\rho}\rho}(\nabla\rho\cdot\nabla)\boldsymbol{B}_u+\frac{1}{\sqrt{\mu_0\rho}}\nabla^2\boldsymbol{B}_u \tag{B.9}$$

$$= \frac{1}{\sqrt{\mu_0\rho}}\left[\nabla^2\boldsymbol{B}_u-\frac{\nabla^2\rho}{2\rho}\boldsymbol{B}_u+\frac{3}{4\rho^2}\nabla\rho\cdot\nabla\rho+\frac{1}{\rho}(\nabla\rho\cdot\nabla)\boldsymbol{B}_u\right].$$

The remaining viscous term gives

$$\begin{aligned}\nabla(\nabla\cdot\boldsymbol{u}) &= \frac{1}{\sqrt{\mu_0}}\nabla(-\frac{\nabla\rho}{2\rho^{\frac{3}{2}}}\cdot\boldsymbol{B}_u+\frac{1}{\sqrt{\rho}}\nabla\cdot\boldsymbol{B}_u) \\ &= -\frac{1}{2\sqrt{\mu_0\rho}\rho}\nabla(\nabla\rho\cdot\boldsymbol{B}_u)+\frac{3}{4\sqrt{\mu_0\rho}\rho^2}(\nabla\rho\cdot\boldsymbol{B}_u)\nabla\rho+\frac{1}{\sqrt{\mu_0\rho}}\nabla(\nabla\cdot\boldsymbol{B}_u)-\frac{1}{2\sqrt{\mu_0\rho}\rho}(\nabla\cdot\boldsymbol{B}_u)\nabla\rho \\ &= \frac{1}{2\sqrt{\mu_0\rho}}\left[-\frac{1}{\rho}\nabla(\nabla\rho\cdot\boldsymbol{B}_u)+\frac{3}{2\rho^2}(\nabla\rho\cdot\boldsymbol{B}_u)\nabla\rho+2\nabla(\nabla\cdot\boldsymbol{B}_u)-\frac{1}{\rho}(\nabla\cdot\boldsymbol{B}_u)\nabla\rho\right].\end{aligned} \tag{B.10}$$

Putting the velocity field evolution altogether produces

$$\begin{aligned}\dot{\boldsymbol{B}}_u =& -\frac{1}{\sqrt{\rho\mu_0}}(\frac{1}{2}\boldsymbol{B}_u\nabla\cdot\boldsymbol{B}_u+\boldsymbol{B}_u\cdot\nabla\boldsymbol{B}_u-\frac{1}{4\rho}\boldsymbol{B}_u(\nabla\rho\cdot\boldsymbol{B}_u)-(\nabla\times\boldsymbol{B})\times\boldsymbol{B}+\frac{2T\mu_0}{m_i}\nabla\rho) \\ &+\nu\left[\nabla^2\boldsymbol{B}_u-\frac{\nabla^2\rho}{2\rho}\boldsymbol{B}_u+\frac{3\boldsymbol{B}_u}{4\rho^2}\nabla\rho\cdot\nabla\rho+\frac{1}{\rho}(\nabla\rho\cdot\nabla)\boldsymbol{B}_u)-\frac{1}{6\rho}\nabla(\nabla\rho\cdot\boldsymbol{B}_u)\right] \\ &+\nu\left[\frac{1}{4\rho^2}(\nabla\rho\cdot\boldsymbol{B}_u)\nabla\rho+\frac{1}{3}\nabla(\nabla\cdot\boldsymbol{B}_u)-\frac{1}{6\rho}(\nabla\cdot\boldsymbol{B}_u)\nabla\rho\right].\end{aligned} \tag{B.11}$$

In order for this to be quadratic in the temporal dependence, the density must be approximately time-independent. If that is the case then it can be seen by inspection that all of the nonlinearities are quadratic in the magnetic and velocity fields. Then Equations (B.5) and (B.11) can be summarized as in Eq. (3.11). Substituting in the expansion of $\boldsymbol{q}$ in the POD basis, multiplying both sides by the spatial POD modes $\boldsymbol{\chi}_j$, and integrating over space leads straightforwardly to the Galerkin model in Eq. (3.12).

The time-independent density limit produces

$$\frac{1}{2\rho}(\nabla\rho\cdot\boldsymbol{B}_u) = -\nabla\cdot\boldsymbol{B}_u. \tag{B.12}$$

The velocity field evolution can now be summarized:

$$\dot{\boldsymbol{B}}_u = -\frac{1}{\sqrt{\rho\mu_0}}(\frac{1}{2}\boldsymbol{B}_u\nabla\cdot\boldsymbol{B}_u+\boldsymbol{B}_u\cdot\nabla\boldsymbol{B}_u-\frac{1}{4\rho}\boldsymbol{B}_u(\nabla\rho\cdot\boldsymbol{B}_u)-(\nabla\times\boldsymbol{B})\times\boldsymbol{B}+\frac{2T\mu_0}{m_i}\nabla\rho) \tag{B.13}$$
$$+\nu\left[\nabla^2\boldsymbol{B}_u-\frac{\nabla^2\rho}{2\rho}\boldsymbol{B}_u+\frac{3\boldsymbol{B}_u}{4\rho^2}\nabla\rho\cdot\nabla\rho+\frac{1}{\rho}(\nabla\rho\cdot\nabla)\boldsymbol{B}_u)-\frac{1}{6\rho}\nabla(\nabla\rho\cdot\boldsymbol{B}_u)\right]$$
$$+\nu\left[\frac{1}{4\rho^2}(\nabla\rho\cdot\boldsymbol{B}_u)\nabla\rho+\frac{1}{3}\nabla(\nabla\cdot\boldsymbol{B}_u)-\frac{1}{6\rho}(\nabla\cdot\boldsymbol{B}_u)\nabla\rho\right].$$

