Unconstrained learning of networked nonlinear systems via free parametrization of stable interconnected operators

Leonardo Massai, Danilo Saccani, Luca Furieri and Giancarlo Ferrari-Trecate

Abstract— This paper characterizes a new parametrization of nonlinear networked incrementally L_2-bounded operators in discrete time. The distinctive novelty is that our parametrization is *free* – that is, a sparse large-scale operator with bounded incremental L_2 gain is obtained for any choice of the real values of our parameters. This property allows one to freely search over optimal parameters via unconstrained gradient descent, enabling direct applications in large-scale optimal control and system identification. Further, we can embed prior knowledge about the interconnection topology and stability properties of the system directly into the large-scale distributed operator we design. Our approach is extremely general in that it can seamlessly encapsulate and interconnect state-of-the-art Neural Network (NN) parametrizations of stable dynamical systems. To demonstrate the effectiveness of this approach, we provide a simulation example showcasing the identification of a networked nonlinear system. The results underscore the superiority of our free parametrizations over standard NN-based identification methods where a prior over the system topology and local stability properties are not enforced.

I. INTRODUCTION

Interconnected systems are composed of multiple subsystems interacting over physical or cyber coupling channels and encompass diverse engineering applications like intelligent buildings, power grids, and transportation networks. Typically, effective control and monitoring of interconnected systems hinge on precise models of the subsystems. Failure to accurately and promptly identify the dynamics of a single subsystem can trigger network-wide instability, primarily because of uncertainty propagation through the coupling channels. This renders conventional system identification methods tailored for single-agent systems unsuitable for this scenario [1].

To improve on this matter, we want to exploit the prior knowledge about the model that we want to identify in terms of its structure and stability properties to define a class of distributed operators that capture such a structure while also ensuring some stability properties. More in detail, in this work we focus on the stability property known as incremental L_2-boundedness, which is a particular case of incremental dissipativity.

When dealing with linear models, incremental L_2-boundedness [2] can be enforced by imposing Linear Matrix Inequality (LMI) constraints. This approach has been

extended to certain families of non-linear systems via the Linear Parameter Varying framework [2]. In system identification, this means that if we want to enforce incremental L_2-boundedness we end up with a constrained optimization problem. However, this is not desirable for several reasons such as poor scalability, feasibility issues, and computational efficiency.

With this in mind, in this paper, we introduce a new parametrization of distributed incremental L_2-bounded operators. The parameterization is free in the sense that for any value of the parameters, the resulting distributed operator is incrementally L_2-bounded. Consequently, this parametrized set of operators can be applied to system identification without the need to impose additional constraints for incremental L_2-boundedness, hence allowing one to solve an unconstrained optimization problem.

Related works: Dissipativity theory has found applications in a wide range of control problems [3], [4]. It offers a unified framework for analyzing stability and performance in nonlinear systems [3]. Several frameworks exist to analyze these properties [5]–[7] and ensure global stability of nonlinear systems [8], [9]. However, the dissipativity and stability properties of the subsystems can be preserved or altered when the systems are combined [10]. Different works have derived sufficient conditions to provide guarantees on input-output properties for generic interconnected systems [11], [12]. This framework has recently been exploited to enforce robustness in Neural Networks (NNs) by analyzing them from a dynamical system perspective. Recurrent Equilibrium Networks (RENs) [13]–[15] achieve incremental L_2-boundedness through a specific parameterization, enabling unconstrained training while guaranteeing stability. While NNs have found attractive applications in various fields, including the identification of dynamical systems [16]–[18], their full potential is not yet fully understood. While certain NN architectures can offer dissipative properties, it remains a challenge to develop a parameterization method that ensures these properties are maintained when the network has a specific interconnection topology.

Contributions: The main contribution of this paper is the development of a free parametrization for interconnected incremental L_2-bounded operators. Specifically, we derive an explicit mapping from a set of freely chosen parameters to a set of parametrized sub-operators such that their interconnection through a desired topology is also incrementally L_2-bounded. This result is fully compatible with any free parametrization of the single sub-operators, as long as these parametrizations ensure a finite incremental L_2-bound for

each sub-operator as well. This allows for the exploitation of pre-existing state-of-the-art parameterizations of incrementally L_2-bounded sub-operators (such as RENs [13]), thus opening the door to unconstrained optimization over distributed operators that also enjoy such a stability property. This is key as it enables the use of distributed NN operators in large-scale nonlinear system identification as well as control applications [19] while ensuring stability properties throughout the whole training phase.

