Evaluating Error Bound for Physics-Informed Neural Networks on Linear Dynamical Systems

Anonymous Author(s)
Affiliation
Address
email

1 Introduction

Differential equations play an essential role in a wide range of mathematical modeling processes. In cases where analytical solutions are nonexistent, numerical methods (finite difference, finite volume, finite element, spectral method) have been studied. Recently, massive attention has been paid to solving differential equations with neural networks. These networks are trained to minimize the squared residual of some differential equations. However, there is little interpretation for the loss functions (squared residuals) except that it should be as close to zero as possible. Little effort has been made to quantify the error of network solution based on the residuals. As a result, the reliability of neural network solutions remain questionable.

With mathematical proof, we propose here an algorithm for fast error bound evaluation based only on residuals of linear ODEs. The method makes no assumption on the network architecture or whether the network is sufficiently trained. Apart from the characteristics and structure of the dynamical systems in question, the error bound yielded by the algorithm ($O(\varepsilon t^m)$) only depends on time t and the largest differential equation residual (ε) (infinity norm) over the domain of interest. We further present that, for strictly stable systems, one can derive a bound ($O(\varepsilon)$) that is independent of time t. Finally, we present a technique to tighten the error bound by dividing the time domain into subintervals and evaluating the maximum residual on each one.

2 Background and Previous Work

Lagaris et al. [1] first proposed solving differential equations using neural networks [2] due to differentiability of neural networks with appropriate activation functions. To train a network solution $\mathbf{u}(t)$ for a differential equation $\mathcal{L}\mathbf{u} = \mathbf{f}$, one essentially minimizes an approximation of the L_2 norm of differential equation residual on a domain Ω

$$\int_I (\mathcal{L}\mathbf{u} - \mathbf{f})^2 \,\mathrm{d}t \approx \frac{|I|}{N} \sum_{\substack{i=1 \\ t_i \in I}}^{N} (\mathcal{L}\mathbf{u}(t_i) - \mathbf{f}(t_i))^2 := \mathrm{Loss}.$$

where $\mathcal{L}$ is a (possibly nonlinear) differential operator.

Little effort has been to study the failure modes and absolute error of network solutions until recent years [3] [4] [5]. In [6], Ryck and Mishra established a foundation and rationale for error of PINNs in approximating PDEs. Making Kolmogorov PDEs as an example, they have shown that there exists PINNs, approximating these PDEs such that the resulting generalization error and the total error can be made arbitrarily small. However, the existence of such neural networks does not guarantee network training converges in practice. A more practical concern is how to evaluate the error given any network (possibly illy trained) on certain equations. In our work, we derive the error bound for a class of linear ODEs, which can be efficiently computed using only ODE residuals.

Submitted to the DLDE Workshop in the 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

Equation	Forcing $f(t)$	$u(0)$	$u'(0)$	Exact Solution $u(t)$
$u''+u=f$	$2e^t$	2.0	2.0	$\sin t+\cos t+e^t$
$u''+u=f$	t^2+t+3	2.0	2.0	$\sin t+\cos t+t^2+t+1$
$u''+u=f$	$\ln(t+1)-(t+1)^{-2}$	1.0	2.0	$\sin t+\cos t+\ln(t+1)$
$u''+u=f$	$2\cos t^2+(1-4t^2)\sin t^2$	1.0	1.0	$\sin t+\cos t+\sin t^2$
$u''+4u'+3u=f$	$8e^t$	3.0	-3.0	$e^{-t}+e^{-3t}+e^t$
$u''+4u'+3u=f$	$3t^2+11t+9$	3.0	-3.0	$e^{-t}+e^{-3t}+t^2+t+1$
$u''+4u'+3u=f$	$3\ln(t+1)+4(t+1)^{-1}-(t+1)^{-2}$	2.0	-3.0	$e^{-t}+e^{-3t}+\ln(t+1)$
$u''+4u'+3u=f$	$6\cos t-2\sin t$	3.0	-3.0	$e^{-t}+e^{-3t}+\sin t+\cos t$

Table 1: Experiment Setup for Section 4.1

3 Approach for Evaluating Error Bound

For any ODE (or system of ODEs) discussed in Appendix A, we are able to bound the error of any network by simply evaluating its infinite norm (maximum residual). This is true for any neural network solution, regardless of how well it is trained or trained at all. Tthe process is straightforward.

First, we compute the residual of the network over sampled points $\{t_1, t_2, \dots\}$ (usually 1k – 10k points will suffice) from a domain I using automatic differentiation. This only take milliseconds due to GPU's power of parallel computation. Then, we compute the maximum absolute value of the residuals evaluated, which we denote as ε. In the case of ODE systems, ε will be the maximum norm of residual vectors.

Finally, a loose error bound is given by $\left\|u_{\text{net}} - u_{\text{true}}\right\| \le K\varepsilon t^m$ assuming the initial condition is met (i.e., no error at $t=0$), where t is the time and m and K are constants depending on the structure of the ODE. For strictly stable ODE systems, $m=0$, and the error bound is proportional to ε. For a single linear ODE, constant $K=\prod_k \left|\lambda_k\right|$ where $\lambda_k + i\omega_k$ are roots to its characteristic polynomial with nonzero λ_k. For a system of linear ODEs, constant K is determined by its eigenvectors and eigenvalues (as well as their multiplicity). An exact formula is given by Eq. 37 in Appendix A.3.

Note that the above bound $K\varepsilon t^m$ is a loose estimation which can be tightened when domain I is bounded as shown in Appendix A. Furthermore, an even tighter bound is discussed in Appendix A.5 by partitioning domain I into subintervals $I = I_1 \cup I_2 \cup \dots$ and compute an ε_k for each I_k.

4 Experimental Results

We run the following experiments with the NeuroDiffEq library [7], which provides a convenient and flexible framework for training neural networks to solve differential equations. Unless otherwise specified, we use an Adam optimizer with a learning rate of 1.0×10^{-3} and $(\beta_1, \beta_2) = (0.9,\ 0.999)$ for training the networks. The neural networks are simple fully-connected neural networks, with two 32-unit hidden layers and $\tanh$ activation function. The loss function we use is the L_2-norm of the ODE residuals at sampled points in the domain. The solution we choose is the one from the epoch with the lowest validation loss. We apply the reparametrization $\mathbf{u}(t) = \mathbf{u}_0 + \left(1-e^{-(t-t_0)}\right)\mathrm{ANN}(t)$ to enforce the initial conditions $\mathbf{u}(t_0) = \mathbf{u}_0$ and, where required, $\mathbf{u}(t) = \mathbf{u}_0 + (t-t_0)\,\mathbf{u}_0' + \left(1-e^{-(t-t_0)^2}\right)\mathrm{ANN}(t)$ to enforce $\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{u}(t_0) = \mathbf{u}_0'$ in addition to $\mathbf{u}(t_0) = \mathbf{u}_0$.

4.1 Higher-Order Linear ODE with Constant Coefficients

Here we consider two types of second-order differential equation, $u''+u=f$ and $u''+4u'+3u=f$ where the solution space of the the associated homogeneous solution has basis $\{\sin t, \cos t\}$ and $\{e^{-t}, e^{-3t}\}$ respectively. By Eq. 23 and 20 the error bounds for a single interval are $\varepsilon t^2/2$ and $\varepsilon\left(2+e^{-3t}-3e^{-t}\right)/6$ respectively, where ε is the largest absolute residual over the interval.