If the density is also uniform in space, corresponding to the incompressible limit, the equations drastically simplify to:

$$\dot{\boldsymbol{B}} = \frac{1}{\sqrt{\rho\mu_0}}\nabla\times\left[\boldsymbol{B}_u\times\boldsymbol{B}-d_i(\nabla\times\boldsymbol{B})\times\boldsymbol{B}\right]+\frac{\eta}{\mu_0}\nabla^2\boldsymbol{B}, \tag{B.14}$$
$$\dot{\boldsymbol{B}}_u = -\frac{1}{\sqrt{\rho\mu_0}}\left[(\boldsymbol{B}_u\cdot\nabla)\boldsymbol{B}_u-(\nabla\times\boldsymbol{B})\times\boldsymbol{B}\right]+\nu\nabla^2\boldsymbol{B}_u,$$
$$\nabla\cdot\boldsymbol{B}=0,\quad \nabla\cdot\boldsymbol{B}_u=0.$$

This is in fact the incompressible, viscoresistive Hall-MHD equations written in a form equivalent to the Elsässer formulation [45, 133]. Notice the symmetries between the evolutions of the magnetic and velocity fields (broken largely by the inclusion of the Hall-term). In the limit of zero dissipation and $d_i \to 0$, one obtains

$$\dot{\boldsymbol{B}} = \frac{1}{\sqrt{\rho\mu_0}}\left[(\boldsymbol{B}\cdot\nabla)\boldsymbol{B}_u-(\boldsymbol{B}_u\cdot\nabla)\boldsymbol{B}\right], \tag{B.15}$$
$$\dot{\boldsymbol{B}}_u = -\frac{1}{\sqrt{\rho\mu_0}}\left[(\boldsymbol{B}_u\cdot\nabla)\boldsymbol{B}_u-(\boldsymbol{B}\cdot\nabla)\boldsymbol{B}+\nabla\left(\frac{B^2}{2}\right)\right],$$
$$\nabla\cdot\boldsymbol{B}=0,\quad \nabla\cdot\boldsymbol{B}_u=0.$$

These are the incompressible ideal MHD equations.

Appendix C

CLASSIFICATION AND PREDICTION OF PLASMA STRUCTURES AND INSTABILITIES WITH NEURAL NETWORKS

This chapter briefly describes collaborative work that attempts to predict a class of plasma modes from a large DIII-D tokamak database. Much of this work was produced by other researchers, so this section only provides some background information, some promising results from Jalalvand et al. [187], and initial progress towards improved models in future work.

While ROMs and deep learning predictors in the plasma physics community have already been motivated elsewhere in this thesis, it is worth a quick review of why these methods are so critical. Most future magnetic-confinement-fusion reactor designs require steady-state operation for economic viability. In the context of high-performance tokamaks, steady-state operation necessitates active real-time control of a number of complex instabilities including edge-localized modes (ELMs) [239, 158], Alfvén eigenmodes (AEs) [86, 424], and more general disruptions [363, 129]. Even fundamentally steady-state devices such as stellarators will require sophisticated real-time control for modulating gas puff, divertor dynamics, and transport [50]. Furthermore, instabilities can occur on time scales of milliseconds or even microseconds. Subsequently, if plasma control schemes are built and updated in real-time, they are limited to simple models such as those based on 1D transport [271], linearization or local-expansions [227, 6, 302, 419, 207, 314, 441, 10, 11, 301], heuristics (based on prior experimental knowledge) [145], the biorthogonal decomposition [246, 132, 131, 211, 209], and so forth. Many of these models have been successfully employed for real-time control in operational scenarios but the variety and complexity of Alfvén eigenmodes in toroidal

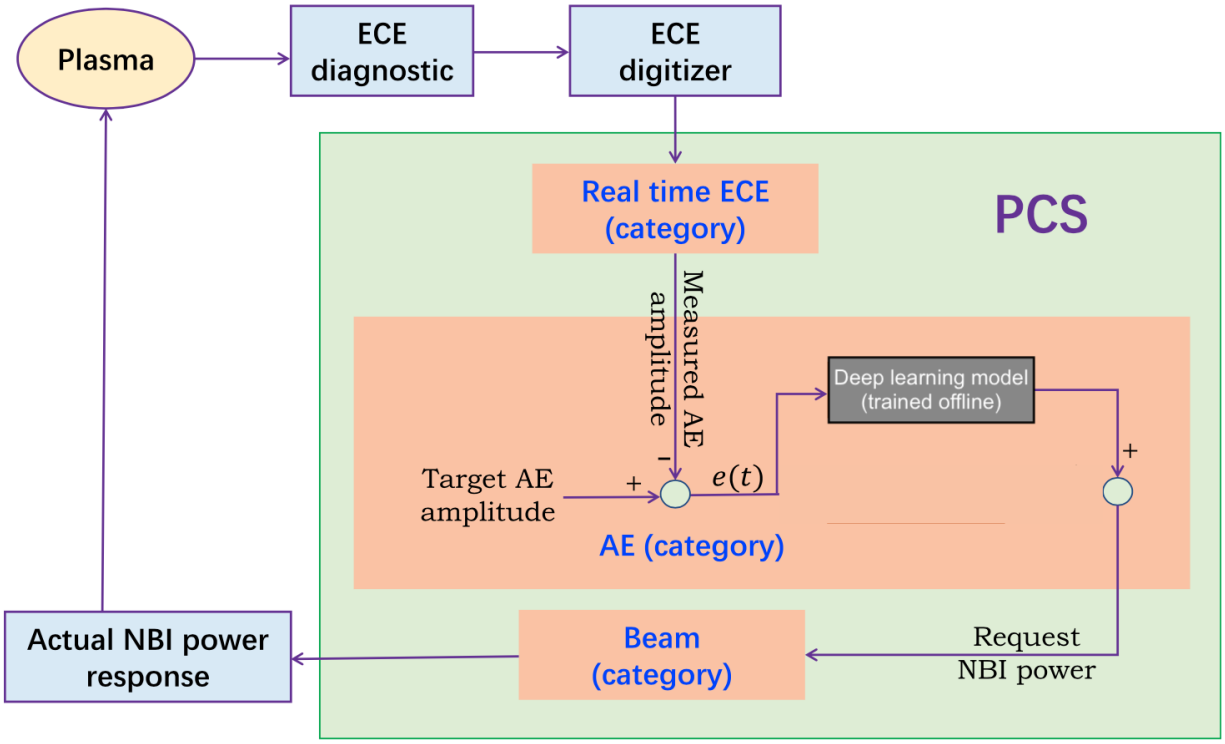


Figure C.1: Possible future implementation illustrating how a capable deep learning model could be used for real-time control of AEs. Figure adapted with permission from Hu, Wenhui, et al. (2018). Active real-time control of Alfvén eigenmodes by neutral beam and electron cyclotron heating in the DIII-D tokamak. Nuclear Fusion, 58(12), 124001.

devices pose many challenges for simple models, generalization to new datasets, and analytic methods. The eventual goal of the following work is to train deep learning models with sufficient performance to be integrated into the DIII-D PCS real-time control system as in Fig. C.1.

