Identification of nonlinear circuits as port-Hamiltonian systems

Judy Najnudel, Remy Müller, Thomas Hélie, David Roze

To cite this version:

Judy Najnudel, Remy Müller, Thomas Hélie, David Roze. Identification of nonlinear circuits as port-Hamiltonian systems. 24th International Conference on Digital Audio Effects (DAFx-21), Sep 2021, Vienna, Austria. ⟨hal-03327522⟩

HAL Id: hal-03327522
https://hal.science/hal-03327522v1

Submitted on 4 Oct 2021

IDENTIFICATION OF NONLINEAR CIRCUITS AS PORT-HAMILTONIAN SYSTEMS

Judy Najnudel, Rémy Müller, Thomas Hélie, and David Roze

S3AM team, STMS laboratory
IRCAM - CNRS - SU
Paris, France
`firstname.lastname@ircam.fr`

ABSTRACT

This paper addresses identification of nonlinear circuits for power-balanced virtual analog modeling and simulation. The proposed method combines a port-Hamiltonian system formulation with kernel-based methods to retrieve model laws from measurements. This combination allows for the estimated model to retain physical properties that are crucial for the accuracy of simulations, while representing a variety of nonlinear behaviors. As an illustration, the method is used to identify a nonlinear passive peaking EQ.

1. INTRODUCTION

Virtual analog modeling is an active research field, in particular within the audio community. Indeed, vintage analog audio effects are still appreciated among musicians, but original devices are now rare and delicate to maintain. A virtual replica then becomes a compelling alternative.

Historically, modeling methods have been roughly classified into two categories, white box and black box. White box modeling relies heavily on physics, and requires extensive knowledge on the circuit, from components datasheets to circuit schematics. Such modeling include Wave Digital Filters [1, 2, 3] and State-Space representations [4, 5, 6]. Black box modeling on the other hand is more concerned with reproducing a global input-output behavior, without necessarily capturing local phenomena taking place inside the system. It is usually more adaptable, but less physically interpretable. Neural Networks [7, 8] are popular black box models. Volterra-based methods can be either white box oriented [9, 10], or black box oriented [11].

A crucial counterpart of modeling is system identification, that is, retrieving model laws and parameters from measurements. Obviously, one designs an identification method with a specific model —white box or black box— in mind (see e.g. [12, 13] for Volterra, or [14] for WDF). In recent years nonetheless, hybrid modeling, or grey box, has gained considerable momentum for identification. Indeed, it often successfully combines desirable features from both paradigms. Such in-between modeling associates State-Space representations and polynomial models [15], State-Space representations and Neural Networks [16], digital filters and Neural Networks [17] to list a few.

In this paper, we consider an identification method relying on the State-Space representation known as port-Hamiltonian systems (PHS) [18, 19] on one hand, and kernel-based methods [20, 21] on the other hand. PHS are built as interconnected components with physical constitutive laws, so that the power balance of the system is structurally satisfied. This energy-based formulation can be associated with numerical methods that preserve the power balance and passivity in the discrete-time domain, for both linear and nonlinear systems. The PHS approach has proved relevant to simulate audio systems [22]. Although extensive work is concerned with linear PHS identification [23, 24], nonlinear PHS identification is still little explored (see e.g. [25] for an overview). To address nonlinear behavior, we rely on reproducing kernels. Reproducing kernels have long proven to be a valuable and flexible tool for function approximation and solving of differential equations [26]. As such, they already have succesfully been implemented for audio circuit modeling [27]. Reproducing kernels are privileged here over deep learning approaches such as in [28], because the number of parameters to infer is much smaller, and because they are generally more interpretable. The specificity of our method resides in choosing a kernel and tailoring the regression so that key physical properties of the system, such as power balance and passivity, are retained.

This paper is organized as follows. In section 2, we give a brief overview of PHS. In section 3, we propose a parametrization of the PHS interconnection matrix. In section 4, we present an energy modeling based on reproducing kernels. Section 5 describes an optimization procedure to retrieve PHS parameters from measurements. In section 6, the complete method is tested on a virtual passive peaking EQ. Finally, some conclusions and perspectives are given in section 7.

2. PORT-HAMILTONIAN FORMULATION AND WORKING ASSUMPTIONS

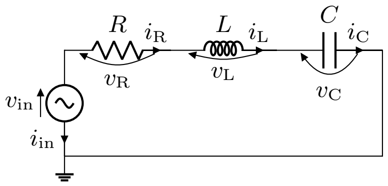
2.1. Port-Hamiltonian formulation

The identification method described in this paper relies on port-Hamiltonian systems [29, 19], under a differential-algebraic formulation (PHS). This formulation allows the representation of a dynamical system as a network of:

1. storage components of state $\boldsymbol{x}$ and energy $\mathcal{H}(\boldsymbol{x})$, storing the power $P_{\text{stored}} = \nabla\mathcal{H}(\boldsymbol{x})^\intercal \dot{\boldsymbol{x}}$;
2. dissipative components of flow $\boldsymbol{w}$ and effort $z(\boldsymbol{w})$, such as the dissipated power $P_{\text{diss}} = z(\boldsymbol{w})^\intercal \boldsymbol{w}$ is non-negative for all $\boldsymbol{w}$;
3. connection ports conveying the *outgoing* power $P_{\text{ext}} = \boldsymbol{u}^\intercal \boldsymbol{y}$ where $\boldsymbol{u}$ are inputs and $\boldsymbol{y}$ are outputs.

The flows $\boldsymbol{f}$ and efforts $\boldsymbol{e}$ of all components are coupled through a skew-symmetric interconnection matrix $\boldsymbol{S} = -\boldsymbol{S}^\intercal$ (possibly de-

---



(a) RLC circuit in series.

$$\begin{array}{cc|cc|c|c|}
 & & \multicolumn{2}{c|}{\nabla\mathcal{H}(\boldsymbol{x})} & z(\boldsymbol{w}) & \boldsymbol{u} \\
 & & v_{\mathrm{C}} & i_{\mathrm{L}} & v_{\mathrm{R}} & v_{\mathrm{in}} \\
\dot{\boldsymbol{x}} & i_{\mathrm{C}} & . & 1 & . & . \\
 & v_{\mathrm{L}} & -1 & . & -1 & 1 \\
\hline
\boldsymbol{w} & i_{\mathrm{R}} & . & 1 & . & . \\
\hline
\boldsymbol{y} & i_{\mathrm{in}} & . & -1 & . & .
\end{array}$$

(b) Corresponding PHS. Dots indicate zeros.

$$\begin{array}{cc|cc|c|}
 & & \multicolumn{2}{c|}{\nabla\mathcal{H}(\boldsymbol{x})} & \boldsymbol{u} \\
 & & v_{\mathrm{C}} & i_{\mathrm{L}} & v_{\mathrm{in}} \\
\dot{\boldsymbol{x}} & i_{\mathrm{C}} & . & 1 & . \\
 & v_{\mathrm{L}} & -1 & -R & 1 \\
\hline
\boldsymbol{y} & i_{\mathrm{in}} & . & -1 & .
\end{array}$$

(c) Corresponding PH-ODE. Dots indicate zeros.