NOTATION

Throughout the paper, we denote with $\mathbb{N}$ the set of non-negative integers. $I_{a \times b}$ is the identity matrix with dimensions $a \times b$ and with $0_{a \times b}$ we denote a zero matrix of size $a \times b$, while I_a and 0_a are square matrices with size a. Positive semidefinite matrices A are denoted as $A \succeq 0$ and $A_{[ij]}$ denotes the element of matrix A corresponding to row i and column j. We denote with $\text{blkdiag}(A, B)$ a block diagonal matrix created by aligning the matrices A and B along the diagonal and with $\text{col}_{j \in \mathcal{N}}(v_j)$, with $\mathcal{N} \subseteq \mathbb{N}$ a vector which consists of the stacked subvectors v_j where $j \in \mathcal{N}$ while $v_{0:T}$ denotes a sequence of $v(t)$ with t ranging from 0 to T. The set of functions or sequences from $\mathbb{X}$ to $\mathbb{Y}$ is denoted by $\mathbb{Y}^{\mathbb{X}}$.

II. PRELIMINARIES

The L_2-gain is a fundamental concept in the stability analysis and control of dynamical systems and it has been applied to ensure stability properties in networks of systems [11], [20]. We leverage this concept to derive trainable incrementally L_2-bounded distributed operators. Now we recap some basic definitions regarding L_2-gain based on results presented in [2], [21].

We consider parametrized discrete-time systems Σ_θ in the form:

$$\Sigma_\theta : \begin{cases} x(t+1) & = f_\theta(x(t), u(t)), \quad x \in \mathbb{R}^n, u \in \mathbb{R}^m \\ y(t) & = h_\theta(x(t), u(t)), \quad y \in \mathbb{R}^p \\ x(0) & = x_0, \qquad\qquad\quad t = 0, \ldots, T \end{cases}$$
(1)

where $x(t) \in \mathbb{R}^n, u(t) \in \mathbb{R}^m, y(t) \in \mathbb{R}^p, \theta \in \mathbb{R}^q$ are the system state, input, output and parameter respectively. $f_\theta : \mathbb{R}^q \times \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}^n$ and $h_\theta : \mathbb{R}^q \times \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}^p$ are, instead, assumed to be Lipschitz continuous functions depending on θ, such that $f_\theta(0, 0) = 0$ and $h_\theta(0, 0) = 0$ for all θ, and such that for all initial conditions $x_0 \in \mathbb{R}^n$ there is a unique solution $(x, u, y) \in (\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p)^{\mathbb{N}}$.
We define the set of solutions of (1) as

$$\mathcal{B} := \{(x, u, y) \in (\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p)^{\mathbb{N}} | (x, u, y) \text{ satisfies (1)}\}.$$

Notice that the system Σ_θ given by (1) can be seen as a parametrized operator that, for a fixed value of the parameter, takes as input an initial condition $x(0) \in \mathbb{R}^{n_i}$ as well as a sequence $u_{0:T}$ and outputs a sequence $y_{0:T}$ for any given $T \in \mathbb{N}$. In this sense, we will also use the notation $y_{0:T} = \Sigma_\theta(x_0, u_{0:T})$ to highlight the mapping operated by Σ_θ. Notably, recursive operators, such as (1), can be viewed

as deep neural networks unrolled in time. Consequently, these operators can serve as operators in learning problems and be optimized through Backpropagation Through Time (BPTT) [19]. We can provide the following definitions.

Definition 1: (L_2-gain) [4] The operator Σ_θ given by (1) has a finite L_2-gain$\leq \gamma$ if it is dissipative with respect to the supply rate $s(u, y) = \gamma^2 \|u\|^2 - \|y\|^2$; that is, there exists a storage function $V : \mathbb{R}^n \to \mathbb{R}^+$ such that for all $t \in \mathbb{N}$

$$V(x(t+1)) - V(x(t)) \leq s(u(t), y(t)), \qquad (2)$$

The storage function V can be interpreted as a model of the stored "energy" in the system w.r.t. a single point of neutral storage (minimum energy).

Definition 2: (Incremental L_{i2} gain [2]) The operator Σ_θ in the form (1) is said to have a finite incremental L_2-gain γ, denoted as L_{i2}-gain, if for all $t \in \mathbb{N}$ and $(x, u, y), (\hat{x}, \hat{u}, \hat{y}) \in \mathcal{B}$ it is dissipative with respect to the supply rate $s(u, \hat{u}, y, \hat{y}) = \gamma^2 \|u - \hat{u}\|^2 - \|y - \hat{y}\|^2$; that is, there exists an *incremental storage function* $V : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^+$ with $V(0, 0) = 0$ such that for all $t \in \mathbb{N}$:

$$V(x(t+1), \hat{x}(t+1)) - V(x(t), \hat{x}(t)) \leq \\ s(u(t), \hat{u}(t), y(t), \hat{y}(t)). \quad (3)$$

We can now introduce the following set of parameters θ:

$$\Phi_\gamma = \{\theta \in \mathbb{R}^q \mid \Sigma_\theta \text{ has finite } L_{i2}\text{-gain } \gamma\}. \qquad (4)$$