We pick the forcing terms and initial conditions as described in Table 1. We train the network on $I=[0,3]$ for 100 and 1000 epochs. The ODE residual and error bound with $n=1,10,100$ subintervals are plotted in Figures 1 and 2.

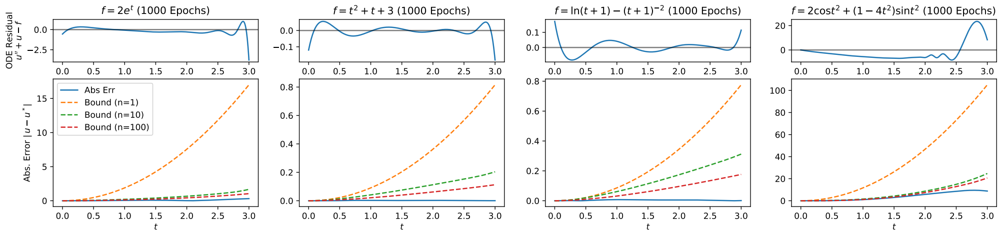

Figure 1: Residuals and Error Bounds of Harmonic Oscillator Under Various Forcing

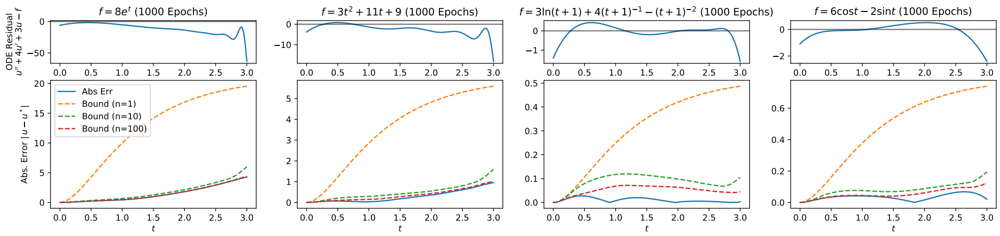

Figure 2: Residuals and Error Bounds of Stable 2nd Order Equation Under Various Forcing

4.2 System of First-Order Linear ODEs with Constant Coefficients

We consider a system of linear ODEs, $\mathbf{u}' + MJM^{-1}\mathbf{u} = \mathbf{f}$, under the initial condition $\mathbf{u}(t_0) = M\begin{pmatrix}1 & 1 & 1 & 1 & 1 & 1\end{pmatrix}^T$ where $J \in \mathbb{R}^{6\times 6}$ is the Jordan canonical form $\begin{pmatrix}J_1 & & \\ & J_2 & \\ & & J_3\end{pmatrix}$ with Jordan blocks $J_1 = \begin{pmatrix}4 & 1 & \\ & 4 & 1 \\ & & 4\end{pmatrix}$, $J_2 = \begin{pmatrix}3 & 1 \\ & 3\end{pmatrix}$, and $J_3 = 2$, and the forcing is determined by

$$\mathbf{f}(t) = M\begin{pmatrix}\cos t + 4\sin t + e^t - 1 & 5e^t - 4 + t^2 & 4t^2 + 2t & 3t^3 + 3t^2 + e^{2t} - 1 & 5e^{2t} - 3 & \frac{1}{t+1} + 2\ln(t+1)\end{pmatrix}^T$$

We randomly sample orthogonal 6×6 matrices $M = M^{-T}$ to ensure $\mathrm{cond}(M) = 1$. By Eq. 34 and 37, the error bound is $\varepsilon\sqrt{H_3^2(t;4) + H_2^2(t;4) + H_1^2(t;4) + H_2^2(t;3) + H_1^2(t;3) + H_1^2(t;2)} \leq \sqrt{6}\varepsilon/2$ for a single interval where ε is the largest residual norm over the interval. The system is solved for $t \in I = [0, 3]$ for 1000 epochs with 1024 uniformly resampled points from the expanded domain $[-1, 4]$ at each epoch. We use networks with two 512-unit (instead of 32-unit) hidden layers due to the coupled nature of the system. However, it should be pointed out that, the error bound holds true regardless of the network size or how well the network is trained. Again, we divide I into $n = 1, 10, 100$ subintervals for increasingly tighter bounds. Figure 3 shows the system residual norm, network solution, as well as error bounds.

4.3 First-Order Linear ODE with Nonconstant Coefficients

In this section, we consider linear ODE with time-dependent coefficients, with $p(t)$, $f(t)$, initial condition and derived error bound for a single interval according to Eq. 40 tabulated in Table 2.

$$u' + p(t)u = f(t) \quad t \in I = [0, 3] \tag{1}$$

We train the network for 1000 epochs with 1024 points uniformly resampled from $I = [0, 3]$. According to Appendix A.4, the absolute error bound is $O(\varepsilon t)$. By evenly dividing I into $n = 1, 10, 100$ subintervals, we obtain the error bounds in Figure 4.

5 Conclusion

In this work, we have ascertained the link between ODE residuals and error bound. We have proven, that for stable ODE systems discussed above, the bound only depends on characteristics of the ODE

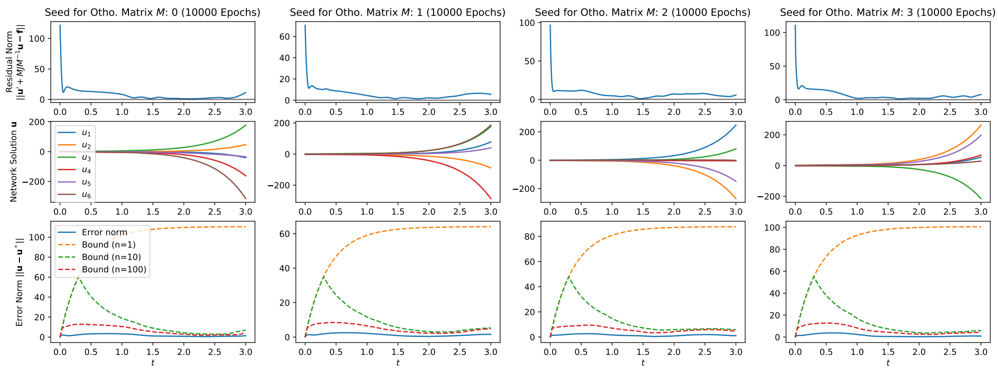

Figure 3: ODE Residuals and Absolute Error of ANN solution (System of Linear ODEs)

Forcing $f(t)$	Coefficient $p(t)$	IC $u(0)$	Exact Solution $u(t)$	Bound
$(2t+1)(t+1)^{-1}\cos t - t\sin t$	$(t+1)^{-1}$	1.0	$(t+1)^{-1} + t\cos t$	$\varepsilon\frac{t^2/2+t}{t+1}$
$(t+1)^2\left(t^2+1\right)^{-1}e^t$	$2t\left(t^2+1\right)^{-1}$	2.0	$(t^2+1)^{-1} + e^t$	$\varepsilon\frac{t^3/3+t}{t^2+1}$
$2t + t^2\cos t\,(1+\sin t)^{-1}$	$\cos t\,(1+\sin t)^{-1}$	1.0	$(1+\sin t)^{-1} + t^2$	$\varepsilon\frac{t-\cos t}{\sin t+1}$
$(1+(t+2)\ln(t+1))\,(t+1)^{-1}$	$(t+2)(t+1)^{-1}$	1.0	$e^{-t}\,(t+1)^{-1} + \ln(t+1)$	$\varepsilon\frac{e^t}{e^t+1}$

Table 2: Experiment Setup $u' + p(t)u = f(t)$ for Section 4.2, where ε in derived bounds is the largest absolute residual over a single interval

(or system of ODEs) and the maximum residual norm. There are efficient ways to evaluate the bound over the interval, as we did in Section 4.

