

# Technical Report for Dissipativity Learning in Reproducing Kernel Hilbert Space

Xiuzhen Ye, Wentao Tang

November 3, 2025

This work presents a nonparametric framework for *dissipativity learning* in reproducing kernel Hilbert spaces (RKHS), which enables data-driven certification of stability and performance properties for unknown nonlinear systems without requiring an explicit dynamic model. Dissipativity is a fundamental system property that generalizes Lyapunov stability, passivity, and finite  $\mathcal{L}_2$ -gain conditions through an energy-balance inequality between a storage function and a supply rate. Unlike prior parametric formulations that approximate these functions using quadratic forms with fixed matrices, the proposed method represents them as Hilbert–Schmidt operators acting on canonical kernel features, thereby capturing nonlinearities implicitly while preserving convexity and analytic tractability. The resulting operator optimization problem is formulated in the form of a one-class support vector machine (OC-SVM) and reduced, via the representer theorem, to a finite-dimensional convex program expressed through kernel Gram matrices. Furthermore, statistical learning theory is applied to establish generalization guarantees, including confidence bounds on the dissipation rate and the  $\mathcal{L}_2$ -gain. Numerical results demonstrate that the proposed RKHS-based dissipativity learning method effectively identifies nonlinear dissipative behavior directly from input–output data, providing a powerful and interpretable framework for model-free control analysis and synthesis.

## 0.1 Introduction

Data-driven techniques, where statistical and machine learning methods are utilized for modelling and extracting system properties, are the main trending of the research on control systems. This stems from the fact that the potential of easing the derivation and maintenance of dynamics models [1]. In model-based data-driven approaches, research effort has been dedicated to the approaches of state-space model construction [2], parameter estimation [3], Koopman operators [4] [5], or neural networks [6], etc. In these approaches, challenges persist in developing models that accurately capture the underlying system dynamics, which consequently leaves the subsequent control design performance unguaranteed. In general, the state-of-the-art model-based approaches still involve model identification scheme and are not truly model independent.

With this consideration, instead of learning system dynamics, we are motivated to develop a *model free* approach that seeks for essential control-relevant property. The rationale of learning system properties lies in its simplicity and sophistication compared to learning a full dynamic model, while maintaining a more direct connection to the resulting control performance. Specifically, the model-free system property learning approach that this work focuses on is *dissipativity learning*, proposed in the author’s earlier works [7] [8] [9]. Dissipativity is a fundamental property of nonlinear dynamics [10], describing system-wide behavior in the form of a Lyapunov-like inequality. Particularly, it refers to the existence of a state-dependent *storage function* whose accumulation is upper-bounded by a *supply rate* that depends on the system’s inputs and outputs [11]. Since transport phenomena are governed by the second law of thermodynamics, dissipativity is known to be a common characteristic of process systems [12].

In the author’s previous works [8] [9], dissipativity learning control (DLC) was proposed as a data-driven control framework, where the dissipativity property of unknown nonlinear

dynamics is learned from input-output trajectories. In these works, the supply rate to be estimated from data is parameterized as a quadratic form, so that the matrix defining the quadratic form is to be learned. Naturally, quadratic forms are suitable for systems that are linear or approximately linear, which restricts their applicability to processes that exhibit severe nonlinearity. In addition, the representation of system-wide behavior by the collected trajectories often rely on restrictive assumptions and the complexity of sampling input-output trajectories is prohibitive.

Therefore, in this work, we formulate dissipativity learning in nonparametric framework in reproducing kernel Hilbert spaces (RKHS) [13] [14]. The information of inputs, outputs, and states is expressed in terms of their canonical features corresponding to the kernel functions in RKHS via kernel feature embeddings [15]. Such canonical features maps data samples nonlinearly into corresponding RKHSs, thereby accounting for the nonlinearity of the dynamics implicitly. Hence, nonlinear functions in the original state-space are represented by the Hilbert-Schmidt (H-S) operators acting on kernel feature maps. This is a generalization of the previous works [8] [9] that adopt quadratic forms of dissipativity and the matrices parameterizing these quadratic forms are learned. With kernel feature embeddings, the functions remains quadratic in the features to preserve convexity and analytic tractability, while nonlinearity is captured for the dissipative behavior.

The dissipativity learning problem is hence formulated as an operator convex optimization problem with constraints that enforce dissipativity and stability properties (e.g., the  $L_2$  gain). Formally, by extending inner product and norm concepts from matrices to operators, the learning of H-S operators preserves the formulation from finite-dimensional spaces and has the form of one-class support vector machine (OC-SVM) [16]. Due to the representation theorem, our formulation is reducible to a tractable, finite-dimensional problem, despite the infinite-dimensional nature. Therefore, the proposed approach enables flexible learning of dissipative properties in general nonlinear forms, without requiring explicit knowledge of the system dynamics.

For a theoretical analysis, we adopt statistical learning theory to establish bounds on generalized performance, namely a confidence lower bound on the dissipation rate and a confidence upper bound on the guaranteed  $L_2$  gain. This work exemplifies the use of operators and kernel tricks in RKHS as generic tools for learning the system behaviors.

The remainder of this paper is organized as follows. In Section 0.2, we first introduce the preliminaries of dissipative systems theory. Then in Section 0.3, data-driven modelling and estimation of dissipativity property via OC-SVM are discussed. Then, in Section 0.4.1, we derived the optimization formulations for learning in RKHS that is solved by OC-SVM, followed with generalized error analysis in Section 0.5. The proposed method is examined with application to systems with complex nonlinear dynamics in Section 0.6. Finally conclusions are given in Section 0.7.

*Notations.* Upper case letters are used to represent matrices and linear operators. Lower case letters are for scalars and column vectors. For a matrix  $A \in \mathbb{R}^{n \times n}$  whose entries are written as  $A_{i,j}, i, j \in \{1, 2, \dots, n\}$ , its trace and Frobenius norm are denoted by  $\text{tr}(A) = \sum_{i=1}^n A_{i,i}$  and  $\|A\|_F = \left(\sum_{i,j=1}^n |A_{i,j}|^2\right)^{1/2}$ , respectively. We use  $\langle A, B \rangle \triangleq \text{tr}(A^\top B)$  to denote the inner product between two matrices  $A, B \in \mathbb{R}^{n \times n}$ , where  $A^\top$  is the transpose of the matrix  $A$ . Identity matrix is denoted by  $I$ . Calligraphic letters are used for sets (e.g.,  $\mathcal{X}$ ) or Hilbert

Space (e.g.,  $\mathcal{H}$ ). Moreover, a Reproducing Kernel Hilbert Space (RKHS) generated by kernel  $K$  is denoted by  $\mathcal{H}_K$ . Capitalized Greek letters (e.g.,  $\Pi$ ) are used to denote operators, and its Hilbert-Schmidt norm is  $\|\Pi\|_{\text{HS}}$ . We use also  $a \otimes b$  to denote Kronecker product between  $a$  and  $b$ . Fourier transform of a function  $\cdot$  is  $\mathcal{F}\{\cdot\}$  and its inverse Fourier transform is  $\mathcal{F}^{-1}\{\cdot\}$ .

## 0.2 Dissipative Systems

Consider an unknown system governed by discrete time nonlinear dynamics

$$\begin{cases} x_{t+1} = f(x_t, u_t) \\ y_t = h(x_t, u_t), \end{cases} \quad (1)$$

where  $x_t \in \mathbb{R}^{n_x}$ ,  $u_t \in \mathbb{R}^{n_u}$ , and  $y_t \in \mathbb{R}^{n_y}$  denote the state, input and the output variables of the system at time  $t \geq 0$ , respectively. The maps  $f : \mathcal{X} \times \mathcal{U} \rightarrow \mathbb{R}^{n_x}$  is the drift and  $h : \mathcal{X} \times \mathcal{U} \rightarrow \mathbb{R}^{n_y}$  is the output function. It is assumed that  $\mathcal{X}$  is assumed to be compact and  $f, h$  are locally Lipschitz in  $(x, u) \in \mathcal{X} \times \mathcal{U}$ . We also assume  $f(0) = 0$ , that is, the origin is an equilibrium point of (1). For an initial condition  $x_0 \in \mathbb{R}^{n_x}$ , we assume the existence and uniqueness of the solution of (1), and denote the solution at time  $t$  by  $x(t; x_0)$ .

Through out this paper, the system dynamics  $f$  is unknown, whereas, the data samples  $x \in \mathcal{X} \subset \mathbb{R}^{n_x}$ ,  $u \in \mathcal{U} \subset \mathbb{R}^{n_u}$ , and  $y \in \mathcal{Y} \subset \mathbb{R}^{n_y}$  are available. The goal is to learn the properties of the system. Therefore, the learning approach relies only on the data sample of the system. Specifically, we focus on the dissipativity property of a system. Dissipativity is a universal property of physical and engineering systems, capturing the balance between storage function  $V : \mathbb{R}^{n_x} \rightarrow \mathbb{R}$  and external supply rate  $s : \mathbb{R}^{n_y} \times \mathbb{R}^{n_u} \rightarrow \mathbb{R}$ . From a control-theoretic perspective, dissipativity links directly to stability, performance, and controller synthesis.

Specifically, defined by Willems in [17], the change of the state dependent function (or *storage function*)  $V(x)$  can not exceed the accumulation of an input and output dependent function (or *supply rate*)  $s(z)$ . More precisely, the dissipativity property is stated in the following.

**Definition 1.** A dynamical system of the form in (1) is said to be *dissipative* in the storage function  $V(x)$  with respect to the supply rate  $s(y, u)$  if

$$V(x_{t+1}) - V(x_t) \leq s(y_t, u_t) \quad (2)$$

holds for all input  $u_t$  and state  $x_t$ ,  $0 \leq t \leq T$ .

From the definition, dissipativity is a fundamental property that connects the internal energy balance of a system with its input–output behavior. If inequality (2) holds, the storage function  $V(x)$  can be interpreted as a generalized energy of the system, while the supply rate  $s(y, u)$  captures the net inflow of energy. This perspective provides a powerful framework for *stability analysis*, since a nonincreasing  $V(x)$  implies bounded state trajectories and can certify potential convergence to the equilibrium for controlled systems. Moreover, dissipativity generalizes passivity and small-gain conditions, thereby enabling *controller synthesis* for performance objectives such as finite  $L_2$ -gain and robust regulation. For partially known

or unknown systems, learning the dissipativity property offers a model-agnostic approach for certifying desirable closed-loop properties, and provides a framework for synthesizing controllers directly from data. If the supply rate has a quadratic form

$$s(y, u) = [y^\top \ u^\top] \begin{bmatrix} \Pi_{yy} & \Pi_{yu} \\ \Pi_{yu}^\top & \Pi_{uu} \end{bmatrix} \begin{bmatrix} y \\ u \end{bmatrix}, \quad (3)$$

where  $\Pi_{yy}$ ,  $\Pi_{yu}$ ,  $\Pi_{uu}$  are symmetric matrices, then the system is said to be  $(\Pi_{yy}, \Pi_{yu}, \Pi_{uu})$ -dissipative. This concept of dissipativity is closely related to input-output  $L_2$  stability [18] [19]. More generally, a tighter upper bound of  $L_2$ -gain is given in the following theorem [7].

