

Orthogonal projection-based regularization for efficient model augmentation

Bendegúz M. Györök¹

Jan H. Hoekstra²

Johan Kon³

Tamás Péni^{1,4}

Maarten Schoukens²

Roland Tóth^{1,2}

GYOROKBENDE@SZTAKI.HUN-REN.HU

J.H.HOEKSTRA@TUE.NL

J.J.KON@TUE.NL

PENI@SZTAKI.HUN-REN.HU

M.SCHOUKENS@TUE.NL

R.TOTH@TUE.NL

¹*Systems and Control Lab, HUN-REN Institute for Computer Science and Control, Budapest, Hungary*

²*Control Systems Group, Eindhoven University of Technology, The Netherlands*

³*Control Systems Technology Group, Eindhoven University of Technology, The Netherlands*

⁴*National Laboratory for Health Security, Budapest, Hungary*

Editors: N. Ozay, L. Balzano, D. Panagou, A. Abate

Abstract

Deep-learning-based nonlinear system identification has shown the ability to produce reliable and highly accurate models in practice. However, these black-box models lack physical interpretability, and a considerable part of the learning effort is often spent on capturing already expected/known behavior of the system, that can be accurately described by first-principles laws of physics. A potential solution is to directly integrate such prior physical knowledge into the model structure, combining the strengths of physics-based modeling and deep-learning-based identification. The most common approach is to use an additive model augmentation structure, where the physics-based and the *machine-learning* (ML) components are connected in parallel, i.e., additively. However, such models are overparametrized, training them is challenging, potentially causing the physics-based part to lose interpretability. To overcome this challenge, this paper proposes an orthogonal projection-based regularization technique to enhance parameter learning and even model accuracy in learning-based augmentation of nonlinear baseline models.

Keywords: Nonlinear system identification, Model augmentation, Hybrid modeling, Physics-based learning

1. Introduction

Continuously rising performance requirements and the growing complexity of systems resulted in high demand for nonlinear models that can accurately describe complex behavior. For example, reliable path planning and motion control of autonomous vehicles require accurate dynamic models. *First-principle* (FP) models for these systems can usually be obtained from physical knowledge and practical engineering insight. While FP modeling approaches offer models that can describe the main dynamic behavior of systems on a wide operating range, they often provide approximative descriptions of complex dynamical aspects such as aerodynamic forces, tire dynamics, or friction characteristics, which can become dominant under high-performance operation of the system.

As an alternative approach, data-driven nonlinear modeling methods (identification) have been developed, capable of providing highly accurate and consistent models of the system behavior. In particular, recent approaches utilizing *state-space* (SS) models based on deep *artificial neural networks* (ANNs) have shown exceptional results (Masti and Bemporad, 2021). However, the use of such models in practical applications, such as trajectory planning and control raised some significant concerns, since they lack physical interpretation (Ljung, 2010). Furthermore, such ANN-based

black-box models typically extrapolate inaccurately from training data and are more prone to overfitting when trained on small datasets. Another drawback of black-box identification is that often a considerable time of the learning effort is spent on capturing already expected behavior due to first-principles-based understanding of some aspects of the system.

To address these challenges, *physics-informed neural networks* (PINNs) and *physics-guided neural networks* (PGNNs) were introduced in [Raissi et al. \(2019\)](#), and [Daw et al. \(2022\)](#), respectively. Both branches of model learning apply an additional term in the cost function, penalizing when the ANN predictions deviate from priori-selected physical laws during model training ([Karpatne et al., 2017](#)). A more traditional approach for combining physics-based knowledge and data-driven identification techniques is (light) grey-box modeling. A wide range of methods has been developed and applied successfully in practice for years ([Bohlin, 2006](#)), mainly for *linear time invariant* (LTI) system representations. Light grey-box models utilize physics-based system descriptions and data-driven parameter estimation ([Schoukens and Ljung, 2019](#)).

Model augmentation, i.e., hybrid modeling, is a promising direction of utilizing physical knowledge in nonlinear system identification ([Schön et al., 2022](#)). By combining FP models with flexible learning components, (i) faster convergence and (ii) better accuracy can be achieved compared to black-box learning methods while producing (iii) physically interpretable models ([Djeumou et al., 2022](#)). Furthermore, existing approaches can be incorporated into the model augmentation framework, e.g., physics-based loss functions can be utilized during model optimization ([Daw et al., 2022](#)) and physical parameters can be co-estimated with the parameters of the learning component ([Psichogios and Ungar, 1992](#)).

This paper investigates learning for hybrid model augmentation when the model is composed of an FP model with an additive neural network part, and the objective is to jointly estimate the ANN parameters and the physical parameters based on measured data from the system. This formulation can produce models with properties (i-iii), however simultaneous tuning of the learning-based and physical parameters causes overparametrization and results in "competing" submodels. Instead of only compensating for the unmodeled terms, the ML part can learn some of the known dynamics of the FP model as well. This can generate unrealistic parameters in the FP model, resulting in less interpretable models and even hurting extrapolation properties. Recent studies have addressed similar challenges in the context of PGNN-based feedforward control methods ([Bolderman et al., 2022, 2024](#)), yet, this issue remains unsolved for nonlinear system identification. Orthogonalization is an attractive approach that we now generalize for nonlinear model learning, inspired by [Kon et al. \(2022\)](#). With this modification, the learning component is penalized for learning the known dynamics of the FP model, by forcing a particular form of orthogonality between the ML- and physics-based layers. The original approach is limited to FP models that are linear in their parameters; therefore, a generalized version of the method is also presented in this paper.