C.1 Instability prediction and classification within plasma physics and fusion

There has already been remarkable success in machine learning for disruption identification and real-time control in tokamaks [74, 364, 363, 309, 214, 129], including high-performance models that are not limited to a specific device [300]. There has also been recent deep learning work for AE activity, which utilized manually-labeled spectrogram data from the

TJ-II stellarator [102] and COMPASS tokamak [403] for automated identification of these modes in diagnostic data from a single magnetic probe. The former paper focuses on a binary classification of the spectrogram pixels, indicating whether each pixel corresponds to Alfvénic MHD activity or not. The latter focuses specifically on identifying a useful feature space for unstable reversed-shear Alfvén eigenmodes (RSAEs), which exhibit a unique frequency-sweeping behavior. A recent paper also showed that AE "mode character" (i.e. whether the activity is chirping, avalanching, fixed frequency, or quiescent) can be effectively classified [444]. All three papers indicate promising avenues for future work. Two places for improvement stem from these initial studies using 1) inputs that are single spectrograms from magnetic probes, meaning there is no ability to use spatial correlations or identify internal modes that do not appear near the device walls, and 2) there was no attempt made at discrimination between different kinds of plasma dynamics. Prior collaborative work improved on these studies substantially by utilizing high-resolution spatiotemporal data from a large database, described further below.

C.1.1 Alfvén eigenmodes

AEs are a class of common instabilities observed in tokamaks and other plasma devices. Unfortunately, some types of AE instability, such as energetic particle resonance, can lead to confinement loss and damage to plasma-facing components of the device. The database used in this work (described in Section C.2) distinguishes between several types of AE activity: Low-frequency modes (LFMs $\lesssim 50$ kHz, these "Christmas light" patterns have been formerly characterized as BAAE modes [166]), Beta-induced Alfvén eigenmodes (BAE $\sim 30 - 150$ kHz), Ellipticity Alfvén eigenmodes (EAE $\sim 150 - 200$ kHz), Reversed-Shear Alfvén eigenmodes (RSAE $\sim 50 - 200$ kHz), and Toroidal Alfvén eigenmodes (TAE $\sim 90 - 200$ kHz) [166, 268]. The quoted frequency ranges for each type are approximate, specific to DIII-D, and can vary significantly in differing DIII-D parameter regimes such as L-mode or H-mode. The AE modes are further described in Table C.1, where references to the relevant theoretical and experimental manuscripts can also be found.

Acronym	**Name**	**Cause**
BAE [429, 168, 170]	Beta	Compressibility
EAE [38, 124]	Ellipticity	m_θ and $m_\theta + 2$
LFM [166]	Low-frequency modes	Hot electrons, $q_{\min} \sim$ rational
RSAE [397, 219]	Reversed-shear	$q_{\min}$
TAE [88, 87, 169, 443]	Toroidal	m_θ and $m_\theta + 1$

Table C.1: Description of the AE activity considered in this work, adapted from Heidbrink [167]. Recall that the poloidal wave number is denoted m_θ and the minimum value of the safety factor is denoted $q_{\min}$.

Lastly, AEs are an excellent choice for training predictive models, because there are a wide range of experimental actuators that can be used for real-time control of different AE activity. Recent work indicates TAE suppression by resonant magnetic perturbations (RMPs) in the EAST tokamak [248] and AE stabilization in DIII-D via a controlled energetic ion density ramp [418]. For a review of potential AE control avenues, see Garcia-Muñoz et al. [134].

C.1.2 Electron cyclotron emission

Electron cyclotron emission (ECE) provides direct local measurements of the electron temperature for thermal DIII-D plasmas [15], and as such, can provide spatiotemporally-localized information about AE activity. The electron temperature and all ECE data is reported in keV throughout the paper. The DIII-D ECE diagnostic data is obtained at 500 kHz, in a single toroidal cross-section, at 40 different channels corresponding to varying radial locations, as shown in Fig. C.2. Each ECE channel spans an approximately $1 - 2$ cm radial extent, which is small compared to AE structure (most of the time, an AE mode can be seen across several channels).

Properly capturing the spatial correlations is difficult because the ECE radial positions change with the magnetic field, and therefore can vary substantially during startup operation. The first few ECE channels regularly view data that is outside the last closed flux surface (LCFS); this data is not a trustworthy measurement. In fact, any signal from outside the

LCFS is not blackbody emission. At such locations, the measured emissions are typically a mix of downshifted X-mode radiation from the core, scrambled O-mode radiation, and other "background" emissions (see most plasma physics textbooks for definitions of the X and O waves [412]). Although some plasma instabilities or features can sometimes be seen on these channels, the change in emissions means the measurements can no longer be interpreted as local. Despite this spatial variability and data corruption, initial work used the full, raw, unprocessed ECE data, i.e. only the ECE channel indices corresponding to the relative radial positions of the measurements. This has the advantage that the magnetic field evolution is not required for this analysis.

In summary, machine learning models are directly trained on the full, raw, unprocessed ECE time series data, including rare measurements below the cutoff and corrupted measurements from outside the LCFS. Despite these simplifications, high classification and prediction performance is obtained in initial work [187], and it is interesting that high performance is accessible with minimal data processing.

C.2 The 2009-2017 DIII-D AE energetic particle database

A high-resolution, informative, and properly-labeled database is typically required for effectively training advanced machine learning models. In that regard, a current ramp AE dataset based on Heidbrink et al. [166] is used for training the models in this work. The labels in this DIII-D database cover many years of operation and a very broad parameter space. As is described below, these manual labels are fairly imprecise, but this degradation in label quality is somewhat ameliorated by the data quantity.