**Theorem 1.** *For a  $(\Pi_{yy}, \Pi_{yu}, \Pi_{uu})$ -dissipative system, if there exists positive real numbers  $\alpha$  and  $\beta$ , such that*

$$I + \alpha \Pi_{yy} \prec 0, \quad \begin{bmatrix} I + \alpha \Pi_{yy} & \alpha \Pi_{yu} \\ \alpha \Pi_{yu}^\top & \alpha \Pi_{uu} - \beta I \end{bmatrix} \preceq 0, \quad (4)$$

*then the system has an  $L_2$ -gain not greater than  $\beta^{1/2}$ .*

The above result highlights how the dissipativity property inherently encodes the Lyapunov asymptotic stability of the system as proven in [20]. This connection motivates the learning of dissipativity directly from data such that the inequality holds approximately and generalizes to unseen trajectories, as it enables data-driven certification of robustness without requiring an explicit dynamic model.

## 0.3 Dissipativity parameters and its learning: quadratic case

### 0.3.1 Dissipativity parameters

Note that the supply rate function  $s(y, u)$  in (2) is both input and output dependent. To simplify the analysis and learning of the function  $s(y, u)$ , we start with joining two variables into a single combined variable  $z$  such that

$$z \triangleq \begin{bmatrix} y \\ u \end{bmatrix} \in \mathbb{R}^{n_y+n_u}. \quad (5)$$

We assume a quadratic form for supply function and a Lyapunov function candidate for the storage function in (2), that is,

$$s(z) = z^\top Q z, \quad V(x) = x^\top P x, \quad (6)$$

where  $0 \prec Q = Q^\top \in \mathbb{R}^{(n_y+n_u) \times (n_y+n_u)}$ ,  $0 \prec P = P^\top \in \mathbb{R}^{x_n \times x_n}$ . Hence, the dissipative property in (2) is captured by

$$z_t^\top Q z_t - (x_{t+1} - x_t)^\top P (x_{t+1} - x_t) \geq 0. \quad (7)$$

**Remark 1.** The system is passive if  $n_u = n_y$ , and  $Q = \begin{bmatrix} M & S \\ S^\top & R \end{bmatrix}$  with  $M, R = 0$  and  $S = I$ .

**Definition 2.** The matrix  $Q$  and  $P$  are called the dissipativity parameters that are to be estimated.

More generally, considering a *dynamic multiplier*  $\Omega \in \mathcal{RH}_\infty^{z_n \times (y_n+u_n)} : (y_k, u_k) \rightarrow z_k$  in [21] acting on inputs and outputs separately, it yields that

$$\tilde{z}_t \triangleq \Omega(q^{-1}) \begin{bmatrix} y_t \\ u_t \end{bmatrix} = \begin{bmatrix} \Omega_y(q^{-1}) & 0 \\ 0 & \Omega_u(q^{-1}) \end{bmatrix} \begin{bmatrix} y_t \\ u_t \end{bmatrix} = \begin{bmatrix} z_{y_t} \\ z_{u_t} \end{bmatrix}. \quad (8)$$

Under the general dynamic multiplier  $\Omega$ , the system in (1) is said to be *dissipative* with respect to the quadratic supply rate  $s(\tilde{z}_k) = \tilde{z}_k^\top \tilde{Q} \tilde{z}_k$ , if for any trajectory of the aggregated dynamics  $(f, \Omega)$  starting from the origin ( $x(t_0) = 0$ ) and for any  $t > t_0$ , we have  $\tilde{z}_t^\top \tilde{Q} \tilde{z}_t - (x_{t+1} - x_t)^\top P(x_{t+1} - x_t) \geq 0$ .

**Remark 2.** A trivial dynamic multiplier  $\Omega = I$  yields a quadratic form of supply rate function

$$s(y, u) = [y^\top u^\top] Q \begin{bmatrix} y \\ u \end{bmatrix}. \quad (9)$$

From the above remark, for simplicity, we assume that  $\Omega = I$  in this paper. The resulting dissipativity coincides with (7) where  $z$  is defined in (5). The dissipativity parameters  $Q$  and  $P$  established in (7) are to be estimated.

By exploiting the system model in (1), the following theorem provides the linear matrices inequality (LMI) for model-based dissipative certification for the linearized system.

**Theorem 2.** *The linearized system around the equilibrium point of (1), i.e.,  $x_{t+1} = Ax_t + Bu_t, y_t = Cx_t$  is dissipative if there exist  $\rho > 0$ ,  $0 \prec P = P^\top, 0 \prec Q = Q^\top$  such that*

$$\begin{bmatrix} I & 0 \\ 0 & C^\top \end{bmatrix} Q \begin{bmatrix} I & 0 \\ 0 & C \end{bmatrix} - \begin{bmatrix} B^\top PB & B^\top PA \\ A^\top PB & A^\top PA - P \end{bmatrix} \succeq \rho I \quad (10)$$

holds.

*Proof.* The proof is obtained by applying  $y_t = Cx_t$  and elementary operations to (7), and enforce strong positive definiteness by introducing  $\rho$ .  $\square$

### 0.3.2 OC-SVM for dissipativity parameters

By linearizing and exploiting the system model in (1), the dissipative certificate can be obtained according to (10). More realistically, in data-driven, or model-free approach, the dynamic modeling procedure can be skipped. Instead, system properties such as the discussed dissipativity are estimated from data [7, 22, 23]. For the estimation of the dissipativity parameters  $Q$  and  $P$  as defined in (6), we suppose that  $m$  trajectories  $\{y^i, u^i\}_{i=1}^m$  and  $\{x^i\}_{i=1}^m$  are sampled independently, from a distribution of random input excitations such that the magnitude of the input satisfies a Gaussian distribution with zero mean and each magnitude lasts for a fixed duration. Formally, we denote this distribution as

a probability measure  $\mathbb{P}$ . The goal is therefore to determine valid matrices  $Q$  and  $P$  such that for all  $i$  and  $t$ , that is, from (7), the following inequality holds approximately:  $(z_t^i)^\top Q z_t^i - (x_{t+1}^i - x_t^i)^\top P (x_{t+1}^i - x_t^i) \geq 0$ . Hence, after elementary operations, the dissipativity inequality becomes  $\langle Q, z_t z_t^\top \rangle + \langle P, x_t x_t^\top - x_{t+1} x_{t+1}^\top \rangle \geq 0$ , where  $\langle \cdot, \cdot \rangle$  is inner product.

**Definition 3.** An input-output trajectory  $\{y_t, u_t\}_{t=1}^T$  that determines  $z_t = [y_t \ u_t]^\top$ , the corresponding dual dissipativity parameters refer to

$$\begin{cases} \Gamma_t \triangleq z_t z_t^\top, \\ \Delta_t \triangleq x_t x_t^\top - x_{t+1} x_{t+1}^\top, \end{cases} \quad (11)$$

for all  $t$ . Hence, with the dual dissipativity parameters of the sampled trajectories  $\{\Gamma^i\}_{i=1}^n$  and  $\{\Pi^i\}_{i=1}^n$ , we seek  $Q$  and  $P$  such that for all  $t$ ,

$$\langle Q, \Gamma_t^i \rangle - \langle P, \Delta_t^i \rangle \geq 0, \quad (12)$$

approximately. This problem is amenable to one-class support vector machine (OC-SVM), where we maximize the margin of the inequality, i.e., a nonnegative value  $\rho > 0$  such that  $\langle Q, \Gamma_t^i \rangle - \langle P, \Delta_t^i \rangle \geq \rho$  for all  $i$  and  $t$ , but penalizing the norm of the slope, i.e.,  $\|Q\|_F + \|P\|_F$ . Equivalently, the problem is to minimize  $\|Q\|_F^2 + \|P\|_F^2$  while rewarding  $\rho$ . To ensure a strong positive definiteness, we impose the constraint  $P \succeq \epsilon I$  for some scalar  $\epsilon > 0$ . Similarly, we impose supply rate to be  $\leq -\|y\|^2 + \alpha \|u\|^2$  for an  $\mathcal{L}_2$ -gain, that is  $Q \prec \begin{bmatrix} \alpha I & 0 \\ 0 & -I \end{bmatrix}$ . Therefore, the learning of the dissipativity parameters in (12) is cast as the following optimization problem:

$$\min_{P, Q, \rho, \alpha} \left\| \begin{bmatrix} Q & 0 \\ 0 & P \end{bmatrix} \right\|_F^2 - \rho + \lambda \alpha \quad (13)$$

$$s.t. \quad P \succeq \epsilon I, \quad (14)$$

$$Q \prec \begin{bmatrix} \alpha & 0 \\ 0 & -I \end{bmatrix}, \alpha > 0, \quad (15)$$

$$\left\langle \begin{bmatrix} Q & 0 \\ 0 & P \end{bmatrix}, \begin{bmatrix} \Gamma_t & 0 \\ 0 & \Delta_t \end{bmatrix} \right\rangle \geq \rho, 1 \leq t \leq T, \quad (16)$$

where  $\Gamma_t$  and  $\Delta_t$  are in (11),  $\lambda$  serves as a weighting parameter that characterizes the trade-off between the dissipativity margin and the  $\mathcal{L}_2$ -gain. A larger  $\lambda$  penalizes higher  $\alpha$  more strongly, thereby enforcing a stricter  $\mathcal{L}_2$ -gain, that favors systems with smaller input-output amplification. Conversely, a smaller  $\lambda$  relaxes this restriction, prioritizing the feasibility margin  $\rho$  and fit the dissipativity inequalities closely to the data. Hence,  $\lambda$  controls the trade-off between conservativeness and data adaptivity. Large  $\lambda$  yields a more conservative estimate of dissipativity parameters, while smaller  $\lambda$  leads to tighter data fitting at the risk of weaker generalization.

## 0.4 Dissipativity learning in RKHS

The parametric dissipativity learning framework in section 0.3 constrains the supply and storage function  $s$  and  $V$ , to have quadratic forms as in (6). Although this yields a tractable formulation, it limits expressiveness and may inadequately capture the nonlinear behavior of complex systems. To address this limitation, the framework is extended to *Reproducing Kernel Hilbert Spaces* (RKHSs) in this section, where  $s$  and  $V$  are sought in rich, infinite-dimensional function spaces. The RKHS structure defines functions through inner products with kernel-induced feature maps, enabling flexible and data-driven representations of dissipativity properties while preserving convexity and computational tractability of the learning problem.

Hence, we first discuss various preliminary results necessary for learning dissipativity properties in RKHS in section 0.4.1. The nonparametric dissipativity learning in RKHS is introduced in section 0.4.2.

### 0.4.1 Reproducing Kernel Hilbert Spaces

We first introduce reproducing kernel Hilbert spaces with its properties.

**Definition 4.** [24] A reproducing kernel Hilbert Space (RKHS) is a Hilbert Space  $\mathcal{H}_k$  of functions  $f : \mathcal{X} \rightarrow \mathbb{R}$  with a reproducing kernel  $k : \mathcal{X}^2 \rightarrow \mathbb{R}$  where  $k(x, \cdot) \in \mathcal{H}$  and  $f(x) = \langle k(x, \cdot), f \rangle$ .

The RKHS is discussed in more details in the following. Consider a kernel function  $k(x, \cdot)$  with two variables  $x, \cdot \in \mathcal{X} \times \mathcal{X}$ . RKHS is a function space consisting functions  $f$  such that  $\mathcal{H}_k \triangleq \{f(\cdot) = \sum_{i=1}^n \alpha_i k(x_i, \cdot)\}$ . This construction equips  $\mathcal{H}_k$  with an inner product  $\langle f, g \rangle_{\mathcal{H}_k}$  such that the reproducing property  $f(x) = \langle k(x, \cdot), f \rangle_{\mathcal{H}_k}$  holds for all  $f \in \mathcal{H}_k$  and  $x \in \mathcal{X}$ . This property implies that the evaluation of any function at a point  $x$  can be expressed as an inner product with the kernel section  $k(x, \cdot)$ , that is crucial for learning problems to be carried out entirely in terms of kernel evaluations.

