The structure of global conservation laws in Galerkin plasma models

Alan A. Kaptanoglu
Department of Physics, University of Washington, Seattle, WA, 98195, USA

Kyle D. Morgan
Department of Aeronautics and Astronautics, University of Washington, Seattle, WA, 98195, USA

Chris J. Hansen
Department of Aeronautics and Astronautics, University of Washington, Seattle, WA, 98195, USA and Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY, 10027, USA

Steven L. Brunton
Department of Mechanical Engineering, University of Washington, Seattle, WA, 98195, USA

Plasmas are highly nonlinear and multi-scale, motivating a hierarchy of models to understand and describe their behavior. However, there is a scarcity of plasma models of lower fidelity than magnetohydrodynamics (MHD). Galerkin models, obtained by projection of the MHD equations onto a truncated modal basis, can furnish this gap in the lower levels of the model hierarchy. In the present work, we develop low-dimensional Galerkin plasma models which preserve global conservation laws by construction. This additional model structure enables physics-constrained machine learning algorithms that can discover these types of low-dimensional plasma models directly from data. This formulation relies on an energy-based inner product which takes into account all of the dynamic variables. The theoretical results here build a bridge to the extensive Galerkin literature in fluid mechanics, and facilitate the development of physics-constrained reduced-order models from plasma data.

I. INTRODUCTION

There are a tremendous number of known plasma models of varying model complexity, from magnetohydrodynamics (MHD) to the Klimontovich equations, but a large gap exists in the lower rungs of this hierarchy between simple circuit models and the many MHD variants. This is a valuable place for improvement because higher fidelity models often require computationally intensive and high-dimensional simulations [1–3], obfuscating the dynamics and precluding model-based real-time control. Additionally, many high-dimensional nonlinear systems tend to evolve on low-dimensional attractors [4], defined by spatio-temporal coherent structures that characterize the dominant behavior of the system. A number of studies in the plasma physics community indicate that the vast majority of the total plasma energy can be explained by fewer than ten low-dimensional modes, across a large range of parameter regimes, geometry, and degree of nonlinearity [5–13]. In these cases, the evolution of only a few coherent structures, obtained from systematic model-reduction techniques [14, 15], can closely approximate the full evolution of the high-dimensional physical system.

In the present work and our companion paper [16], we provide a theoretical framework for physics-constrained, low-dimensional plasma models which furnish this important gap in the hierarchy of plasma models. This is a breakthrough in principled and interpretable model reduction, offering an alternative to deep learning methods which often require vast quantities of data and produce opaque models. The present work focuses on the important technical details for constructing and constraining these models, while the accompanying publication contains an overview of our high-level contributions and initial results on 3D plasma simulations.

II. LOW-DIMENSIONAL MODELS

Although there are many ways to obtain low-dimensional models, Galerkin methods and their extensions have seen remarkable success in fluid mechanics; careful development of a dimensionalized inner product enabled the extension of the proper orthogonal decomposition (POD) from incompressible to compressible fluid flows [17]. It is also common in fluid mechanics to obtain nonlinear reduced-order models by Galerkin projection of the Navier-Stokes equations onto POD modes, making it possible to enforce known symmetries and conservation laws, such as conservation of energy [18, 21].

The present work adapts and extends these innovations for plasmas, enabling a wealth of advanced modeling and control machinery. The POD is already used extensively for interpreting plasma physics data across a range of parameter regimes [22, 26]. For clarity of presentation and robust connection with the Galerkin literature in fluid
mechanics, we present results for MHD models which are at most quadratic in nonlinearity. This includes ideal MHD, incompressible Hall-MHD, or compressible Hall-MHD with a slowly time-varying density, which together describe the dynamics of a fairly broad class of space and laboratory plasmas [27, 32].

A. An MHD energy inner product

Traditional use of the POD on the MHD fields (velocity, magnetic, and temperature) requires separate decompositions for $u$, $B$, and $T$, or an arbitrary choice of dimensionalization. However, separate decompositions of the fields obfuscates the interpretability and increases the complexity of a low-dimensional model, and choosing of the fields obfuscates the interpretability and increases the complexity of the SVD is typically chosen to balance accuracy and complexity [33]. The computational complexity of the SVD is $\mathcal{O}(DM^2 + M^3)$ [34], although there are randomized singular value decompositions [35, 37] for extremely large problems that can be as fast as $\mathcal{O}(DM \log(r))$. This computational speed enables the SVD to be computed online, for updating a model in a real-time control application, or offline, for the decomposition of very large data, an examination of the physics, or development of a more generic model describing the dynamics of an amalgam of discharges. A well-defined SVD requires that the measurements in $X$ have the same physical dimensions. With a dimensionalized measurement vector $q$, the matrix $X^*X$ may be approximated via

$$X^*X \approx \langle q(t_k), q(t_m) \rangle. \quad (5)$$

When the number of snapshots is far fewer than the number of measurements, $M \ll D$, we can use the method of snapshots. Substitution of Eq. [4] into $X^*X$ produces