The dataset consists of 1139 discharges collected between years 2009-2017. Figure C.3 depicts the database AE labels for the DIII-D discharge 170670 superimposed on several spectrograms of the more illustrative ECE signals. Special marks are added to better visually indicate the various AEs. The database was constructed by using a few different experimental diagnostics to cross-validate the AE label choices, especially when concurrent AEs were present. To ensure a variety of safety factor profiles and to facilitate mode classification,

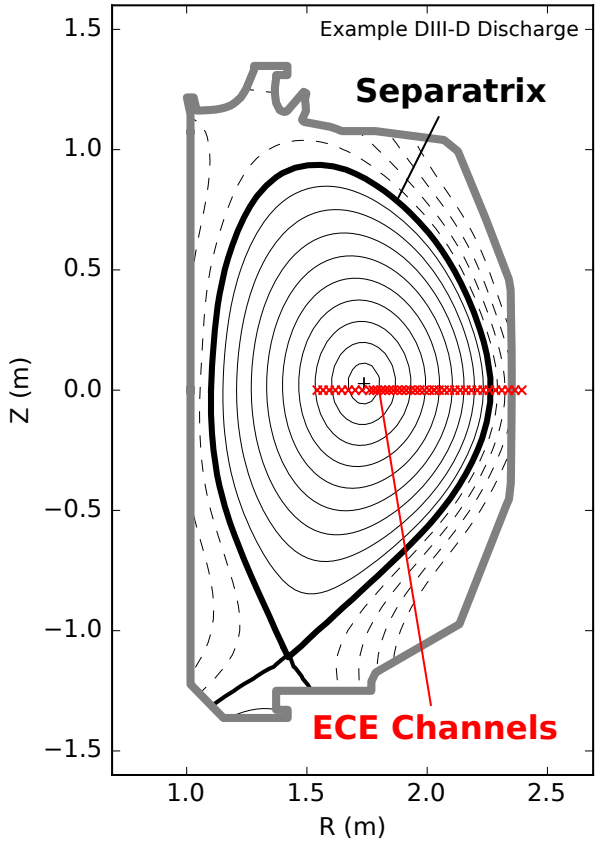


Figure C.2: Illustration of the 40 radial ECE measurement locations alongside the closed (solid) and open (dashed) flux surfaces for an example DIII-D discharge. The ECE radial locations can vary significantly in each discharge, and measurements outside the last closed flux surface are not local or accurate.

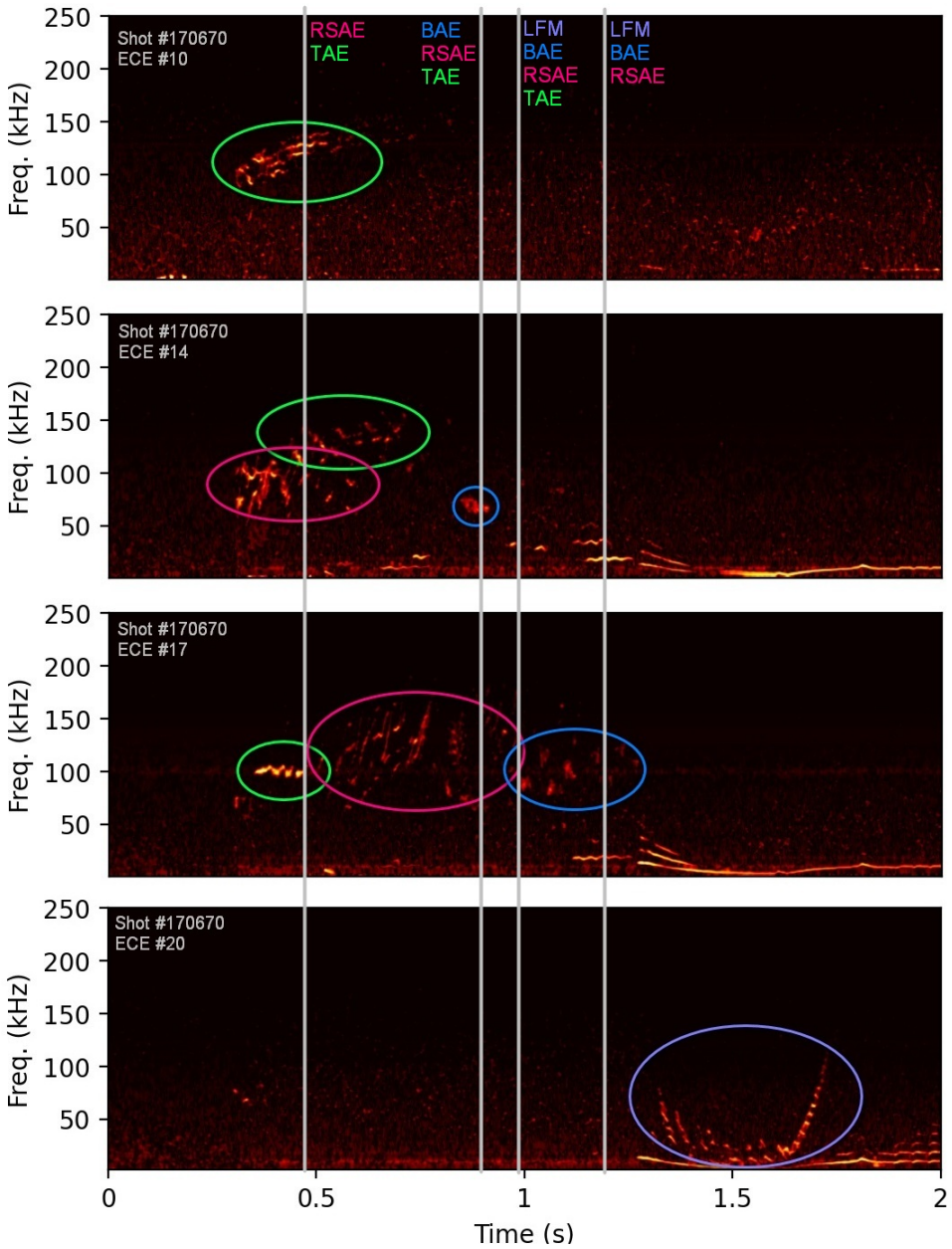


Figure C.3: Illustration of several post-processed (denoised) ECE spectrograms for discharge 170670. The vertical white lines and labels indicate the database timestamps and corresponding instabilities that are used for training the model. The labels indicate only approximate occurrence and there can be substantial regions of unlabeled AE activity. For the reader, some extra colored circles are added to this image to better visualize the different plasma modes.