Figure 1: RLC circuit in series and its PHS representations.

pendent on $\boldsymbol{x}$):

$$\underbrace{\begin{bmatrix} \dot{\boldsymbol{x}} \\ \boldsymbol{w} \\ \boldsymbol{y} \end{bmatrix}}_{\boldsymbol{f}} = \boldsymbol{S} \underbrace{\begin{bmatrix} \nabla\mathcal{H}(\boldsymbol{x}) \\ z(\boldsymbol{w}) \\ \boldsymbol{u} \end{bmatrix}}_{\boldsymbol{e}}. \tag{1}$$

Here, flows can either be currents (e.g. for capacitors) or voltages (e.g. for inductors), and vice versa for efforts. Such systems satisfy the power balance $P_{\mathrm{stored}} + P_{\mathrm{diss}} + P_{\mathrm{ext}} = 0$. Indeed, $P_{\mathrm{stored}} + P_{\mathrm{diss}} + P_{\mathrm{ext}} = \boldsymbol{e}^{\intercal}\boldsymbol{f} = \boldsymbol{e}^{\intercal}\boldsymbol{S}\boldsymbol{e}$ is zero since $\boldsymbol{e}^{\intercal}\boldsymbol{S}\boldsymbol{e} = (\boldsymbol{e}^{\intercal}\boldsymbol{S}\boldsymbol{e})^{\intercal} = -(\boldsymbol{e}^{\intercal}\boldsymbol{S}\boldsymbol{e})$ due to the skew-symmetry of $\boldsymbol{S}$.

Under some additional assumptions (not detailed here), this formulation can be reduced to the following PH-ODE representation [19]:

$$\begin{bmatrix} \dot{\boldsymbol{x}} \\ \boldsymbol{y} \end{bmatrix} = (\boldsymbol{J} - \boldsymbol{R}) \begin{bmatrix} \nabla\mathcal{H}(\boldsymbol{x}) \\ \boldsymbol{u} \end{bmatrix}, \tag{2}$$

where $\boldsymbol{J} = -\boldsymbol{J}^{\intercal}$ is skew-symmetric, and $\boldsymbol{R} = \boldsymbol{R}^{\intercal}$ is positive semi-definite. Both matrices are possibly dependent on $\boldsymbol{x}$. In particular, formulation (2) is possible when the dissipative law z is linear.

Throughout this paper, we adopt the *passive sign convention* for all components, including ports and external sources. This means that the current is defined positive when entering the component through the positive voltage terminal [30].

2.2. Example

As an illustration, consider the linear RLC circuit in series (Fig. 1a). The resistor obey Ohm's law, with flow i_{R} (current through the resistor) and effort $z(i_{\mathrm{R}}) = R\ i_{\mathrm{R}} = v_{\mathrm{R}}$ (voltage across the resistor). The inductor has state Φ_{L} (magnetic flux linkage through the inductor), energy $\mathcal{H}_{\mathrm{L}}(\Phi_{\mathrm{L}}) = \Phi_{\mathrm{L}}^2/2\,L$ and effort $\mathcal{H}'_{\mathrm{L}}(\Phi_{\mathrm{L}}) = \Phi_{\mathrm{L}}/L = i_{\mathrm{L}}$ (current through the inductor). The capacitor has state q_{C} (electric charge in the capacitor), energy $\mathcal{H}_{\mathrm{C}}(q_{\mathrm{C}}) = q_{\mathrm{C}}^2/2\,C$ and effort $\mathcal{H}'_{\mathrm{C}}(q_{\mathrm{C}}) = q_{\mathrm{C}}/C = v_{\mathrm{C}}$ (voltage across the capacitor). Kirchhoff's laws in receptor convention yield the PHS and PH-ODE representations in Fig. 1b and Fig. 1c, respectively.

2.3. Working assumptions

In the following, we limit ourselves to circuits verifying that:

1. The dissipation law z is linear.
2. The coefficients of $\boldsymbol{J} - \boldsymbol{R}$ are constant.
3. Separability: the energy law is a separable function of the state, i.e. it takes the form $\mathcal{H}(\boldsymbol{x}) = \sum_{k=1}^{N_x} \mathcal{H}_k(x_k)$, $N_x = \dim(\boldsymbol{x})$.
4. Smoothness: $\mathcal{H}_k$ is at least $\mathcal{C}^2$ $\forall k \in \{1, ..., N_x\}$.
5. Convexity: the energy law is convex, i.e. $\mathcal{H}_k^{(2)}(x_k) \geq 0$ $\forall k, x_k$, where $\mathcal{H}_k^{(2)}$ denote the second derivative of $\mathcal{H}_k$ (a formal definition can be found in Appendix A).

Assumptions 1 to 3 cover electronic circuits constituted of one-ports such as (possibly nonlinear, see [31]) inductors, (possibly nonlinear, see [32]) capacitors, linear conductors and linear resistors, which admit a PH-ODE formulation such as Eq. (2). This deliberately excludes nonlinear dissipative components such as diodes, transistors and vacuum tubes, which will be the object of future work. Assumptions 4 and 5 are stricter than necessary: actually, a sufficient condition to ensure passivity of storage components is for the energy to have an inferior bound [19]. Nevertheless, enforcing convexity results in a desirable asymptotic behavior, and most energy laws are convex anyway (a notable exception concerns meta-stable ferromagnetic cores [31]).

In the next sections, we assume that we can measure $\boldsymbol{x}$, $\boldsymbol{u}$ and $\boldsymbol{y}$, and look for an estimation of the reduced matrix $\boldsymbol{J} - \boldsymbol{R}$ and an approximant of $\mathcal{H}$, verifying both Eq. (2) and our working assumptions.