The main contributions of this work can be summarized as:

- C1 Generalization of an orthogonal projection-based regularization method ([Kon et al., 2022](#)) for FP models that are nonlinear in the parameters, and integrating the approach to learning-based model augmentation.
- C2 Efficient initialization and normalization schemes for the additive model augmentation.

The remainder of the paper is organized as follows: Sect. 2 introduces the identification problem and the additive model augmentation structure. The applied cost function, normalization, and initialization schemes are also discussed. Then, Sect. 3 gives the orthogonal projection-based regularization term for FP models that are linear in the parameters, which is generalized in Sect. 4 to FP

models that can be nonlinear in the parameters. Sect. 5 provides a simulation study to demonstrate the effectiveness of the methodology. Finally, in Sect. 6, the conclusions on the presented work are drawn.

2. Additive model augmentation

2.1. Problem formulation

The dynamics of the data-generating system are considered to be defined by a *discrete-time* (DT) nonlinear SS representation:

$$x_{k+1} = f(x_k, u_k), \quad (1a)$$

$$y_k = h(x_k, u_k) + e_k, \quad (1b)$$

where $k \in \mathbb{Z}$ is the discrete time index, $x_k \in \mathbb{R}^{n_x}$ is the state, $u_k \in \mathbb{R}^{n_u}$ is the control input, $y_k \in \mathbb{R}^{n_y}$ is the measured output, $f : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \rightarrow \mathbb{R}^{n_x}$ is the DT state transition function, $h : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \rightarrow \mathbb{R}^{n_y}$ is the output map, and $e_k \in \mathbb{R}^{n_y}$ is an i.i.d. white noise process with finite variance, representing measurement noise.

The exact dynamics of (1) are not known, but we assume that based on prior knowledge, a physics-based approximative model (baseline model) is available in the form of

$$\hat{x}_{k+1} = f_\theta(\hat{x}_k, u_k), \quad (2a)$$

$$\hat{y}_k = h_\theta(\hat{x}_k, u_k), \quad (2b)$$

where $\hat{x}_k \in \mathbb{R}^{n_x}$ is the model state, $\hat{y}_k \in \mathbb{R}^{n_y}$ is the model output, $f_\theta : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \rightarrow \mathbb{R}^{n_x}$ is the FP state transition function, and $h_\theta : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \rightarrow \mathbb{R}^{n_y}$ is the output function, that both depend on physical parameters $\theta \in \mathbb{R}^{n_\theta}$. Nominal parameter values θ_0 can often be obtained directly from manufacturer datasheets, as they typically provide specifications and coefficient values for many subcomponents. In cases where such information is unavailable, prior coefficient estimation may be required to determine suitable nominal values.

As discussed in Sect. 1, even by optimizing the parameters of FP models, they can often only provide an approximative representation of the true underlying dynamics of the system, hence we can use deep-learning-based model augmentation to enhance their modeling capabilities. One of the most straightforward ways to augment the baseline first-principle dynamics is to introduce an additive term into the state transition (Sun et al., 2020; Sohlberg and Jacobsen, 2008). The FP output function h_θ is assumed to contain simple relations (it can even be an identity map), thus, the aim is to identify the system dynamics in the form of

$$\hat{x}_{k+1} = f_\theta(\hat{x}_k, u_k) + f_\eta^{\text{ANN}}(\hat{x}_k, u_k), \quad (3a)$$

$$\hat{y}_k = h_\theta(\hat{x}_k, u_k), \quad (3b)$$

where f_η^{ANN} here represents a fully connected, feedforward neural network with $\eta \in \mathbb{R}^{n_\eta}$ collecting its parameters, but can be replaced by any function approximator without loss of generality.

Although nominal values are typically available, still certain physical parameters are often only approximately known. Hence, to get the best prediction performance, the first-principle model parameters are estimated simultaneously with the ANN parameters. The goal is to find the underlying physical parameters θ_* that resemble the actual system properties as closely as possible while finding η_* parameter values that maximize the data fit.

2.2. Model training

To achieve co-estimation efficiently, the T -step ahead prediction cost on multiple overlapping subsections of the training data, used in the SUBNET method (Beintema et al., 2023), is chosen to estimate the augmented model. In this way, model estimation using data sequence $\mathcal{D}_N = \{(y_i, u_i)\}_{i=1}^N$ acquired from (1), corresponds to an optimization problem with objective function $V_{\mathcal{D}_N}^{(\text{sec})}$, as

$$V_{\mathcal{D}_N}^{(\text{sec})}(\eta, \theta) = \frac{1}{N_{\text{sec}}} \sum_{k \in \mathcal{I}_{N_{\text{sec}}}} \frac{1}{T} \sum_{l=0}^{T-1} \|y_{k+l} - \hat{y}_{k+l|k}\|_2^2, \quad (4a)$$

$$\hat{x}_{k+l+1|k} = f_{\theta}(\hat{x}_{k+l|k}, u_{k+l}) + f_{\eta}^{\text{ANN}}(\hat{x}_{k+l|k}, u_{k+l}), \quad (4b)$$

$$\hat{y}_{k+l|k} = h_{\theta}(\hat{x}_{k+l|k}, u_{k+l}) \quad (4c)$$

$$\hat{x}_{k|k} = x_k, \quad (4d)$$

where the starting times of the sections collected in $\mathcal{I}_{N_{\text{sec}}}$ are randomly selected from $\{1, \dots, N - T + 1\}$, $N_{\text{sec}} = |\mathcal{I}_{N_{\text{sec}}}|$. Here, $\hat{x}_{k+l|k}$ denotes the simulated state at $k + l$, starting from the measured initial state $\hat{x}_{k|k} = x_k$ at time k . Training with the T -step ahead prediction cost brings two main benefits compared to traditional simulation error-based training: (i) forward simulation on the subsections can be calculated in parallel, therefore requiring less computational load, and allowing the use of batch optimization algorithms; (ii) using overlapping sections increases cost function smoothness (Ribeiro et al., 2020), which further enhances the efficiency of (stochastic) gradient-based optimization methods.