$$X^*XV_r \approx V_r \Sigma_r^2, \quad (6)$$

an eigenvalue equation for $V_r$; therefore we can obtain $V_r$ by diagonalizing $X^*X \in \mathbb{R}^{M \times M}$ instead of computing the SVD of $X$. The chronos are the temporal SVD modes, i.e. the columns of $V_r$, denoted $v_j$. The topos are the spatial modes forming the columns of $U_r$, denoted $x$. In the present work, we scale the normalized matrix of chronos, $a$, through $a_{jk} = v_{jk}/\sum_{j=1}^r \max_k |v_{jk}|$. Finally, the reconstruction can be written

$$q(x_i,t_k) \approx \bar{q}(x_i) + \sum_{j=1}^r \chi_j(x_i) a_{j}(t_k), \quad (7)$$

where $U \in \mathbb{R}^{D \times D}$ and $V \in \mathbb{R}^{M \times M}$ are unitary matrices, and $\Sigma \in \mathbb{R}^{D \times M}$ is a diagonal matrix containing non-negative and decreasing entries $\sigma_{ij}$ called the singular values of $X$. $V^*$ denotes the complex-conjugate transpose of $V$. The singular values indicate the relative importance of the corresponding columns of $U$ and $V$ for describing the spatio-temporal structure of $X$. Although varying terminology is used in different fields (in the plasma physics community this method is often referred to as the biorthogonal decomposition, or BOD), in practice the SVD, BOD, and POD are synonymous.

$$B_u = \sqrt{\rho \mu_0 u}, \quad B_T = 2 \sqrt{\rho \mu_0 k_b T/m_i (\gamma - 1)}, \quad (1)$$

where $u$ is the fluid velocity, $\rho$ is the mass density, $k_b$ is Boltzmann’s constant, $\mu_0$ is the permeability of free space, $T$ is the plasma temperature, $m_i$ is the ion mass, $\gamma$ is the adiabatic index, and $p = 2\rho T/m_i$ is the plasma pressure. $B_u$ and $B_T$ are defined so that the following scaled inner product yields the total energy $W$,

$$W = \frac{1}{2} \rho \mu_0 B^2 + \frac{B^2}{2 \rho \mu_0} + \frac{p}{\gamma - 1} x^2. \quad (2)$$

B. Review of the POD

For the POD, measurements at time $t_k$ are arranged in a vector $q_k \in \mathbb{R}^D$, called a snapshot, where the dimension $D$ is the product of the number of spatial locations and the number of variables measured at each point. The data is sampled at times $t_1, t_2, ..., t_M$, arranged in a matrix $X \in \mathbb{R}^{D \times M}$, and the average in time $\bar{q}$ is subtracted off. The singular value decomposition (SVD) provides a low-rank approximation of the subsequent data matrix

$$X = \begin{bmatrix} q_1(t_1) & q_1(t_2) & \cdots & q_1(t_M) \\ q_2(t_1) & q_2(t_2) & \cdots & q_2(t_M) \\ \vdots & \vdots & \ddots & \vdots \\ q_D(t_1) & q_D(t_2) & \cdots & q_D(t_M) \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{M,M} \end{bmatrix}, \quad \chi_j = \begin{bmatrix} \chi_j(t_1) \\ \chi_j(t_2) \\ \vdots \\ \chi_j(t_M) \end{bmatrix}, \quad U = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_M \end{bmatrix}, \quad \Sigma^* \Sigma = U V^*, \quad (3)$$

where $U \in \mathbb{R}^{D \times D}$ and $V \in \mathbb{R}^{M \times M}$ are unitary matrices, and $\Sigma \in \mathbb{R}^{D \times M}$ is a diagonal matrix containing non-negative and decreasing entries $\sigma_{ij}$ called the singular values of $X$. $V^*$ denotes the complex-conjugate transpose of $V$. The singular values indicate the relative importance of the corresponding columns of $U$ and $V$ for describing the spatio-temporal structure of $X$. Although varying terminology is used in different fields (in the plasma physics community this method is often referred to as the biorthogonal decomposition, or BOD), in practice the SVD, BOD, and POD are synonymous.

The truncation rank $r$ is typically chosen to balance accuracy and complexity [33]. The computational complexity of the SVD is $\mathcal{O}(DM^2 + M^3)$ [34], although there are randomized singular value decompositions [35, 37] for extremely large problems that can be as fast as $\mathcal{O}(DM \log(r))$. This computational speed enables the SVD to be computed online, for updating a model in a real-time control application, or offline, for the decomposition of very large data, an examination of the physics, or development of a more generic model describing the dynamics of an amalgam of discharges. A well-defined SVD requires that the measurements in $X$ have the same physical dimensions. With a dimensionalized measurement vector $q$, the matrix $X^*X$ may be approximated via

$$X^*X \approx \langle q(t_k), q(t_m) \rangle. \quad (5)$$

When the number of snapshots is far fewer than the number of measurements, $M \ll D$, we can use the method of snapshots. Substitution of Eq. [4] into $X^*X$ produces

$$X^*XV_r \approx V_r \Sigma_r^2, \quad (6)$$

an eigenvalue equation for $V_r$; therefore we can obtain $V_r$ by diagonalizing $X^*X \in \mathbb{R}^{M \times M}$ instead of computing the SVD of $X$. The chronos are the temporal SVD modes, i.e. the columns of $V_r$, denoted $v_j$. The topos are the spatial modes forming the columns of $U_r$, denoted $x$. In the present work, we scale the normalized matrix of chronos, $a$, through $a_{jk} = v_{jk}/\sum_{j=1}^r \max_k |v_{jk}|$. Finally, the reconstruction can be written

$$q(x_i,t_k) \approx \bar{q}(x_i) + \sum_{j=1}^r \chi_j(x_i) a_{j}(t_k). \quad (7)$$