3. INTERCONNECTION MATRIX MODELING

3.1. Decomposition of J

Since $\boldsymbol{J} = -\boldsymbol{J}^{\intercal}$ is skew-symmetric, it can be written as:

$$\boldsymbol{J} = \boldsymbol{J}(\boldsymbol{\theta}^J) := \sum_{k=1}^{N_J} \theta_k^J\, \boldsymbol{J}_k, \quad \boldsymbol{\theta}^J = \left[\theta_1^J, \ldots, \theta_{N_J}^J\right] \in \mathbb{R}^{N_J}, \tag{3}$$

where $\{\boldsymbol{J}_k\}$ is the canonical base of skew-symmetric matrices, and $N_J = \dim(\boldsymbol{J})\left(\dim(\boldsymbol{J}) - 1\right)/2$.

For the example of section 2.2, we have:

$$\{\boldsymbol{J}_k\} = \left\{ \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} \right\},$$
$$\boldsymbol{\theta}^J = [1, 0, 1].$$

3.2. Decomposition of R

Since $\boldsymbol{R}$ is positive semi-definite, it admits a Cholesky decomposition [33, 34] and can be written as:

$$\begin{aligned} \boldsymbol{R} &= \boldsymbol{R}(\boldsymbol{\theta}^R) = \boldsymbol{T}(\boldsymbol{\theta}^R)\boldsymbol{T}(\boldsymbol{\theta}^R)^\intercal, \\ \boldsymbol{T}(\boldsymbol{\theta}^R) &:= \sum_{k=1}^{N_R} \theta_k^R \, \boldsymbol{T}_k, \quad \boldsymbol{\theta}^R = \left[\theta_1^R, \dots, \theta_{N_R}^R\right] \in \mathbb{R}^{N_R}, \end{aligned} \tag{4}$$

where $\{\boldsymbol{T}_k\}$ is the canonical base of lower triangular matrices, $N_R = \dim(\boldsymbol{R})\left(\dim(\boldsymbol{R}) + 1\right)/2$, and diagonal coefficients are non-negative. For convenience, we choose the first $\dim(\boldsymbol{R})$ coefficients to be the diagonal coefficients.

For the example of section 2.2, we have:

$$\begin{aligned} \{\boldsymbol{T}_k\} = &\left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \right. \\ &\left. \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \right\} \\ \boldsymbol{\theta}^R = &\left[0, \sqrt{R}, 0, 0, 0, 0\right]. \end{aligned}$$

4. ENERGY LAW MODELING

4.1. Optimal approximant

We consider the reproducing kernel theory to build an approximant of $\mathcal{H}$. We refer to [35, 26] for a complete theoretical overview of Reproducing Kernel Hilbert Spaces (RKHS). In this section, we only recall practical results.

Consider the classic scattered data interpolation problem, which is finding a function f verifying:

$$f(x_i) = y_i \quad \forall i \in \{1, ..., N\}, \tag{5}$$

for some given data $\{(x_1, y_1), ..., (x_N, y_N)\}$, $[x_1, ..., x_N] \in \Omega^N$. Assuming that f belongs to a RKHS $\mathcal{K}$, the optimal approximant of f in $\mathcal{K}$ is the function $\tilde{f}$ defined as:

$$\tilde{f} : x \mapsto \sum_{j=1}^{N} \tilde{\lambda}_j K(x, x_j), \tag{6}$$

with $K : \Omega \times \Omega \mapsto \mathbb{R}$ the positive definite kernel inducing $\mathcal{K}$. In Eq. (6), the vector $\tilde{\boldsymbol{\lambda}}$ is the solution of the linear system $\boldsymbol{K}\boldsymbol{\lambda} = \boldsymbol{y}$, where $K_{i,j} = K(x_i, x_j)$. For instance, the kernel defined as:

$$K(x_i, x_j) = \frac{a}{\pi} \text{sinc}\left(a(x_i - x_j)\right),$$

where $0 < a < +\infty$ and $\text{sinc}(u) = \sin(u)/u$, is the reproducing kernel of the space of bandlimited continuous functions of bandwith in $[-a, a]$.

In practice, the RKHS $\mathcal{K}$ of interest for a given problem is often unknown. Therefore what motivates the choice of kernel boils down to the properties one wishes to attach to the approximant, such as locality[1], smoothness, interpolation behavior, sensitivity to noise, etc. However, there is an inherent tradeoff between these

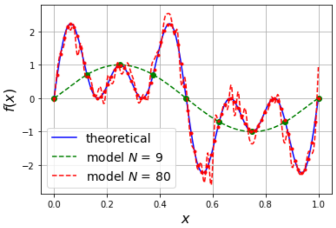


Figure 2: Two different approximants of the same function $f(x) = \sin(2\pi x) + \sin(6\pi x) + \sin(10\pi x)$ with noisy data.

properties. Indeed, if locality is desirable (typically, because f contains high frequencies), choosing a local kernel is necessary. Nonetheless good reconstruction with a highly local kernel necessitates a high number of samples N. But a high number of samples can in turn lead to overfitting, especially with noisy data. Therefore, N is usually chosen low enough so that the approximant generalizes well and small changes in the given data do not impact the reconstruction too much. Concomitantly, the kernel support is adjusted through a scaling parameter $\rho > 0$, so that the approximant is reasonably smooth for the chosen N.

Once N is fixed, several strategies are available in order to find optimal parameters (the N samples and the scaling parameter ρ) for a given kernel, see e.g. [36, 37, 38]. In this work, as the input space dimension is only 1 due to the separability of the energy law, we adopt a grid-based approach: the samples are distributed evenly along the measured data range, and the scaling parameter is adjusted so that some test function is approximated with an arbitrary degree of accuracy. The advantage of this approach resides in its simplicity. Should the energy law not be separable though, a more sophisticated strategy would be preferable. Figure 2 shows approximations from noisy data (SNR = 14 dB with a normal distribution) for different values of N, and K defined as:

$$K(x_i, x_j) = \begin{cases} (1 - r)^4 (1 + 4r) & r = \|x_i - x_j\| < 1, \\ 0 & \text{otherwise.} \end{cases}$$

4.2. Choice of kernel

The only assumption we have on the energy $\mathcal{H}$ is the convexity assumption. Therefore, any kernel reproducing the convexity of $\mathcal{H}$ should be relevant. A simple way of enforcing the convexity of $\mathcal{H}$ is to construct a positive approximant of $\nabla^2\mathcal{H}$. An approximant of $\mathcal{H}$ can then be obtained by integration. To ensure the positivity of each $\mathcal{H}_k^{(2)}$, we look for approximants of the form:

$$\mathcal{H}_k^{(2)}(x_k) \approx \sum_{i=1}^{N_\mathcal{H}} \theta_{k,i}^\mathcal{H} \, K(x_k, x_{k,i}), \text{ with } \theta_{k,i}^\mathcal{H} \geq 0 \quad \forall k, i, \tag{7}$$

where K is positive definite and continuous. Additionally, we choose K radial, of the form $K(x_i, x_j) = \phi\left(\|x_i - x_j\|\right)$. Indeed, a radial kernel is local by construction. We also choose K compactly supported, so that interpolation matrices are sparse and computation is efficient. A possible choice of kernel verifying these properties is one of the Wendland functions [39, 26] defined as:

$$\phi(\rho; r) = \begin{cases} \dfrac{1}{\rho}\left(1 - \dfrac{r}{\rho}\right)^2 & r = \|x_i - x_j\| < \rho, \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

---

[1] that is, how much influence a data point has over its neighborhood.

where the scaling parameter ρ is different for each component.