Cost-function (4) uses the initial state of each subsection for calculating the loss value, as visible in (4d). When full-state measurements are not applicable, i.e, when $y_k \neq x_k + e_k$, an encoder network is used to estimate the initial states from past input and output values, similarly as in Beintema et al. (2023); Hoekstra et al. (2024).

2.3. Normalization and initialization scheme

To efficiently train artificial neural networks, *input-output* (IO) normalization is essential to avoid exploding and vanishing gradients, which would potentially cause worse estimation results (LeCun et al., 1998). A transformation for data normalization is defined according to Schoukens and Tóth (2020):

$$T_u = \text{diag}(1/\sigma_{u,i}), \quad (5)$$

$$T_x = \text{diag}(1/\sigma_{x,i}), \quad (6)$$

where σ notes the standard deviation, and the transformation matrix $T_u \in \mathbb{R}^{n_u \times n_u}$ is a diagonal matrix containing the inverse of the standard deviation of u_k for each channel i . The transformation matrix T_x can be defined similarly. The standard deviations are calculated based on the training data, or, in case when state measurements are not available, transformation matrices are obtained by forward simulating the FP model on the training data.

Based on (5) and (6), the state transition of the augmented model can be expressed as

$$\hat{x}_{k+1} = f_{\theta}(\hat{x}_k, u_k) + T_x^{-1} f_{\eta}^{\text{ANN}}(T_x \hat{x}_k, T_u u_k). \quad (7)$$

Besides IO normalization, a reliable parameter initialization scheme is also important for efficient model training to improve accuracy and potentially decrease convergence time. A similar

initialization method is applied as in [Ramkannan et al. \(2023\)](#). This guarantees that the augmented state transition function behaves like f_{θ_0} at the beginning of training. A fully random initialization can increase convergence time and even cause stability loss at the beginning of the optimization. To prevent this, the weight and bias values of the last layer in the ANN¹ are initialized as zero matrices, while the rest of the parameters are initialized randomly, e.g., by the Xavier approach ([Glorot and Bengio, 2010](#)).

3. Orthogonalization-based regularization

Assuming that system (1) is part of the model set spanned by the parametrization of (3), there should exist a choice of parameter pairs (θ, η) for the additive model augmentation structure that realizes exactly the dynamical relations of the data-generating system. However, since f_{η}^{ANN} and f_{θ} are connected in parallel, i.e., additively, the optimal θ_* , η_* parameters that minimize (4) are not unique, i.e., different θ , η parameters can result in the same input-state behavior. This is generally called non-identifiability. For ease of understanding, Example 1 illustrates this non-uniqueness.

Example 1 Consider an FP model with a state-transition function of $f_{\theta}(\hat{x}, u) = \theta \hat{x}$, and f_{η}^{ANN} as an ANN. By the universal approximation properties, there exist choices of weights and biases such that $f_{\eta}^{\text{ANN}}(\hat{x}, u) \approx W_{L+1} \hat{x}$. Consequently, for a true state-transition map $f(x, u) = \theta_* x$, any (θ, W_{L+1}) pair that satisfies $W_{L+1} + \theta = \theta_*$ is a global minimizer of (4).

A straightforward approach would be to define a $\Theta_* \subset \mathbb{R}^{n_{\theta} + n_{\eta}}$ set that collects all parameter pairs that achieve an equivalent system representation with the data-generating system. Then, during model training, the aim would be to find any $(\theta_*, \eta_*) \in \Theta_*$. However, if the parameterization is non-identifiable, certain gradient directions become indistinguishable from one another, increasing the risk of convergence to local minima. While stochastic gradient descent-based optimizers inherently provide some degree of implicit regularization that helps to mitigate this issue ([Zhang et al., 2017](#)), another critical concern arises: the physical parameters may converge to unrealistic values due to the non-uniqueness of θ_* . As a result, the estimated FP model cannot reliably serve as a standalone physical model of the system. Additionally, physics-based models are often capable of extrapolating accurately beyond the range of observed data, while ANNs extrapolate poorly outside of the training data set. Hence, the augmented model can lose this advantageous extrapolation capability due to the indistinguishability of the model contributions in the cost during training.

We aim to address the above-mentioned problems by adding a new term to the cost function based on [Kon et al. \(2022\)](#) that forces f_{η}^{ANN} to prioritize the learning of the unknown dynamics. It is achieved by regularizing the output of f_{η}^{ANN} in the subspace where f_{θ} can generate outputs. The original approach has been derived for feedforward control, but it can be easily modified for system identification problems if the FP model can be represented as being *linear in the parameters*:

$$f_{\theta}(\hat{x}_k, u_k) = \phi(\hat{x}_k, u_k) \theta. \quad (8)$$

By taking a further assumption of full state-measurement, i.e., $y_k = x_k + e_k$, and fixing the learning component to zero, due to the linear parameterization, fitting of the FP model parameters

¹Consider a simple feedforward neural network as $f_{\eta}^{\text{ANN}}(\hat{x}_k, u_k) = W_{L+1} \sigma(W_L \dots \sigma(W_0 [\hat{x}_k^{\top} u_k^{\top}]^{\top} + b_0) + b_L) + b_{L+1}$, where $\sigma(\cdot)$ is an element-wise (nonlinear) activation function, W_i represents the weight values, b_i notes the biases, and L notes the number of hidden layers.