We have absorbed the normalization of $a_{jk}$ and the singular values into the definition of $\chi_j(x_i)$. By construction $(\chi_i, \chi_j) \propto \delta_{ij}$. Note that, in principle, we could have expanded $q$ in any set of modes, although orthonormal modes are preferred because this property facilitates the analysis in Section [I]. The advantage of the POD basis is
that the modes are ordered by energy content; a truncation of the system still captures the vast majority of the dynamics. 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, our approach captures all the fields simultaneously, resulting in a single set of modes $a_i(t)$ in Eq. (7).

An example of this decomposition is illustrated in Fig. 1 for a 3D plasma simulation that is dominated by harmonics of a driving frequency. In general, examining the structure and symmetry in the spatial and temporal POD modes can inform physical understanding. For instance, in Fig. 1, the short-wavelength structures exhibited in the 3D spatial modes derive both from dispersive whistler waves via the Hall term and the small characteristic scale associated with the driving actuator. The steep fall-off in the singular values also indicates that the modes are ordered by energy content; a truncation of the system still captures the vast majority of the dynamics. A separate POD of each of the MHD fields would involve significantly more complicated nonlinear terms from mixing and a lack of orthonormality $\langle \chi^a_i, \chi^B_j \rangle \neq \delta_{ij}$ between the different POD modes associated with each field.

### III. CONSTRAINTS ON MODEL STRUCTURE

Local and global MHD conservation laws are retained in this low-dimensional basis. Vanishing $\nabla \cdot B$ and the orthonormality of the temporal POD modes produce

$$\nabla \cdot \chi^B_i = 0, \quad \forall i. \quad (9)$$

In other words, the orthonormality of the temporal modes guarantees that the local divergence constraint is satisfied by each of the $\chi^B_i$ by construction. In contrast, global energy conservation will produce strong constraints on the structure of the coefficients in Eq. (8).

#### A. Global conservation of energy

For an examination of the global conservation laws, we consider isothermal Hall-MHD with the assumption that the density is slowly-varying in time. This model reduces to ideal MHD and incompressible resistive or Hall MHD in the appropriate limits, and produces (Galtier [39] Eq. 3.22)

$$\frac{\partial W}{\partial t} = -\int \left[ \frac{\nu}{\mu_0} (\nabla \times B)^2 + \frac{\eta}{\mu_0} \nabla \cdot B^2 + \frac{4}{3} \frac{\nu}{\mu_0} (\nabla \times u)^2 \right] d^3 x \quad (10)$$

$$-\int \left[ \frac{1}{2} \rho u^2 + p \right] u + P - \frac{3}{4} \frac{\nu}{\mu_0} (\nabla \times u) \cdot u \times (\nabla \times u) \cdot \hat{n} dS.$$  

Here $\hat{n}$ is a unit normal vector to the boundary, and $P = \frac{1}{\mu_0} E \times B - \frac{u_e}{\mu_0} (B^2 I - BB) - \frac{\eta}{\mu_0} (\nabla \times B) \times B. \quad (11)$

is the Poynting vector ($E$ is the electric field), which is often a known function of space and time. Omission of the Hall term changes $u_e$ to $u$. Even with a Hall-MHD model that evolves the temperature, the electron diamagnetic contribution to $P$ does not change the energy balance if Dirichlet boundary conditions are used for $\rho$ and $T$. To simplify, we assume that $u \cdot \hat{n} = u \times \hat{n} = 0$, $J \cdot \hat{n} = 0$, and $\mathbf{B} \cdot \hat{n} = 0$ at the wall, consistent with the simulations used in the accompanying work [10]. Moreover, we now assume that we have a steady-state, define a constant $a_0(t) = 1$, and substitute Eq. (7) into Eq. (10).

$$0 \approx \frac{\partial W}{\partial t} = \int \left[ \frac{\nu}{\mu_0} (\nabla \times B_a)^2 - \frac{\nu}{2 \rho} \mathbf{B}_a (\nabla \cdot \mathbf{B}) \right] d^3 x \quad (12)$$

$$= W_C + \sum_{i=1}^r W_i^L a_i + \sum_{i,j=1}^r W_{ij}^Q a_i a_j = W_C + \sum_{i=1}^r a_i (W_i^L + \sum_{j=1}^r W_{ij}^Q a_j) = W_C + \sum_{i=1}^r \sum_{j=0}^r W_{ij}^Q a_i a_j = \sum_{i=0}^r \sum_{j=0}^r W_{ij}^Q a_i a_j,$$
With some algebra, we can compute $a_\omega$ where we have padded the matrix in the last step so that $W_{0i}^Q=0$, $W_{0i}^Q=W_i^L$ for $i\in\{1,\ldots,r\}$, and $W_{00}^Q=W^C$. It immediately follows from Eq. (12) that $W_{0i}^Q$ is an antisymmetric matrix. Writing out the coefficients we have