4.3. Energy law model

To simplify notations in the following, we denote

$$\phi_{k,i} : x \mapsto \phi\left(\rho_k; \|x - x_{k,i}\|\right). \tag{9}$$

Finally, the energy law approximant we look for has the form:

$$\mathcal{H}\left(\boldsymbol{\theta}^{\mathcal{H}}; \boldsymbol{x}\right) := \sum_{k=1}^{N_x} \sum_{i=1}^{N_{\mathcal{H}}} \theta^{\mathcal{H}}_{k,i}\, \phi^{(-2)}_{k,i}(x_k)\,,\ \theta^{\mathcal{H}}_{k,i} \geq 0 \quad \forall k, i, \tag{10}$$

where $\phi^{(-2)}_{k,i}$ denotes the second antiderivative of $\phi_{k,i}$ (formal definition in Appendix A). The $\{\phi_{k,i}\}$ constitute a *compactly supported radial basis* of $\nabla^2\mathcal{H}$. The $\{x_{k,i}\}$ are called the *centers* of the radial basis [40].

5. PARAMETER ESTIMATION

5.1. Objective function

We denote $\boldsymbol{f}[j]$ the measured average flows at sample j:

$$\boldsymbol{f}[j] = \begin{bmatrix} \delta\boldsymbol{x}[j] f_s \\ \boldsymbol{y}[j] \end{bmatrix}, \tag{11}$$

where $\delta\boldsymbol{x}[j] = \boldsymbol{x}[j+1] - \boldsymbol{x}[j]$ and f_s is the sample rate. Similarly we denote $\tilde{\boldsymbol{f}}(\boldsymbol{\Theta})[j]$ the estimated flows at sample j:

$$\tilde{\boldsymbol{f}}(\boldsymbol{\Theta})[j] = \left(\boldsymbol{J}(\boldsymbol{\theta}^J) - \boldsymbol{R}(\boldsymbol{\theta}^R)\right) \begin{bmatrix} \overline{\nabla}\mathcal{H}\left(\boldsymbol{\theta}^{\mathcal{H}}, \boldsymbol{x}[j], \delta\boldsymbol{x}[j]\right) \\ \boldsymbol{u}[j] \end{bmatrix}, \tag{12}$$

where $\boldsymbol{\theta}^{\mathcal{H}} \succeq 0, \operatorname{diag}\left(\boldsymbol{T}(\boldsymbol{\theta}^R)\right) \succeq 0$, and $\overline{\nabla}\mathcal{H}$ is the discrete gradient [41] defined component-wise as:

$$\overline{\nabla}\mathcal{H}_k\left(\boldsymbol{\theta}^{\mathcal{H}}, \boldsymbol{x}[j], \delta\boldsymbol{x}[j]\right) = \begin{cases} \displaystyle\sum_{i=1}^{N_{\mathcal{H}}} \theta^{\mathcal{H}}_{k,i} \frac{\phi^{(-2)}_{k,i}\left(x_k[j] + \delta x_k[j]\right) - \phi^{(-2)}_{k,i}\left(x_k[j]\right)}{\delta x_k[j]} & |\delta x_k[j]| > \epsilon, \\ \displaystyle\sum_{i=1}^{N_{\mathcal{H}}} \theta^{\mathcal{H}}_{k,i}\, \phi^{(-1)}_{k,i}\left(x_k[j] + \frac{\delta x_k[j]}{2}\right) & \text{otherwise.} \end{cases} \tag{13}$$

Denoting $\boldsymbol{F}$ (resp. $\tilde{\boldsymbol{F}}(\boldsymbol{\Theta})$) the $\dim(\boldsymbol{J}) \times n$ matrix of measured flows (resp. estimated flows) at all n samples, we define the error $\boldsymbol{\epsilon}(\boldsymbol{\Theta}) = \tilde{\boldsymbol{F}}(\boldsymbol{\Theta}) - \boldsymbol{F}$, and the objective function $\mathcal{E}(\boldsymbol{\Theta}) = \left\|\tilde{\boldsymbol{F}}(\boldsymbol{\Theta}) - \boldsymbol{F}\right\|^2$, where $\|\|$ is the Frobenius norm. Finally we look for the optimal:

$$\begin{aligned} \boldsymbol{\Theta}^\star &= \arg\min \mathcal{E}(\boldsymbol{\Theta}) \\ \text{subject to } \boldsymbol{\theta}^{\mathcal{H}} &\succeq 0, \\ \operatorname{diag}\left(\boldsymbol{T}(\boldsymbol{\theta}^R)\right) &\succeq 0. \end{aligned} \tag{14}$$

Here $\operatorname{diag}\left(\boldsymbol{T}(\boldsymbol{\theta}^R)\right)$ denotes the diagonal elements of $\boldsymbol{T}(\boldsymbol{\theta}^R)$.

The sample rate f_s is considered high enough so that the numerical error is negligible. Therefore we consider that we identify the continuous system, and the specific contribution of the discretization scheme to the global error [42] is not addressed here.

5.2. Constrained optimization

To perform a constrained minimization of $\mathcal{E}$, we rely on the Interior Point Method [43]. As this method is well documented, in this section we only provide a basic mathematical layout as a reminder.