becomes a *linear regression* (LR) problem:

$$\underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}}_{X^+} = \underbrace{\begin{bmatrix} \phi(x_0, u_0) \\ \phi(x_1, u_1) \\ \vdots \\ \phi(x_{N-1}, u_{N-1}) \end{bmatrix}}_{\Phi(X,U)} \theta + \underbrace{\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}}_E, \quad (9)$$

where $X = [x_0^\top \ x_1^\top \ \cdots \ x_{N-1}^\top]^\top \in \mathbb{R}^{Nn_x}$, and U is defined similarly. The E matrix collects the residuals of the FP state transition. Given dataset \mathcal{D}_N , an explicit base for the output space of f_θ can be calculated by taking the reduced *singular value decomposition* (SVD) of $\Phi(X, U) \in \mathbb{R}^{Nn_x \times n_\theta}$, as

$$\Phi(X, U) = Q_{X,U} \Sigma_{X,U} V_{X,U}^\top, \quad (10)$$

in which we assume $Nn_x > n_\theta$ such that $Q_{X,U} \in \mathbb{R}^{Nn_x \times n_\theta}$, $\Sigma_{X,U} \in \mathbb{R}^{n_\theta \times n_\theta}$, and $V_{X,U} \in \mathbb{R}^{n_\theta \times n_\theta}$. Thus, (9) can be reformulated as

$$X^+ = Q_{X,U} \Sigma_{X,U} V_{X,U}^\top \theta + E. \quad (11)$$

The columns of $Q_{X,U}$ form a basis for the output space of f_θ for the particular state and input values in X, U . The projection matrix onto the subspace spanned by these basis vectors can be expressed as

$$\Pi_{X,U} = Q_{X,U} Q_{X,U}^\top. \quad (12)$$

Thus, the component of f_η^{ANN} that lies in the subspace spanned by $Q_{X,U}$ can be given by $\Pi_{X,U} f_\eta^{\text{ANN}}(X, U)$, where $f_\eta^{\text{ANN}}(X, U) = [f_\eta^{\text{ANN}}(x_0, u_0)^\top \cdots f_\eta^{\text{ANN}}(x_{N-1}, u_{N-1})^\top]^\top \in \mathbb{R}^{Nn_x}$. The aim is to penalize such η parameters that result in significant contributions in this subspace, by adding $\|\Pi_{X,U} f_\eta^{\text{ANN}}(X, U)\|_2^2$ to (4a), as an orthogonality-promoting term. Thus, it can be interpreted as a targeted ℓ_2 regularization of the ANN that only penalizes the directions of η which generate output in the subspace of the model f_θ . The modified cost function is given as

$$V_{\mathcal{D}_N}^{(\text{orth})}(\eta, \theta) = V_{\mathcal{D}_N}^{(\text{sec})}(\eta, \theta) + \beta \|\Pi_{X,U} f_\eta^{\text{ANN}}(X, U)\|_2^2, \quad (13)$$

where $\beta \in \mathbb{R}_{\geq 0}$ is the orthogonalization coefficient. The value of β can be interpreted as a trade-off between orthogonality and performance. With large β values, f_η^{ANN} is restricted too much, and the optimizer focuses more on promoting orthogonality between the FP model and the ANN part, rather than producing an accurate model. On the other hand, too small β values can lead to a near unregularized situation by diminishing the penalization of subspace contributions.

Remark 1 *In practice, often only a small amount of orthogonal regularization is needed to achieve the desired complementarity. A wide range of β values can be chosen, usually spanning several orders of magnitude. Later we will illustrate this in our example (in Figure 3). A similar observation has been made in Kon et al. (2023).*

Since $Nn_x \gg n_\theta$, matrix multiplication with $\Pi_{X,U} \in \mathbb{R}^{Nn_x \times Nn_x}$ is computationally demanding in $\Pi_{X,U} f_\eta^{\text{ANN}}(X, U)$. However, since $\Pi_{X,U} = Q_{X,U} Q_{X,U}^\top$, and $Q_{X,U}$ is orthonormal, it holds that

$$\|\Pi_{X,U} f_\eta^{\text{ANN}}(X, U)\|_2^2 = \|Q_{X,U}^\top f_\eta^{\text{ANN}}(X, U)\|_2^2. \quad (14)$$

Calculating (14) is less intensive, as it is composed of n_θ numbers of dot product calculation between f_η^{ANN} and vectors of length Nn_x . Another advantage of the decomposition shown in (11)

is that the projection matrix $\Pi_{X,U}$ is independent of the physical parameters since Φ depends only on the state and input values. This practically means that the SVD can be calculated before the optimization process begins, thus lowering the computational demand of the method. When full-state measurements are not available, an approximate \hat{X} data set can be constructed by forward simulating the FP model on the training data, and using \hat{X} to compute $\Pi_{\hat{X},U}$.

Remark 2 The current model estimates of \hat{X} can be also used to update the projection matrix at each iteration step. This modification increases computational demand, as the SVD needs to be recomputed at the start of each epoch. However, the benefit of it is that the regularization term is applied for the accurate states, potentially leading to faster convergence and better model accuracy.