In a learning problem, such as in system identification, we aim to fit a set of data by tuning the parameters of an operator Σ_θ by minimizing a certain, differentiable, loss function. However, it's often desirable to also preserve the stability properties of the underlying system, obtaining a constrained optimization problem. By considering a constraint on the incremental L_2-boundedness of the system, we can write the identification problem as follows:

$$\min_\theta \quad \mathcal{L}(y_{0:T}, \tilde{y}_{0:T})$$

$$\text{s.t.} \quad y_{0:T}^i = \Sigma_\theta(\tilde{x}_0^i, \tilde{u}_{0:T}^i), \quad i = 1, \ldots n_s \qquad (5a)$$

$$\theta \in \Phi_\gamma, \qquad (5b)$$

where $(\tilde{u}_{0:T}^i, \tilde{y}_{0:T}^i, \tilde{x}_0^i)$ denote finite sequences of input-output measurements and the initial condition, and n_s is the number of trajectories considered. For instance, for system identification purposes, one might want to minimize the sum of the squared residuals over all the sequences, i.e.,

$$\mathcal{L}(y_{0:T}, \tilde{y}_{0:T}) = \sum_{i=1}^{n_s} \|y_{0:T}^i - \tilde{y}_{0:T}^i\|^2 \qquad (6)$$

The problem (5) is a constrained optimization problem that, when Σ_θ is a linear operator, features linear convex constraints (5b) that can be written as LMIs [22]. This is not desirable because of poor computational efficiency as well as scalability and feasibility issues. Notice that when Σ_θ is a non-linear operator, all these issues are exacerbated even further, as the general problem tends to become non-convex and non-linear even in the constraints.

For these reasons, one would like to cast the problem (5) as an unconstrained optimization problem by getting rid of (5b), even at the cost of restricting the set of feasible parameters. This is key as it will open the door for the utilization of parametrized families of highly non-linear operators in the form (1), such as neural networks, which can be optimized very efficiently by solving an unconstrained optimization problem with standard backpropagation and unconstrained gradient descent methods. All of this can be obtained by freely parametrizing the set Φ_γ.

Definition 3: Given a non-empty set $\Psi \subseteq \mathbb{R}^a$, a parametrization of Ψ is a mapping $\psi : \Theta \subseteq \mathbb{R}^q \mapsto \Psi$. the parametrization ψ is called free if $\Theta = \mathbb{R}^q$.

We will consider a free parametrization of the set Φ_γ, that is, a mapping $\psi(\xi, \gamma) : \mathbb{R}^q \mapsto \Phi_\gamma$ where ξ are the free parameters. In other words, $\theta = \psi(\xi, \gamma) \in \Phi_\gamma$, $\forall \xi \in \mathbb{R}^q$ or, equivalently, $\Sigma_{\psi(\xi,\gamma)}$ has L_{i2}-gain γ, $\forall \xi \in \mathbb{R}^q$. Notice that here γ is fixed but could also be treated as an additional free parameter of ψ through a proper mapping such as $\gamma = z^2$ where $z \in \mathbb{R}$ is the new parameter.

Notice that if one can exhibit such a free parametrization, then that can be used to reformulate (5) as an unconstrained problem by optimizing over the new parameter ξ as the constraint (5b) is satisfied for any value of ξ by definition.

Remark 1: One remarkable example of such free parameterization is given by operators called Recurrent Equilibrium Networks (RENs) in [14], which result from the closed-loop interconnection of a discrete-time linear dynamical system with a static non-linearity. RENs extend various non-linear operators, including Recurrent Neural Networks (RNNs) and standard feed-forward neural networks. A significant contribution of [14] lies in presenting a free parametrization for a class of RENs with stability and dissipativity properties, such as L_{i2}-boundedness. Because of such features, REN operators can be efficiently optimized in a learning problem using unconstrained techniques.

Remark 2: There is usually a "price to pay" for transforming the problem (5) into an unconstrained one via a free parametrization. This stems from the fact that, usually, the map ψ is not surjective and it effectively describes only a subset of Φ_γ, which in turn means that we are restricting ourselves to optimize over a subset of L_{i2}-bounded operators. This is also true for REN operators. However, as it is shown in [14], the advantages of having a free parametrization greatly outweigh such a drawback.

Notice that so far we have considered only one single centralized operator Σ_θ but, as mentioned in the introduction, we would like to use a distributed operator to better capture the interconnectedness of real systems and we want to do that while preserving the L_{i2}-boundedness property. The free parametrization approach outlined earlier has never been generalized to distributed operators built by several interconnected sub-operators. It is worth mentioning that simply connecting L_{i2}-bounded sub-operators will not work, in general, as the resulting operator could fail to be L_{i2}-bounded.
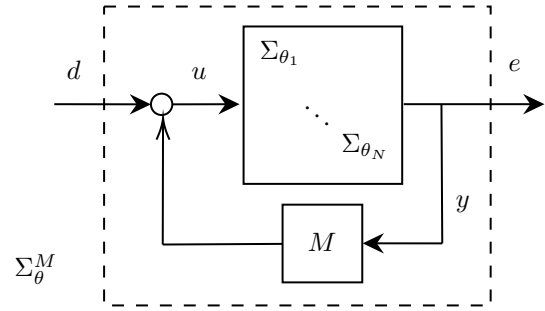

Fig. 1: Interconnection of N operators Σ_{θ_i}.