In our experiments, we have shown that while these bounds are sometimes too loose using only the global maximum residual norm, they are usually asymptotically bounded by some constant. One can further tighten the bound by dividing the domain into smaller subintervals. In our experiments, the subintervals are linearly divided, but one can also use an adaptive quadrature for this task [8].

6 Future Work

In this work, we tie linear ODEs residuals to the absolute error and showed the error can be bounded by a function of residuals. A subsequent research topic is what strategy can be used to ensure a low residual, which in turn guarantees a low absolute error.

As another future extension, our proposed method may be generalizable to system of linear ODEs with time-dependent coefficients. It is also interesting to study if this method is applicable to local linear approximation of nonlinear ODEs. Furthermore, spatial or spatiotemporal PDEs with Dirichlet boundary conditions is also worth exploring.

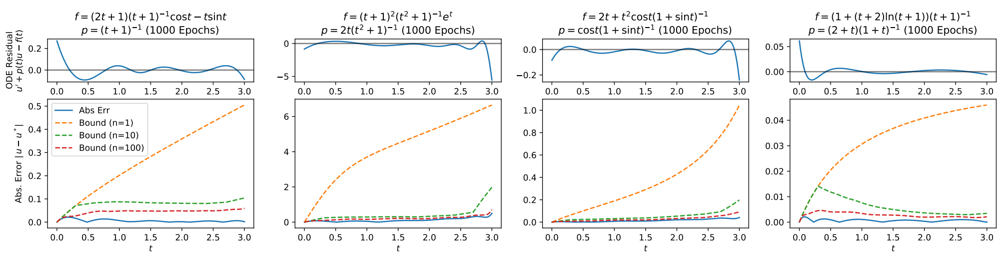

Figure 4: ODE Residuals and Absolute Error of ANN solution (Nonconstant Coefficients)

References

[1] Isaac E Lagaris, Aristidis Likas, and Dimitrios I Fotiadis. "Artificial neural networks for solving ordinary and partial differential equations". In: *IEEE transactions on neural networks* 9.5 (1998), pp. 987–1000.

[2] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. "Multilayer feedforward networks are universal approximators". In: *Neural networks* 2.5 (1989), pp. 359–366.

[3] Olga Graf et al. "Uncertainty Quantification in Neural Differential Equations". In: *arXiv preprint arXiv:2111.04207* (2021).

[4] Mengwu Guo and Ehsan Haghighat. "An energy-based error bound of physics-informed neural network solutions in elasticity". In: *arXiv preprint arXiv:2010.09088* (2020).

[5] Aditi Krishnapriyan et al. "Characterizing possible failure modes in physics-informed neural networks". In: *Advances in Neural Information Processing Systems* 34 (2021).

[6] Tim De Ryck and Siddhartha Mishra. "Error analysis for physics informed neural networks (PINNs) approximating Kolmogorov PDEs". In: *arXiv preprint arXiv:2106.14473* (2021).

[7] Feiyu Chen et al. "NeuroDiffEq: A Python package for solving differential equations with neural networks". In: *Journal of Open Source Software* 5.46 (2020), p. 1931.

[8] William Marshall McKeeman. "Algorithm 145: Adaptive numerical integration by Simpson's rule". In: *Communications of the ACM* 5.12 (1962), p. 604.

[9] Friedrich L Bauer and Alston S Householder. "Absolute norms and characteristic roots". In: *Numerische Mathematik* 3.1 (1961), pp. 241–246.

[10] Friedrich L Bauer. "Optimally scaled matrices". In: *Numerische Mathematik* 5.1 (1963), pp. 73–87.

[11] Richard D Braatz and Manfred Morari. "Minimizing the Euclidean condition number". In: *SIAM Journal on Control and Optimization* 32.6 (1994), pp. 1763–1768.

A Error Bound Proof and Methodology

Throughout this section, we use $\mathbf{u} : I \to \mathbb{C}^n$ to denote the neural network solution to $\mathcal{L}\mathbf{u} = \mathbf{f}$ where I can be any of the forms $(t_0, t_1]$, (t_0, t_1) or (t_0, ∞), and $\mathcal{L}$ is a linear differential operator. In the one-dimensional case, we use non-bold font u and f instead of $\mathbf{u}$ and $\mathbf{f}$. The solution residual is defined as $R\mathbf{u}(t) := \mathcal{L}\mathbf{u}(t) - \mathbf{f}(t)$. The exact solution $\mathbf{u}^*(t)$ satisfies $R\mathbf{u}^*(t) \equiv \mathbf{0}$ and the exact natural response $\mathbf{u}_n^*(t)$ is defined to satisfy the associated homogeneous equation $\mathcal{L}\mathbf{u}_n^*(t) = \mathbf{0}$. Both $\mathbf{u}^*$ and $\mathbf{u}_n^*$ satisfy the same initial condition $\mathbf{u}^*(t_0) = \mathbf{u}_n^*(t_0) = \mathbf{u}_0^*$.

A.1 First-Order Linear ODE with Constant Coefficients

It is well known that the most general form of first-order linear ODE with constant coefficients is $u'(t) + cu(t) = f(t)$ where $c \in \mathbb{C}$ is a constant and u' is the derivative of u.

Proposition If the residual $Ru(t)$ of equation $u' + (\lambda + i\omega)u = f$, where $\lambda, \omega \in \mathbb{R}$, is bounded by $\varepsilon \geq 0$ on I, namely,

$$|u' + (\lambda + i\omega)u - f| \leq \varepsilon \quad \forall t \in I, \tag{2}$$

and the network solution u satisfies initial condition with $u(t_0) = u_0^* \neq 0$, then,

a) The absolute error is bounded by $|u - u^*| \leq \dfrac{\varepsilon}{\lambda} \leq O(\varepsilon)$ on I if the natural response u_n^* is convergent ($\lambda > 0$);

b) The relative error w.r.t. u_n^* is bounded by $\left|\dfrac{u - u^*}{u_n^*}\right| \leq \dfrac{\varepsilon}{-\lambda |u_0^*|} \leq O(\varepsilon)$ on I if the natural response u_n^* is divergent ($\lambda < 0$); and

c) The absolute and relative errors are bounded by $|u - u^*| \leq O(\varepsilon t)$ and $\left|\dfrac{u - u^*}{u_n^*}\right| \leq O(\varepsilon t)$ on I if $\lambda = 0$.