Poloidal plasma current, I_θ	$I_\theta \leq 1.6$ MA
Toroidal magnetic field, B_ϕ	$0.5 \leq B_\phi \leq 2.1$ T
Normalized Beta, β_N	$0.1 \leq \beta_N \leq 3.2$
Elongation, κ_E	$1.1 \leq \kappa_E \leq 2.2$
Triangularity, δ_T	$-0.4 \leq \delta_T \leq 1.0$
Chord-averaged density, $\overline{n}_e$	$0.4 \leq \overline{n}_e/10^{19} \leq 5.0$ m^{-3}
Central electron temperature, T_e	$T_e \leq 7.6$ keV
Central ion temperature, T_i	$T_i \leq 11.4$ keV
Operational modes	L and H modes
Neutral beam injection	Deuterium
Dominant impurity	Carbon

Table C.2: Tabulation of the broad parameter regime that is spanned by the DIII-D AE database.

selected times in the discharge are all during the first 1.9 s of the discharge, when the safety factor profile steadily evolves. Selected shots had a wide variety of purposes but nearly all dedicated energetic particle experiments are included. Time slices for the labels are chosen to sample either different plasma conditions or different types of mode activity. Subsequently, a given discharge may have only a single AE label or as many as nine labels. In total, the database spans the conditions in Table C.2. More details about the labeling process can be found in [166].

C.3 Summary of results from prior work

Table C.3 lists the performance from initial work in Jalalvand et al. [187] of a 2-layer reservoir computing network (RCN) with 8K-500 nodes per layer on the validation set. The table shows the imbalanced nature of the problem; in total, only 0.02% of the validation set are labeled as an AE (the training set exhibits a similar percentage). This fact is why true-positive-ratio (TPR) and false-positive-ratio (FPR) are the primary metrics. Reporting the accuracy of the model would be profoundly misleading; a model that *never* predicts AE activity would report accuracy above 99%. The strong performance with RCNs is encouraging for continued machine learning work in the future.

Table C.3: Performance of a 2-layer RCN with 8K-500 nodes per layer on the validation set. A threshold of 0.2 has been applied to binarize the model output. There are 566 labeled AEs in total.

AE	TP	FP	TN	FN	TPR	FPR
BAE	75	46,982	470,368	17	0.82	0.09
EAE	17	11,976	566,324	9	0.65	0.02
LFM	8	4,102	587,088	3	0.73	0.01
RSAE	167	48,319	417,211	13	0.93	0.10
TAE	248	76,133	330,057	9	0.97	0.19
Total	515	187,512	2,371,048	51	0.91	0.07

C.4 Recent attempts at spatiotemporally-localized AE classification

There are two primary improvements that can facilitate more advanced machine learning predictors for AE activity, (1) improved database labels that are unique to each ECE channel, facilitating spatially-localized machine learning predictors, and (2) data cleaning by mapping the channels to the normalized plasma radius, removing data outside the LCFS, and removing corrupt ECE data. First, the original database labels are used with the data cleaning in (2).

C.4.1 First attempt: spatiotemporal-local convolutional neural networks

A new 2-layer convolutional neural network (CNN) was designed for classification. In order to fix corrupted data, and capture spatial correlations, a number of steps were taken: mapping the channels to normalized radius ρ_n, zeroing out channels outside the LCFS, and arranging the full 40-channel ECE data as an image in (ρ_n, t) space for each discharge. Only 901 of the database discharges contain the time-resolved data required to perform this mapping from channels to ρ_n. The ECE measurement ρ_n locations change a bit over the discharge as the magnetic field evolves, so a series of ρ_n values are computed every 50 ms for each discharge and then interpolated onto the ECE time-base for mapping the channels at every time. It is common for ECE channels to move by a few cm during a discharge, and occasionally an ECE channel can even swap relative positions with a neighboring ECE channel. As Fig. C.4

illustrates, ECE channels 1-5 are also frequently outside the LCFS, and should be discarded.

To summarize the details, the channel data should be preprocessed by performing the following set of steps:

1 Extract the ECE time series and radial location data. The ECE data is initially a matrix of size $[N, M, 40]$.

2 Every 50 ms, extract the effective plasma radius ρ_n from EFIT [240] or other real-time equilibrium reconstruction software. If need be, obtain density and magnetic field profiles for estimating ECE frequency cutoffs [49] for discarding corrupt data. This is an exceedingly rare occurrence in the dataset so one can safely ignore the cutoffs.

3 At each 50 ms time slice, map the channels to ρ_n. Organize channels from smallest to largest ρ_n, and zero out channels beyond the LCFS, $|\rho_n| > 1$. These channels effectively are ignored by becoming part of the zero padding in the next step. Interpolate the new ρ_n values to the original timebase consisting of M slices.

4 Since ECE measurements observe different radial regions in different discharges, pad and interpolate the rest of the "image" in the $(t, \rho_n) \in [0, 1.9] \times [-1.2, 1.2]$ plane with zeros, with final size $[N, M, N_{\rho_n}]$ (i.e. there is only "vertical" padding in the ρ_n direction), where N_{ρ_n} is the number of interpolated points in the ρ_n direction. Previous studies have shown that padding images had minor effects on the network performance [164]. Notice that the range $-1.2 \leq \rho_n \leq 1.2$ is slightly enlarged past the LCFS so that the border is comfortably zero-padded. The zero-padding standardizes the inputs into a rectangular shape, a requirement for input into a standard CNN architecture.