Specifically, instead of assuming quadratic forms as in the parametric setting, we lift the data points  $z$  and  $x$  into (possibly infinite-dimensional) feature spaces  $\mathcal{H}_z$  and  $\mathcal{H}_x$  using feature map in the following assumption.

**Assumption 1.** Suppose that feature maps that takes data into RKHS are  $\phi : \mathcal{Z} \rightarrow \mathcal{H}_z$  and  $\psi : \mathcal{X} \rightarrow \mathcal{H}_x$ , respectively, where  $\mathcal{H}_z$  and  $\mathcal{H}_x$  are the corresponding RKHSs.

**Remark 3.** The feature map enables linear operation in RKHSs to represent nonlinear functions in the original space. From the defining property of RKHSs, inner products between features  $\phi(z)$  or  $\psi(x)$  define similarities through the kernel as follows

$$\langle \phi(z), \phi(z') \rangle_{\mathcal{H}_z} = \kappa(z, z'). \quad (17)$$

Similarly to (12) and (11), we redefine the dual dissipativity parameters in the following.

**Definition 5.** An input-output trajectory  $\{y_t, u_t\}_{t=1}^T$  that determines  $z_t = [y_t \ u_t]^\top$ , the corresponding dual dissipativity parameters refer to

$$\begin{cases} \Gamma_t : \phi(z_t) \otimes \phi(z_t) \\ \Delta_t : \psi(x_t) \otimes \psi(x_t) - \psi(x_{t+1}) \otimes \psi(x_{t+1}), \end{cases} \quad (18)$$

where  $\Gamma_t$  and  $\Delta_t$  are Hilbert-Schmidt operators in  $\mathcal{H}_z$  and  $\mathcal{H}_x$ , respectively.

**Remark 4.**  $\Gamma_t$  is a rank one operator in the space of Hilbert-Schmidt operators on  $\mathcal{H}_z$ .  $\Delta_t$  is at most rank two operator.

**Definition 6.** The supply rate and storage functions are defined as

$$\begin{cases} s(z) \triangleq \langle \phi(z), \Pi \phi(z) \rangle \\ V(x) \triangleq \langle \psi(x), P \psi(x) \rangle, \end{cases} \quad (19)$$

where  $\Pi : \mathcal{H}_z \rightarrow \mathcal{H}_z$  and  $P : \mathcal{H}_x \rightarrow \mathcal{H}_x$  are Hilbert-Schmidt operators and new dissipativity parameters.

With the above assumption,  $s(z)$  and  $V(x)$  can be expressed as inner products with operators acting on these features, and regularization can be performed via the Hilbert–Schmidt norms of these operators. This construction preserves convexity of the learning problem while allowing highly nonlinear, data-adaptive representations of dissipativity, which leads directly to the formulation in Section 0.4.2.

#### 0.4.2 OC-SVM for Nonparametric Dissipativity Learning

The RKHS framework generalizes the parametric quadratic forms by lifting the variables into high-dimensional feature spaces. The following theorem provides the dissipativity learning problem in the RKHS framework.

**Theorem 3.** *Suppose Assumption 1 holds. The dissipativity in (2) is captured by the following inequality with margin  $\rho$  such that*

$$\langle \Pi, \Gamma_t \rangle + \langle P, \Delta_t \rangle \geq \rho, \quad (20)$$

for all sampled data points.

*Proof.* From (18), (19) and Assumption 1, it follows that for all  $t$ ,  $s(z_t) + V(x_t) - V(x_{t+1}) \geq 0$ , that is

$$\begin{aligned} & \langle \phi(z_t), \Pi \phi(z_t) \rangle - (\langle \psi(x_{t+1}), P \psi(x_{t+1}) \rangle - \langle \psi(x_t), P \psi(x_t) \rangle) \\ &= \text{tr}(\Pi \phi(z_t) \otimes \phi(z_t)) + \text{tr}(P \psi(x_t) \otimes \psi(x_t)) - \text{tr}(P \psi(x_t) \otimes \psi(x_{t+1})), \end{aligned} \quad (21)$$

which can be written as in (21) from the definition of  $\Gamma_t$  and  $\Delta_t$  in (18). This completes the proof.  $\square$

Theorem 3 provides an inequality to be satisfied for dissipativity. We posed certain constraints on the operators to ensure the robustness of the dissipativity learning.

Firstly, the operator  $P$  that corresponds to the storage function  $V(x) = \langle \psi(x), P\psi(x) \rangle_{\mathcal{H}_x}$ , must be a positive operator in  $\mathcal{H}_x$ , i.e.,  $P \in \text{HS}_+(\mathcal{H}_x)$ . This is enforced by constraining  $P - \epsilon P_0 \geq 0, \epsilon > 0$ , where  $P_0$  can be an identity operator. This ensures  $P$  to be strictly positive. Secondly, we introduce two operators  $\Upsilon_0$  and  $\Upsilon_1$ , that act on the feature map  $\phi(z)$  and jointly upper bound the supply rate as  $s(z) \leq \langle \phi(z), \Upsilon_0 \phi(z) \rangle + \alpha \langle \phi(z), \Upsilon_1 \phi(z) \rangle, \alpha > 0$ . Note that in this bound, we can specify that  $\Upsilon_1$  acts on the output feature  $\phi(y)$ . To this end,  $\alpha$  characterizes the bound for  $\mathcal{L}_2$ -gain.

The representer theorem makes the operators  $\Pi, P$  lie in the finite span of the training data  $\{\varphi(z), \psi(x)\}$ , thus reducing the infinite-dimensional optimization problem to a finite-dimensional convex program. This makes nonparametric dissipativity learning both expressive and tractable, allowing us to capture highly nonlinear energy-exchange properties of the system while maintaining computational feasibility.

Follow the structure of OC-SVM for learning dissipativity parameters in section 0.3.2, we introduce a margin  $\rho$  of the inequality of dissipativity,  $\rho > 0$  and it yields an optimization problem for  $1 \leq t \leq T$ :

$$\begin{aligned} \min_{\Pi, P, \rho, \alpha} \quad & \frac{1}{2} (\|\Pi\|_{\text{HS}}^2 + \|P\|_{\text{HS}}^2) - \rho + \lambda \alpha \\ \text{s.t.} \quad & P - \epsilon P_0 \geq 0, \\ & -\Pi - \Upsilon_0 + \alpha \Upsilon_1 \geq 0, \alpha > 0 \\ & \langle \Pi, \Gamma_t \rangle + \langle P, \Delta_t \rangle \geq \rho, \end{aligned} \tag{22}$$

where  $\Upsilon_0$  and  $\Upsilon_1$  act on the features of  $y$  and  $u$ , respectively.

#### 0.4.3 The Finite-dim Optimization

The dissipativity learning problem in (22) is posed in terms of operators  $P, \Pi$  acting on infinite-dimensional Hilbert spaces, which makes the problem intractable for direct computation. To bridge this gap, we leverage the representer property of RKHS operators and restrict  $P$  and  $\Pi$  to finite-rank forms spanned by the available data, thereby expressing all operator norms and dissipativity constraints through kernel Gram matrices. This reduction replaces inner products in feature space with kernel evaluations, transforming trace and bilinear operator terms into quadratic forms involving  $\Phi, \Psi, \Phi_y$ , and  $\Phi_u$ . As a result, the original operator-based OCSVM formulation is equivalently rewritten as a finite-dimensional constrained optimization problem over the coefficient matrices  $p$  and  $\pi$ , together with the slack margin  $\rho$  and gain parameter  $\alpha$ . This step ensures that the infinite-dimensional dissipativity learning problem is now amenable to practical computation, while retaining the expressive power of the RKHS-based formulation. That is, the operators play a central role in both the cost function and the dissipativity constraints, and their treatment in RKHS allows one to bridge the infinite-dimensional formulation with a finite-dimensional optimization problem.

Specifically, to obtain a finite-dimensional optimization formulation of (22), we interpret both the objective and the constraints in the RKHS setting.

## Operators in the cost

We begin with the Hilbert-Schmidt (H-S) norm of an operator  $\Pi : \mathcal{H}_z \rightarrow \mathcal{H}_z$ .

**Definition 7.** For a Hilbert-Schmidt operator  $\Pi$  on a Hilbert space  $\mathcal{H}_z$ , its norm is defined as  $\|\Pi\|_{\text{HS}}^2 \triangleq \sum_k \|\Pi e_k\|^2$ , where  $e_k$  is an orthonormal basis of the Hilbert space. Equivalently, if  $\Pi$  admits a decomposition  $\Pi = \sum_r u_r \otimes v_r$  with  $u_r, v_r \in \mathcal{H}_z$ , then

$$\|\Pi\|_{\text{HS}}^2 = \sum_{r,s} \langle u_r, u_s \rangle \langle v_r, v_s \rangle. \quad (23)$$

In practice, we restrict  $\Pi$  and  $P$  to a finite-dimensional subspace of  $\mathcal{H}_z$  and  $\mathcal{H}_x$  spanned by the training samples. That is,  $\Pi$  and  $P$  are assumed to be finite-rank operators of rank at most  $T$ , expressed as  $\Pi = \sum_{i,j}^T \pi_{i,j} \phi_i \otimes \phi_j$ ,  $P = \sum_{i,j}^T p_{i,j} \psi_i \otimes \psi_j$ , where  $\pi_{i,j}$  and  $p_{i,j}$  are coefficients that correspond to the basis  $\phi_i \otimes \phi_j$  and  $\psi_i \otimes \psi_j$ , respectively. Hence, by construction, the coefficients  $\pi_{i,j}$  and the feature vectors  $\phi(z_i)$  define  $\Pi$  completely on the subspace spanned by  $\{\phi(z_1), \dots, \phi(z_T)\}$ . Therefore, the Hilbert-Schmidt norm of  $\Pi$  is as follows:

$$\|\Pi\|_{\text{HS}}^2 = \text{tr} (\pi^\top \Phi \pi \Phi^\top), \quad (24)$$

where  $\Phi$  is the Gram matrix constructed by data;  $\Phi_{i,j}$  is the element on  $i$ -th row and  $j$ -th column, i.e.  $\Phi_{i,j} \triangleq \kappa(\phi(z_i), \phi(z_j)) = \langle \phi(z_i), \phi(z_j) \rangle$  as shown in (17); (24) holds from the symmetry of  $\Phi$ . Similarly, the coefficients  $p_{i,j}$  and the feature vectors  $\phi(x_i)$  define  $P$  completely on the subspace spanned by  $\{\phi(x_1), \dots, \phi(x_T)\}$  and it follows that

$$\|P\|_{\text{HS}}^2 = \text{tr} (p^\top \Psi p \Psi^\top). \quad (25)$$

## Operators in the dissipativity inequality constraint

For the dissipativity inequality  $\langle \Pi, \Gamma_t \rangle + \langle P, \Delta_t \rangle \geq \rho$  as in Theorem 3, we restrict the operators  $\Gamma_t$  and  $\Delta_t$  to be spanned by the training samples. From (18), it follows that

$$\langle \Pi, \Gamma_t \rangle = \sum_{i,j}^T \pi_{i,j} \Phi_{j,t} \Phi_{i,t}, \quad (26)$$

where (26) holds from the trace of a rank one operator equals the inner product of its factors, i.e.,  $\text{tr}(\phi(z_i) \otimes \phi(z_t)) = \langle \phi(z_t), \phi(z_i) \rangle$ . Similarly,

$$\langle P, \Delta_t \rangle = \sum_{i,j}^T p_{i,j} (\Psi_{j,t} \Psi_{i,t} - \Psi_{j,t+1} \Psi_{i,t+1}). \quad (27)$$