We introduce a free parametrization for distributed operators in the next Section.

III. Setting, Models, and Problem formulation

We consider N parametrized sub-operators indexed by $i \in \mathcal{N}$ with $\mathcal{N} = \{1, \dots, N\}$

$$\Sigma_{\theta_i} : \begin{cases} x_i(t+1) & = f_{\theta_i}(x_i(t), u_i(t)), \\ y_i(t) & = h_{\theta_i}(x_i(t), u_i(t)), \quad t = 0, \dots, T, \\ x_i(0) & = x_{0i} \end{cases}$$
(7)

where $x_i(t) \in \mathbb{R}^{n_i}, u_i(t) \in \mathbb{R}^{m_i}, y_i(t) \in \mathbb{R}^{p_i}, \theta \in \mathbb{R}^q$ are the system state, input, output and parameter respectively. These operators are interconnected, as shown in Fig. 1[1], to accommodate an exogenous input $d \in \mathbb{R}^m$ and performance output $e \in \mathbb{R}^p$. More in detail, the interconnection topology is described by the matrix $\bar{M}$ that relates the sub-operators inputs and performance outputs as:

$$\begin{bmatrix} u \\ e \end{bmatrix} = \bar{M} \begin{bmatrix} y \\ d \end{bmatrix} = \begin{bmatrix} M & I_m \\ I_p & 0_m \end{bmatrix} \begin{bmatrix} y \\ d \end{bmatrix},$$
(8)

where $u = \mathrm{col}_{\mathcal{N}}(u_i) \in \mathbb{R}^m$, $y = \mathrm{col}_{\mathcal{N}}(y_i) \in \mathbb{R}^p$. The matrix $\bar{M}$ describes how each sub-operator input u_i is coupled with each sub-operator output y_i as well as with the exogenous input d_i. More in detail, by developing (8) we get $u = My + d$ and $e = y$. The first equation states that the input u of each sub-operator is a fully generic linear combination of the sub-operator's output y via the matrix M plus the exogenous input d, which is an additive to each input. As we indicated before, this choice constrains d to have the same dimension of u, which is a hypothesis that we aim to relax in future work by considering a more general form for the matrix $\bar{M}$.

The second equation $e = y$ describes the fact that the performance output of the interconnected operator coincides with all the outputs of each sub-operator.

The coupling of each sub-operator Σ_{θ_i} via (8) produces an operator Σ_θ^M that, with a slight abuse of notation, can be written as

$$\Sigma_\theta^M : \begin{cases} x(t+1) & = f_\theta\left(x(t), My(t) + d(t)\right) \\ y(t) & = h_\theta\left(x(t), My(t) + d(t)\right), \\ x(0) & = x_0, \quad\quad t = 0, \dots, T \end{cases}$$
(9)

---

[1] Notice that this is a special case of the framework introduced in [11].

where $x(t) = \mathrm{col}_{i \in \mathcal{N}}(x_i(t)) \in \mathbb{R}^n, \theta = \mathrm{col}_{i \in \mathcal{N}}(\theta_i) \in \mathbb{R}^q$ and the functions f_θ and h_θ are the stacked sub-functions f_{θ_i} and h_{θ_i} applied block-wise. We assume that the coupling, and therefore the operator Σ_θ^M, is well-defined, i.e., the equation $y(t) = h_\theta(x(t), My(t) + d(t))$ admits a unique solution $y(t)$, given $x(t)$ and $d(t)$, for $t = 0, \ldots, T$. Now we are ready to state the problem we aim to solve:

Problem 1: Consider the operator Σ_θ^M as defined in (9). We aim to find a free parametrization of the set

$$\Phi_{\gamma_M} = \{\theta_i, i \in \mathcal{N} \mid \Sigma_\theta^M \text{ has finite } L_{i2}\text{-gain } \gamma_M\}. \quad (10)$$

In other words, we want to find some mapping from a set of free parameters to a set of parameters for each sub-operator such that their coupling Σ_θ^M has finite L_{i2}-gain. This enables free tuning of sub-operators while preserving L_{i2}-gain for the interconnected operator, hence allowing one to solve problem (5) with any unconstrained optimization algorithm.