$$0=W_{0i}^Q=\frac{\eta}{\mu_0}\int [\nabla \times B] \times \bar{B} \, d\Omega S - \int \left[ \nu (\nabla \times B) \cdot \frac{\nabla \rho}{2\rho}\bar{B} \right] d^3x,$$

$$0=W_{0j}^Q=\frac{\eta}{\mu_0}\int \left[ (\nabla \times B) \cdot \chi_i^B + (\nabla \times \chi_i^B) \times \bar{B} \right] \, d\Omega S - 2\int \left[ \nu (\nabla \times B) \cdot \frac{\nabla \rho}{2\rho}\bar{B} \right] \left[ \nabla \chi_i^B \times \frac{\nabla \rho}{2\rho} \chi_i^B \right] d^3x,$$

$$W_{ij}^Q=\frac{\eta}{\mu_0}\int \left[ (\nabla \chi_i^B) \times \frac{\nabla \rho}{2\rho} \chi_j^B \right] \, d\Omega S - \int \left[ \nu (\nabla \times \chi_i^B) \cdot \frac{\nabla \rho}{2\rho} \chi_j^B \right] \left[ \nabla \chi_i^B \times \frac{\nabla \rho}{2\rho} \chi_j^B \right] d^3x.$$

With some algebra, we can compute $a_i\hat{a}_i$ for $i\in\{1,\ldots,r\}$,

$$a_i\hat{a}_i = \sum_{i,j=1}^{r} a_i \frac{\partial a_j}{\partial t} \int \chi_i \chi_j \, d^3x = \int \frac{1}{2} \frac{\partial q^2}{\partial t} \, d^3x = \frac{\partial W}{\partial t} \quad (14)$$

FIG. 1: The first seven POD modes for an isothermal, compressible Hall-MHD simulation of the HIT-SI device [38]: (a) Feature space trajectories of every mode pair and the singular values; (b) 3D spatial modes visualized in the $Z=0$ midplane and normalized to $\pm 1$; (c) Normalized temporal modes and corresponding Fourier transforms produce harmonics of the driving frequency, labeled 1-5 in the Fourier space.
for all \(i \in \{1,...,r\}\). In other words, there are no constant terms in the Galerkin model; data-driven methods can implement this constraint by simply searching for models that do not have constant terms. This a physical consequence of our assumption that \(\bar{q}\) is steady-state because nonzero constant terms in the Galerkin model 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 \(a^T \cdot a\),

\[
a^T L^0 a \approx 0. \tag{15}
\]

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 \(P \propto \eta J \times B\), must be balancing these losses. Finally, because of the boundary conditions there are no cubic terms in the energy at all, meaning

\[
a^T Q^0 a a = 0. \tag{16}
\]

This is the condition for a system to have energy-preserving quadratic nonlinearities; this conclusion does not rely on any assumption of steady-state and energy-preserving structure in other quadratic nonlinearities is well-studied in fluid mechanics [20, 40].

B. Global conservation of cross-helicity

An analogous derivation can be done to further constrain the model-building for systems which conserve cross-helicity, although this is not appropriate for the simulation results in the accompanying work [16]. Consider the local form of cross-helicity \(H_c = u \cdot B\). Using Galtier [39] Eq. (3.36),

\[
\frac{\partial H_c}{\partial t} = -\nabla \cdot \left[ \left( \frac{u^2}{2} + \frac{\gamma p}{\gamma - 1} \right) B + u \times (u \times B) - \frac{d_i}{\sqrt{\mu_0}} u \times (\nabla \times B) \right] + \nu \nabla \cdot \left( B \times (\nabla \times u) + \frac{4}{3} (\nabla \cdot u) B \right) - \frac{d_i}{\sqrt{\mu_0}} (\nabla \times u) \cdot ((\nabla \times B) \times B) - (\eta + \nu)(\nabla \times B) \cdot (\nabla \times u). \tag{17}
\]

Consider again the simplifying case \(J \cdot \hat{n} = 0, B \cdot \hat{n} = 0\), and \(u \cdot \hat{n} = u \times \hat{n} = 0\). Then the integral form is

\[
0 \approx \int \frac{\partial H_c}{\partial t} d^3 x = \int \left[ \nu \frac{\nabla^2}{\rho} \left( B \times (\nabla \times u) + \frac{4}{3} (\nabla \cdot u) B \right) - \frac{d_i}{\sqrt{\mu_0}} (\nabla \times u) \cdot ((\nabla \times B) \times B) - (\eta + \nu)(\nabla \times B) \cdot (\nabla \times u) \right] d^3 x. \tag{18}
\]

Substituting in Eq. (7) produces terms up to cubic in the temporal POD modes,

\[
\int \frac{\partial H_c}{\partial t} d^3 x = \frac{\partial}{\partial t} (a_{i,j}) \int \frac{1}{\sqrt{\mu_0}} \chi^i_{\beta} \cdot \chi^j_{\beta} d^3 x = \int A_{ij} \left( C^0_j a_i \right) + A_{ij} L^0_{kj} a_i a_k + A_{ij} Q^0_{ijkl} a_i a_k a_l. \tag{19}
\]

Note that if the system is energy-preserving, \(C^0_j = 0\) for all \(j\), so the first equality is already satisfied. The second inequality gives \(A_{ij} L^0_{jk}\) anti-symmetric under swapping \(i\) and \(k\), and energy-preservation gave anti-symmetry under swapping \(j\) and \(k\) (Eq. 15). The most straightforward solution is \(L^0_{jk} = 0\) for all \(j,k\); this solution is precisely the ideal limit corresponding to \(\eta = \nu = 0\). Since \(A_{ij}\) is not symmetric by construction, this constraint can also apply to systems which conserve cross-helicity despite finite dissipation.