Therefore, the dissipativity inequality is

$$\sum_{i,j}^T \pi_{i,j} \Phi_{j,t} \Phi_{i,t} + \sum_{i,j}^T p_{i,j} (\Psi_{j,t} \Psi_{i,t} - \Psi_{j,t+1} \Psi_{i,t+1}) \geq \rho. \quad (28)$$

### Operators in the constraints for bounds of the storage and $L_2$ gain

In learning dissipativity parameters in (13), it is assumed that the storage is quadratic with baseline  $V(x) \geq \epsilon V_0(x)$  with  $V_0(x) \stackrel{\Delta}{=} \|x\|^2$ . Similarly, let  $V(x) \stackrel{\Delta}{=} \langle \psi(x), P\psi(x) \rangle$  and  $P_0 \stackrel{\Delta}{=} \sum_i \sum_j c_i c_j \psi(x_i) \otimes \psi(x_j) \in \mathcal{H}_x$  such that  $P \geq \epsilon P_0$ . From (19), it follows that for all  $t$ , the constraint that corresponds to the energy storage is

$$\begin{aligned} & \langle \psi(x_t), P\psi(x_t) \rangle - \langle \psi(x_t), P_0\psi(x_t) \rangle \\ &= \text{tr}(P\psi(x_t) \otimes \psi(x_t)) - \text{tr}(P_0\psi(x_t) \otimes \psi(x_t)) \\ &= \text{tr}\left(\sum_{i,j} p_{i,j} \psi(x_i) \otimes \psi(x_j) \psi(x_t) \otimes \psi(x_t)\right) - \text{tr}\left(\sum_{i,j} c_i c_j \psi(x_i) \otimes \psi(x_j) \psi(x_t) \otimes \psi(x_t)\right) \\ &= \text{tr}\left(\sum_{i,j} p_{i,j} \Psi_{j,t} \psi(x_i) \otimes \psi(x_t)\right) - \text{tr}\left(\sum_{i,j} c_i c_j \Psi_{j,t} \psi(x_i) \otimes \psi(x_t)\right) \\ &= \sum_{i,j} (p_{i,j} - c_i c_j) \Psi_{j,t} \Psi_{i,t}. \end{aligned} \quad (29)$$

To obtain the linear coefficients  $c_i, i \in \{1, 2, \dots, T\}$  in the operator  $P_0$ , we assume that

$$\sqrt{V_0(x)} = \left\langle \sum_{i=1}^T c_i \psi(x_i), \psi(x) \right\rangle, \quad (30)$$

that is, the function  $V_0$  is approximated by a linear summation of  $(\sqrt{V_0})(\cdot) = \sum_{i=1}^T c_i \Psi(x_i, \cdot)$ . Therefore, with (30), we admit the following assumption.

**Assumption 2.** For all  $t = \{1, 2, \dots, T\}$ , the following holds for all  $c_i$

$$\sqrt{V_0(x_t)} = \sum_{i=1}^T c_i \Psi_{i,t} = \|x_t\|. \quad (31)$$

The coefficients  $c \stackrel{\Delta}{=} [c_1, c_2, \dots, c_T]^\top \in \mathbb{R}^T$  can be estimated as

$$\hat{c} = \Psi^{-1} [\|x_1\|, \|x_2\|, \dots, \|x_T\|]^\top. \quad (32)$$

For the supply function  $s(z)$ , let  $\Upsilon_0, \Upsilon_1 \in \mathcal{H}_z$  be operators applying on the outputs  $y$  and inputs  $u$ , respectively. Similar to (31), we assume that

$$\|y_t\| \approx \sum_{i=1}^T v_i^y \Phi_{i,t}^y, \quad \|u_t\| \approx \sum_i v_i^u \Phi_{i,t}^u. \quad (33)$$

Consequently, we make the following kernel interpolation approximation scheme. For all  $t = \{1, 2, \dots, T\}$ ,

$$\|y_t\|^2 = \sum_{i,j}^T v_i^y v_j^y \Phi_{i,t}^y \Phi_{j,t}^y, \quad \|u_t\|^2 = \sum_{i,j}^T v_i^u v_j^u \Phi_{i,t}^u \Phi_{j,t}^u. \quad (34)$$

The coefficient estimates are

$$\hat{v}^y = (\Phi^y)^{-1} [\|y_1\|, \|y_2\|, \dots, \|y_T\|]^T, \quad (35)$$

$$\hat{v}^u = (\Phi^u)^{-1} [\|u_1\|, \|u_2\|, \dots, \|u_T\|]^T, \quad (36)$$

where  $\Phi_{ij}^y \triangleq \kappa_y(y_i, y_j)$  and  $\Phi_{ij}^u \triangleq \kappa_u(y_i, y_j)$ . Hence, the estimate  $\hat{\Upsilon}_0$  and  $\hat{\Upsilon}_1$  are both finite-rank Hilbert-Schmidt operators in the corresponding RKHS such that

$$\hat{\Upsilon}_0 = \sum_{i=1}^T \sum_{j=1}^T \hat{v}_i^y \hat{v}_j^y \phi(y_i) \otimes \phi(y_j), \quad (37)$$

$$\hat{\Upsilon}_1 = \sum_{i=1}^T \sum_{j=1}^T \hat{v}_i^u \hat{v}_j^u \phi(u_i) \otimes \phi(u_j). \quad (38)$$

Note that the estimated  $\hat{\Upsilon}_0$  and  $\hat{\Upsilon}_1$  are constrained to the available data, which causes estimation error with respect to  $\Upsilon_0$  and  $\Upsilon_1$ , respectively.

The  $L_2$  gain-like inequality enforces that  $-\hat{\Pi} + \hat{\Upsilon}_0 + \alpha \hat{\Upsilon}_1 \geq 0$ , and for all  $t$ ,

$$-\langle \phi(y_t), \hat{\Upsilon}_0 \phi(y_t) \rangle + \alpha \langle \phi(u_t), \hat{\Upsilon}_1 \phi(u_t) \rangle \geq \langle \phi(z_t), \hat{\Pi} \phi(z_t) \rangle. \quad (39)$$

Hence, the following holds

$$\sum_{i,j}^T \pi_{i,j} \Phi_{j,t} \Phi_{i,t} \leq - \sum_{i,j}^T \hat{v}_i^y \hat{v}_j^y \Phi_{i,t}^y \Phi_{j,t}^y + \alpha \sum_{i,j}^T \hat{v}_i^u \hat{v}_j^u \Phi_{i,t}^u \Phi_{j,t}^u. \quad (40)$$

Based on the finite-rank operator expansions the Hilbert–Schmidt operators  $P$  and  $\Pi$  are restricted to the finite spans of the training data, so that their actions and norms can be expressed entirely in terms of the kernel Gram matrices  $\Phi$ ,  $\Phi^y$ ,  $\Phi^u$  and  $\Psi$ . Specifically, the Hilbert–Schmidt norms of these operators become quadratic forms of the coefficient matrices  $p$  and  $\pi$ , yielding the trace expressions in (24) and (25), respectively.

The dissipativity inequalities involving  $\Pi$ ,  $P$ ,  $\Gamma_t$ , and  $\Delta_t$  reduce to algebraic constraints represented through kernel evaluations as in (26) and (27). The operator bounds enforcing the positivity of the storage function and the  $\mathcal{L}_2$ -gain condition are similarly reformulated through data-dependent terms as shown in (29) and (37), where the reference operators  $\hat{\Upsilon}_0$  and  $\hat{\Upsilon}_1$  are constructed from the regression coefficients  $\hat{v}^y$  and  $\hat{v}^u$ , and the baseline storage operator is determined by  $\hat{c}$ .

From these results, the infinite-dimensional operator optimization problem in the RKHS

is reduced to the following finite-dimensional convex optimization problem:

$$\begin{aligned}
& \min_{p, \pi, \rho, \alpha} \frac{1}{2} (\text{tr}(\pi^\top \Phi \pi \Phi^\top) + \text{tr}(p^\top \Psi p \Psi^\top)) - \rho + \lambda \alpha \\
& \text{s.t. } \rho \geq 0, \alpha \geq 0, \\
& \sum_{i,j}^T (p_{i,j} - \hat{c}_i \hat{c}_j) \Psi_{j,t} \Psi_{i,t} \geq 0, \\
& \sum_{i,j}^T \alpha \hat{v}_i^u \hat{v}_j^u \Phi_{i,t}^u \Phi_{j,t}^u - \hat{v}_i^y \hat{v}_j^y \Phi_{i,t}^y \Phi_{j,t}^y \geq \sum_{i,j}^T \pi_{i,j} \Phi_{j,t} \Phi_{i,t}, \\
& \sum_{i,j}^T \pi_{i,j} \Phi_{j,t} \Phi_{i,t} + p_{i,j} (\Psi_{j,t} \Psi_{i,t} - \Psi_{j,t+1} \Psi_{i,t+1}) \geq \rho.
\end{aligned} \tag{41}$$

Note that the parameters  $\hat{c}$ ,  $\hat{v}^u$ , and  $\hat{v}^y$  are obtained from the training data samples according to (32) and (35).

## 0.5 Error Analysis for Learning in RKHS

In KRR, we assume that data samples  $\bar{z}_i, i \in \{1, 2, \dots, m\}$  are drawn from a probability measure  $\rho$  on  $\tilde{\mathcal{Z}} = \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}$ . The regression function to be solved is

$$f_\rho(x) = \int_Y y d\rho(y|x), x \in \bar{\mathcal{X}}, \tag{42}$$

where  $\rho(y|x)$  is the conditional distribution at  $x$  induced by  $\rho$ .

Let  $K : \bar{\mathcal{X}} \times \bar{\mathcal{X}} \rightarrow \mathbb{R}$  be a *Mercer kernel*, that is, the matrix formed by  $K_{i,j} = K(\bar{x}_i, \bar{x}_j), i, j \in \{1, 2, \dots, m\}$  is positive semidefinite for any finite set of distinct data points  $\bar{x}_i, i \in \{1, 2, \dots, m\}$ . We denote the generated RKHS by kernel  $K$  as  $\mathcal{H}_K$ . Hence, the regression algorithm we investigate is a Tikhonov regularization scheme associated with  $K$ , that is,

$$f_{z,\lambda} \triangleq \arg \min_{f \in \mathcal{H}_K} \left\{ \frac{1}{m} \sum_{i=1}^m (f(x_i) - y_i)^2 + \lambda \|f\|_K^2 \right\} \tag{43}$$

The following theorem provides an upper bound for approximating  $f_\rho$  through the learning scheme in (43) under the condition that  $f_\rho$  is in the range of  $L_K^r$ .