Proof Multiply the integrating factor $e^{\lambda t+i\omega t}$ on both sides of Eq. 2 and evaluate the integral on $(t_0, t) \subseteq I$,

$$\left|\int_{t_0}^{t} e^{\lambda\tau+i\omega\tau}\Big(u'(\tau)+(\lambda+\omega i)u(\tau)-f(\tau)\Big)\mathrm{d}\tau\right| \\ \leq \int_{t_0}^{t}\left|e^{\lambda\tau+i\omega\tau}\Big(u'(\tau)+(\lambda+\omega i)u(\tau)-f(\tau)\Big)\right|\mathrm{d}\tau \leq \int_{t_0}^{t}\left|e^{\lambda\tau+i\omega\tau}\right|\varepsilon\mathrm{d}\tau \quad (3)$$

The first part of inequality holds because modulus of integral is smaller than integral of modulus. The second part holds by multiplying $e^{\lambda t+i\omega t}$ on both sides of Eq. 2 and taking the integral on (t_0, t), both of which preserve inequality property.

$$\left|e^{\lambda t+i\omega t}u(t)-e^{\lambda t_0+i\omega t_0}u(t_0)-\int_{t_0}^{t}e^{\lambda\tau+i\omega\tau}f(\tau)\mathrm{d}\tau\right| \leq \varepsilon\int_{t_0}^{t}e^{\lambda\tau}\mathrm{d}\tau \quad (4)$$

L.H.S. is reduced using $\int_{t_0}^{t} e^{\lambda\tau+i\omega\tau}\left(u'+(\lambda+i\omega)u\right)\mathrm{d}\tau = \int_{t_0}^{t}\mathrm{d}\left(e^{\lambda\tau+i\omega\tau}u(\tau)\right) = \left[e^{\lambda\tau+i\omega\tau}u(\tau)\right]_{t_0}^{t}$ and R.H.S. is reduced using $\left|e^{\lambda\tau+i\omega\tau}\right| \equiv e^{\lambda\tau}$.

$$\left|u(t)-e^{\lambda(t_0-t)+i\omega(t_0-t)}u(t_0)-e^{-\lambda t-i\omega t}\int_{t_0}^{t}e^{\lambda\tau+i\omega\tau}f(\tau)\mathrm{d}\tau\right| \leq \varepsilon e^{-\lambda t}\int_{t_0}^{t}e^{\lambda\tau}\mathrm{d}\tau \quad (5)$$

Both sides are divided by $\left|e^{\lambda t+i\omega t}\right|$.

Notice that the analytical solution is given by

$$u^*(t) = e^{\lambda(t_0-t)+i\omega(t_0-t)}u_0^* + e^{-\lambda t-i\omega t}\int_{t_0}^{t}e^{\lambda\tau+i\omega\tau}f(\tau)\mathrm{d}\tau.$$

Define the alternative solution to Eq. 2 under perturbed initial condition, $u(t_0)$ as

$$\widetilde{u}(t) := e^{\lambda(t_0-t)+i\omega(t_0-t)}u(t_0) + e^{-\lambda t-i\omega t}\int_{t_0}^{t}e^{\lambda\tau+i\omega\tau}f(\tau)\mathrm{d}\tau.$$

With this, Eq. 4 can be rewritten as

$$|u(t)-\widetilde{u}(t)| \leq \varepsilon e^{-\lambda t}\int_{t_0}^{t}e^{\lambda\tau}\mathrm{d}\tau. \quad (6)$$

By the triangle inequality,

$$|u(t)-u^*(t)| \leq |u(t)-\widetilde{u}(t)| + |\widetilde{u}(t)-u^*(t)| \leq \varepsilon e^{-\lambda t}\int_{t_0}^{t}e^{\lambda\tau}\mathrm{d}\tau + |\widetilde{u}(t)-u^*(t)|. \quad (7)$$

As $\widetilde{u} = u^*$ when $u(t_0) = u_0^*$, Eq. 7 is reduced to

$$|u(t)-u^*(t)| \leq \varepsilon e^{-\lambda t}\int_{t_0}^{t}e^{\lambda\tau}\mathrm{d}\tau. \quad (8)$$

If $\lambda > 0$, Eq. 8 gives rise to the absolute error bound

$$|u(t)-u^*(t)| \leq \varepsilon\frac{1-e^{\lambda(t_0-t)}}{\lambda} \leq \frac{\varepsilon}{\lambda} = O(\varepsilon) \quad (\lambda > 0). \quad (9)$$

If $\lambda < 0$, dividing Eq. 8 by $|u_n^*(t)| = \left|e^{\lambda(t_0-t)+i\omega(t_0-t)}u_0^*\right| = e^{\lambda(t_0-t)}|u_0^*|$ yields the relative error bound

$$\left|\frac{u(t)-u^*(t)}{u_n^*(t)}\right| \leq \varepsilon\frac{e^{-|\lambda|(t-t_0)}-1}{-|\lambda||u_0^*|} = \varepsilon\frac{1-e^{-|\lambda|(t-t_0)}}{|\lambda||u_0^*|} \leq \frac{\varepsilon}{|\lambda||u_0^*|} = O(\varepsilon) \quad (\lambda < 0). \quad (10)$$

If $\lambda = 0$, the integral on R.H.S. of Eq. 8 is reduced to $(t-t_0)$, and therefore the absolute error bound is

$$|u(t)-u^*(t)| \leq \varepsilon(t-t_0) = O(\varepsilon t) \quad (\lambda = 0). \quad (11)$$

Since the natural response has constant modulus $|u_n^*(t)| = \left|e^{i\omega(t_0-t)}u_0^*\right| \equiv |u_0^*|$ when $\lambda = 0$, the relative error with respect to the natural response is bounded by $O(\varepsilon t)$ as well.

A.2 Higher-Order Linear ODE with Constant Coefficients

Proposition Let the residual $Ru(t)$ of the higher-order equation $u^{(n)}+a_{n-1}u^{(n-1)}+\cdots+a_0 u = f$ be bounded by some $\varepsilon \geq 0$ on I, where $u^{(n)}$ is the n-th order derivative of u, namely,

$$\left|u^{(n)} + a_{n-1}u^{(n-1)} + \cdots + a_0 u - f\right| \leq \varepsilon \quad \forall t \in I. \tag{12}$$

Let the network solution u satisfy initial conditions $u^{(k)}(t_0) = u_0^{*(k)}$ ($k = 0, \ldots, n-1$). By the fundamental theorem of algebra, the characteristic polynomial $p_c(x)$ can be uniquely factorized as

$$p_c(x) := x^n + a_{n-1}x^{n-1} + \cdots + a_0 = \prod_{k=0}^{n-1}(x + \lambda_k + i\omega_k). \tag{13}$$

It is well-known that the exact solution has the form $u^*(t) = u_p^*(t) \sum_{k=0}^{n-1} c_k \exp(\lambda_k t + i\omega_k t)$, where u_p^* is any particular solution to the original equation and $c_0, \ldots, c_{n-1}$ are constants chosen to satisfy the initial conditions.

Let m be the total number of k in Eq. 13 such that $\lambda_k = 0$, then the absolute error is bounded by:

$$|u - u^*| \leq O(\varepsilon t^m) \ \text{ if } \ \lambda_k \geq 0 \ \text{ for all } \ k. \tag{14}$$

Proof For brevity, we prove the second-order case here to provide an intuition of the complete proof, which is presented in Appendix C.