Lastly, \(A_{ij} Q^0_{ijkl}\), containing only the contribution from the Hall-term, exhibits the same structure as (and is compatible with) our constraint on the energy-preserving nonlinearities in Eq. (16). The simplest solution is \(A_{ij} Q^0_{ijkl} = 0\) 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 if the Hall-terms have this special energy-preserving structure, nonzero Hall contributions can still conserve cross-helicity. Enforcing other invariants of Hall-MHD may require alternative formulations to the one presented here, since derived fields like the vector potential are involved.

C. Conservation laws with velocity units

In closer analogy to fluid dynamics [17], we could have alternatively used \(q = [u, u_A, u_s]\),

\[
u_i^2 = \frac{4T}{m_i(\gamma - 1)}, \quad u_A = \frac{B}{\sqrt{\mu_0 \rho}} \tag{20}
\]

\[
\frac{1}{2} q, q = \frac{1}{2} \int \left( u^2 + u_A^2 + u_s^2 \right) d^3 x. \tag{21}
\]
We have defined a scaled plasma sound speed, $u_s$. If $\rho$ is uniform $\rho(q,q)/2=W$. The isothermal and time-independent density assumptions allow us to derive another quadratic model in $q$, for which a POD-Galerkin model is readily available (the form is identical to Eq. [8] but the coefficients have changed). Once again, assume $u \cdot n = u \times n = 0$, and $B \cdot n = 0$ on the boundary, so that

$$\int \frac{\rho}{2} \frac{d^2q}{dt^2} d^3x = \frac{\partial W}{\partial q}.$$  \hspace{1cm} (22)

This is equivalent to Eq. (11) in the particular case of time-dependent density. Without this assumption, an extra term appears, proportional to $\int u \cdot \nabla (u^2 + u_s^2) d^3x$. 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 $u \cdot \nabla u_s^2$ and $u \cdot \nabla u^2$ 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 $q=[u, u_A, u_s]$ 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. Now that we have illustrated how global conservation laws manifest as structure in Galerkin models, we could compute these coefficients and evolve the subsequent model. However, For an explicit calculation of the model coefficients, the first and second order spatial derivatives for the MHD fields must be evaluated. This is equivalent to Eq. (14) in the particular case of linear dynamics (SINDy) algorithm \[47\] to identify nonlinearities sparse regression techniques, optimizing for the sparsest set of equations that produces an accurate fit of the data. The SINDy optimization problem is

$$\min_\Xi ||\hat{a} - \Theta(a)\Xi||_2^2 + \lambda R(\Xi),$$

subject to $D\Xi[:]=d$.

where $R(\Xi)$ is some regularizer, like the $L_0$ or $L_1$ norm, which promotes sparsity in $\Xi$. Here $a, \hat{a} \in \mathbb{R}^{M \times r}$, $\Theta(a) \in \mathbb{R}^{M \times N}$, $\Xi \in \mathbb{R}^{N \times r}$, $D \in \mathbb{R}^{N_s \times rN}$, $\Xi[:]=\begin{bmatrix} \xi_1 \cdots \xi_r \cdots \xi_N \cdots \xi_N \end{bmatrix}$.

B. Derivation of the SINDy constraints

In Sec. III we derived model constraints from global conservation laws; our goal here is to rewrite these constraints to be compatible with the notation in Eq. [21]. 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 $\Xi$, and 3) a more complicated energy-preserving structure in the quadratic coefficients. Consider a quadratic library in a set of $r$ modes, ordered as $\Theta(a)=[a_1, \ldots, a_r]$, where $N_s$ is the number of candidate terms, $N_r$ is the number of constraints, and $\Xi[:]=\begin{bmatrix} \xi_1^a \cdots \xi_r^a \cdots \xi_N^a \cdots \xi_N^a \end{bmatrix}$. Note that this arrangement of the polynomials in $\Theta$ differs from Loiseau et al. [40], so the indexing and subscripts are different here. First we will consider the constraint on the linear part of the Galerkin model in Eq. [8] or equivalently that the quadratic term $a^T L_0 a=0$. We can rewrite this in the SINDy notation as

$$0=[a_1 \cdots a_r] \begin{bmatrix} \xi_1^a & \cdots & \xi_r^a \\ \vdots & \ddots & \vdots \\ \xi_r^a & \cdots & \xi_r^a \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_r \end{bmatrix}.$$

