Turnpike properties in nonlinear system identification

Julian D. Schiller, **Matthias A. Müller**

Leibniz University Hannover, Institute of Automatic Control, Appelstr. 11, 30167 Hannover,
E-Mail: {schiller,mueller}@irt.uni-hannover.de

Learning nonlinear state-space models from input-output data is a central problem in system identification. If accurate multi-step predictions are of interest, a natural approach is simulation error minimization (SEM), where the parameters of a dynamical model are obtained by minimizing the discrepancy between measured and simulated outputs over a finite dataset [1]. For nonlinear models and long datasets, however, solving the resulting optimization problem can become computationally demanding. This is particularly relevant for recurrent neural networks and related nonlinear state-space models, where training requires forward simulation and corresponding sensitivity computations over long time horizons. Practical approaches therefore frequently perform SEM over shorter subsequences and fix their initial states, possibly after a finite burn-in phase [4]. This raises the question of when the resulting solution is nevertheless representative of the original SEM problem in which the initial state is optimized jointly with the model parameters.

We study this question using *turnpike* properties, which classically describe the phenomenon that optimal trajectories remain close to an optimal steady state or reference trajectory for most of the time [2, 3]. Specifically, we consider a discrete-time nonlinear state-space model

$$x^+ = f(x, u; \theta), \qquad y = h(x, u; \theta),$$

with states x, inputs u, outputs y, the parameter vector θ to be identified, and the SEM objective

$$J(x_0, \theta) = \sum_{j=0}^{N} \beta_j l(y_j(x_0, \theta), y_j^{\mathrm{d}}) + r(\theta).$$

Here, the stage cost l penalizes the difference between model outputs $y_j(x_0, \theta)$ and given measurement data y_j^{d}, while r denotes a regularization term and β_j is a user-defined scalar weighting. We distinguish the unconstrained SEM problem, where both x_0 and θ are optimized, from a computationally simpler constrained problem in which the initial state x_0 is fixed to some $\bar{x}$ and only θ remains a decision variable.

In contrast to classical optimal control, the turnpike is naturally formulated here in terms of *outputs*. Moreover, optimal model parameterizations and even optimal output sequences need not be unique, as is particularly relevant for overparameterized black-box models. The corresponding *cumulative output turnpike* property requires the existence of $\alpha \in \mathcal{K}_\infty$ and $E > 0$, independent of the training dataset length N, such that

$$\sum_{j=0}^{N} \alpha\left(\|y_j(\bar{x}, \theta') - y_j(x_0^*, \theta^*)\|\right) \leq E$$

for every constrained optimum θ' and every associated closest unconstrained optimum (x_0^*, θ^*); the corresponding optimal output sequence $y_j(x_0^*, \theta^*)$, $j \in [0, N]$, is referred to as the *turnpike*. Hence, the average output discrepancy vanishes as $N \to \infty$, while individual discrepancies at all time instants remain uniformly bounded.

Our main result provides a system-theoretic characterization of this property [5]. Under a mild cost-reachability condition, we establish the equivalence

$$\text{cumulative output turnpike} \quad \Longleftrightarrow \quad \text{coercivity} \quad \Longleftrightarrow \quad \text{strict dissipativity.} \qquad (\star)$$

The dissipativity notion is tailored to system identification: its supply rate compares the cost generated by a constrained solution with that of its associated optimal output sequence, while the storage function may depend on the finite dataset length. This extends classical connections between turnpike properties and strict dissipativity in optimal control [2, 3] to SEM problems with potentially non-unique and horizon-dependent turnpikes.

We additionally investigate the weaker *cardinality turnpike* property, which bounds the number of time indices at which a constrained optimal output lies outside a neighborhood of its turnpike. We show that cumulative turnpike behavior always implies cardinality turnpike behavior. The converse, however, is false in general, which we prove by constructing a corresponding counterexample.

Finally, we derive sufficient conditions under which turnpike behavior arises in nonlinear system identification. Specifically, we show that cost reachability follows from incremental output stability of the model, while coercivity of the value function follows under uniform strict convexity of the stage cost and a suitable optimality condition. Together with the equivalence result ($\star$), these conditions establish cumulative turnpike behavior.

The theory is illustrated for a scalar Elman-type recurrent neural network, for which the relevant conditions can be verified analytically. The example also explicitly exhibits a non-unique optimal solution set for a short training horizon and illustrates how the cumulative output discrepancy remains bounded as the dataset size N increases.

Overall, these results provide a theoretical explanation for why computationally convenient SEM formulations with fixed initial states can yield output sequences close to those of the corresponding unconstrained identification problem, and establish connections between system identification, stability, dissipativity, and turnpike theory.

References

[1] L. Ljung, System Identification: Theory for the User, 2 ed., Prentice Hall, Upper Saddle River, NJ, USA, 1999. 14th printing.

[2] T. Faulwasser, L. Grüne, Turnpike properties in optimal control, in: E. Trélat, E. Zuazua (Eds.), Handbook of Numerical Analysis, volume 23, Elsevier, Amsterdam, The Netherlands, 2022, pp. 367–400.

[3] L. Grüne, M. A. Müller, On the relation between strict dissipativity and turnpike properties, Syst. Control Lett. 90 (2016) 45–53.

[4] J. D. Schiller, M. Heinrich, V. G. Lopez, M. A. Müller, Tuning the burn-in phase in training recurrent neural networks improves their performance, in: Proc. Int. Conf. Learn. Represent. (ICLR), 2026.

[5] J. D. Schiller, M. A. Müller, Turnpike properties in nonlinear system identification, under preparation, 2026.