In the second-order case, Eq. 12 can be reduced to

$$|u'' + (\lambda_1 + i\omega_1 + \lambda_2 + i\omega_2)\, u' + (\lambda_1 + i\omega_1)\,(\lambda_2 + i\omega_2)\, u - f| \leq \varepsilon \quad (\lambda_1 \geq \lambda_2), \tag{15}$$

or, equivalently,

$$\left|\Big(u' + (\lambda_1 + i\omega_1)\, u\Big)' + (\lambda_2 + i\omega_2)\,\Big(u' + (\lambda_1 + i\omega_1)\, u\Big) - f\right| \leq \varepsilon. \tag{16}$$

Let $v = u' + (\lambda_1 + i\omega_1)u$, Eq. 16 is then reduced to a first-order inequality w.r.t. v

$$|v' + (\lambda_2 + i\omega_2)\, v - f| \leq \varepsilon. \tag{17}$$

By Eq. 8,

$$|v(t) - v^*(t)| \leq \varepsilon e^{-\lambda_2 t}\int_{t_0}^{t} e^{\lambda_2 \tau}\mathrm{d}\tau, \tag{18}$$

where $v^*(t) = {u^*}'(t) + (\lambda_1 + i\omega_1)u^*(t)$. Substituting $v = u' + (\lambda_1 + i\omega_1)u$ into Eq. 18 yields

$$|u'(t) + (\lambda_1 + i\omega_1)\, u(t) - v^*(t)| \leq \varepsilon e^{-\lambda_2 t}\int_{t_0}^{t} e^{\lambda_2 \tau}\mathrm{d}\tau = \varepsilon\frac{1 - e^{\lambda_2(t_0 - t)}}{\lambda_2} \tag{19}$$

Multiplying Eq. 19 by $e^{\lambda_1 t + i\omega_1 t}$, taking the integral on $(t_0, t) \subseteq I$, and dividing by $\left|e^{\lambda_1 t + i\omega_1 t}\right|$, we have

$$|u(t) - u^*(t)| \leq \varepsilon\frac{1}{\lambda_1\lambda_2}\left(1 - \frac{\lambda_1 e^{-\lambda_2 t} - \lambda_2 e^{-\lambda_1 t}}{\lambda_1 - \lambda_2}\right) =: \varepsilon\phi(t; \lambda_1, \lambda_2) \tag{20}$$

If $\lambda_1, \lambda_2 > 0$, it can be verified that $\phi(t; \lambda_1, \lambda_2)$ is strictly increasing on I and is bounded by $\left[0, \dfrac{1}{\lambda_1\lambda_2}\right)$. Therefore

$$|u(t) - u^*(t)| \leq \frac{\varepsilon}{\lambda_1\lambda_2} = O(\varepsilon) \tag{21}$$

If $\lambda_1 > \lambda_2 = 0$, taking the limit $\lambda_2 \to 0$ in Eq. 20, there is

$$|u(t) - u^*(t)| \leq \lim_{\lambda_2 \to 0} \varepsilon\phi(t; \lambda_1, \lambda_2) = \frac{\varepsilon}{\lambda_1^2}\left(e^{-\lambda_1 t} + \lambda_1 t - 1\right) \leq \frac{\varepsilon t}{\lambda_1} = O(\varepsilon t). \tag{22}$$

If $\lambda_1 = \lambda_2 = 0$, taking the double limit $\lambda_1, \lambda_2 \to 0$ in Eq. 20, there is

$$|u(t) - u^*(t)| \leq \lim_{\lambda_1, \lambda_2 \to 0} \varepsilon \phi(t; \lambda_1, \lambda_2) = \frac{\varepsilon t^2}{2} = O(\varepsilon t^2). \tag{23}$$

A detailed derivation of Eq. 22 and Eq. 23 can be found in Appendex B.

A.3 System of First-Order Linear ODEs with Constant Coefficients

Proposition Let the p-norm of the residual $\|R\mathbf{u}(t)\|$ of the linear system $\mathbf{u}' + A\mathbf{u} = \mathbf{f}$ ($\mathbf{u}, \mathbf{f} \in \mathbb{C}^n$ and $A \in \mathbb{C}^{n\times n}$) be bounded by some $\varepsilon \geq 0$ on I, namely,

$$\|\mathbf{u}' + A\mathbf{u} - \mathbf{f}\| \leq \varepsilon \quad \forall t \in I, \tag{24}$$

and the network solution satisfy the initial condition $\mathbf{u}(t_0) = \mathbf{u}_0^*$. Denote the Jordan canonical form of A as

$$J = M^{-1}AM = \begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_m \end{pmatrix} \text{ where } J_k = \begin{pmatrix} \lambda_k + i\omega_k & 1 & & \\ & \ddots & \ddots & \\ & & \lambda_k + i\omega_k & 1 \\ & & & \lambda_k + i\omega_k \end{pmatrix} \quad k = 1, \dots, m \tag{25}$$

where M is composed of generalized eigenvectors and J_k ($1 \leq k \leq m \leq n$) is a $n_k \times n_k$ Jordan block ($n_1 + \cdots + n_m = n$). Then, the absolute error is bounded by $\|\mathbf{u} - \mathbf{u}^*\| \leq O(\varepsilon)$ if $\lambda_k > 0$ for all k.

Proof With the substitution $\mathbf{v} := M^{-1}\mathbf{u}$, $\mathbf{g} := M^{-1}\mathbf{f}$, Eq. 24 can be transformed into

$$\|\mathbf{v}' + J\mathbf{v} - \mathbf{g}\| = \left\|M^{-1}\mathbf{u}' + M^{-1}A\mathbf{u} - M^{-1}\mathbf{f}\right\| \leq \left\|M^{-1}\right\| \|\mathbf{u}' + A\mathbf{u} - \mathbf{f}\| \leq \left\|M^{-1}\right\| \varepsilon \tag{26}$$

where $\left\|M^{-1}\right\|$ is the induced p-norm of M^{-1}. Each entry in $(\mathbf{v}' + J\mathbf{v} - \mathbf{g})$ must be no greater than $\left\|M^{-1}\right\| \varepsilon$ in order for Eq. 26 to hold. To bound the error for each Jordan chain, we first define two auxiliary sequence of functions $\{h_k\}$ and $\{H_k\}$, which will be useful in following derivations.

$$h_k(t; \lambda) := \frac{1}{\lambda^k}\left(1 - \sum_{j=0}^{k-1} \frac{\lambda^j (t - t_0)^j}{j!} e^{\lambda(t_0 - t)}\right) \quad \text{and} \quad H_k(t; \lambda) := \sum_{j=1}^{k} h_k(t; \lambda). \tag{27}$$

Notice the property that, if $\lambda > 0$

$$0 \leq h_k(t; \lambda) < \frac{1}{\lambda^k} \quad 0 \leq H_k(t; \lambda) < \sum_{j=1}^{k} \frac{1}{\lambda^j} \quad \forall t \in I.$$

Now, consider the first Jordan chain,

$$|v_1' + (\lambda_1 + i\omega_1)v_1 + v_2 - g_1| \leq \left\|M^{-1}\right\| \varepsilon \tag{28}$$

$$\vdots$$

$$\left|v_{n_1-1}' + (\lambda_1 + i\omega_1)v_{n_1-1} + v_{n_1} - g_{n_1-1}\right| \leq \left\|M^{-1}\right\| \varepsilon \tag{29}$$

$$\left|v_{n_1}' + (\lambda_1 + i\omega_1)v_{n_1} - g_{n_1}\right| \leq \left\|M^{-1}\right\| \varepsilon \tag{30}$$

If $\lambda_1 > 0$, Eq. 30 implies (by section A.1) the absolute error bound on v_{n_1}

$$\left|v_{n_1} - v_{n_1}^*\right| \leq \left\|M^{-1}\right\| \varepsilon \frac{1 - e^{\lambda_1(t_0 - t)}}{\lambda_1} = H_1(t; \lambda_1) \left\|M^{-1}\right\| \varepsilon \tag{31}$$

Plugging Eq. 29 and Eq. 31 into the following triangle inequality yields

$$\begin{aligned}\left|v_{n_1-1}' + (\lambda_1 + i\omega_1)v_{n_1-1} + v_{n_1}^* - g_{n_1-1}\right| &\leq \left|v_{n_1-1}' + (\lambda_1 + i\omega_1)v_{n_1-1} + v_{n_1} - g_{n_1-1}\right| + \left|v_{n_1}^* - v_{n_1}\right| \\ &\leq \left\|M^{-1}\right\| \varepsilon + H_1(t; \lambda_1) \left\|M^{-1}\right\| \varepsilon \end{aligned} \tag{32}$$