We define the loss function $\mathcal{L}$:

$$\mathcal{L} : \boldsymbol{\Theta} \mapsto \mathcal{E}(\boldsymbol{\Theta}) - \underbrace{\frac{1}{t}\left(\sum_{k=1}^{N_x}\sum_{i=1}^{N_{\mathcal{H}}} \ln\theta^{\mathcal{H}}_{k,i} + \sum_{k=1}^{\dim(\boldsymbol{J})} \ln\theta^R_k\right)}_{\text{logarithmic barrier}}. \tag{15}$$

The *logarithmic barrier* [43] penalizes the minimization if all coefficients are not strictly positive. The parameter t is set by the user to enforce or relax the constraint[2]. A necessary condition to minimize $\mathcal{L}$ is finding a solution to:

$$\boldsymbol{\mathcal{F}}(\boldsymbol{\Theta}, \boldsymbol{\mu}) = \begin{bmatrix} \partial\boldsymbol{\epsilon}(\boldsymbol{\Theta})^\intercal \boldsymbol{\epsilon}(\boldsymbol{\Theta}) - \boldsymbol{\mu} \\ \boldsymbol{\mu} \odot \boldsymbol{\Theta} - \boldsymbol{1}/t \end{bmatrix} = \boldsymbol{0}, \tag{16}$$

where $\partial\boldsymbol{\epsilon}$ is the Jacobian of $\boldsymbol{\epsilon}$, $\boldsymbol{\mu}$ is the derivative of the logarithmic barrier w.r.t $\boldsymbol{\Theta}$, and $\odot$ denotes the element-wise product. A solution is estimated with a damped Gauss-Newton iteration [44]. Starting from a particular set of parameters $\boldsymbol{\Theta}^0$ meeting the constraints, the set is improved iteratively using:

$$\left[\boldsymbol{\Theta}^{k+1}, \boldsymbol{\mu}^{k+1}\right]^\intercal = \left[\boldsymbol{\Theta}^{k}, \boldsymbol{\mu}^{k}\right]^\intercal - \alpha\, \boldsymbol{\mathcal{J}}^{-1}\left(\boldsymbol{\Theta}^{k}, \boldsymbol{\mu}^{k}\right) \boldsymbol{\mathcal{F}}\left(\boldsymbol{\Theta}^{k}, \boldsymbol{\mu}^{k}\right) \tag{17}$$

where $\alpha \in [0, 1]$ is a damping coefficient computed with a line search [43] and $\boldsymbol{\mathcal{J}}$ is the Jacobian of $\boldsymbol{\mathcal{F}}$ defined as:

$$\boldsymbol{\mathcal{J}}(\boldsymbol{\Theta}, \boldsymbol{\mu}) = \begin{bmatrix} \partial\boldsymbol{\epsilon}(\boldsymbol{\Theta})^\intercal \partial\boldsymbol{\epsilon}(\boldsymbol{\Theta}) & -\boldsymbol{I} \\ \operatorname{diag}(\boldsymbol{\mu}) & \operatorname{diag}(\boldsymbol{\Theta}) \end{bmatrix}. \tag{18}$$

Here $\operatorname{diag}(\boldsymbol{\mu})$ (resp. $\operatorname{diag}(\boldsymbol{\Theta})$) denotes the square diagonal matrix with the elements of $\boldsymbol{\mu}$ (resp. $\boldsymbol{\Theta}$) on its diagonal. The iteration is stopped when the error is sufficiently low, or, since the objective function is non convex, if the error starts increasing.

6. RESULTS FOR A VIRTUAL PASSIVE PEAKING EQ

6.1. Circuit parameters and data generation

We consider a passive peaking EQ [45] (Fig. 3a). The potentiometer wiper position is parametrized by $\gamma \in [0, 1]$, where γ = 0 corresponds to the lowest position, and γ = 1 to the highest. This parameter determines the shape of the frequency response (Fig. 3b). The potentiometer, resistor and capacitor are all considered to be linear. The inductor is saturating with an effort law of the type [31]:

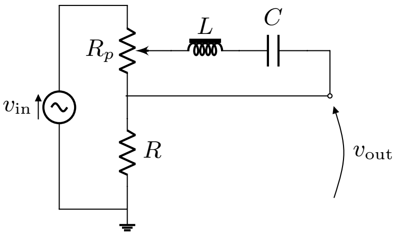
$$i_{\mathrm{L}} = I_0\left(\frac{\Phi_{\mathrm{L}}}{\Phi_{\mathrm{sat}}} - \tanh\left(\frac{\Phi_{\mathrm{L}}}{\eta\,\Phi_{\mathrm{sat}}}\right)\right), \tag{19}$$

where I_0, Φ_{sat}, and η are model parameters (hysteresis is neglected here). Circuit parameters are set so that the center frequency is 50 Hz and the quality factor is 1. They are shown in Table 1.

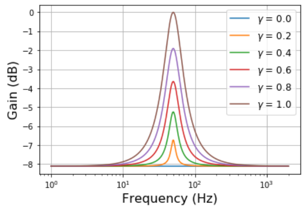
Synthetic measurement data are artificially generated for an input voltage of the form $v_{\mathrm{in}} = U_0\cos(2\pi f_0 t)$. The values of f_0 and γ are chosen so that the circuit is maximally resonant. This

---

[2] it can be increased dynamically during iteration.



(a) Passive peaking EQ circuit.



(b) Frequency response for several wiper positions.

Figure 3: Passive peaking EQ and its frequency responses.

way, nonlinearities of the inductor can be accurately captured for a plausible U_0. The theoretical PH-ODE of the passive peaking EQ is found to be that in Appendix B. The generation is achieved discretizing Eq. (2), and performing a standard Newton-Raphson iteration at each sample (a detailed numerical scheme can be found in [46]). Finally, some noise (SNR from 38 to 98 dB, with a normal distribution) is added to the data to test the robustness of the identification method.

6.2. Choice of kernel parameters and initialization

A minima, the approximant must reproduce a linear effort on the measured range of each state. We fix $N_{\mathcal{H}} = 6$. For this $N_{\mathcal{H}}$, we determine (for each component) the smallest ρ so that the relative error $\left|(g(x) - \tilde{g}(x))/g(x)\right| \times 100$ on the test effort $g(x) = x$ stays within some arbitrary bounds, chosen to be 10 %. Here, $\rho_{\text{C}} = 4.10^{-5}$ and $\rho_{\text{L}} = 4.10^{-3}$ yield satisfying results (Figs 4c-4d, for the inductor). Figure 4a (resp. 4b) shows the resulting basis for the approximation of $\nabla^2\mathcal{H}$ (resp. $\nabla\mathcal{H}$).