5 Uniformly slice the images into windows, so that each image is now of size $[N, M_W, N_{\rho_n}]$, i.e. a piece of the original matrix of size $[N, M, N_{\rho_n}]$.

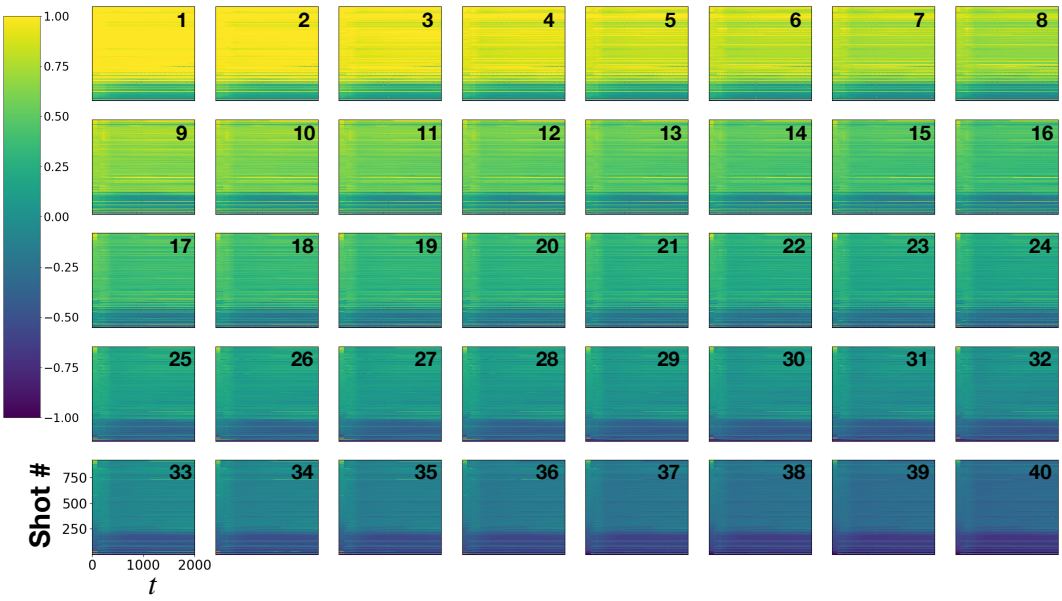


Figure C.4: Contours of ρ_n as a function of the time t (x-axis) and the shot number (y-axis) for the 901 database discharges containing the required data to map each of the 40 channels to ρ_n. Negative ρ_n values correspond to the inner radii values past the magnetic axis. Contour rows are sorted by the first temporal value for a color gradient that improves the visualization. The primary conclusion is that ECE channels 1-4 are often outside the LCFS ($|\rho_n| > 1$), while channels 4-8 and 35-40 occasionally dip outside as well. During pre-processing, any ECE data outside the LCFS is set to zero.

6 Optionally apply a normalization. In previous work, it was found that global normalization to zero-mean-unit-variance signals worked well, so this normalization is used for the following CNN model. The mean and variance are computed only over the training set.

Inputs to the CNN are now organized as a normalized stack of temporally and spatially ordered "partial images" of each discharge. In particular, the spatial correlations have been unscrambled from the raw data, and the temporal correlations are only preserved within each window. The number of windows to use is now a model hyperparameter, and has the effect of changing how much temporal memory is available in a given input image, since temporal correlations outside each window are lost.

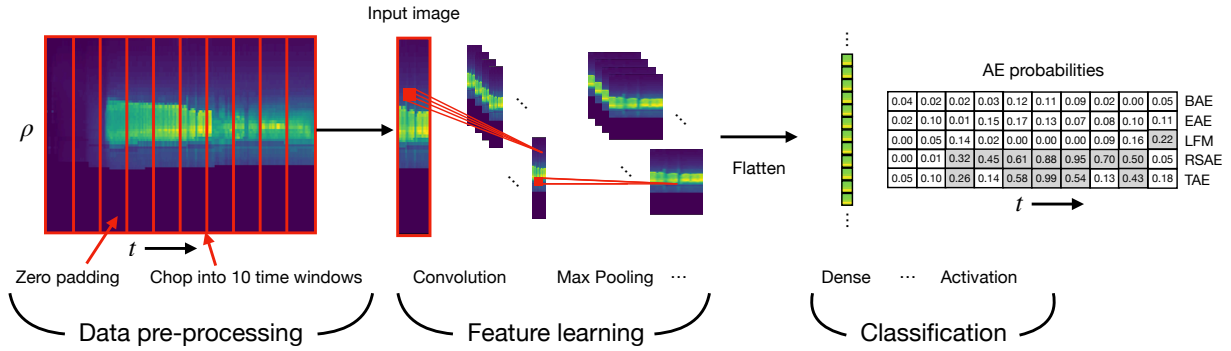


Figure C.5: After mapping every ECE measurement to (t, ρ_n), each ECE channel is zero-padded and sliced into a number of time windows. A basic CNN is composed of a series of convolution and pooling layers, typically followed by a small feed-forward network. The convolutional layers learn the important features of the data. The sigmoid activation layer outputs the probability for each class between 0 and 1. The probabilities above a threshold, here equal to 0.2, are highlighted in grey and register as a true positive if they are within t_{label} of the correct AE label.

However, it was found that CNNs trained with these database labels struggle to find performance greater than $\sim 70\%$ TPR or so, and very quickly overfit to the data. The loss, TPR, and other metrics are summarized in Fig. C.6. This poor training performance is true despite varying all of the hyperparameters by some orders of magnitude, providing a hint that the fundamental issue is the database labels, not the model. This is a sensible result – the database labels are not unique to each ECE channel so the model is inevitably being trained on many false positives and many false negatives, providing an upper bound on the performance achievable with any deep learning model. But this investigation was valuable because it motivated a high-precision labeling of AEs on a small subset of the database.

C.4.2 High precision database labels

As was discussed, the inadequacy of the AE labels in the full database prevent deep learning models from high performance TPR/FPR and prevent accurate spatially-localized predictions, since ground-truth labels are not specific to ECE channel.