**Theorem 4.** [25] Let  $\bar{z}$  be randomly drawn according to  $\rho$  and satisfy  $|\bar{y}| \leq M$  almost surely. Assume that, for some  $0 < r \leq 1$ ,  $f_\rho$  is in the range of  $L_K^r$ . Take the regularization parameter as  $\lambda = (3\kappa M / \|L_K^{-r} f_\rho\|_\rho)^{2/(1+2r)} m^{-1/(1+2r)}$ , where  $\kappa \triangleq \sqrt{\sup_{x \in \bar{\mathcal{X}}} K(x, x)}$ . Then, for any  $0 < \delta < 1$ , with confidence  $1 - \delta$ ,

$$\begin{aligned}
& \|f_{z,\lambda} - f_\rho\|_K \\
& \leq 4 \log\left(\frac{2}{\delta}\right) (3\kappa M)^{\frac{2r-1}{2r+1}} \|L_K^{-r} f_\rho\|_\rho^{2/(1+2r)} m^{-\frac{2r-1}{4r+2}}.
\end{aligned} \tag{44}$$

The above Theorem provides a quantitative learning rate for the KRR approximation in the RKHS. In our context, the operators  $\Upsilon_0, \Upsilon_1$  are estimated through such regressions as in (37).

The function  $f_\rho$  that we are approximating is  $f(\cdot) = \|\cdot\|^2$  as in (45). In the proposed dissipativity learning in RKHS, we estimate the Hilbert-Schmidt operators  $\Upsilon_0$  and  $\Upsilon_1$  by Kernel ridge regression (KRR) so that  $\langle \phi(y), \Upsilon_0 \phi(y) \rangle = \|y\|^2$  and  $\langle \phi(u), \Upsilon_1 \phi(u) \rangle = \|u\|^2$ . Let the learned operators be  $\hat{\Upsilon}_0$  and  $\hat{\Upsilon}_1$  and suppose that, for some norms on functions of  $y$  and  $u$

$$\left\| \left\langle \phi(y), \hat{\Upsilon}_0 \phi(y) \right\rangle - \|y\|^2 \right\| \leq \epsilon_0, \quad (45)$$

$$\left\| \left\langle \phi(u), \hat{\Upsilon}_1 \phi(u) \right\rangle - \|u\|^2 \right\| \leq \epsilon_1. \quad (46)$$

Note that  $\Gamma_t \triangleq \phi(z_t) \otimes \phi(z_t) \in \text{HS}(\mathcal{H}_z)$  and  $\Delta_t \triangleq \psi(x_t) \otimes \psi(x_t) - \psi(x_{t+1}) \otimes \psi(x_{t+1}) \in \text{HS}(\mathcal{H}_x)$ . Denote by  $(\hat{\Pi}, \hat{P}, \hat{\rho})$  any feasible solution of the optimization so that

$$\langle \hat{\Pi}, \Gamma_t \rangle + \langle \hat{P}, \Delta_t \rangle \geq \hat{\rho}, \forall t. \quad (47)$$

Then, when generalized to the entire population,  $\Gamma$  and  $\Delta$  are random operators in  $\mathcal{H}_z$  and  $\mathcal{H}_x$ , respectively. We assume that with probability  $1 - \delta$ ,

$$\langle \hat{\Pi}, \Gamma \rangle + \langle \hat{P}, \Delta \rangle \geq \hat{\rho} - \epsilon_\rho, \quad (48)$$

for some generalization error  $\epsilon_\rho \geq 0$ . Therefore, from (39) and (45), the propagation of operator errors to the supply bound is

$$\begin{aligned} & \left\langle \phi(z_t), \hat{\Pi} \phi(z_t) \right\rangle \\ & \leq - \left\langle \phi(y_t), \hat{\Upsilon}_0 \phi(y_t) \right\rangle + \alpha \left\langle \phi(u_t), \hat{\Upsilon}_1 \phi(u_t) \right\rangle \leq -\|y\|^2 + \alpha\|u\|^2 + \epsilon_0 + \epsilon_1. \end{aligned} \quad (49)$$

Assume  $\|\Gamma\| \geq 1$  and (48) holds with probability  $1 - \delta$ . Let

$$\Pi = \hat{\Pi} + \max(\hat{\rho} - \epsilon_\rho, 0)I, \quad (50)$$

where  $I$  is the identity on  $\mathcal{H}_z$ . Then,

$$\langle \Pi, \Gamma \rangle + \langle \hat{P}, \Delta \rangle \geq 0. \quad (51)$$

Combining (51) with (49) yields that, for any  $(y, u)$ ,

$$\langle \phi(z), \Pi \phi(z) \rangle \leq -\|y\|^2 + \alpha\|u\|^2 + \bar{\epsilon}, \quad (52)$$

where  $\bar{\epsilon} \triangleq \epsilon + \max(\hat{\rho} - \epsilon_\rho, 0)$ .

**Remark 5.** The total error  $\bar{\epsilon}$  in (52) decomposes into: (i) approximation errors  $\epsilon$  from approximating via KRR, and (ii) the generalization error  $\epsilon_\rho$  from empirical margin  $\hat{\rho}$  in (47) to a population guarantee.

The following theorem provides the total error bound for  $\bar{\epsilon}$ .

**Theorem 5.** *Let Assumption 3, 4, 5 holds. Let  $\bar{z}$  be randomly drawn according to  $\rho$  and satisfy  $|\bar{y}| \leq M$  almost surely. For any  $1 < \delta < 1$ , with confidence  $1 - \delta$ , the total approximation error of the proposed dissipativity learning in RKHS is bounded as*

$$\bar{\epsilon} \leq \inf_{\epsilon > 0} \left\{ 2\epsilon + 4 \log \left( \frac{2}{\delta} \right) (3\kappa M)^{\frac{1}{3}} c_K c_{f,s} \|l_{\pi,s,\theta}\|_{L^2} m^{-\frac{1}{6}} \right\} + \frac{1}{m} \mathcal{O} \left( \log \frac{1}{\delta} + \epsilon^{\frac{1}{2}} \log m + \frac{1}{\epsilon^2} \log \frac{1}{\epsilon} \right).$$

*Proof.* The proof follows from characterizing the error bound for KRR and the generalization error in Section 0.5.1 and in Section 0.5.2, respectively.  $\square$

### 0.5.1 Approximation Error from KRR

**Assumption 3.** Assume that  $K_\theta(x, y)$  is a radial kernel and  $\int_{\mathbb{R}^d} K(x) dx = 1$

**Assumption 4.** Suppose that (1)  $|\mathcal{F}\{f\}(\xi)| \leq \frac{c_{f,s}}{(1+|\xi|^2)^{s/2}}$  almost everywhere  $\forall \xi \in \mathbb{R}^n$ ; (2) there exists  $l_{\pi,s,\theta} \in L^2(\mathbb{R}^d)$ , such that  $\int_{\mathbb{R}^d} \frac{|\mathcal{F}\{\pi^\theta\}(\xi-\eta)|}{(1+\|\eta\|^2)^{s/2}} d\eta \leq |\mathcal{F}\{K\}(\xi)| \cdot |l_{\pi,s,\theta}(\xi)|$  holds; (3)  $\sup_{\xi \in \mathbb{R}^d} |\mathcal{F}\{K_\theta\}(\xi)| = c_K < \infty$ .

**Assumption 5.** Suppose that (1)  $\int_{|y-x|>\eta_\theta} |K_\theta(x-y)| dy < \frac{\epsilon}{6M_f M_\pi^\theta}$ , where  $M_f \triangleq \sup_{x \in \text{supp } \pi} |f(x)|$ ,  $M_\pi^\theta = \sup_{x \in \mathbb{R}^d} |\pi(x)|$ ; (2) when  $\theta$  is sufficiently small and  $|y-x| \leq \eta_\theta$ ,  $|f(y) - f(x)| \leq L_f \eta_\theta$  holds for all  $x \in \text{supp } \pi$  and  $|(\pi(y))^\theta - (\pi(x))^\theta| \leq L_\pi \eta_\theta$  holds for all  $x \in \text{supp } f$ ; (3)  $\int_{\mathbb{R}^d} |K(t)| dt \leq M_K$ ; (4)  $\forall x \in \text{supp } f$ ,  $|\pi(x)| \in [m_\pi, M_\pi]$ , and  $\theta \leq \{\frac{\log(1-\frac{\epsilon}{3M})}{\log m_\pi}, \frac{\log(1+\frac{\epsilon}{3M})}{\log M_\pi}\}$ ,  $M \geq |f(x)|$ .

The following theorem provides the approximation error from KRR.

**Theorem 6.** *Let Assumption 3, 4, 5 holds. Let  $\bar{z}$  be randomly drawn according to  $\rho$  and satisfy  $|\bar{y}| \leq M$  almost surely. Then,  $\forall \epsilon > 0$ ,  $\exists \theta_\epsilon \in (0, 1)$  such that  $f_{\theta_\epsilon} \in \text{Range}(L_K)$  and  $\lim_{\theta_\epsilon \rightarrow 0} \|f_{\theta_\epsilon} - f\|_\infty = \epsilon$  hold. Moreover, for any  $0 < \delta < 1$ , with confidence  $1 - \delta$ , the KRR approximation error is bounded as follows:*

$$\begin{aligned} & \|f - \hat{f}^{\text{KRR}}\|_\infty \\ & \leq \inf_{\epsilon > 0} \left\{ 2\epsilon + 4 \log \left( \frac{2}{\delta} \right) (3\kappa M)^{\frac{1}{3}} c_K c_{f,s} \|l_{\pi,s,\theta}\|_{L^2} m^{-\frac{1}{6}} \right\}, \end{aligned} \tag{53}$$

where  $\kappa \triangleq \sqrt{\sup_{x \in \mathcal{X}} K(x, x)}$ .

*Proof.* The function that we are approximating is  $f(\cdot) = \|\cdot\|^2$  as in (45). Suppose  $f(x) \in \text{Range}(L_K^r)$ . Then, there exists  $\varphi \in L_\rho^2$  such that  $f(x) = \int_{\mathbb{R}^d} K(x, y) \varphi(y) d\rho(y)$ . Equivalently,  $f \in \mathcal{C}^s$  follows from  $K(\cdot, \cdot) \in \mathcal{C}^s$ . Note that  $f(x) = \|x\|^2 \in \mathcal{C}^1$  holds for all  $K(\cdot, \cdot)$ . Hence,  $f(x) \notin \text{Range}(L_K^r)$ . Consequently, the induced approximation error  $\epsilon_0, \epsilon_1$  when approximating  $\Upsilon_0$  and  $\Upsilon_1$  in (45) are not fully aligned with Theorem 4. To this end, we break the proof into two parts. Firstly, we construct a parameterized function  $f_\theta$  such that  $f_\theta \in \text{Range}(L_K^r)$ ,

and then, ensure that the function  $f_\theta$  satisfy  $\lim_{\theta \rightarrow 0} \|f_\theta - f\|_\infty = \epsilon, \forall \epsilon > 0$ . From the definition of an integral operator  $L_K : \mathcal{H}_K \rightarrow \mathcal{H}_K$

$$L_K(f_\rho)(x) \triangleq \int_{\mathcal{X}} K(x, y) f_\rho(y) d\rho_X(y), \quad (54)$$

the range of  $L_K^r$  is

$$\text{Range}(L_K^r) \triangleq \{ f = L_K^r g \mid g \in \mathcal{H}_K \}. \quad (55)$$

Thus, we aim to construct

$$\begin{aligned} f_\theta &\triangleq L_{K_\theta} f \\ \text{s.t. } f_\theta &\in \text{Range}(L_K), \\ \lim_{\theta \rightarrow 0} \|f_\theta - f\|_\infty &= \epsilon, \forall \epsilon > 0 \end{aligned} \quad (56)$$

where  $L_{K_\theta} f \triangleq \int K_\theta(x, y) f(y) d\rho(y)$ ,  $d\rho(x) = (\pi(x))^\theta dx, \theta \in (0, 1]$ , and  $\pi(x) \geq 0$  is measurable on  $\mathbb{R}^d$ . Then, the approximation error is bounded as follows:

$$\|f - \hat{f}^{\text{KRR}}\|_\infty \leq \|f - f_\theta\|_\infty + \|\hat{f} - \hat{f}_\theta\|_\infty + \|f_\theta - \hat{f}_\theta\|_\infty. \quad (57)$$

Particularly, the assumptions such that  $f_\theta \in \text{Range}(L_K)$  and  $\lim_{\theta \rightarrow 0} \|f_\theta - f\|_\infty = 0$  hold in Section 0.5.1 and Section 0.5.1, respectively. Then, we jointly evaluate a resulting bound for (57) in (53). This completes the proof.  $\square$

The following Corollary is for the case when the Kernel for KRR approximation is a Gaussian kernel.

