9th Asia-Pacific Conference on Plasma Physics, 21-26 Sep, 2025 at Fukuoka

Identification of reduced-order models by sparse regression with oracle property

Y. Shirasawa[1], Y. Ida[2], R. Yoneda[1], T. Kojima[1], M. Takahashi[1]

[1]NTT Space Environment and Energy Laboratories, [2]NTT Computer and Data Science Laboratories

e-mail (speaker): yuita.shirasawa@ntt.com

Various mathematical models have been proposed to explain the phenomena of magnetized plasma. On the other hand, real-time plasma control requires identifying reduced-order models (ROMs) from data, which are simplifications of the original governing equations.

Sparse regression is an attempt to approximate observed information with fewer basis functions. Sparse solutions can be obtained by solving optimization problems with regularization terms. A commonly used sparse regression is the least absolute shrinkage and selection operator (Lasso, L1) [1]. On the other hand, Lasso has been shown not to have oracle property [2]. Oracle property consists of (a) consistency of variable selection and (b) asymptotic normality. Using a regularization with oracle property, two things can be expected that (a) the true non-zero elements can be obtained, and (b) the coefficients of the true non-zero elements can be estimated correctly as the number of samples increases. The smoothly clipped absolute derivation (SCAD) was designed by Fan and Li [2] to have this oracle property.

Sparse Identification of Nonlinear Dynamics (SINDy) [3, 4] is a data-driven method that exploits the variable selectivity of sparse regression. This method identifies differential equations from time-series data by sparse regression under the assumption that the governing equation of a system is composed of few active terms. However, the regularization included in SINDy does not have oracle property. Therefore, plasma prediction may fail because the truly correct equation cannot be identified.

To enhance the accuracy of governing equations identification, we applied SCAD, which has oracle property, to the optimizer SR3 of SINDy. Our research aims to identify ROMs of plasma transport phenomena from time-series data. The train data were generated using the integrated code TASK [5]. First, we assumed that the ROMs can be represented by ordinary differential equations (0D models). For the identified ROMs by SINDy, we mainly compared the performance of SCAD and Lasso. As shown in Figure 1, we found that SCAD has a higher identification accuracy for ROMs than Lasso, and the range of Pareto solutions for the regularization parameter λ is wider. Additionally, the results suggest that the oracle property allows us to identify the governing equations without shrinking to zero the terms that contribute to dynamics. In the presentation, the effect of oracle property and its application to the transport control will be reported. In addition, we will also report the results of attempts to identify partial differential equations (1D models).

References

[1] R. Tibshirani, Journal of the Royal Statistical Society Series B: Statistical Methodology **58**, 267 (1996).

[2] J. Fan and R. Li, Journal of the American statistical Association **96**, 1348 (2001).

[3] S. L. Brunton *et al*, Proceedings of the National Academy of Sciences **113**, 3932 (2016).

[4] J. D. Lore *et al*, Nuclear Fusion **63**, 046015 (2023).

[5] A. Fukuyama *et al*, Plasma Physics and Controlled Fusion **37**, 611 (1995).

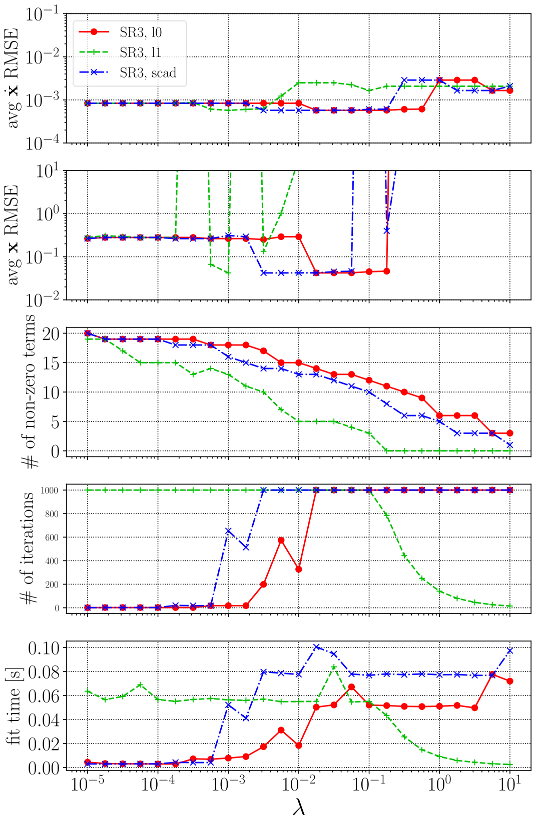

Figure 1: SCAD (blue) has a higher identification accuracy for ROMs than Lasso, and the range of Pareto solutions for the regularization parameter λ is wider. The state variable is $\mathbf{x}(t) = (x_1(t), x_2(\mathrm{t})) = (\langle n_e(t)\rangle, \langle T_e(t)\rangle)$, where $\langle n_e\rangle$ [m^{-3}] and $\langle T_e\rangle$ [keV] are volume-averaged electron density and tempareture. The control variable is $u_1(t) = P_{\mathrm{ECH}}(t)$, where P_{ECH} [MW] is electron cyclotron heating power.