4. Generalization for nonlinear first-principle models

A serious limitation of the orthogonalization-based method derived in Sect. 3 is that it can only be applied to FP models that are *linear-in-the-parameters*. To extend the method to the nonlinear case, the Taylor series expansion of the FP state transition function is taken w.r.t. the physical parameters:

$$f_\theta(\hat{x}_k, u_k) \approx f_\theta(\hat{x}_k, u_k)|_{\theta=\bar{\theta}} + \underbrace{\frac{\partial f_\theta(\hat{x}_k, u_k)}{\partial \theta}}_{\Phi_{\bar{\theta}}(\hat{x}_k, u_k)}|_{\theta=\bar{\theta}} (\theta - \bar{\theta}), \quad (15)$$

where $\bar{\theta}$ is the linearization point, and $\partial f_\theta(\hat{x}_k, u_k)/\partial \theta \in \mathbb{R}^{n_x \times n_\theta}$ is the Jacobian matrix of $f_\theta(\hat{x}_k, u_k)$ with respect to the θ parameters. Next, expanding (15), we arrive to the expression of

$$f_\theta(\hat{x}_k, u_k) \approx \Phi_{\bar{\theta}}(\hat{x}_k, u_k)\theta + f_{\bar{\theta}}(\hat{x}_k, u_k) - \Phi_{\bar{\theta}}(\hat{x}_k, u_k)\bar{\theta}, \quad (16)$$

where the first term is linear in θ , while the second and third terms are a θ -independent offset given by the linearization point $\bar{\theta}$. Similar to (9), by vectorizing the approximate FP model responses,

$$\underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}}_{X^+} = \underbrace{\begin{bmatrix} \Phi_{\bar{\theta}}(x_0, u_0) \\ \Phi_{\bar{\theta}}(x_1, u_1) \\ \vdots \\ \Phi_{\bar{\theta}}(x_{N-1}, u_{N-1}) \end{bmatrix}}_{\Phi_{\bar{\theta}}(X,U)} \theta + \underbrace{\begin{bmatrix} f_{\bar{\theta}}(x_0, u_0) - \Phi_{\bar{\theta}}(x_0, u_0)\bar{\theta} \\ f_{\bar{\theta}}(x_1, u_1) - \Phi_{\bar{\theta}}(x_1, u_1)\bar{\theta} \\ \vdots \\ f_{\bar{\theta}}(x_{N-1}, u_{N-1}) - \Phi_{\bar{\theta}}(x_{N-1}, u_{N-1})\bar{\theta} \end{bmatrix}}_{\Gamma_{\bar{\theta}}(X,U)} + \underbrace{\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}}_E. \quad (17)$$

To consider all components of the baseline model in promoting orthogonality, and to enable adjustments to the linearization point during model learning, an extended parameter vector is introduced as $\tilde{\theta} \in \mathbb{R}^{n_\theta+1}$. Then, the two terms in (17) can be combined:

$$X^+ = \underbrace{\begin{bmatrix} \Phi_{\bar{\theta}}(X, U) & \Gamma_{\bar{\theta}}(X, U) \end{bmatrix}}_{\tilde{\Phi}_{\bar{\theta}}(X,U)} \underbrace{\begin{bmatrix} \theta \\ 1 \end{bmatrix}}_{\tilde{\theta}} + E. \quad (18)$$

It is evident that (18) has a structure similar to (9), hence the same orthogonalization-based penalization term can be derived for it, but instead of taking the reduced SVD of $\Phi(X, U)$, now it should be calculated for $\tilde{\Phi}_{\bar{\theta}}(X, U)$. The projection matrix $\tilde{\Pi}_{X,U}$ can then be calculated as in (12). Similarly to (13), the orthogonality-promoting term is added to the cost function, as

$$V_{\mathcal{D}_N}^{(\text{orth})}(\eta, \theta) = V_{\mathcal{D}_N}^{(\text{sec})}(\eta, \theta) + \beta \left\| \tilde{\Pi}_{X,U} f_\eta^{\text{ANN}}(X, U) \right\|_2^2. \quad (19)$$

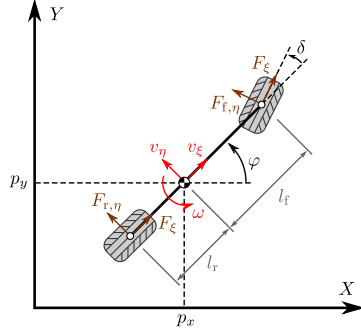


Figure 1: The single track model.

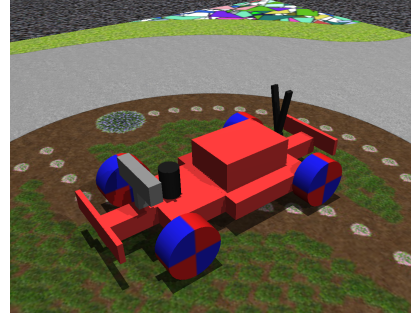


Figure 2: Digital twin of the F1Tenth vehicle.

The linearization point $\bar{\theta}$ can be updated at each iteration using the current estimate of θ , but this requires recomputing the SVD and projection matrix every time the cost function is evaluated, increasing computational demand. Alternatively, $\bar{\theta}$ can be approximated by the nominal parameter values θ_0 , allowing the projection matrix to be precomputed at the start of training, which significantly reduces the computational cost of the regularization method. As θ_0 is derived from physical principles and typically aligns well with the actual system properties, the approximation remains valid and the basis for the response of the physical model is accurately maintained.

5. Identification study

To demonstrate the advantages of the proposed orthogonal projection-based regularization for learning-based model augmentation, we aim to identify the dynamics of a small-scale electric vehicle (F1Tenth (Agnihotri et al., 2020)), by augmenting with an ANN component a simplified mechanical model that describes the main characteristics of the vehicle².