Before performing the optimization procedure, an initial guess $\boldsymbol{\Theta}^0$ has to be estimated. To this end, the problem is linearized around the desired solution. Measurement data are generated with an input voltage small enough to observe a quasi-linear effort, so we have:

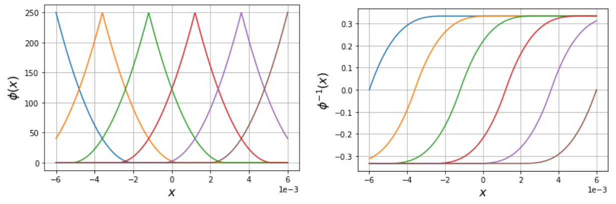
$$\boldsymbol{F} \approx \underbrace{(\boldsymbol{J} - \boldsymbol{R})\,\boldsymbol{D}}_{\boldsymbol{M}}\,\boldsymbol{X}, \tag{20}$$

where $\boldsymbol{D}$ is a diagonal matrix with positive coefficients, and $\boldsymbol{X}$ is the $\dim(\boldsymbol{J}) \times n$ matrix of average states $\bar{\boldsymbol{x}}[i] = (\boldsymbol{x}[i] + \boldsymbol{x}[i+1])/2$ and inputs at all n samples. Since $n \gg \dim(\boldsymbol{J})$, the matrix $\boldsymbol{M}$ is extracted using $\boldsymbol{M} = \boldsymbol{F}\boldsymbol{X}^{\dagger}$, where † denotes the pseudo-inverse. Denoting $\boldsymbol{M_s} = -(\boldsymbol{M} + \boldsymbol{M}^{\intercal})/2$ the opposite of the symmetric part of $\boldsymbol{M}$, the matrix $\boldsymbol{R}$ is initialized to a positive-definite matrix close to $\boldsymbol{M_s}$. To this end, $\boldsymbol{M_s}$ is decomposed as:

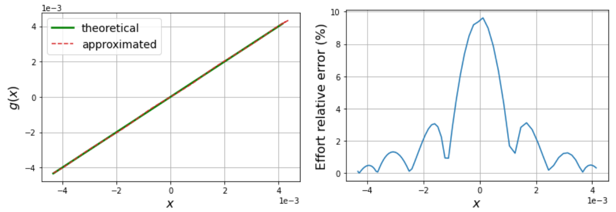
$$\boldsymbol{M_s} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\intercal}. \tag{21}$$

Then the matrix $\tilde{\boldsymbol{R}}$ defined as:

$$\tilde{\boldsymbol{R}} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^{\intercal} \tag{22}$$



(a) Chosen radial basis for $\nabla^2\mathcal{H}$. (b) Antiderivatives of the chosen radial basis for $\nabla^2\mathcal{H}$.



(c) Resulting approximation of the effort $g(x) = x$. (d) Relative error on the effort.

Figure 4: Chosen radial basis for the inductor and resulting approximation of the test effort $g(x) = x$.

is positive-definite, and $\boldsymbol{R}$ is initialized to:

$$\boldsymbol{R}^0 = \tilde{\boldsymbol{R}}/\left\|\tilde{\boldsymbol{R}}\right\|, \tag{23}$$

to account for the (unknown) contribution of $\boldsymbol{D}$. Finally, all θ^J_k and $\theta^{\mathcal{H}}_{k,i}$ are initialized to 1.

6.3. Results

The optimization procedure returns a set of estimated parameters after less than 50 iterations. Here, constraint enforcement is privileged over convergence speed as there are no real-time requirements. Still, the estimation is faster compared to deep neural networks methods (mainly because there are far less parameters to estimate).

New simulations are computed with the parameters estimated for each SNR. Figures 5a-5c show that the simulated states match very closely with the "measured" states (here with measured state SNR = 50 dB). Figures 5b-5d show the estimated effort laws. The estimated effort law for the capacitor is linear as expected. The saturating behavior of the inductor is accurately captured within the range of measured data. Figure 6 shows the simulated state Normalized Mean Square Error (NMSE = $20\log\left(\|x - \tilde{x}\|/\|x\|\right)$) vs the measured state SNR. The NMSE stays low (around -60 dB) regardless of the SNR.

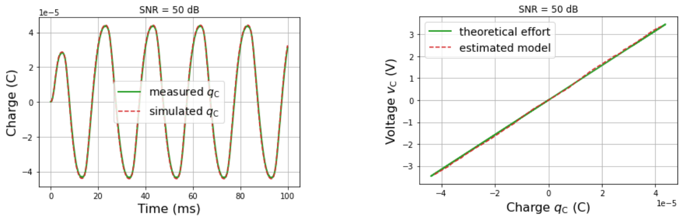
Finally, to evaluate the robustness of the estimated model, simulations and measurements are also run with different input amplitudes and frequencies than those used for the estimation. Figures 7a-7d show that the simulated states match closely with the measurements in that case also.

7. CONCLUSION

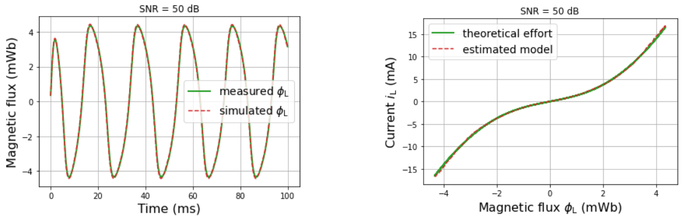
In this paper, we have presented an identification method to retrieve parameters of a circuit modeled as a port-Hamiltonian system, given measurements of state $\boldsymbol{x}$, input $\boldsymbol{u}$, and output $\boldsymbol{y}$. This

Table 1: Data generation parameters for the virtual passive peaking EQ.

R_p (Ω)	R (Ω)	γ	C (F)	I_0 (A)	Φ_{sat} (Wb)	η	U_0 (V)	f_0 (Hz)	f_s (Hz)
251.3	502.6	1	12.6×10^{-6}	50×10^{-3}	4×10^{-3}	1.1	0.5 & 10	50	10×10^{3}



(a) Capacitor simulated state. (b) Capacitor estimated effort.



(c) Inductor simulated state. (d) Inductor estimated effort.

Figure 5: Estimation results for the virtual passive peaking EQ, with measured state SNR = 50 dB.

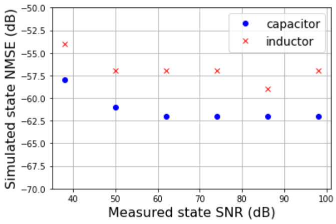


Figure 6: Simulated state Normalized Mean Square Error with respect to measured state SNR.

method allows the joint estimation of constitutive laws of storage components, as well as the interconnection matrix encoding the circuit topology. In turn, the estimated model may be used for passive guaranteed simulations.

The method has been tested on a virtual peaking EQ, with accurate results. However, the method needs to be more thoroughly assessed against real measurements. In particular, we should control that the discretization scheme in the simulation error does not introduce too much numerical dispersion, which would alter the optimization process. In that case, the discretization error would have to be taken into account [47], or the sample rate would have to be increased.

Another immediate perspective for this work is to extend the method in order to include nonlinear dissipative components. This would allow the inclusion of transistors and vacuum tubes, which are an important part of audio circuits.