In response, ECE-channel-specific labels for 20 very active AE discharges were produced

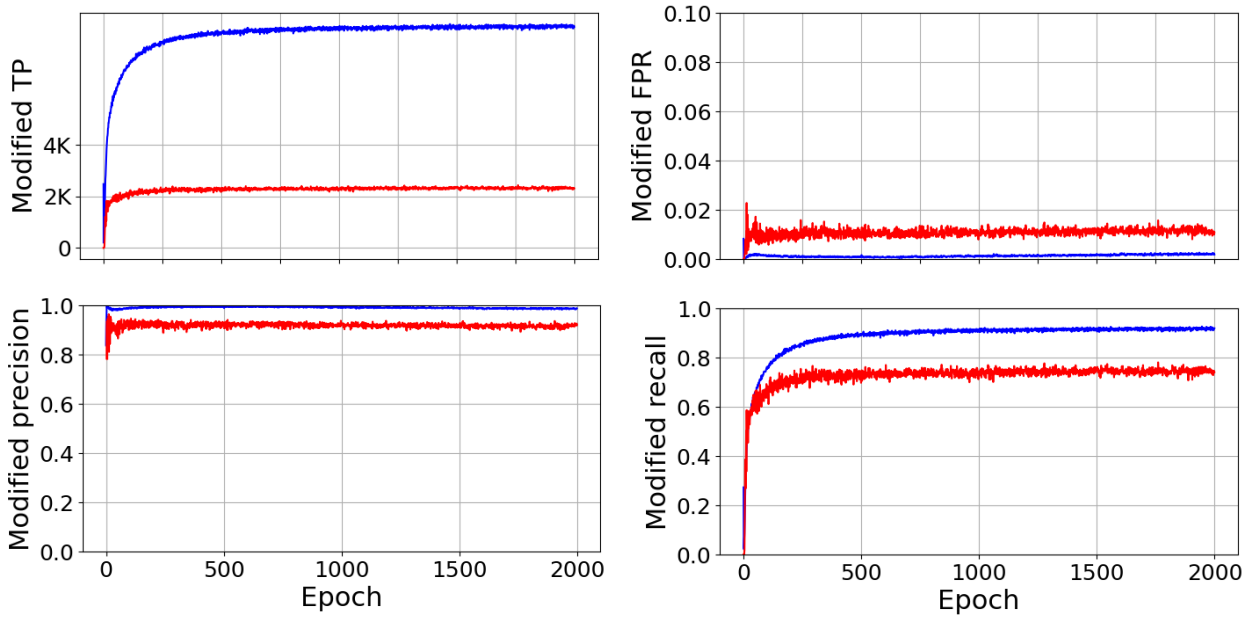


Figure C.6: CNN training indicates the model quickly overfits, although the baseline validation performance is reasonable. These trends are quite insensitive to hyperparameter scans, providing evidence that the fundamental issues are the data and associated labels.

by manually examining all $20 \times 40 = 800$ ECE channel spectrograms and providing specific time and frequency windows for the labels (a box is drawn on the spectrogram where the mode appears). This manual examination is made significantly easier by first using image processing techniques to provide denoised spectrograms that highlight the AE activity [4]. Although this is a very small subset of the original database, each spectrogram contains $\sim 950,000$ data points, meaning the data can still be used to train deep learning models on a large percentage of ~ 760 million data points. Because EAE is quite rare in this new, small database, the focus is on prediction and classification of LFM, BAE, TAE, and RSAE. The new database labels and predictions are illustrated for two active ECE channels during ECE discharge 132710 in Fig. C.7 and the percentage of AE activity in the discharges, broken down by ECE channel, is illustrated in Fig. C.8. Overall, 93.4% of the data has no AE activity whatsoever. Although this data is still quite imbalanced, this is a substantial improvement over the original database where it was estimated that over 99% of the data contained no

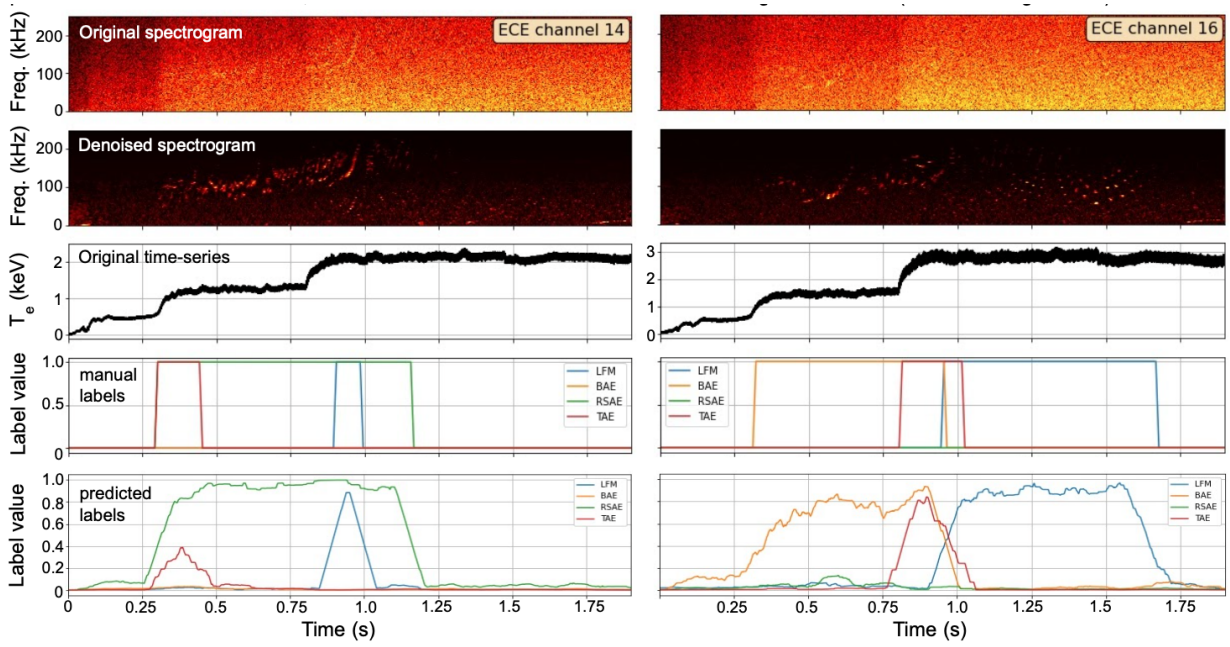


Figure C.7: In order to facilitate more accurate NN predictions, a small dataset of 20 discharges was developed and contains high-precision labels unique for each ECE channel. This illustration shows two ECE channel time series for DIII-D discharge 132710, along with the pre-processing of the data into denoised spectrograms, the manual labels corresponding to each ECE channel, and the predictions on this training data using a simple neural network.