Apply the integrating factor technique again, there is

$$\left|v_{n_1-1} - v^*_{n_1-1}\right| \leq H_2(t;\lambda) \left\|M^{-1}\right\| \varepsilon \tag{33}$$

Repeating the above procedure, there is

$$|v_1 - v_1^*| \leq H_{n_1}(t;\lambda_1) \left\|M^{-1}\right\| \varepsilon, \quad |v_2 - v_2^*| \leq H_{n_1-1}(t;\lambda_1) \left\|M^{-1}\right\| \varepsilon, \quad \ldots \quad , \left|v_{n_1} - v^*_{n_1}\right| \leq H_1(t;\lambda_1) \left\|M^{-1}\right\| \varepsilon \tag{34}$$

If $\lambda_1 = 0$, it can be proven (see Appendix D) that

$$|v_1 - v_1^*| \leq \left\|M^{-1}\right\| \varepsilon \sum_{j=1}^{n_1} \tfrac{(t-t_0)^j}{j!}, \quad |v_2 - v_2^*| \leq \left\|M^{-1}\right\| \varepsilon \sum_{j=1}^{n_1-1} \tfrac{(t-t_0)^j}{j!}, \quad \ldots \quad , \left|v_{n_1} - v^*_{n_1}\right| \leq \left\|M^{-1}\right\| \varepsilon (t-t_0) \tag{35}$$

Similarly, if $\lambda_k > 0$ for the k-th Jordan chain, then

$$\begin{aligned}
\left|v_{n_1+\cdots+n_{k-1}+1} - v^*_{n_1+\cdots+n_{k-1}+1}\right| &\leq H_{n_k}(t;\lambda) \left\|M^{-1}\right\| \varepsilon \\
\left|v_{n_1+\cdots+n_{k-1}+2} - v^*_{n_1+\cdots+n_{k-1}+2}\right| &\leq H_{n_k-1}(t;\lambda) \left\|M^{-1}\right\| \varepsilon \\
&\vdots \\
\left|v_{n_1+\cdots+n_{k-1}+n_k} - v^*_{n_1+\cdots+n_{k-1}+n_k}\right| &\leq H_1(t;\lambda) \left\|M^{-1}\right\| \varepsilon
\end{aligned}$$

It can be shown that, if $\lambda_k > 0$ for all k, then

$$\|\mathbf{v} - \mathbf{v}^*\| \leq \sqrt[p]{n} \left(\max_k \sum_{j=1}^{n_k} \frac{1}{\lambda_k^j} \right) \left\|M^{-1}\right\| \varepsilon. \tag{36}$$

Substituting $\mathbf{u} = M\mathbf{v}$ into Eq. 36, we have the absolute error bound on $\mathbf{u}$,

$$\|\mathbf{u} - \mathbf{u}^*\| = \|M\mathbf{v} - M\mathbf{v}^*\| \leq \|M\| \, \|\mathbf{v} - \mathbf{v}^*\| \leq \sqrt[p]{n} \left(\max_k \sum_{j=1}^{n_k} \frac{1}{\lambda_k^j} \right) \operatorname{cond}(M) \varepsilon = O(\varepsilon) \tag{37}$$

where $\operatorname{cond}(M) = \|M\| \left\|M^{-1}\right\|$ is the condition number of M. Note that the matrix of generalized eigenvectors, M, can be replaced with MD where $D \in \mathbb{C}^{n\times n}$ is a diagonal matrix. The infimum of condition number under right multiplication

$$\operatorname{cond}^{\mathrm{R}}(M) := \inf_{D \text{ diagonal}} \operatorname{cond}(MD) = \inf_{D \text{ diagonal}} \|MD\| \left\|D^{-1}M^{-1}\right\|$$

has been studied for induced 1-norm, 2-norm, and ∞-norm in [9], [10], and [11].

A.4 First-Order Linear ODE with Nonconstant Coefficients

Proposition Let the residual $|Ru(t)|$ of $u' + (p(t) + iq(t))\,u = f(t)$ ($p, q : I \to \mathbb{R}, f : I \to \mathbb{C}$) be bounded by some $\varepsilon \geq 0$ on I, namely,

$$|u' + (p(t) + iq(t))\,u - f(t)| \leq \varepsilon \quad \forall t \in (t_0, \infty), \tag{38}$$

and the network satisfy the initial condition $u(t_0) = u_0^*$, then the absolute error is bounded by

$$|u - u^*| \leq O(\varepsilon t) \tag{39}$$

if $p(t) \geq 0$ for sufficiently large t on I.

Proof Denote the antiderivatives of $p(t)$ and $q(t)$ as

$$P(t) = \int_{t_0}^t p(\tau)\mathrm{d}\tau \quad Q(t) = \int_{t_0}^t q(\tau)\mathrm{d}\tau.$$

Applying the integrating factor technique again, there is

$$\left|\int_{t_0}^{t} e^{P(\tau)+iQ(\tau)}\Big(u'(\tau)+\big(p(\tau)+iq(\tau)\big)u(\tau)-f(\tau)\Big)\mathrm{d}\tau\right|$$
$$\leq \int_{t_0}^{t}\left|e^{P(\tau)+iQ(\tau)}\right|\left|u'(\tau)+\big(p(\tau)+iq(\tau)\big)u(\tau)-f(\tau)\right|\mathrm{d}\tau$$

$$\left|e^{P(t)+iQ(t)}u(t)-u(t_0)-\int_{t_0}^{t} e^{P(\tau)+iQ(\tau)}f(\tau)\mathrm{d}\tau\right| \leq \varepsilon\int_{t_0}^{t} e^{P(\tau)}\mathrm{d}\tau$$
$$\left|u(t)-e^{-P(t)-iQ(t)}u_0^*-e^{-P(t)-iQ(t)}\int_{t_0}^{t} e^{P(\tau)+iQ(\tau)}f(\tau)\mathrm{d}\tau\right| \leq \varepsilon e^{-P(t)}\int_{t_0}^{t} e^{P(\tau)}\mathrm{d}\tau$$
$$|u(t)-u^*(t)| \leq \varepsilon e^{-P(t)}\int_{t_0}^{t} e^{P(\tau)}\mathrm{d}\tau. \quad (40)$$

Rewriting the R.H.S. of of Eq. 40, there is

$$|u(t)-u^*(t)| \leq \varepsilon t\left(1+\frac{\phi(t)}{te^{P(t)}}\right), \quad (41)$$

where

$$\phi(t)=\int_{t_0}^{t} e^{P(\tau)}\mathrm{d}\tau - te^{P(t)} = \int_{t_0}^{t}\left(e^{P(\tau)}-e^{P(t)}\right)\mathrm{d}\tau. \quad (42)$$

Let $p(t)\geq 0$ for $t>t'$. Subsequently, $P(t)$ is nondecreasing for $t>t'$. Therefore,

$$\phi(t)=\int_{t_0}^{t'}\left(e^{P(\tau)}-e^{P(t)}\right)\mathrm{d}\tau+\int_{t'}^{t}\left(e^{P(\tau)}-e^{P(t)}\right)\mathrm{d}\tau \leq \int_{t_0}^{t'}\left(e^{P(\tau)}-e^{P(t)}\right)\mathrm{d}\tau=\phi(t') \quad t>t'. \quad (43)$$

Consequently,

$$\frac{\phi(t)}{te^{P(t)}} \leq \max_{\tau\in[t_0,t']}\left[\frac{\phi(\tau)}{\tau e^{P(\tau)}}\right] =: M, \quad (44)$$

and finally,

$$|u(t)-u^*(t)| \leq \varepsilon t\,(1+M) = O(\varepsilon t). \quad (45)$$

A.5 Dividing the Intervals for a Tightened Error Bound

In Sections A.1 to A.4, we only consider the global maximum residual norm ε on I. However, one can also partition I into subintervals $I = I_1 \cup I_2 \cup \dots$ and consider the local maximum residual norm ε_k on I_k. This leads to an even tighter error bound since $\varepsilon_k \leq \varepsilon$ for all k.