5.1. First-principle model of the F1Tenth vehicle

To develop an approximate baseline model of the F1Tenth platform, the so-called single-track model has been used (Paden et al., 2016). The model is illustrated in Figure 1, and expressed as

$$\dot{p}_x = v_\xi \cos \varphi - v_\eta \sin \varphi, \quad \dot{v}_\xi = \frac{1}{m} (F_\xi + F_\xi \cos \delta - F_{f,\eta} \sin \delta + m v_\eta \omega), \quad (20a)$$

$$\dot{p}_y = v_\xi \sin \varphi + v_\eta \cos \varphi, \quad \dot{v}_\eta = \frac{1}{m} (F_{r,\eta} + F_\xi \sin \delta - F_{f,\eta} \cos \delta - m v_\xi \omega), \quad (20b)$$

$$\dot{\varphi} = \omega, \quad \dot{\omega} = \frac{1}{J_z} (F_{f,\eta} l_f \cos \delta + F_\xi l_f \sin \delta - F_\xi l_r), \quad (20c)$$

where (p_x, p_y) is the position of the *center-of-gravity* (CoG) in the (X, Y) plane, φ is the orientation of the vehicle, which is measured from the X axis. The variables v_ξ and v_η denote the longitudinal and lateral velocities of the vehicle, respectively, while ω is the yaw rate. Furthermore, l_r and l_f are the distances of the rear and front axis from the CoG, δ is the steering angle, m is the mass of the vehicle, and J_z is the inertia along the vertical axis.

The longitudinal tire force component F_ξ is expressed with an empirical drivetrain model. For more details, refer to Floch et al. (2024). Moreover, the linearized Magic Formula (Pacejka, 2012) is utilized to model the lateral tire forces. Thus, the tire force components are expressed as

$$F_\xi = C_{m1}d - C_{m2}v_\xi - \text{sign}(v_\xi) C_{m3}, \quad F_{r,\eta} = C_r \frac{-v_\eta + l_r \omega}{v_\xi}, \quad F_{f,\eta} = C_f \left(\delta - \frac{v_\eta + l_f \omega}{v_\xi} \right), \quad (21)$$

²Code, data available: <https://github.com/AIMotionLab-SZTAKI/orthogonal-augmentation>

where C_{m1} , C_{m2} , and C_{m3} are the drivetrain constants, d is the motor PWM percentage, $F_{r,\eta}$ and $F_{f,\eta}$ are the rear and front lateral tire forces, while C_r and C_f are the rear and front cornering stiffness values, respectively. The applied tire models (particularly the empirical drivetrain model) are highly approximative and are the primary sources of inaccuracy in the FP model.

The state of the baseline model is $x = [p_x \ p_y \ \varphi \ v_\xi \ v_\eta \ \omega]^\top$. The control inputs are δ and d . As discussed in Sect. 2, we have assumed that the FP model is given in DT form, thus (20) is discretized with the forward Euler scheme. The parameters of the baseline model are

$$\theta = [m \ J_z \ l_r \ l_f \ C_{m1} \ C_{m2} \ C_{m3} \ C_r \ C_f]^\top, \quad (22)$$

and their nominal values have been determined in Floch (2022). However, to achieve accurate representation of the true dynamics, θ in (22) is tuned jointly with the ANN parameters.

As one might see, in (20), the differential equations for the position and orientation values contain simple kinematic relations and can be separated from the rest of the equations. As proposed by Szécsi et al. (2024), to simplify the neural network structure, we only augment the velocity states of the system, then p_x , p_y , and φ can be determined by numerical integration.

5.2. Data acquisition

To generate data for training, a high-fidelity multi-body simulator, the MuJoCo physics engine is utilized (Todorov et al., 2012). In this simulator, the digital twin model of the car was assembled using parameters that were identified based on measurements with the real car. Figure 2 shows the digital twin model. Multiple experiments were performed in the simulator to acquire sufficient data. The MuJoCo model operates with motion states in terms of joint positions and velocities, utilizing complex contact models and friction characteristics. The logged signals have been determined based on the states of the FP model. All states are directly measured, as it would be for the real F1Tenth. Two different trajectories (a lemniscate and a circular path), with 12 different velocity references have been utilized to achieve robust identification results. With a sampling frequency of $f_s = 40$ Hz, altogether $N = 15985$ data points have been recorded. Half of the measurements were separated for model training, while the other half were used for testing. Trajectories with alternating reference velocities have been included in both. Lastly, 20% of each training trajectory have been randomly selected to form a validation data set, achieving an 80%-20% ratio of the training and validation data points. The gathered data does not contain any noise, but to test the methodology with noise levels that are typical for this application type, i.i.d. white Gaussian noise is added to the output signals in the training and validation data sets to reach *Signal-to-noise ratio*³ (SNR) values of 30 dB and 25 dB. To ensure a clear comparison of the resulting models, the test data is kept noise-free.

5.3. Model training and hyperparameter selection

All models have been trained with the Adam optimizer (Kingma and Ba, 2015), applying a learning rate of 10^{-3} and a batch size of 256. For the truncation length, $T = 15$ has been used, corresponding to roughly 1.5 times the largest characteristic time scale of the data-generating system. The hyperparameters of f_η^{ANN} have been determined empirically: 2 hidden layers with 64 nodes per layer have been applied, using the *hyperbolic tangent* (tanh) activation function. The regularization coefficient β was also tuned empirically, as illustrated in Figure 3, with $\beta = 10^{-7}$ selected for the noiseless case. For scenarios with output noise, β was chosen using a similar approach. Based on our experience, values of β in the range $[10^{-7}, 10^{-5}]$ yielded the best results for the considered modeling task.

³ $\text{SNR}_{\text{dB}} = 10 \log_{10} (P_s/P_n)$, where P_s and P_n are the sample mean signal power and noise power, respectively.