A. The constrained SINDy method

The goal of SINDy is to identify a low-dimensional model for $a(t)$, the vector of POD amplitudes, as a sparse linear combination of elements from a library of candidate terms $\Theta=[\theta_1(a) \ \theta_2(a) \ \cdots \ \theta_p(a)]$:

$$\frac{da}{dt} = f(a) \approx \Theta(a)\Xi.$$  \hspace{1cm} (23)

To address this combinatorically hard problem, it leverages sparse regression techniques, optimizing for the sparsest set of equations that produces an accurate fit of the data. The SINDy optimization problem is

IV. GLOBAL CONSERVATION LAWS IN MACHINE LEARNING MODEL DISCOVERY

Increasingly, machine learning techniques are allowing scientists to extract a system’s governing equations of motion directly from data. Here we show how these algorithms can directly incorporate global conservation laws during the search for low-dimensional models in plasma datasets. We use the sparse identification of nonlinear dynamics (SINDy) algorithm \[17\] to identify nonlinear reduced-order models for plasmas in the accompanying work \[16\].
We conclude $\xi_{ij} = -\xi_{ji}$ for $i,j \in \{1, \ldots, r\}$ and identify $\xi_{a}^{(j)}$ by accessing the $(i-1)r+j$ index in $\Xi[i]$. Note we are only accessing the first $r^2$ elements of $\Xi[i]$. For models limited to linear and quadratic polynomials, $N = (r^2 + 3r)/2$ and the number of constraints from anti-symmetry of the linear coefficients is $N_L = (r^2 + r)/2$. Thus there are now only $rN - N_L = r(r^2 + 2r - 1)/2$ free parameters. Since the constrained SINDy algorithm solves linear equality constraints of the form $D\Xi[i] = d$, we can write this out explicitly for $r = 3$.

\[
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \cdots \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & \cdots \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \cdots \\
\end{bmatrix}
\begin{bmatrix}
\xi_1 \\
\xi_2 \\
\xi_3 \\
\xi_4 \\
\xi_5 \\
\xi_6 \\
\end{bmatrix}
= 
\begin{bmatrix}
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
\end{bmatrix}.
\]

(26)

The boundary conditions $u \cdot \hat{n} = 0$, $J \cdot \hat{n} = 0$, $B \cdot \hat{n} = 0$ guaranteed that the quadratic nonlinearities were energy-preserving, and thus that cubic terms in Eq. (10) vanish. In this case, regardless of the steady-state assumption, we have that

\[
\sum_{i,j,k=0}^{r} Q_{ijk}^{0} a_i a_j a_k = 0.
\]

(27)

This constraint is significantly more involved to reformat. Written in SINDy notation, this is equivalent to

\[
0 = [a_1 \cdots a_r]
\begin{bmatrix}
\xi_1^{a_1} & \xi_1^{a_2} & \cdots & \xi_N^{a_1} \\
\xi_2^{a_1} & \xi_2^{a_2} & \cdots & \xi_N^{a_2} \\
\vdots & \vdots & \ddots & \vdots \\
\xi_r^{a_1} & \xi_r^{a_2} & \cdots & \xi_N^{a_r} \\
\end{bmatrix}
\begin{bmatrix}
a_1 a_2 \\
a_1 a_3 \\
\vdots \\
a_1 a_r \\
a_2 a_1 \\
\vdots \\
a_r a_1 \\
a_2 a_2 \\
\vdots \\
a_r a_r \\
\end{bmatrix}.
\]

(28)

Expand this all out and group the like terms, i.e. terms which look like $a_1^3$, $a_1 a_2^2$ or $a_1 a_2 a_3$, $i,j,k \in \{1, \ldots, r\}$, $i \neq j \neq k$. All of the like terms can be straightforwardly shown to be linearly independent, so we 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 $r(r^2 + 1)(r+2)/6 = N_Q$ constraints. With both constraints, we have $rN - N_L - N_Q = r(r-1)(2r+5)/6$ free parameters, and $N_c = N_L + N_Q$ constraints. Further considering the quadratic case, we find that coefficients which adorn $a_1^3$ must vanish, $\xi_{r+1}\xi_{r+2} = 0$. Now define

\[
\tilde{\xi}_{ijk} = \xi_{r+1}^{a_i} (2r-j-3)+k-1.
\]

(29)

The second type of constraint, with $i \neq j$, produces

\[
\xi_{N-r+j}^{a_i} = \left\{ \begin{array}{ll}
\tilde{\xi}_{iij} & i < j \\
\tilde{\xi}_{jji} & i > j,
\end{array} \right.
\]

(30)

while the third type of constraint produces

\[
\tilde{\xi}_{ijk} + \tilde{\xi}_{jik} + \tilde{\xi}_{kij} = 0.
\]

(31)

This relation is equivalent to the energy-preserving conditions in Schlegel et al. [20], but the indexing is not straightforward, even after fully expanding Eq. (28). This equation is an arbitrary $r$ generalization to the $r = 3$ constraint used in Loiseau et al. [10]. For the specific case where the plasma system is Hamiltonian (for instance in ideal [48], Hall [49], and extended [50] 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 [51] and subsequently derive the equations of motion. Lastly, cross-helicity constraints can be straightforwardly implemented from these results. If the constraint on the quadratic terms in $\Xi$ is written $D_{jk} \Xi_k = 0$, then the quadratic cross-helicity constraint can be written $D_{jk} A_{kl} \Xi_l = 0$.