AE activity (or at least no AE labels).

As a first attempt, a simple two-layer perceptron (MLP) (also called a feedforward network) was trained directly on the denoised spectrogram data from this new database. Each input is a single time slice from the denoised spectrogram, consisting of a feature vector of length equal to the number of frequency points in the spectrogram. Although the inputs contain the frequency information in the spectrogram, each input corresponds to a single time slice and therefore there is no temporal memory.

The initial results are promising and the strength of the predictions appear to come from both the data label quality and the denoising; for instance, the spectrogram denoising appears to significantly improve performance over unprocessed spectrograms and suggests additional data pre-processing may be worthwhile. In order to decide when the model has

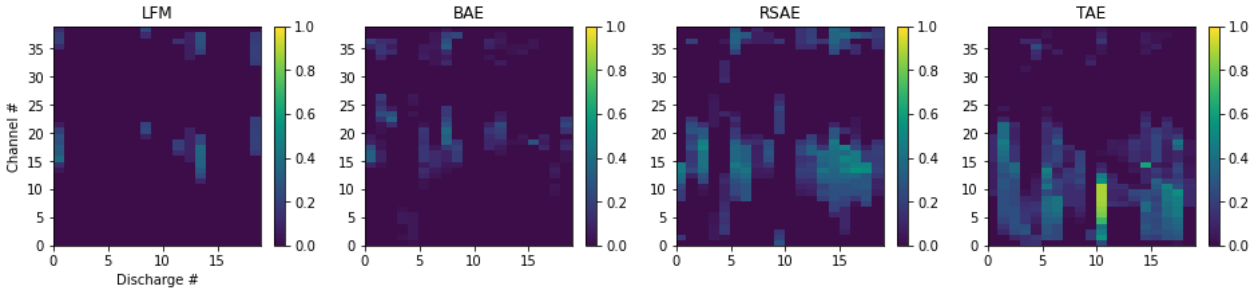


Figure C.8: A summary of the total percentage of AE activity in each ECE time series, plotted against discharge and channel number. TAE and RSAE are present in most of the discharges, and TAE is almost always between channels 1-22 while RSAE tends to be exhibited between channels 7-22 and 35-40. BAE and LFM are significantly more rare in this dataset, and tend to occur over smaller timescales.

started overfitting to the training data, the 20 discharges were randomly split into 15 training discharges and 5 validation discharges. The MLP accuracy, precision, TPR and FPR are plotted as a function of the number of training epochs in Fig. C.9. For each AE type, the training performance is quite high but the validation performance is significantly lower. TAE and RSAE show reasonable validation performance but BAE and LFM indicate low performance. This low performance may be a symptom of the simplicity of this MLP. The MLP is only two layers, and the input data contains no temporal memory for the MLP to extract. LFMs in particular exhibit a complex "Christmas-light" pattern in time, meaning one might expect that identifying these modes would be particularly sensitive to temporal memory. Both of these drawbacks to this first attempt are clear places for further improvement.

C.4.3 Conclusion

Prior collaborative work was furthered by the construction of a spatially-localized set of labels for AE activity identification with machine learning methods. Initial work with this smaller database shows that simple feedforward networks can already produce reasonable spatiotemporally-localized AE classification from denoised spectrograms of individual DIII-

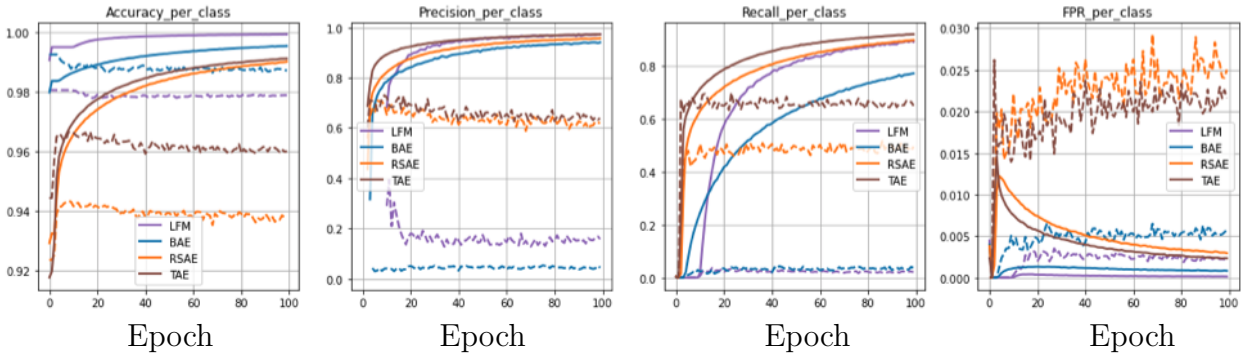


Figure C.9: The MLP accuracy, precision, TPR (recall), and FPR as the neural network trains. One can see high performance on the training data (solid), but only reasonable validation (dashed) performance for TAE and RSAE. LFM and BAE are significantly underrepresented in the data and the validation performance is low.

D ECE channel data.

Although this research indicates that accurate, spatiotemporally localized AE classification is possible from ECE diagnostic data, there remains a lot more work to develop models that are capable enough and compatible with real-time control systems in fusion reactors such as DIII-D. Integrated models, for instance obtained by enriching the input data by a suite of diagnostics including ECE, beam-emission spectroscopy, and magnetics, may improve predictions. Similar gains may be found by investigating more complex deep learning methods and the role of temporal memory in the ML predictions. Future work with multi-machine datasets could also investigate building universal AE detection models for application across toroidal plasma devices.