It would also be interesting to adapt the method to a co-energy variables formulation. This way, measurements would only need to be voltages and currents, instead of charges and magnetic fluxes, which are much more difficult to obtain in practice. Identification from partial measurements (incomplete state, or input and output only) could be studied as well. This will be the object of future work.

8. REFERENCES

[1] Alfred Fettweis, “Wave digital filters: Theory and practice,” *Proceedings of the IEEE*, vol. 74, no. 2, pp. 270–327, 1986.

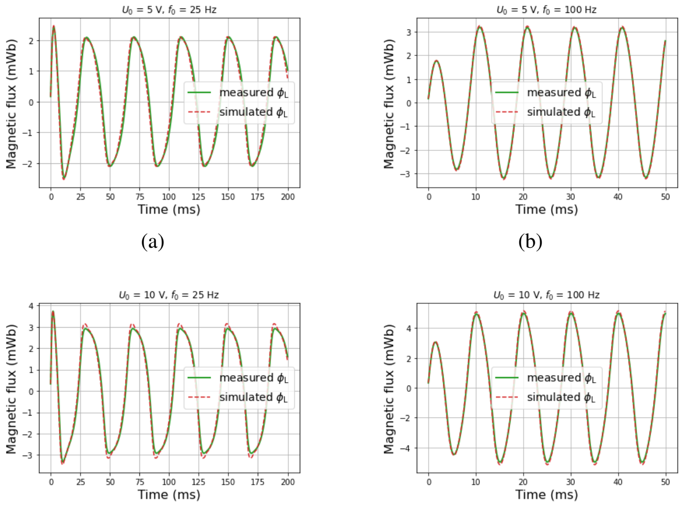
[2] Kurt James Werner, Alberto Bernardini, Julius O. Smith, and Augusto Sarti, “Modeling circuits with arbitrary topologies and active linear multiports using wave digital filters,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 65, no. 12, pp. 4233–4246, 2018.

[3] Alberto Bernardini, Alessio E. Vergani, and Augusto Sarti, “Wave digital modeling of nonlinear 3-terminal devices for virtual analog applications,” *Circuits, Systems, and Signal Processing*, pp. 1–31, 2020.

[4] Ivan Cohen and Thomas Hélie, “Real-time simulation of a guitar power amplifier,” in *Proceedings of the 13th International Conference on Digital Audio Effects (DAFx-10)*, 2010.

[5] Jaromir Macak, Jiri Schimmel, and Martin Holters, “Simulation of Fender type guitar preamp using approximation and state space model,” in *Proceedings of the 15th International Conference on Digital Audio Effects (DAFx-12)*, 2012.

[6] Martin Holters and Udo Zölzer, “A generalized method for the derivation of non-linear state-space models from circuit



(a) (b)

(c) (d)

Figure 7: Inductor simulated state for different input amplitudes U_0 and input frequencies f_0.

schematics," in *23rd European Signal Processing Conference (EUSIPCO)*, 2015, pp. 1073–1077.

[7] Eero-Pekka Damskägg, Lauri Juvela, Vesa Välimäki, et al., "Real-time modeling of audio distortion circuits with deep learning," in *Proceedings of the International Sound and Music Computing Conference (SMC-19), Malaga, Spain*, 2019, pp. 332–339.

[8] M. A. Martínez Ramírez, Emmanouil Benetos, and Joshua D. Reiss, "Deep learning for black-box modeling of audio effects," *Applied Sciences*, vol. 10, no. 2, pp. 638, 2020.

[9] Thomas Hélie and David Roze, "Sound synthesis of a nonlinear string using Volterra series," *Journal of Sound and Vibration*, vol. 314, no. 1-2, pp. 275–306, 2008.

[10] Thomas Hélie, "Volterra series and state transformation for real-time simulations of audio circuits including saturations: Application to the Moog ladder filter," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 4, pp. 747–759, 2009.

[11] Stephen Boyd, Y. S. Tang, and Leon Chua, "Measuring Volterra kernels," *IEEE Transactions on Circuits and Systems*, vol. 30, no. 8, pp. 571–577, 1983.

[12] Simone Orcioni, Alessandro Terenzi, Stefania Cecchi, Francesco Piazza, and Alberto Carini, "Identification of Volterra models of tube audio devices using multiple-variance method," *Journal of the Audio Engineering Society*, vol. 66, no. 10, pp. 823–838, 2018.

[13] Damien Bouvier, Thomas Hélie, and David Roze, "Phase-based order separation for Volterra series identification," *International Journal of Control*, pp. 1–11, 2019.

[14] Akanksha Sondhi, Richa Barsainya, and Tarun Kumar Rawat, "Lattice Wave Digital Filter based IIR system identification with reduced coefficients," in *The International Symposium on Intelligent Systems Technologies and Applications*. Springer, 2016, pp. 695–706.

[15] Johan Paduart, Lieve Lauwers, Jan Swevers, Kris Smolders, Johan Schoukens, and Rik Pintelon, "Identification of nonlinear systems using polynomial nonlinear state space models," *Automatica*, vol. 46, no. 4, pp. 647–656, 2010.

[16] Julian D. Parker, Fabián Esqueda, and André Bergner, "Modelling of nonlinear state-space systems using a deep neural network," in *Proceedings of the 22rd International Conference on Digital Audio Effects (DAFx-19), Birmingham, UK*, 2019, pp. 2–6.

[17] Shahan Nercessian, Andy Sarroff, and Kurt James Werner, "Lightweight and interpretable neural modeling of an audio distortion effect using hyperconditioned differentiable biquads," *arXiv preprint arXiv:2103.08709*, 2021.

[18] B. M. Maschke, A. J. van der Schaft, and P. Breedveld, "An intrinsic Hamiltonian formulation of network dynamics: Non-standard Poisson structures and gyrators," *Journal of the Franklin Institute*, pp. 923–966, 1992.

[19] A. J. van der Schaft, Dimitri Jeltsema, et al., "Port-Hamiltonian systems theory: An introductory overview," *Foundations and Trends in Systems and Control*, vol. 1, no. 2-3, pp. 173–378, 2014.

[20] Jooyoung Park and Irwin W. Sandberg, "Universal approximation using radial-basis-function networks," *Neural Computation*, vol. 3, no. 2, pp. 246–257, 1991.

[21] Vladimir Vapnik, Steven E. Golowich, Alex Smola, et al., "Support vector method for function approximation, regression estimation, and signal processing," *Advances in Neural Information Processing Systems*, pp. 281–287, 1997.