IV. MAIN RESULTS

To tackle the problem, first of all, we consider sub-operators verifying $\Sigma_{\theta_i} \in \Phi_{\gamma_i}, i \in \mathcal{N}$ and such that they are dissipative with a positive definite, continuously differentiable incremental storage function $V_i(\cdot, \cdot)$ and the quadratic supply rate:

$$s_i(u_i, \hat{u}_i, y_i, \hat{y}_i) = \\ \begin{bmatrix} u_i - \hat{u}_i \\ y_i - \hat{y}_i \end{bmatrix}^\top \underbrace{\begin{bmatrix} \gamma_i^2 I_{m_i} & 0_{m_i \times p_i} \\ 0_{p_i \times m_i} & -I_{p_i} \end{bmatrix}}_{X_i} \begin{bmatrix} u_i - \hat{u}_i \\ y_i - \hat{y}_i \end{bmatrix}. \quad (11)$$

The goal is now to certify the validity of a prescribed L_{i2} gain γ_M for the interconnected operator Σ_θ^M, or in other words, its dissipativity with respect to the quadratic supply rate:

$$\begin{bmatrix} d - \hat{d} \\ e - \hat{e} \end{bmatrix}^\top \underbrace{\begin{bmatrix} \gamma_M^2 I_m & 0_{m \times p} \\ 0_{p \times m} & -I_p \end{bmatrix}}_{W} \begin{bmatrix} d - \hat{d} \\ e - \hat{e} \end{bmatrix}. \quad (12)$$

Now, let us consider the following candidate incremental storage function, defined as the weighted sum of the local storage functions

$$V(x, \hat{x}) = \alpha_1 V_1(x_1, \hat{x}_1) + \cdots + \alpha_N V_N(x_N, \hat{x}_N), \quad (13)$$

with $\alpha_i > 0$. Thus, following the results in [11], we want to satisfy the following inequality

$$\sum_{i=1}^N \alpha_i [V_i(x_i(t+1), \hat{x}_i(t+1)) - V_i(x_i(t), \hat{x}_i(t))] \leq \\ \sum_{i=1}^N \alpha_i \begin{bmatrix} u_i - \hat{u}_i \\ y_i - \hat{y}_i \end{bmatrix}^\top X_i \begin{bmatrix} u_i - \hat{u}_i \\ y_i - \hat{y}_i \end{bmatrix}. \quad (14)$$

By rewriting the right-hand side of (14), we obtain:

$$\begin{bmatrix} u_1 - \hat{u}_1 \\ \vdots \\ u_N - \hat{u}_N \\ y_1 - \hat{y}_1 \\ \vdots \\ y_N - \hat{y}_N \end{bmatrix}^\top \boldsymbol{X}(\alpha_i X_i) \begin{bmatrix} u_1 - \hat{u}_1 \\ \vdots \\ u_N - \hat{u}_N \\ y_1 - \hat{y}_1 \\ \vdots \\ y_N - \hat{y}_N \end{bmatrix}, \quad (15)$$

where

$$\boldsymbol{X}(\alpha_i X_i) = \begin{bmatrix} \Gamma_{N,m} & 0 \\ 0 & -A_{N,p} \end{bmatrix}, \quad (16)$$

with $\Gamma_{N,m} = \mathrm{blkdiag}\{\alpha_1 \gamma_1^2 I_{m_1}, \ldots, \alpha_N \gamma_N^2 I_{m_N}\}$ and $A_{N,p} = \mathrm{blkdiag}\{\alpha_1 I_{p_1}, \ldots, \alpha_N I_{p_N}\}$.

The storage function (13), serves as a storage function for the interconnection and, if the right-hand side of (14) is negative semidefinite in y, when u is eliminated by considering $u = My$, then the interconnection is Lyapunov stable [11, Thm. 3.1]. The right-hand side of (14), rewritten as in (15), is dominated by the performance supply rate (12) if

$$\begin{bmatrix} u - \hat{u} \\ y - \hat{y} \\ d - \hat{d} \\ e - \hat{e} \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{X}(\alpha_i X_i) & 0 \\ 0 & -W \end{bmatrix} \begin{bmatrix} u - \hat{u} \\ y - \hat{y} \\ d - \hat{d} \\ e - \hat{e} \end{bmatrix} \preceq 0, \quad (17)$$

when the variables u and e are constrained by (8). Substituting

$$\begin{bmatrix} u \\ y \\ d \\ e \end{bmatrix} = \begin{bmatrix} M & I_m \\ I_p & 0_{p \times m} \\ 0_{m \times p} & I_m \\ I_p & 0_{p \times m} \end{bmatrix} \begin{bmatrix} y \\ d \end{bmatrix}, \quad (18)$$

in (17) we obtain the following performance condition [11]:

$$\begin{bmatrix} M & I \\ I & 0 \\ 0 & I \\ I & 0 \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{X}(\alpha_i X_i) & 0 \\ 0 & -W \end{bmatrix} \begin{bmatrix} M & I \\ I & 0 \\ 0 & I \\ I & 0 \end{bmatrix} \preceq 0 \quad (19)$$

where $\boldsymbol{X}(\alpha_i X_i)$ is as defined in (16). For the sake of simplicity, from now on we omit the size of identity and zero matrices. The previous computations prove the following proposition adapted from [11].