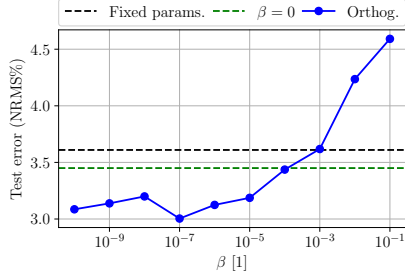


Figure 3: Illustrating the empirical process of tuning β (with no noise).

Model	No noise	30 dB	25 dB
Initial baseline model with θ_0	37.91%	37.91%	37.91%
Baseline model with $\hat{\theta}$ (no reg.)	118.01%	80.49%	68.18%
Baseline model with $\hat{\theta}$ (orth. reg.)	36.07%	36.64%	30.69%
Augm. model (fixed baseline θ_0)	3.61%	4.18%	5.11%
Augm. model (co-estim., no reg.)	3.45%	4.17%	4.92%
Augm. model (co-estim., orth. reg.)	3.00%	3.84%	4.76%
Black-box (Beintema et al., 2023)	2.12%	2.31%	3.06%

Table 1: Test NRMS errors with various models. SNR values refer to the noise level in the training and validation sets.

5.4. Results

Model performance is evaluated using the *Normalized Root Mean Square* (NRMS) error, see, e.g., Beintema et al. (2021), with results summarized in Table 1. As shown, augmenting the baseline model while keeping the physical parameters fixed leads to a significant improvement in accuracy compared to the standalone FP model. Jointly optimizing the physical and ANN parameters has further reduced the test error, while using the orthogonalization-based cost function led to additional improvements in model accuracy across all noise conditions. With increasing noise levels, the effect of orthogonal regularization on model accuracy is less noticeable compared to the noiseless scenario, but this is expected, as it is harder to separate the baseline model dynamics and the noise.

When the physical and learning-based parameters are co-estimated, we acquire a $(\hat{\theta}, \hat{\eta})$ parameter pair as a result of the optimization. This allows for evaluating the FP model separately using the estimated physical parameters $\hat{\theta}$, providing insight into the interpretability of the augmented model. Table 1 also includes these test results with the FP model, using $\hat{\theta}$. Without regularization, the physics-based component of the augmentation structure shows a notable drop in accuracy compared to the nominal baseline model, supporting our earlier statement that, in the absence of the proposed regularization, physical parameters may be tuned to unrealistic values, potentially undermining the interpretability of the augmented model. In contrast, applying orthogonal regularization not only enhances the overall accuracy of the augmented model but also improves the performance of the physics-based part relative to the nominal FP model. This demonstrates that the proposed method effectively promotes the desired complementarity between the baseline and learning components.

Comparing the results to a state-of-the-art black-box method, it is visible that even with the orthogonal projection-based approach, black-box models have performed slightly better. However, as we have demonstrated, model augmentation generates physically interpretable models in contrast to black-box approaches. Consequently, a slight reduction in model accuracy is an acceptable trade-off for the added interpretability.

6. Conclusion

This paper introduced an orthogonal projection-based regularization method that extends the approach in Kon et al. (2022), originally limited to linear-in-the-parameter FP models, to support models that are nonlinear in the parameters and improve the efficiency of estimating additive model augmentation structures. The proposed method improved model performance across various output noise levels. Although the impact of online recalculation of the projection matrix has not been evaluated, its potential benefits, despite the increased computational cost, will be explored in future studies. Further theoretical analysis and application in other model augmentation settings will be addressed in future research.

Acknowledgments

This project has received funding from the European Defence Fund programme under grant agreement number No 101103386 and has also been supported by the Air Force Office of Scientific Research under award number FA8655-23-1-7061 and by the EKÖP-24-2-BME-123 University Research Scholarship Programme of the Ministry for Culture and Innovation from the source of the National Research, Development, and Innovation Fund. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the granting authority can be held responsible for them. The study was also partly funded by the National Research, Development and Innovation Office in Hungary (RRF-2.3.1-21-2022-00006).

References

- Abhijeet Agnihotri, Matthew O’Kelly, Rahul Mangharam, and Houssam Abbas. Teaching Autonomous Systems at 1/10th-scale: Design of the F1/10 Racecar, Simulators and Curriculum. In *Proc. of the ACM Technical Symposium on Computer Science Education*, pages 657–663, 2020.
- Gerben Beintema, Roland Tóth, and Maarten Schoukens. Nonlinear state-space identification using deep encoder networks. In *Proc. of the 3rd Conference on Learning for Dynamics and Control*, volume 144 of *Proceedings of Machine Learning Research*, pages 241–250, 2021.
- Gerben I. Beintema, Maarten Schoukens, and Roland Tóth. Deep subspace encoders for nonlinear system identification. *Automatica*, 156:111210, 2023.
- Torsten Bohlin. *Practical Grey-box Process Identification*. Advances in Industrial Control. Springer London, 1st edition, 2006.
- Max Bolderman, Mircea Lazar, and Hans Butler. On feedforward control using physics–guided neural networks: Training cost regularization and optimized initialization. In *Proc. of the European Control Conference*, pages 1403–1408, 2022.
- Max Bolderman, Hans Butler, Sjirk Koekebakker, Eelco Van Horssen, Ramidin Kamidi, Theresa Spaan-Burke, Nard Strijbosch, and Mircea Lazar. Physics-guided neural networks for feedforward control with input-to-state-stability guarantees. *Control Engineering Practice*, 145:105851, 2024.
- Arka Daw, Anuj Karpatne, William D. Watkins, Jordan S. Read, and Vipin Kumar. Physics-Guided Neural Networks (PGNN): An Application in Lake Temperature Modeling. In *Knowledge Guided Machine Learning*. Chapman and Hall/CRC, 2022.
- Franck Djeumou, Cyrus Neary, Eric Goubault, Sylvie Putot, and Ufuk Topcu. Neural networks with physics-informed architectures and constraints for dynamical systems modeling. In *Proc. of the 4th Annual Learning for Dynamics and Control Conference*, volume 168 of *Proceedings of Machine Learning Research*, pages 263–277, 2022.
- Kristóf Floch. *Model-based motion control of the FITENTH autonomous electrical vehicle*. Bachelor’s thesis, Budapest University of Technology and Economics, 2022. URL <https://eprints.sztaki.hu/10544/>.