[22] Antoine Falaize and Thomas Hélie, "Passive guaranteed simulation of analog audio circuits: A port-Hamiltonian approach," *Applied Sciences*, vol. 6, no. 10, pp. 273, 2016.

[23] Silviu Medianu, *Identification for port-controlled Hamiltonian systems*, Ph.D. thesis, Universitatea politehnica (Bucarest), 2017.

[24] Peter Benner, Pawan Goyal, and Paul Van Dooren, "Identification of port-Hamiltonian systems from frequency response data," *Systems & Control Letters*, vol. 143, pp. 104741, 2020.

[25] Karim Cherifi, "An overview on recent machine learning techniques for port-Hamiltonian systems," *Physica D: Nonlinear Phenomena*, p. 132620, 2020.

[26] Robert Schaback and Holger Wendland, "Kernel techniques: From machine learning to meshless methods," *Acta Numerica*, vol. 15, pp. 543, 2006.

[27] Daniel J. Gillespie and Daniel P. W. Ellis, "Modeling nonlinear circuits with linearized dynamical models via kernel regression," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2013, pp. 1–4.

[28] Michael Lutter, Christian Ritter, and Jan Peters, "Deep lagrangian networks: Using physics as model prior for deep learning," *arXiv preprint arXiv:1907.04490*, 2019.

[29] Vincent Duindam, Alessandro Macchelli, Stefano Stramigioli, and Herman Bruyninckx, *Modeling and control of complex physical systems: the port-Hamiltonian approach*, Springer Science & Business Media, 2009.

[30] Timothy A. Bigelow, "Power and energy in electric circuits," in *Electric Circuits, Systems, and Motors*, pp. 105–121. Springer, 2020.

[31] Judy Najnudel, Rémy Müller, Thomas Hélie, and David Roze, "A power-balanced dynamic model of ferromagnetic coils," in *Proceedings of the 23rd International Conference on Digital Audio Effects (DAFx-20)*, 2020.

[32] Zdenek Biolek, Dalibor Biolek, Zdenek Kolka, and Viera Biolkova, "Real-world capacitor as a memcapacitive element," in *New Trends in Signal Processing (NTSP)*, 2018, pp. 1–6.

[33] Nicholas J. Higham, *Analysis of the Cholesky decomposition of a semi-definite matrix*, Oxford University Press, 1990.

[34] Commandant Benoit, "Note sur une méthode de résolution des équations normales provenant de l'application de la méthode des moindres carrés à un système d'équations linéaires en nombre inférieur à celui des inconnues. Application de la méthode à la résolution d'un système défini d'équations linéaires (procédé du Commandant Cholesky)," *Bulletin géodésique*, vol. 2, no. 1, pp. 67–77, 1924.

[35] Robert Schaback, "Native Hilbert spaces for radial basis functions I," in *New Developments in Approximation Theory*, pp. 255–282. Springer, 1999.

[36] Oliver Nelles, "Nonlinear system identification," 2002.

[37] Gregory E Fasshauer and Jack G Zhang, "On choosing "optimal" shape parameters for rbf approximation," *Numerical Algorithms*, vol. 45, no. 1-4, pp. 345–368, 2007.

[38] Michael Mongillo, "Choosing basis functions and shape parameters for radial basis function methods," *SIAM undergraduate research online*, vol. 4, no. 190-209, pp. 2–6, 2011.

[39] Holger Wendland, "Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree," *Advances in Computational Mathematics*, vol. 4, no. 1, pp. 389–396, 1995.

[40] Robert Schaback, "A practical guide to radial basis functions," *Electronic Resource*, vol. 11, pp. 1–12, 2007.

[41] Robert I. McLachlan and G. R. W. Quispel, "Discrete gradient methods have an energy conservation law," *arXiv preprint arXiv:1302.4513*, 2013.

[42] Jan L. Cieśliński and Bogusław Ratkiewicz, "Discrete gradient algorithms of high order for one-dimensional systems," *Computer Physics Communications*, vol. 183, no. 3, pp. 617–627, 2012.

[43] Stephen Boyd and Lieven Vandenberghe, *Convex optimization*, Cambridge University Press, 2004.

[44] Peter Deuflhard, *Newton methods for nonlinear problems: Affine invariance and adaptive algorithms*, vol. 35, Springer Science & Business Media, 2011.

[45] Dennis A Bohn, "Operator adjustable equalizers: An overview," in *Audio Engineering Society Conference: 6th International Conference: Sound Reinforcement*. Audio Engineering Society, 1988.

[46] Judy Najnudel, Thomas Hélie, David Roze, and Henri Boutin, "Simulation of an ondes Martenot circuit," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2651–2660, 2020.

[47] Takeru Matsuda and Yuto Miyatake, "Estimation of ordinary differential equation models with discretization error quantification," *SIAM/ASA Journal on Uncertainty Quantification*, vol. 9, no. 1, pp. 302–331, 2021.

A. DERIVATIVE AND ANTIDERIVATIVE NOTATIONS

By convention, $f^{(0)} = f$. Then for some positive integer $k \geq 1$, we denote $f^{(k)}$ the function defined by:

$$f^{(k)} : x \mapsto \frac{\mathrm{d}}{\mathrm{d}x} f^{(k-1)}(x). \tag{24}$$

Similarly, we denote $f^{(-k)}$ a function defined by:

$$f^{(-k)} : x \mapsto \int_0^x f^{(1-k)}(u)\mathrm{d}u + C, \tag{25}$$

where C is a constant. In this work, C is taken so that $f^{(-k)}(0) = 0$.

B. PH-ODE OF THE PASSIVE PEAKING EQ

With

$$\alpha_p = \frac{R}{R_p + R}, \quad G_p = \frac{1}{R_p + R}, \quad R_p \parallel R = \frac{R_p R}{R_p + R}, \tag{26}$$

the PH-ODE of the passive peaking EQ on Fig. 3a is given by:

$$\begin{array}{c|cc|cc} & v_{\mathrm{C}} & i_{\mathrm{L}} & v_{\mathrm{in}} & i_{\mathrm{out}} \\ i_{\mathrm{C}} & \cdot & 1 & \cdot & \cdot \\ v_{\mathrm{L}} & -1 & -\gamma R_p \parallel R \left(\frac{R_p}{R} (1-\gamma) + 1 \right) & \gamma R_p G_p & \gamma R_p \parallel R \\ \hline i_{\mathrm{in}} & \cdot & -\gamma R_p G_p & -G_p & -\alpha_p \\ v_{\mathrm{out}} & \cdot & \gamma R_p \parallel R & \alpha_p & -R_p \parallel R \end{array}. \tag{27}$$