**Corollary 1.** A Gaussian kernel satisfies  $\|l_{\pi, s, \theta_\epsilon}\| = C_1 \theta_\epsilon^{-d} \prod_{i=1}^d \int_{\mathbb{R}} \left| \exp \left( -\frac{\sigma^2}{\theta_\epsilon} (\xi_i - i)^2 + h^2 \xi_i^2 \right) \right| d\xi_i$ , that is  $\|l_{\pi, s, \theta_\epsilon}\| \leq C_2 \frac{1}{\theta_\epsilon^d} \left( \frac{\sigma^2}{\theta_\epsilon} h^2 \right)^{-d/2} \leq C_3 \theta_\epsilon^{-d/2} \leq C_4 \epsilon^{-d/2}$ . The corresponding KRR approximation error is

$$\|f - \hat{f}^{\text{KRR}}\|_\infty \leq \inf_{\epsilon > 0} \left\{ 2\epsilon + \frac{C \log(2/\delta) m^{-1/6}}{\epsilon^{d/2}} \right\}. \quad (58)$$

$$f_\theta \in \text{Range}(L_K)$$

From the definition of integral operator as in (54), similarly, we have

$$L_{K_\theta} f \triangleq \int_{\mathbb{R}^d} K_\theta(x, y) f(y) (\pi(y))^\theta dy, \quad (59)$$

for a parameterized kernel  $K_\theta$ . Given  $f \in L_\rho^2$ , i.e.,  $\int |f(x)|^2 d\rho(x) < \infty$ , then  $\int |f(x)\pi^{\theta/2}(x)|^2 dx < \infty$ , that is,  $f\pi^{\theta/2} \in L^2$  and  $f = \bar{f}\pi^{-\theta/2}, \bar{f} \in L^2$ . We evaluate the Fourier transform of  $L_{K_\theta} f$ , that is,

$$\begin{aligned} \mathcal{F}\{L_{K_\theta} f\} &= \mathcal{F}\left\{ \int_{\mathbb{R}^d} K_\theta(x, y) f(y) (\pi(y))^\theta d\rho(y) \right\} \\ &= \mathcal{F}\{K_\theta \star (f\pi^\theta)\} \\ &= \mathcal{F}\{K_\theta\} \cdot \mathcal{F}\{f\pi^\theta\} \\ &= \mathcal{F}\{K_\theta\} \cdot \mathcal{F}\{\bar{f}\pi^{\theta/2}\}. \end{aligned}$$

Hence  $L_{K_\theta} f = \mathcal{F}^{-1} \{ \mathcal{F}\{K_\theta\} \cdot \mathcal{F}\{\bar{f}\pi^{\theta/2}\} \}$ . Similarly, we obtain

$$L_K^{-1} L_{K_\theta} f = \mathcal{F}^{-1} \left\{ \frac{\mathcal{F}\{L_{K_\theta} f\}}{\mathcal{F}\{K\}} \right\} \cdot \pi^{-\theta/2}, \quad (60)$$

$$= \mathcal{F}^{-1} \left\{ \frac{\mathcal{F}\{K_\theta\} \cdot \mathcal{F}\{\bar{f}\pi^{\theta/2}\}}{\mathcal{F}\{K\}} \right\} \cdot \pi^{-\theta/2}, \quad (61)$$

where (61) holds from plugging  $L_{K_\theta} f = \mathcal{F}^{-1} \{ \mathcal{F}\{K_\theta\} \cdot \mathcal{F}\{\bar{f}\pi^{\theta/2}\} \}$  into (60). The sufficient and necessary condition for  $L_K^{-1} L_{K_\theta} f \in L^2$  is  $L_K^{-1} L_{K_\theta} f \cdot \pi^{\theta/2} \in L^2$ , that is,

$$\mathcal{F}^{-1} \left\{ \frac{\mathcal{F}\{K_\theta\} \cdot \mathcal{F}\{\bar{f}\pi^{\theta/2}\}}{\mathcal{F}\{K\}} \right\} \in L^2. \quad (62)$$

**Proposition 1.** If Assumption 4 holds, then,  $f_\theta \in \text{Range}(L_K)$  and it holds that  $\|L_K^{-1} f_\theta\|_{L^2_\rho} \leq c_K c_{f,s} \|l_{\pi,s,\theta}\|_{L^2}$ ,  $l_{\pi,s,\theta} \in L^2(\mathbb{R}^d)$ .

*Proof.* Given  $\bar{f} = f\pi^{\theta/2}$ , the functions  $\bar{f}\pi^{\theta/2} = f\pi^\theta$  almost everywhere, hence,

$$\begin{aligned} |\mathcal{F}\{\bar{f}\pi^{\theta/2}\}(\xi)| &= |\mathcal{F}\{f\pi^\theta\}(\xi)| \\ &= |\mathcal{F}\{f\} \star \mathcal{F}\{\pi^\theta\}(\xi)| \\ &= \left| \int_{\mathbb{R}^d} \mathcal{F}\{f\}(\eta) \mathcal{F}\{\pi^\theta\}(\xi - \eta) d\eta \right| \\ &\leq \int_{\mathbb{R}^d} |\mathcal{F}\{f\}(\eta)| \cdot |\mathcal{F}\{\pi^\theta\}(\xi - \eta)| d\eta \\ &\leq \int_{\mathbb{R}^d} \frac{c_{f,s}}{(1 + |\xi|^2)^{s/2}} |\mathcal{F}\{\pi^\theta\}(\xi - \eta)| d\eta \end{aligned} \quad (63)$$

where (63) holds from Assumption 4 (1). By moving  $c_{f,s}$  to the left of (63),

$$\frac{1}{c_{f,s}} |\mathcal{F}\{\bar{f}\pi^{\theta/2}\}(\xi)| \leq \int_{\mathbb{R}^d} \frac{1}{(1 + |\xi|^2)^{s/2}} |\mathcal{F}\{\pi^\theta\}(\xi - \eta)| d\eta \leq |\mathcal{F}\{K\}(\xi)| \cdot |l_{\pi,s,\theta}(\xi)|, \quad (64)$$

for a function  $l_{\pi,s,\theta} \in L^2$ , where (64) holds from Assumption 4 (2). Then,  $\left| \frac{\mathcal{F}\{\bar{f}\pi^{\theta/2}\}}{\mathcal{F}\{K\}}(\xi) \right| \leq |l_{\pi,s,\theta}(\xi)| \cdot c_{f,s}$ . From Assumption 4 (3), it follows

$$\left\| \frac{\mathcal{F}\{K_\theta\} \cdot \mathcal{F}\{\bar{f}\pi^{\theta/2}\}}{\mathcal{F}\{K\}} \right\|_{L^2} \leq \sup_{\xi \in \mathbb{R}^d} |\mathcal{F}\{K_\theta(\xi)\}| \cdot \left\| \frac{\mathcal{F}\{\bar{f}\pi^{\theta/2}\}}{\mathcal{F}\{K\}} \right\|_{L^2} \leq c_K c_{f,s} \|l_{\pi,s,\theta}(\xi)\|_{L^2}$$

From Plancherel theorem, the above is equivalent to  $\mathcal{F}^{-1} \left\{ \frac{\mathcal{F}\{K_\theta\} \cdot \mathcal{F}\{\bar{f}\pi^{\theta/2}\}}{\mathcal{F}\{K\}} \right\} < \infty$ , that is, (62) holds, which implies  $L_K^{-1} L_{K_\theta} f \in L^2$  and  $f_\theta \in \text{Range}(L_K)$ . This completes the proof.  $\square$

$$\lim_{\theta \rightarrow 0} \|f_\theta - f\|_\infty = 0$$

**Proposition 2.** If Assumption 5 holds, then  $\lim_{\theta \rightarrow 0} \|f_\theta - f\|_\infty = \epsilon$

*Proof.* From the definition of  $f_\theta$ ,

$$\begin{aligned} & |(L_{K_\theta} f - f)(x)| \\ &= \left| \int_{\mathbb{R}^d} K_\theta(x-y) f(y) (\pi(y))^\theta dy - f(x) \right| \\ &\leq \left| \int_{\mathbb{R}^d} K_\theta(x-y) f(y) (\pi(y))^\theta dy - f(x) (\pi(x))^\theta \right| + |f(x)| |1 - (\pi(x))^\theta| \\ &\leq \int_{\mathbb{R}^d} |K_\theta(x-y)| |f(y) (\pi(y))^\theta - f(x) (\pi(x))^\theta| dy + |f(x)| |1 - (\pi(x))^\theta| \end{aligned} \quad (65)$$

$$\begin{aligned} &\leq \int_{|y-x|>\eta_\theta} |K_\theta(x-y)| |f(y) (\pi(y))^\theta - f(x) (\pi(x))^\theta| dy \\ &\quad + \int_{|y-x|\leq\eta_\theta} |K_\theta(x-y)| |f(y) (\pi(y))^\theta - f(x) (\pi(x))^\theta| dy + |f(x)| |1 - (\pi(x))^\theta|, \end{aligned} \quad (66)$$

where (65) holds from  $\int_{\mathbb{R}^d} K_\theta(x) dx = 1$  in Assumption 3. We start with the first term in (66). From Assumption 5 (1), it follows

$$\begin{aligned} & \int_{|y-x|>\eta_\theta} |K_\theta(x-y)| |f(y) (\pi(y))^\theta - f(x) (\pi(x))^\theta| dy \\ &< \frac{\epsilon}{6M_f M_\pi^\theta} \cdot 2M_f M_\pi^\theta \\ &= \epsilon/3. \end{aligned} \quad (67)$$

For the second term in (66), note that from Assumption 5 (2),

$$\begin{aligned} & |f(y) (\pi(y))^\theta - f(x) (\pi(x))^\theta| \\ &\leq |f(y) (\pi(y))^\theta - f(x) (\pi(y))^\theta| + |f(x) (\pi(y))^\theta - f(x) (\pi(x))^\theta| \\ &\leq |\pi(y)^\theta| |f(y) - f(x)| + |f(x)| \cdot |(\pi(y))^\theta - (\pi(x))^\theta| \\ &\leq (L_f M_\pi^\theta + M_f L_\pi) \eta_\theta. \end{aligned} \quad (68)$$