Proposition 1: If there exist $\alpha_i > 0$, $i \in \mathcal{N}$, such that (19) holds, then the operator Σ_θ^M has a finite incremental L_{i2}-gain γ_M.

Now we have all the tools needed to derive our main result.

Theorem 1: The set Φ_{γ_M} is non-empty and admits a free parametrization such that each sub-operator Σ_{θ_i} is parametrized by $\theta_i = \psi_i(\xi_i, \nu(z_i))$ with $i \in \mathcal{N}$ and with free parameters, $\xi_i \in \mathbb{R}^{q_i}, z_i \in \mathbb{R}$ where $\nu(z_i) : \mathbb{R} \mapsto \mathbb{R}_+$ is given by:

$$\nu(z_i) = \sqrt{\frac{1}{1 + \max_{j \in \mathcal{N}_{p_i}} \left(\sum_k |M_{[kj]}|\right) + z_i^2}} \\ \cdot \sqrt{\frac{\gamma_M^2}{\max_{j \in \mathcal{N}_{u_i}} \left(\sum_k |M_{[jk]}|\right) \gamma_M^2 + 1}}, \quad (20)$$
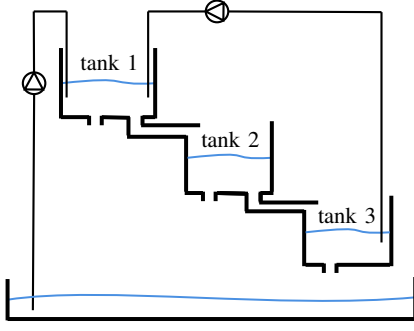
Fig. 2: Triple-tank system with recirculation pump.

where $M_{[ij]}$ is the element in the i-th row and j-th column in matrix M and

$$\mathcal{N}_{u_i} = \text{set of indices of } u \text{ associated with } \Sigma_{\theta_i}, i \in \mathcal{N},$$
$$\mathcal{N}_{y_i} = \text{set of indices of } y \text{ associated with } \Sigma_{\theta_i}, i \in \mathcal{N}. \tag{21}$$

The proof of Theorem 1 can be found in [23].

Theorem 1 allows us to tackle a learning problem, such as system identification, by solving an unconstrained optimization problem using a distributed operator that takes into account the sub-structure of the model that we want to learn while ensuring an L_{i2}-boundedness property of the operator at each step of the optimization process.

Remark 3: It is worth mentioning that incorporating the gain γ_M as an extra free parameter within the proposed parametrization can offer added flexibility when employing this parametrization in a training framework.

Remark 4: Following up Remark 2, we can notice that the free parametrization that we obtained through the mapping $\nu(z)$ is not surjective as we have introduced various degrees of conservatism, for example by using Gershgorin Theorem, which gave us a sufficient but not necessary condition for the positive semi-definiteness of the matrix obtained from (19) (see the proof in [23] for the details). Additional conservatism is present in the parametrization ψ itself. In other words, the free parametrization given by $\psi_i(\xi_i, \nu(z_i))$ allows one to optimize only over a subset of L_{i2}-bounded sub-operators that, once interconnected via the matrix M, define an L_{i2}-bounded distributed operator. As mentioned before, this drawback is greatly outweighed by the advantage of allowing one to solve an unconstrained optimization problem, as shown in the following example.

A detailed description of how Theorem 1 can be leveraged for the training of a network of L_{i2}-bounded operators can be found in [23].

V. NUMERICAL EXAMPLE

In this section, we validate our results on the identification of the interconnected nonlinear dynamical system shown in Fig. 2. The system is composed of three interconnected tanks and a recirculation pump that allows the water to flow from the third tank to the first. The pump itself cannot be controlled and continuously moves water from the last tank to the first. Additionally, an external controlled pump feeds into the first tank. The first-principles model, describing the dynamics of the system, is

$$\dot{h}_1 = -\frac{a_1}{A_1}\sqrt{2gh_1} + k_1\frac{a_3}{A_1}\sqrt{2gh_3} + k_c\frac{v}{A_1}$$
$$\dot{h}_2 = -\frac{a_2}{A_2}\sqrt{2gh_2} + k_2\frac{a_1}{A_2}\sqrt{2gh_1} \tag{22}$$
$$\dot{h}_3 = -\frac{a_3}{A_3}\sqrt{2gh_3} + k_3\frac{a_2}{A_3}\sqrt{2gh_2}$$