For instance, in the case for first-order linear ODE with constant coefficients, the bound in Eq. 9 becomes

$$|u-u^*| \leq e^{-\lambda t}\int_{t_0}^{t} e^{\lambda\tau}\left|Ru(\tau)\right|\mathrm{d}\tau \quad (46)$$

as $\max\limits_{k}\rho(I_k)\to 0$, where $\rho(I_k)$ is the diameter of interval I_k.

B Derivation of Eq. 22 and Eq. 23

Consider the case when $\lambda_1 > 0$, $\lambda_2 \to 0$, we have the following limit

$$
\begin{aligned}
|u-u^*| &\le \lim_{\lambda_2\to 0} \frac{\varepsilon}{\lambda_1\lambda_2}\left(1-\frac{\lambda_1 e^{-\lambda_2}-\lambda_2 e^{-\lambda_1}}{\lambda_1-\lambda_2}\right)\\
&= \lim_{\lambda_2\to 0} \frac{\varepsilon}{\lambda_1\lambda_2}\frac{\lambda_1-\lambda_2-\left(\lambda_1 e^{-\lambda_2}-\lambda_2 e^{-\lambda_1}\right)}{\lambda_1-\lambda_2}\\
&= \lim_{\lambda_2\to 0} \frac{\varepsilon}{\lambda_1\lambda_2}\frac{\lambda_1\left(1-e^{-\lambda_2 t}\right)-\lambda_2\left(1-e^{-\lambda_1 t}\right)}{\lambda_1-\lambda_2}\\
&= \lim_{\lambda_2\to 0} \frac{\varepsilon}{\lambda_1\left(\lambda_1-\lambda_2\right)}\frac{\lambda_1\left(1-e^{-\lambda_2 t}\right)-\lambda_2\left(1-e^{-\lambda_1 t}\right)}{\lambda_2}\\
&= \lim_{\lambda_2\to 0} \frac{\varepsilon}{\lambda_1\left(\lambda_1-\lambda_2\right)}\left(\frac{\lambda_1\left(1-e^{-\lambda_2 t}\right)}{\lambda_2}-\left(1-e^{-\lambda_1 t}\right)\right)\\
&= \frac{\varepsilon}{\lambda_1^2}\lim_{\lambda_2\to 0}\left(\frac{\lambda_1\left(1-e^{-\lambda_2 t}\right)}{\lambda_2}-\left(1-e^{-\lambda_1 t}\right)\right)\\
&= \frac{\varepsilon}{\lambda_1^2}\left(\lambda_1 t-1+e^{-\lambda_1 t}\right) \qquad (47)\\
&\le \frac{\varepsilon}{\lambda_1^2}\left(\lambda_1 t\right)=\frac{\varepsilon t}{\lambda_1}
\end{aligned}
$$

If we take the limit $\lambda_1 \to 0$ on top of $\lambda_2 \to 0$, step 47 can be simplified using Taylor expansion,

$$
\begin{aligned}
|u-u^*| &\le \lim_{\lambda_1\to 0} \frac{\varepsilon}{\lambda_1^2}\left(\lambda_1 t-1+e^{-\lambda_1 t}\right)\\
&= \lim_{\lambda_1\to 0} \frac{\varepsilon}{\lambda_1^2}\left(\lambda_1 t-1+1-\lambda_1 t+\frac{1}{2}\lambda_1^2 t^2+t^3 O\left(\lambda_1^3\right)\right)\\
&= \frac{\varepsilon t^2}{2}
\end{aligned}
$$

C General Case Proof of Section A.2

Define the following sequence of auxiliary functions $\{\phi_n\}_{n=1}^{\infty}$ on I,

$$
\phi_n(t;\lambda_{1:n}) = \frac{1}{\prod_{j=1}^{n}\lambda_j} - \sum_{k=1}^{n}\frac{e^{-\lambda_k(t-t_0)}}{\lambda_k\prod_{j=1,j\neq k}^{n}(\lambda_j-\lambda_k)},
$$

where $\lambda_{1:n}$ is a tuple $(\lambda_1,\lambda_2,\ldots,\lambda_n)$. Note that with $\phi_0(t)=1$, it can be demonstrated that $\{\phi_n\}_{n=1}^{\infty}$ satisfies the recurrence relation

$$
\phi_{n+1}(t;\lambda_{1:n+1}) = e^{-\lambda_{n+1}t}\int_{t_0}^{t} e^{\lambda_{n+1}\tau}\phi_n(\tau;\lambda_{1:n+1})\mathrm{d}\tau \quad \text{for } n\ge 0. \qquad (48)
$$

It can also be proven that $\phi_n(t;\lambda_{1:n})$ is monotonically increasing on I if $\lambda_1,\ldots,\lambda_n \ge 0$ because

$$
\frac{\mathrm{d}}{\mathrm{d}t}\phi_n(t,\lambda_{1:n}) = \sum_{k=1}^{n}\frac{e^{-\lambda_k(t-t_0)}}{\prod_{j=1,j\neq k}^{n}(\lambda_j-\lambda_k)} \ge 0. \qquad (49)
$$

Also, if $\lambda_1,\ldots,\lambda_n > 0$, there is $\lim_{t\to\infty}\phi_n(t,\lambda_{1:n}) = \prod_{j=1}^{n}\lambda_j^{-1}$.

Let $u_0(t) := u(t)$, $u_1(t) := u_0'(t) + (\lambda_n + i\omega_n)u_0(t)$, $u_2(t) := u_1'(t) + (\lambda_{n-1} + i\omega_{n-1})u_1(t), \dots,$ $u_{n-1}(t) = u_{n-2}'(t) + (\lambda_2 + i\omega_2)u_{n-2}(t)$, Eq. 12 can be written as

$$\left|u_{n-1}' + (\lambda_1 + i\omega_1)u_{n-1} - f\right| \leq \varepsilon, \tag{50}$$

which is a first-order inequality in terms of u_{n-1} as discussed in Section A.1. By Eq. 8,

$$\left|u_{n-1} - u_{n-1}^*\right| \leq \varepsilon e^{-\lambda_1 t}\int_{t_0}^{t} e^{\lambda_1 \tau}\mathrm{d}\tau = \varepsilon\phi_1(t;\lambda_1). \tag{51}$$

Substitute $u_{n-1}(t) = u_{n-2}'(t) + (\lambda_2 + i\omega_2)u_{n-2}(t)$ back into Eq. 51, we have

$$\left|u_{n-2} + (\lambda_2 + i\omega_2) - u_{n-1}^*\right| \leq \varepsilon\phi_1(t;\lambda_1),$$

which is a first order inequality in terms of u_{n-2}. Applying the integrating factor trick again, we have

$$\left|u_{n-2} - u_{n-2}^*\right| \leq \varepsilon e^{-\lambda_2 t}\int_{t_0}^{t} e^{\lambda_2 t}\phi_1(\tau,\lambda_1)\mathrm{d}\tau = \varepsilon\phi_2(t;\lambda_{1:2}). \tag{52}$$