From (68) and Assumption 5 (3), it follows that  $\int_{|y-x|\leq\eta_\theta} |K_\theta(x-y)| |f(y) (\pi(y))^\theta - f(x) (\pi(x))^\theta| dy \leq M_K (L_f M_\pi^\theta + M_f L_\pi) \eta_\theta$ . Hence, when  $\eta_\theta \leq \frac{\epsilon}{3M_K (L_f M_\pi + M_f L_\pi)}$ ,

$$\int_{|y-x|\leq\eta_\theta} |K_\theta(x-y)| |f(y) (\pi(y))^\theta - f(x) (\pi(x))^\theta| dy \leq \frac{\epsilon}{3} \quad (69)$$

We now consider the third term in (66). The assumption for  $\theta$  in Assumption 5 (4) yields that  $(1 - \epsilon/(3M))^{1/\theta} \leq m_\pi$  and  $(1 + \epsilon/(3M))^{1/\epsilon} \geq M_\pi$ , that is

$$(1 - \epsilon/(3M))^{1/\theta} < \pi(x) < (1 + \epsilon/(3M))^{1/\theta}, \quad (70)$$

given  $|\pi(x)| \in [m_\pi, M_\pi]$ ,  $\forall x \in \text{supp } f$ . Equivalently, rewriting (70) as  $|1 - (\pi(x))^\theta| < \frac{\epsilon}{3M}$ . Therefore, Assumption 5 (4) ensures

$$|f(x)| |1 - (\pi(x))^\theta| \leq M |1 - (\pi(x))^\theta| < \frac{\epsilon}{3}. \quad (71)$$

Therefore, (67) (69) (71) indicates (66) is upper bounded by  $\epsilon$ , that is,  $|L_{K_\theta} f - f|_\infty \leq \epsilon$ . This completes the proof.  $\square$

In this following, we provide the proof for Theorem 6.

*Proof.* From Proposition 1 and Proposition 2,  $\forall \epsilon > 0$ ,  $\exists \theta_\epsilon \in (0, 1)$ , such that  $f_{\theta_\epsilon}$  defined in (56) is such that  $f_{\theta_\epsilon} \in \text{Range}(L_K)$  and  $\lim_{\theta \rightarrow 0} \|f_{\theta_\epsilon} - f\|_\infty = \epsilon$ . From Theorem 4,

$$\|f_{\theta_\epsilon} - \hat{f}_{\theta_\epsilon}^{\text{KRR}}\|_\infty \leq 4 \log\left(\frac{2}{\delta}\right) (3\kappa M)^{\frac{1}{3}} \|L_K^{-1} L_{K_{\theta_\epsilon}} f\|_{L_\rho^2} m^{-\frac{1}{6}}, \quad (72)$$

where  $\|L_K^{-1} L_{K_{\theta_\epsilon}} f\|_{L_\rho^2} \leq c_K c_{f,s} \|l_{\pi,s,\theta}\|_{L^2}$ ,  $l_{\pi,s,\theta} \in L^2(\mathbb{R}^d)$  from Proposition 1. Therefore,

$$\|f_{\theta_\epsilon} - \hat{f}_{\theta_\epsilon}^{\text{KRR}}\|_\infty \leq 4 \log\left(\frac{2}{\delta}\right) (3\kappa M)^{\frac{1}{3}} c_K c_{f,s} \|l_{\pi,s,\theta}\|_{L^2} m^{-\frac{1}{6}}. \quad (73)$$

The error in (57) is as given in (53). This completes the proof.  $\square$

### 0.5.2 Generalization Error

The probabilistic guarantee on the generalization error of OC-SVM was given in [26].

**Theorem 7.** Suppose that the OC-SVM gives an optimal solution  $(M^*, \rho^*)$  with  $\rho^* > 0$  for all  $1 \leq i \leq m$ . Then for any  $\delta \in (0, 1)$  and  $\epsilon > 0$ , with probability  $1 - \delta$  (over the choice of samples size  $m$ ), we have

$$\mathbb{P}[\langle M, \Gamma \rangle < \rho^* - \epsilon] \leq \frac{2}{m} \left( \lceil \log \tau(\|N\|_F, \epsilon) \rceil + \frac{2}{\delta} \right),$$

where  $\tau(\varrho, \epsilon, 2m)$  is the covering number of the model space  $\{\langle N, \cdot \rangle : \|N\|_F \leq \varrho\}$  by balls of radius of  $\epsilon$  under the metric  $d(N_1, N_2) = \sup_{\Gamma_1, \dots, \Gamma_{2m}} \max_{i=1, \dots, 2m} |\langle N_1 - N_2, \Gamma_i \rangle|$ .

In [27], the conclusion was translated explicitly in terms of the dissipativity learning. Specifically, by the definition of covering number, it could be found that

$$\mathbb{P}[\langle M, \Gamma \rangle < \rho^* - \epsilon] \leq \frac{1}{m} \mathcal{O}\left(\log \frac{1}{\delta} + \epsilon^{\frac{1}{2}} \log m + \frac{1}{\epsilon^2} \log \frac{1}{\epsilon}\right).$$

## 0.6 Numerical Results

For the estimation of the dissipativity parameters  $Q$  and  $P$ , we suppose that  $m$  data points  $(y^i(\cdot), u^i(\cdot))_{i=1}^m$  and  $(x^i(\cdot))_{i=1}^m$  are sampled, from a distribution of random input excitations such that the magnitude of the input satisfies Gaussian distribution with zero mean with a fixed duration for each magnitude. We consider a four-tray binary distillation column where the states are  $\mathbf{x}(t) = [x_1(t) \ x_2(t) \ x_3(t) \ x_4(t)]^\top$  with nonlinear dynamics as follows:

$$\begin{aligned} \dot{x}_1 &= -\frac{V}{H_1} x_1 + \frac{V\beta x_2}{H_1(1 + (\beta - 1)x_2)}, \\ \dot{x}_2 &= \frac{L}{H_2}(x_1 - x_2) + \frac{V}{H_2} \left( \frac{\beta x_3}{1 + (\beta - 1)x_3} - \frac{\beta x_2}{1 + (\beta - 1)x_2} \right), \\ \dot{x}_3 &= -\frac{F}{H_3} x_3 + \frac{L}{H_3}(x_2 - x_3) + \frac{V}{H_3} \frac{\beta x_4}{1 + (\beta - 1)x_4} - \frac{V}{H_3} \frac{\beta x_3}{1 + (\beta - 1)x_3} + \frac{F}{H_3} z_f(1 + \text{sat}(u(t))), \\ \dot{x}_4 &= \frac{F + L}{H_4}(x_3 - x_4) + \frac{V}{H_4} \left( x_4 - \frac{\beta x_4}{1 + (\beta - 1)x_4} \right), \end{aligned}$$

Table 1: Parameters of the distillation column.

Parameter	Symbol	Value
Vapor flowrate	$V$	6.05 (mol/min)
Reflux flowrate	$L$	4.79 (mol/min)
Feed flowrate	$F$	1.70 (mol/min)
Relative volatility constant	$\beta$	1.60
Nominal feed composition	$z_f$	0.56
Liquid holdup	$H_1$	5.25 mol
	$H_2$	0.53 mol
	$H_3$	0.53 mol
	$H_4$	5.25 mol

where  $\text{sat}(u) \triangleq \max\{-0.5, \min\{0.5, u(t)\}\}$ . The measured outputs are the top and bottom compositions  $y(t) = [x_1(t) \ x_4(t)]^\top$ . The parameters are listed in Table 1.

### 0.6.1 Numerical stability of nonparametric and parametric dissipativity learning

In the nonparametric RKHS formulation, we consistently observe that as  $\lambda$  increases, the objective value and the Hilbert–Schmidt norms  $\|\Pi\|_{\text{HS}}$  and  $\|P\|_{\text{HS}}$  grow monotonically, as shown in the top panel of Fig. 1. This behavior aligns with the structure of (22), where a larger  $\lambda$  imposes a stronger penalty on the term  $\lambda\alpha$ . The increased penalty drives the optimizer to reduce  $\alpha$ , which in turn tightens the gain-like constraint  $-\Pi - \Upsilon_0 + \alpha\Upsilon_1 \succeq 0$  and enforces a stricter gain bound. As a result, the feasible set for  $\Pi$  contracts, and the optimizer compensates by increasing the magnitude of  $\Pi$ , leading to a systematic growth in its Hilbert–Schmidt norm.

Similarly, the norm of  $P$  also increases systematically with  $\lambda$ . This trend is coupled with the variation of  $\Pi$  through the sample-wise dissipativity constraints  $\langle \Pi, \Gamma_t \rangle + \langle P, \Delta_t \rangle \geq \rho$ . When  $\alpha$  decreases and  $\Pi$  becomes more negative, the term  $\langle \Pi, \Gamma_t \rangle$  generally decreases for most samples. To preserve the feasibility of the dissipativity inequalities, the optimizer compensates by increasing the contribution from  $P$ , so that  $\langle P, \Delta_t \rangle$  grows accordingly. This adjustment results in a consistent increase in the Hilbert–Schmidt norm  $\|P\|_{\text{HS}}$  as  $\lambda$  increases. In other words, both  $\|\Pi\|_{\text{HS}}$  and  $\|P\|_{\text{HS}}$  must increase jointly to satisfy the tightened dissipativity constraints under a stricter gain bound. Consequently, the overall objective value and the operator norms exhibit a monotone nondecreasing trend with respect to  $\lambda$ , as illustrated in Fig. 1. Furthermore, as expected, increasing  $\lambda$  enforces a stricter gain bound, reflected by the smaller optimal values of  $\alpha$  shown in the top panel of Fig. 2.

Similarly, in the parametric learning setting, the same monotonicity property holds. The objective function is convex and nondecreasing with respect to  $\lambda$ . As in the nonparametric case, increasing  $\lambda$  penalizes large values of  $\alpha$ , thus enforcing a stricter dissipativity gain bound. As  $\lambda$  increases, it leads to a reduction in  $\alpha$  and thereby small feasible region of  $Q$ .

This is enforced by the constraint in (13), i.e.  $Q \prec \begin{bmatrix} \alpha & 0 \\ 0 & -I \end{bmatrix}$ . To maintain the feasibility,

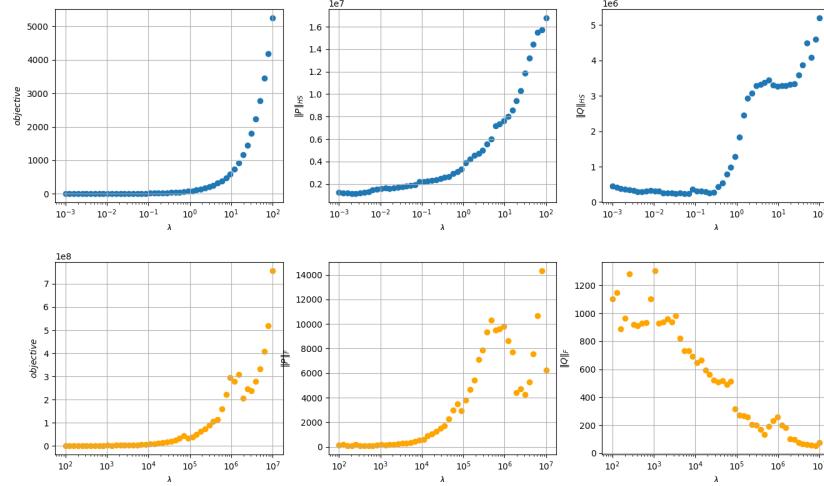


Figure 1: Comparison of the objectives of nonparametric dissipativity learning in RKHS (top) and parametric learning (bottom).