V. CONCLUSION

A hierarchy of models with varying fidelity is essential for understanding and controlling plasmas. The present work, along with the accompanying work [10], provides a principled lower level on this hierarchy — low-dimensional and interpretable plasma models which can be used for physical discovery, forecasting, stability analysis, and real-time control. We have illustrated how Galerkin plasma models retain the global conservation laws of MHD, and how machine learning methods like SINDy can use these constraints directly in an optimization procedure for discovering such models from data. This principled enforcement of global conservation laws is critical for the success of future low-dimensional plasma models.

VI. ACKNOWLEDGEMENTS

The authors would like to extend their grateful thanks to Dr. Uri Shumlak for his input on this work. This work and the companion work were supported by the Army Research Office (ARO W911NF-19-1-0045) and the Air Force Office of Scientific Research (AFOSR FA9550-18-1-0200). Simulations were supported by the U.S. Department of Energy under award numbers DE-SC0016256 and DE-AR0001098 and used resources of the National Energy Research Scientific Computing Center, supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.
[1] J. Candy and R. E. Waltz, Anomalous transport scaling in the DIII-D tokamak matched by supercomputer simulation, Phys. Rev. Lett. 91, 045001 (2003).

[2] 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 (2012).

[3] D. Grosselj, A. Mallet, N. F. Loureiro, and F. Jenko, Fully kinetic simulation of 3D kinetic Alfvén turbulence, Phys. Rev. Lett. 120, 105101 (2018).

[4] K. Taira, S. L. Brunton, S. Dawson, C. W. Rowley, T. Colonius, B. J. McKeon, O. T. Schmidt, S. Gordeyev, V. Theofilis, and L. S. Ukeiley, Modal analysis of fluid flows: An overview, AIAA Journal 55, 4013 (2017).

[5] I. Furno, Fast transient transport phenomena measured by soft x-ray emission in TCV tokamak plasmas, Tech. rep. (2001).

[6] R. Jiménez-Gómez, E. Ascasibar, 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, 20 (2007).

[7] B. P. van Milligen, E. Sánchez, A. Alonso, M. A. Pedrosa, C. Hidalgo, A. M. de Aguilera, and A. L. Fraguas, The use of biorthogonal decomposition for the identification of zonal flows at TJ-II, Plasma Physics and Controlled Fusion 57, 052005 (2014).

[8] M. 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, Ph.D. thesis, Auburn University (2016).

[9] B. Victor, C. Akcay, C. Hansen, T. Jarboe, B. Nelson, and K. Morgan, Development of validation metrics using biorthogonal decomposition for the comparison of magnetic field measurements, Plasma Physics and Controlled Fusion 57, 052010 (2015).

[10] E. Strait, J. King, J. Hanson, and N. Logan, Spatial and temporal analysis of DIII-D 3D magnetic diagnostic data, Review of Scientific Instruments 87, 11D423 (2016).

[11] P. J. Byrne, Study of External Kink Modes in Shaped HBT-EP Plasmas, Ph.D. thesis, Columbia University (2017).

[12] S. Gu, B. Wan, Y. Sun, N. Chu, Y. Liu, T. Shi, H. Wang, M. Jia, and K. He, A new criterion for controlling edge localized modes based on a multi-mode plasma response, Nuclear Fusion 59, 126042 (2019).

[13] A. A. Kapitanoglu, K. D. Morgan, C. J. Hansen, and S. L. Brunton, Characterizing magnetized plasmas with dynamic mode decomposition, Physics of Plasmas 27, 052108 (2020).

[14] P. Benner, V. Mehrmann, and D. C. Sorensen, Dimension reduction of large-scale systems, Vol. 45 (Springer, 2005).

[15] P. Benner, M. Ohlberger, A. Cohen, and K. Willcox, Model reduction and approximation: theory and algorithms (SIAM, 2017).

[16] A. A. Kapitanoglu, K. D. Morgan, C. J. Hansen, and S. L. Brunton, Physics-constrained, low-dimensional models for minhd: First-principles and data-driven approaches, arXiv preprint arXiv:2004.10389 (2020).

[17] C. W. Rowley, T. Colonius, and R. M. Murray, Model reduction for compressible flows using POD and Galerkin projection, Physica D: Nonlinear Phenomena 189, 115 (2004).

[18] B. R. Noack, M. Schlegel, M. Morzynski, and G. Tadmor, Galerkin method for nonlinear dynamics (Springer, 2011).

[19] M. J. Balajewicz, E. H. Dowell, and B. 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).

[20] M. Schlegel and B. R. Noack, On long-term boundedness of Galerkin models, Journal of Fluid Mechanics 765, 325 (2015).

[21] K. Carlberg, M. Barone, and H. Antil, Galerkin v. least-squares Petrov–Galerkin projection in nonlinear model reduction, Journal of Computational Physics 330, 693 (2017).

[22] T. Dudok de Wit, A.-L. Pequet, J.-C. Valette, and R. Lima, The biorthogonal decomposition as a tool for investigating fluctuations in plasmas, Physics of Plasmas 1, 3288 (1994).