where A_i is the cross-section of the i-th tank, a_i is the cross-section of the outlet holes, k_i is the percentage of flow flowing in the next tank, h_i is the water level and v is the inlet flow to tank 1. The parameters employed in the simulation are listed in [23]. The system (22) has been discretized using a Forward Euler scheme with a sampling time $T_s = 0.1$. A dataset of $(\tilde{v}, \tilde{h})$ pairs was collected by exciting the system with a random input signal v with values in $[10, \ 100]$. Note that the simulated output measurements have been corrupted by white noise, i.e., $\tilde{h} = h + w$. The dataset was divided into two subsets: the identification set, which represented 70% of the entire dataset and was used for training, and the remaining portion of the dataset, which was employed to cross-validate the identified model. To identify the system (22), we validate the efficacy of the parametrization provided by Theorem 1 by considering the interconnection of three RENs models [13], [14]. Each REN is parametrized following the method described in [14], ensuring that each resulting sub-operator possesses a finite L_{i2}-gain and, therefore, belongs to the set Φ_γ. Employing the proposed approach, we interconnect the RENs as follows:

$$u = \begin{bmatrix} h_3 \\ v \\ h_1 \\ h_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}}_{M} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} + \underbrace{\begin{bmatrix} 0 \\ v \\ 0 \\ 0 \end{bmatrix}}_{d}. \tag{23}$$

To evaluate the efficacy of the proposed approach, we conducted numerical experiments comparing the system identification results obtained using the proposed approach and different neural networks architecture that ignores the topology of the system. Specifically, we used a Mean Square Error (MSE) as loss function and we compared the proposed approach to a single REN, and a RNN [24], another dynamical operator with no guarantees regarding L_2-boundedness. Fig. 4 shows the obtained loss on the independent validation dataset as a function of the number of tunable parameters in the considered models. For the comparison, all three architectures were implemented in PyTorch with a learning rate of $\eta = 10^{-2}$ and trained for $E = 500$ epochs. The implementation of the proposed approach can be found at: https://github.com/DecodEPFL/DistributedREN. As evident from Figure 4, the proposed architecture, leveraging prior knowledge of the system topology, achieves significantly improved performance in terms of prediction error on the validation dataset with a notably reduced number of tunable parameters, which are essentially optimization variables in the optimization problem. As expected all models exhibit better performance as the number
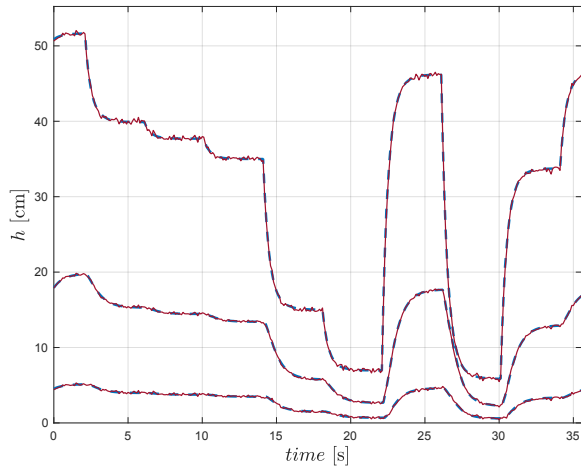
Fig. 3: Comparison of the open-loop prediction of the trained distributed RENs built with the proposed parametrization (blue dashed line) versus ground truth (red continuous line) on an independent validation dataset.
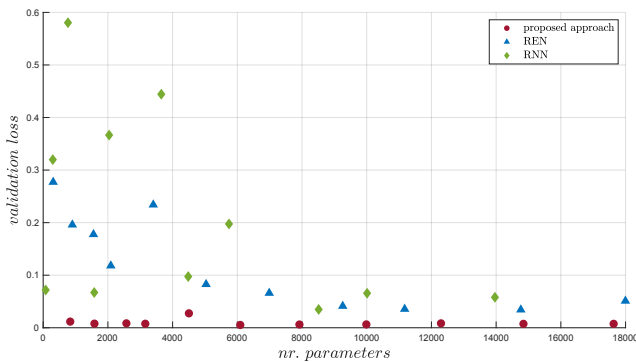

Fig. 4: Comparison of validation loss as a function of tunable parameters for the proposed approach with three RENs (red "○"), the REN (blue "△") and the RNN (green "◇").

of tunable parameters increases. Fig. 3 shows the predicted states of the identified model in blue and the trajectories of the validation dataset in red.

VI. Conclusions

We have introduced a novel free parametrization for incremental L_2 operators for the identification of distributed systems by leveraging prior knowledge of system behavior and sparsity pattern. A natural extension of the approach would involve exploring additional forms of dissipativity for interconnected systems. Moreover, the potential application of these operators as distributed controllers, mirroring the sparsity pattern of the plant, holds promise for enhancing closed-loop properties based on dissipativity theory. The versatility and potential impact of this methodology offer exciting prospects for the continued development and refinement of identification and design control strategies in complex distributed nonlinear systems.