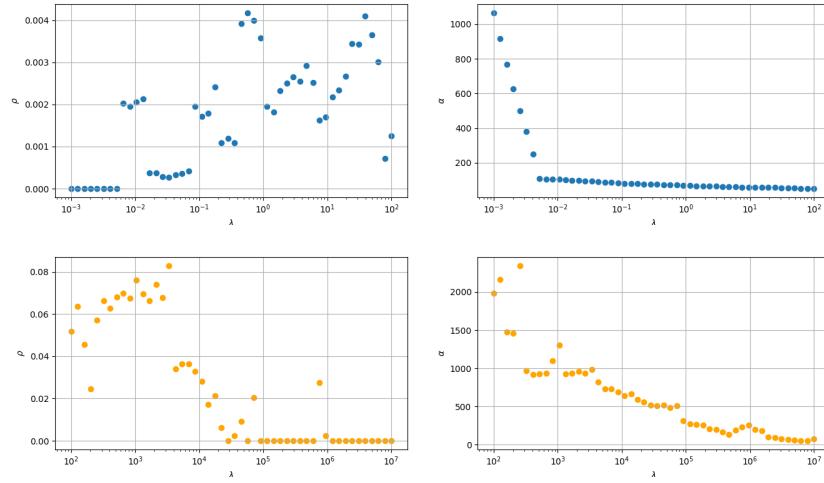


Figure 2: Comparison of the learned dissipativity margin  $\rho$  and  $\alpha$  from nonparametric dissipativity learning in RKHS (top) and parametric learning (bottom), with respect to weighting parameter  $\lambda$ .

the matrix  $Q$  is pushed away from the origin in the matrix space, leading to an increase in its Frobenius norm. This reduces the term  $\langle Q, \Gamma_t \rangle$ . To preserve feasibility and maintain the dissipativity margin  $\rho$  in  $\langle Q, \Gamma_t \rangle + \langle P, \Delta_t \rangle \geq \rho$ , the optimizer compensates by increasing  $\langle P, \Delta_t \rangle$  across samples. Achieving this compensation requires  $P$  to have larger eigenvalues in magnitude, thereby increasing its Frobenius norm. In other words, as the gain constraint on  $Q$  becomes stricter,  $P$  must amplify its effect to balance the tightened inequalities, leading to a systematic growth in  $\|P\|_F$  as  $\lambda$  increases too. This analysis is shown numerically in the bottom panel of Fig. 1. It is worth noting that the nonparametric learning formulation exhibits better numerical stability in comparison with its parametric counterpart.

The verification for the dissipativity learning is shown in Fig. 3, which evaluates the learned dissipativity margin on testing data. For each weighting parameter  $\lambda$ , the sample-wise dissipativity margin was computed using the learned matrices  $P$  and  $Q$ , normalized by  $\|P\|_F + \|Q\|_F$ . The mean and variance of these normalized margins are shown as error bars, together with the learned dissipativity margin  $\rho$  obtained from the training data. The results confirm that the dissipativity inequality remains valid on the verification data for a wide range of  $\lambda$ , as the test margin remains positive and closely follow the trend as the learned margin on training data. As  $\lambda$  increases, the test margin decreases slightly, reflecting the expected trade-off between conservativeness and data fit. A larger  $\lambda$  enforces a stricter gain bound and increases the magnitudes of  $P$  and  $Q$ , leading to a more conservative but stable certificate. The relatively small standard deviations indicate that the learned quadratic storage and supply matrices generalize well to unseen data. Overall, the verification demonstrates that the proposed approach yields a consistent dissipativity characterization that is robust with respect to both the regularization parameter  $\lambda$  and unseen data.

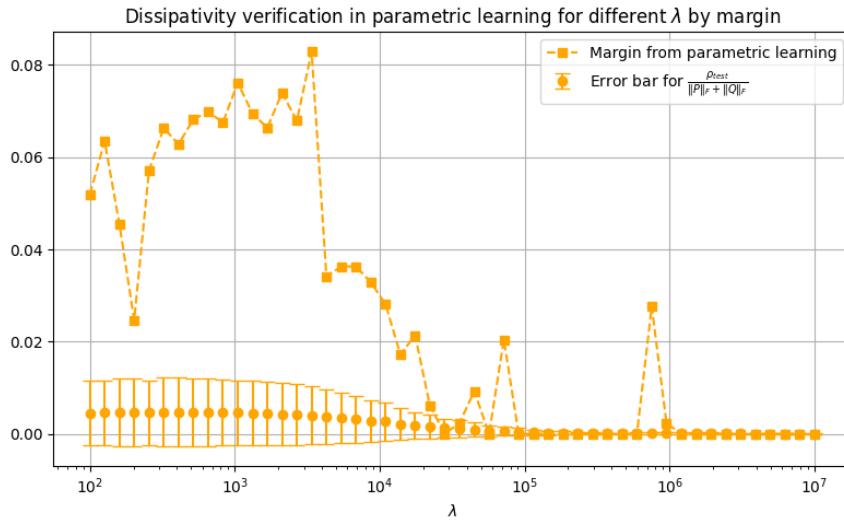


Figure 3: Dissipativity verification for parametric learning.

## 0.7 Conclusion

This paper introduced a nonparametric framework for dissipativity learning in RKHS by casting the problem in an OC-SVM formulation. By leveraging representer properties, we reduced the infinite-dimensional operator problem to a finite-dimensional convex problem expressed solely through kernel Gram matrices, thereby retaining expressiveness while ensuring tractability. We further provided a principled error analysis that separates (i) the kernel regression approximation error used to construct reference operators and (ii) the OC-SVM generalization error on the empirical margin, yielding population-level guarantees on the learned dissipativity certificates and the implied  $\mathcal{L}_2$  gain bound.

Numerical experiments on a nonlinear distillation column demonstrate that the RKHS formulation reliably identifies dissipative behavior directly from input-output trajectories, generalizes to unseen data, and exhibits favorable numerical stability relative to a parametric baseline. As the tradeoff parameter  $\lambda$  increases, we observed a monotone, nondecreasing trend in the objective and operator norms, consistent with the tightening of the gain-like constraint. Concurrently, the verified margins indicates the robustness of the approach.

# Bibliography

- [1] Z.-S. Hou and Z. Wang, “From model-based control to data-driven control: Survey, classification and perspective,” *Information Sciences*, vol. 235, pp. 3–35, 2013.
- [2] L. Ljung, “System identification,” in *Signal analysis and prediction*. Springer, 1998, pp. 163–173.
- [3] M. Krstic, P. V. Kokotovic, and I. Kanellakopoulos, *Nonlinear and adaptive control design*. John Wiley & Sons, Inc., 1995.
- [4] R. Strässer, M. Schaller, K. Worthmann, J. Berberich, and F. Allgöwer, “Koopman-based feedback design with stability guarantees,” *IEEE Transactions on Automatic Control*, 2024.
- [5] V. Kostic, P. Novelli, A. Maurer, C. Ciliberto, L. Rosasco, and M. Pontil, “Learning dynamical systems via koopman operator regression in reproducing kernel hilbert spaces,” *Advances in Neural Information Processing Systems*, vol. 35, pp. 4017–4031, 2022.
- [6] S. S. Ge, C. C. Hang, T. H. Lee, and T. Zhang, *Stable adaptive neural network control*. Springer Science & Business Media, 2013, vol. 13.
- [7] W. Tang and P. Daoutidis, “Input-output data-driven control through dissipativity learning,” in *2019 American Control Conference (ACC)*. IEEE, 2019, pp. 4217–4222.
- [8] ——, “Dissipativity learning control (dlc): A framework of input–output data-driven control,” *Computers & Chemical Engineering*, vol. 130, p. 106576, 2019.
- [9] ——, “Dissipativity learning control (dlc): theoretical foundations of input–output data-driven model-free control,” *Systems & Control Letters*, vol. 147, p. 104831, 2021.
- [10] R. Lozano, B. Brogliato, O. Egeland, and B. Maschke, *Dissipative systems analysis and control: theory and applications*. Springer Science & Business Media, 2013.
- [11] J. Zabczyk, *Mathematical control theory*. Springer, 2020.
- [12] J. Bao and P. L. Lee, *Process control: the passive systems approach*. Springer, 2007.
- [13] A. Berlinet and C. Thomas-Agnan, *Reproducing kernel Hilbert spaces in probability and statistics*. Springer Science & Business Media, 2011.

- [14] V. I. Paulsen and M. Raghupathi, *An introduction to the theory of reproducing kernel Hilbert spaces*. Cambridge university press, 2016, vol. 152.
- [15] B. Ghojogh, A. Ghodsi, F. Karray, and M. Crowley, “Reproducing kernel hilbert space, mercer’s theorem, eigenfunctions, nystr\” om method, and use of kernels in machine learning: Tutorial and survey,” *arXiv preprint arXiv:2106.08443*, 2021.
- [16] M. Amer, M. Goldstein, and S. Abdennadher, “Enhancing one-class support vector machines for unsupervised anomaly detection,” in *Proceedings of the ACM SIGKDD workshop on outlier detection and description*, 2013, pp. 8–15.
- [17] J. C. Willems, “Dissipative dynamical systems part ii: Linear systems with quadratic supply rates,” *Archive for rational mechanics and analysis*, vol. 45, no. 5, pp. 352–393, 1972.
- [18] H. K. Khalil and J. W. Grizzle, *Nonlinear systems*. Prentice hall Upper Saddle River, NJ, 2002, vol. 3.
- [19] O. J. Rojas, J. Bao, and P. L. Lee, “On dissipativity, passivity and dynamic operability of nonlinear processes,” *Journal of Process Control*, vol. 18, no. 5, pp. 515–526, 2008.
- [20] D. Hill and P. Moylan, “The stability of nonlinear dissipative systems,” *IEEE transactions on automatic control*, vol. 21, no. 5, pp. 708–711, 2003.
- [21] W. Tang, “Learning the integral quadratic constraints on plant-model mismatch,” *arXiv preprint arXiv:2502.00976*, 2025.
- [22] T. Martin, T. B. Schön, and F. Allgöwer, “Guarantees for data-driven control of nonlinear systems using semidefinite programming: A survey,” *Ann. Rev. Control*, p. 100911, 2023.
- [23] A. Koch, J. Berberich, and F. Allgöwer, “Provably robust verification of dissipativity properties from data,” *IEEE Trans. Autom. Control*, vol. 67, no. 8, pp. 4248–4255, 2021.
- [24] N. Aronszajn, “Theory of reproducing kernels,” *Trans. the American mathematical society*, vol. 68, no. 3, pp. 337–404, 1950.
- [25] S. Smale and D.-X. Zhou, “Learning theory estimates via integral operators and their approximations,” *Constructive approximation*, vol. 26, no. 2, pp. 153–172, 2007.
- [26] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, “Estimating the support of a high-dimensional distribution,” *Neural computation*, vol. 13, no. 7, pp. 1443–1471, 2001.
- [27] E. J. LoCicero, A. Penne, and L. Bridgeman, “Issues with input-space representation in nonlinear data-based dissipativity estimation,” in *2025 American Control Conference (ACC)*. IEEE, 2025, pp. 5178–5183.