[23] J. Levesque, N. Rath, D. Shiraki, S. Angelini, J. Bialek, P. Byrne, B. DeBono, P. Hughes, M. Mauel, G. Navratil, et al., Multimode observations and 3D magnetic control of the boundary of a tokamak plasma, Nuclear Fusion 53, 073037 (2013).

[24] 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, 114012 (2014).

[25] B. P. van Milligen, E. Sánchez, A. Alonso, M. Pedrosa, C. Hidalgo, A. M. De Aguilera, and A. L. Fraguas, The use of the biorthogonal decomposition for the identification of zonal flows at TJ-II, Plasma Physics and Controlled Fusion 57, 052005 (2014).

[26] C. Hansen, B. Victor, K. Morgan, T. Jarboe, A. Hossock, G. Marklin, B. Nelson, and D. Sutherland, Numerical studies and metric development for validation of magnetohydrodynamic models on the HIT-SI experiment, Physics of Plasmas 22, 056105 (2015).

[27] D. D. Schnack, D. C. Barnes, D. P. Brennan, C. C. Hegna, E. Held, C. C. Kim, S. E. Kruger, A. Y. Pankin, and C. R. Sovinec, Computational modeling of fully ionized magnetized plasmas using the fluid approximation, Physics of Plasmas 13, 058103 (2006).

[28] Z. W. Ma and A. Bhattacharjee, Hall magnetohydrodynamic reconnection: The geospace environment modeling challenge, Journal of Geophysical Research: Space Physics 106, 3773.

[29] V. Krishan and S. M. Mahajan, Magnetic fluctuations and Hall magnetohydrodynamic turbulence in the solar wind, Journal of Geophysical Research: Space Physics 109.

[30] F. Ebrahimi, B. Lefebvre, C. B. Forest, and A. Bhattacharjee, Global Hall-MHD simulations of magnetorotational instability in a plasma Couette flow experiment, Physics of Plasmas 18, 062904 (2011).

[31] N. M. Ferraro, Calculations of two-fluid linear response to non-axisymmetric fields in tokamaks, Physics of Plasmas 19, 056105 (2012).
A. A. Kaptanoglu, T. E. Benedett, K. D. Morgan, C. J. Hansen, and T. R. Jarboe, Two-temperature effects in Hall-MHD simulations of the HIT-SI experiment, Physics of Plasmas 27, 072505 (2020).

S. L. Brunton and J. N. Kutz, Data-driven science and engineering: Machine learning, dynamical systems, and control (Cambridge University Press, 2019).

G. H. Golub et al., Matrix computations, The Johns Hopkins (1996).

A. Frieze, R. Kannan, and S. Vempala, Fast Monte-Carlo algorithms for finding low-rank approximations, Journal of the ACM (JACM) 51, 1025 (2004).

E. Liberty, F. Woolfe, P.-G. Martinsson, V. Rokhlin, and M. Tygert, Randomized algorithms for the low-rank approximation of matrices, Proceedings of the National Academy of Sciences 104, 20167 (2007).

F. Woolfe, E. Liberty, V. Rokhlin, and M. Tygert, A fast randomized algorithm for the approximation of matrices, Applied and Computational Harmonic Analysis 25, 335 (2008).

T. Jarboe, W. Hamp, G. Marklin, B. Nelson, R. O'Neill, A. Redd, P. Sieck, R. Smith, and J. Wrobel, Spheromak formation by steady inductive helicity injection, Physical review letters 97, 115003 (2006).

S. Gallier, Introduction to modern magnetohydrodynamics (Cambridge University Press, 2016).

J.-C. Loiseau, B. R. Noack, and S. L. Brunton, Sparse reduced-order modeling: sensor-based dynamics to full-state estimation, Journal of Fluid Mechanics 844, 459 (2018).

P. Benner, S. Gugercin, and K. Willcox, A survey of projection-based model reduction methods for parametric dynamical systems, SIAM review 57, 483 (2015).

S. Chaturantabut and D. 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 (IEEE, 2009) pp. 4316–4321.

Z. Drmac and S. Gugercin, A new selection operator for the discrete empirical interpolation method—improved a priori error bound and extensions, SIAM Journal on Scientific Computing 38, A631 (2016).

P. Astrid, S. Weiland, K. Willcox, and T. Backx, Missing point estimation in models described by proper orthogonal decomposition, IEEE Transactions on Automatic Control 53, 2237 (2008).

K. Willcox, Unsteady flow sensing and estimation via the gappy proper orthogonal decomposition, Computers & Fluids 35, 208 (2006).

K. Carlberg, C. Farhat, J. Cortial, and D. 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 (2013).

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, 3932 (2016).

P. J. Morrison and J. M. Greene, Noncanonical hamiltonian density formulation of hydrodynamics and ideal magnetohydrodynamics, Physical Review Letters 45, 790 (1980).

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, 335502 (2013).

H. M. Abdelhamid, Y. Kawazura, and Z. Yoshida, Hamiltonian formalism of extended magnetohydrodynamics, Journal of Physics A: Mathematical and Theoretical 48, 235502 (2015).

H. K. Chu and M. Hayashibe, Discovering interpretable dynamics by sparsity promotion on energy and the lagrangian, IEEE Robotics and Automation Letters 5, 2154 (2020).