References

[1] L. Ljung, "System identification," in *Signal analysis and prediction*. Springer, 1998, pp. 163–173.

[2] P. J. Koelewijn and R. Tóth, "Incremental stability and performance analysis of discrete-time nonlinear systems using the lpv framework," *IFAC-PapersOnLine*, vol. 54, no. 8, pp. 75–82, 2021.

[3] J. C. Willems, "Dissipative dynamical systems part i: General theory," *Archive for rational mechanics and analysis*, vol. 45, no. 5, pp. 321–351, 1972.

[4] C. I. Byrnes and W. Lin, "Losslessness, feedback equivalence, and the global stabilization of discrete-time nonlinear systems," *IEEE Transactions on automatic control*, vol. 39, no. 1, pp. 83–98, 1994.

[5] T. Rajpurohit and W. M. Haddad, "Dissipativity theory for nonlinear stochastic dynamical systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 1684–1699, 2016.

[6] C. Verhoek, P. J. Koelewijn, S. Haesaert, and R. Tóth, "Convex incremental dissipativity analysis of nonlinear systems," *Automatica*, vol. 150, p. 110859, 2023.

[7] D. Hill and P. Moylan, "The stability of nonlinear dissipative systems," *IEEE transactions on automatic control*, vol. 21, no. 5, pp. 708–711, 1976.

[8] D. Angeli, "A lyapunov approach to incremental stability properties," *IEEE Transactions on Automatic Control*, vol. 47, no. 3, pp. 410–421, 2002.

[9] A. Pavlov and L. Marconi, "Incremental passivity and output regulation," *Systems & Control Letters*, vol. 57, no. 5, pp. 400–409, 2008.

[10] P. Moylan and D. Hill, "Stability criteria for large-scale systems," *IEEE Transactions on Automatic Control*, vol. 23, no. 2, pp. 143–149, 1978.

[11] M. Arcak, C. Meissen, and A. Packard, *Networks of dissipative systems: compositional certification of stability, performance, and safety*. Springer, 2016.

[12] M. Inoue and K. Urata, "Dissipativity reinforcement in interconnected systems," *Automatica*, vol. 95, pp. 73–85, 2018.

[13] M. Revay, R. Wang, and I. R. Manchester, "Recurrent equilibrium networks: Flexible dynamic models with guaranteed stability and robustness," *IEEE Transactions on Automatic Control*, 2023.

[14] ——, "Recurrent equilibrium networks: Flexible dynamic models with guaranteed stability and robustness," *arXiv preprint arXiv:2104.05942*, 2021.

[15] D. Martinelli, C. L. Galimberti, I. R. Manchester, L. Furieri, and G. Ferrari-Trecate, "Unconstrained parametrization of dissipative and contracting neural ordinary differential equations," *arXiv preprint arXiv:2304.02976*, 2023.

[16] E. Terzi, T. Bonetti, D. Saccani, M. Farina, L. Fagiano, and R. Scattolini, "Learning-based predictive control of the cooling system of a large business centre," *Control Engineering Practice*, vol. 97, p. 104348, 2020.

[17] S. Chen, S. A. Billings, and P. Grant, "Non-linear system identification using neural networks," *International journal of control*, vol. 51, no. 6, pp. 1191–1214, 1990.

[18] M. Zakwan, L. Di Natale, B. Svetozarevic, P. Heer, C. N. Jones, and G. F. Trecate, "Physically consistent neural odes for learning multiphysics systems," *arXiv preprint arXiv:2211.06130*, 2022.

[19] L. Furieri, C. L. Galimberti, and G. Ferrari-Trecate, "Neural system level synthesis: Learning over all stabilizing policies for nonlinear systems," in *2022 IEEE 61st Conference on Decision and Control (CDC)*. IEEE, 2022, pp. 2765–2770.

[20] S. Dashkovskiy, B. S. Rüffer, and F. R. Wirth, "An iss small gain theorem for general networks," *Mathematics of Control, Signals, and Systems*, vol. 19, pp. 93–122, 2007.

[21] A. Van der Schaft, *L2-gain and passivity techniques in nonlinear control*. Springer, 2000.

[22] G. Chesi, "Exact lmi conditions for stability and l2 gain analysis of 2-d mixed continuous–discrete time systems via quadratically frequency-dependent lyapunov functions," *IEEE Transactions on Automatic Control*, vol. 67, no. 3, pp. 1147–1162, 2021.

[23] L. Massai, D. Saccani, L. Furieri, and G. Ferrari-Trecate, "Unconstrained learning of networked nonlinear systems via free parametrization of stable interconnected operators," *arXiv preprint arXiv:2311.13967*, 2023.

[24] L. R. Medsker and L. Jain, "Recurrent neural networks," *Design and Applications*, vol. 5, no. 64-67, p. 2, 2001.