- Kristóf Floch, Tamás Péni, and Roland Tóth. Gaussian-Process-Based Adaptive Trajectory Tracking Control for Autonomous Ground Vehicles. In *Proc. of the European Control Conference*, pages 464–471, 2024.
- Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proc. of the International Conference on Artificial Intelligence and Statistics*, pages 249–256, 2010.
- Jan H. Hoekstra, Chris Verhoek, Roland Tóth, and Maarten Schoukens. Learning-based model augmentation with LFRs. *arXiv preprint arXiv:2404.01901*, 2024.
- Anuj Karpatne, Gowtham Atluri, James H. Faghmous, Michael Steinbach, Arindam Banerjee, Auroop Ganguly, Shashi Shekhar, Nagiza Samatova, and Vipin Kumar. Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data. *IEEE Transactions on Knowledge and Data Engineering*, 29(10):2318–2331, 2017.
- Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. In *Proc. of the International Conference on Learning Representations*, pages 1–15, 2015.
- Johan Kon, Dennis Bruijnen, Jeroen Van De Wijdeven, Marcel Heertjes, and Tom Oomen. Physics-Guided Neural Networks for Feedforward Control: An Orthogonal Projection-Based Approach. In *Proc. of the American Control Conference*, pages 4377–4382, 2022.
- Johan Kon, Naomi de Vos, Dennis Bruijnen, Jeroen van de Wijdeven, Marcel Heertjes, and Tom Oomen. Learning for Precision Motion of an Interventional X-ray System: Add-on Physics-Guided Neural Network Feedforward Control. In *Proc. of the 22nd IFAC World Congress*, pages 7523–7528, 2023.
- Yann LeCun, Leon Bottou, Genevieve B. Orr, and Klaus Robert Müller. Efficient BackProp. In *Neural Networks: Tricks of the Trade*, pages 9–50. Springer, Berlin, Heidelberg, 1998.
- Lennart Ljung. Perspectives on system identification. *Annual Reviews in Control*, 34(1):1–12, 2010.
- Daniele Masti and Alberto Bemporad. Learning nonlinear state–space models using autoencoders. *Automatica*, 129:109666, 2021.
- Hans B. Pacejka. Chapter 4 - Semi-Empirical Tire Models. In *Tire and Vehicle Dynamics (Third Edition)*, pages 149–209. Butterworth-Heinemann, Oxford, 2012.
- Brian Paden, Michal Čáp, Sze Zheng Yong, Dmitry Yershov, and Emilio Frazzoli. A survey of motion planning and control techniques for self-driving urban vehicles. *IEEE Transactions on Intelligent Vehicles*, 1(1):33–55, 2016.
- Dimitris C. Psychogios and Lyle H. Ungar. A hybrid neural network-first principles approach to process modeling. *AIChE Journal*, 38(10):1499–1511, 1992.
- M. Raissi, P. Perdikaris, and G. E. 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.

- Rishi Ramkannan, Gerben I. Beintema, Roland Tóth, and Maarten Schoukens. Initialization Approach for Nonlinear State-Space Identification via the Subspace Encoder Approach. In *Proc. of the 22nd IFAC World Congress*, pages 5146–5151, 2023.
- Antônio H. Ribeiro, Koen Tiels, Jack Umenberger, Thomas B. Schön, and Luis A. Aguirre. On the smoothness of nonlinear system identification. *Automatica*, 121:109158, 2020.
- Johan Schoukens and Lennart Ljung. Nonlinear system identification: A user-oriented road map. *IEEE Control Systems Magazine*, 39(6):28–99, 2019.
- Maarten Schoukens and Roland Tóth. On the Initialization of Nonlinear LFR Model Identification with the Best Linear Approximation. In *Proc. of the 21st IFAC World Congress*, pages 310–315, 2020.
- Oliver Schön, Ricarda-Samantha Götte, and Julia Timmermann. Multi-Objective Physics-Guided Recurrent Neural Networks for Identifying Non-Autonomous Dynamical Systems. In *Proc. of the 14th IFAC Workshop on Adaptive and Learning Control Systems*, pages 19–24, 2022.
- B. Sohlberg and E. W. Jacobsen. GREY BOX MODELLING – BRANCHES AND EXPERIENCES. *IFAC Proceedings Volumes*, 41(2):11415–11420, 2008.
- Bei Sun, Chunhua Yang, Yalin Wang, Weihua Gui, Ian Craig, and Laurentz Olivier. A comprehensive hybrid first principles/machine learning modeling framework for complex industrial processes. *Journal of Process Control*, 86:30–43, 2020.
- M. Szécsi, B. Györök, Á. Weinhardt-Kovács, G. I. Beintema, M. Schoukens, T. Péni, and R. Tóth. Deep learning of vehicle dynamics. In *Proc. of the 20th IFAC Symposium on System Identification*, pages 283–288, 2024.
- Emanuel Todorov, Tom Erez, and Yuval Tassa. MuJoCo: A physics engine for model-based control. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5026–5033, 2012.
- Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning requires rethinking generalization. In *Proc. of the 5th International Conference on Learning Representations*, 2017.