Repeating the above process yields

$$|u - u^*| = |u_0 - u_0^*| \leq \varepsilon\phi_n(t;\lambda_{1:n}) \tag{53}$$

D Proof of Equation 35

Take the limit $\lambda \to 0$ in Eq. 27, and applying Taylor expansions where necessary, we have

$$\begin{aligned}
h_k(t;0) &= \lim_{\lambda\to 0}\frac{1}{\lambda^k}\left(1 - \sum_{j=0}^{k-1}\frac{\lambda^j(t-t_0)^j}{j!}e^{\lambda(t_0-t)}\right) \\
&= \lim_{\lambda\to 0}\frac{e^{\lambda(t_0-t)}}{\lambda^k}\left(e^{\lambda(t-t_0)} - \sum_{j=0}^{k-1}\frac{\lambda^j(t-t_0)^j}{j!}\right) \\
&= \lim_{\lambda\to 0}\frac{e^{\lambda(t_0-t)}}{\lambda^k}\sum_{j=k}^{\infty}\frac{\lambda^j(t-t_0)^j}{j!} \\
&= \lim_{\lambda\to 0}\frac{1}{\lambda^k}\left(\sum_{l=0}^{\infty}\frac{\lambda^l(t_0-t)^l}{l!}\right)\left(\sum_{j=k}^{\infty}\frac{\lambda^j(t-t_0)^j}{j!}\right)
\end{aligned}$$

Notice the lowest order term w.r.t. λ in $\left(\sum_{l=0}^{\infty}\frac{\lambda^l(t_0-t)^l}{l!}\right)\left(\sum_{j=k}^{\infty}\frac{\lambda^j(t-t_0)^j}{j!}\right)$ is λ_k, which is attained only when $l = 0$ and $j = k$. The coefficient for the λ^k term is given by

$$\frac{(t_0-t)^0}{0!}\cdot\frac{(t-t_0)^k}{k!} = \frac{(t-t_0)^k}{k!}.$$

Consequently,

$$\begin{aligned}
h_k(t;0) &= \lim_{\lambda\to 0}\frac{1}{\lambda^k}\left(\frac{(t-t_0)^k}{k!}\lambda^k + O\left(\lambda^{k+1}\right)\right) = \frac{(t-t_0)^k}{k!} \\
H_k(t;0) &= \sum_{j=1}^{k} h_k(t;0) = \sum_{j=1}^{k}\frac{(t-t_0)^j}{j!}
\end{aligned}$$

Eq. 35 is attained by plugging the above equality into Eq. 34.

E Examples of dividing domain into subintervals

In Section A.5, we show that the error bound on $I = [0, t]$ can be further tightened by evaluating the maximum absolute residuals on a sequence of subintervals $I_i = [t_{i-1}, t_i]$. We apply this technique for the experiments in Section 4.

E.1 Second Order Linear Equation with Constant Coefficients

Consider a second-order linear equation with constant coefficients (assuming $\lambda_1, \lambda_2 \geq 0$,

$$u''(t) + (\lambda_1 + i\omega_1 + \lambda_2 + i\omega_2)u'(t) + (\lambda_1 + i\omega_1)(\lambda_2 + i\omega_2)u(t) = f(t) \tag{54}$$

An approximated solution yielded by a neural network does not exactly satisfy the Eq. 54, but instead incurs what we call a residual term $r(t)$

$$u''(t) + (\lambda_1 + i\omega_1 + \lambda_2 + i\omega_2)u'(t) + (\lambda_1 + i\omega_1)(\lambda_2 + i\omega_2)u(t) = f(t) + r(t) \tag{55}$$

Solutions of Eq. 55 and 54 differ by

$$\begin{aligned}\Delta(t) &= e^{-\lambda_1 t}\int_{s=0}^{s=t} e^{\lambda_1 s}e^{-\lambda_2 s}\left(\int_{\tau=0}^{\tau=s} e^{\lambda_2\tau}r(\tau)\mathrm{d}\tau\right)\mathrm{d}s \\ &= e^{-\lambda_1 t}\int_{s=0}^{s=t} e^{(\lambda_1-\lambda_2)s}\left(\int_{\tau=0}^{\tau=s} e^{\lambda_2\tau}r(\tau)\mathrm{d}\tau\right)\mathrm{d}s \\ &= e^{-\lambda_1 t}\int_{s=0}^{s=t}\int_{\tau=0}^{\tau=s} e^{(\lambda_1-\lambda_2)s}e^{\lambda_2\tau}r(\tau)\mathrm{d}\tau\mathrm{d}s\end{aligned}$$

Therefore

$$\begin{aligned}|\Delta(t)| &\leq e^{-\lambda_1 t}\int_{s=0}^{s=t}\int_{\tau=0}^{\tau=s} e^{(\lambda_1-\lambda_2)s}e^{\lambda_2\tau}|r(\tau)|\mathrm{d}\tau\mathrm{d}s \\ &= e^{-\lambda_1 t}\int_{\tau=0}^{\tau=t}\int_{s=\tau}^{s=t} e^{(\lambda_1-\lambda_2)s}e^{\lambda_2\tau}|r(\tau)|\mathrm{d}s\mathrm{d}\tau \\ &= e^{-\lambda_1 t}\int_{\tau=0}^{\tau=t} e^{\lambda_2\tau}|r(\tau)|\left(\int_{s=\tau}^{s=t} e^{(\lambda_1-\lambda_2)s}\mathrm{d}s\right)\mathrm{d}\tau \\ &= e^{-\lambda_1 t}\int_{\tau=0}^{\tau=t} e^{\lambda_2\tau}|r(\tau)|\frac{e^{(\lambda_1-\lambda_2)t} - e^{(\lambda_1-\lambda_2)\tau}}{\lambda_1-\lambda_2}\mathrm{d}\tau \\ &= \int_{\tau=0}^{\tau=t}|r(\tau)|\frac{e^{\lambda_2(\tau-t)} - e^{\lambda_1(\tau-t)}}{\lambda_1-\lambda_2}\mathrm{d}\tau\end{aligned}$$

Notice that $\dfrac{e^{\lambda_2(\tau-t)} - e^{\lambda_1(\tau-t)}}{\lambda_1-\lambda_2} \geq 0$ for $\tau < t$. Let $M(a,b) = \max_{a\leq\tau\leq b}|r(\tau)|$, we have

$$\begin{aligned}|\Delta(t)| &\leq \sum_{i=1}^{n} M(t_{i-1}, t_i)\int_{\tau=t_{i-1}}^{\tau=t_i}\frac{e^{\lambda_2(\tau-t)} - e^{\lambda_1(\tau-t)}}{\lambda_1-\lambda_2}\mathrm{d}\tau \qquad (56) \\ &\leq M_{(0,t)}\int_{\tau=0}^{\tau=t}\frac{e^{\lambda_2(\tau-t)} - e^{\lambda_1(\tau-t)}}{\lambda_1-\lambda_2}\mathrm{d}\tau. \qquad (57)\end{aligned}$$

where $0 = t_0 < t_1 < \cdots < t_n = t$.

Eq. 56 sheds light on how to evaluate the error bound by subdividing interval $[0, t]$ into n subintervals. Namely, we first evaluate the maximum absolute residual $M(t_{i-1}, t_i)$ on $[t_{i-1}, t_i]$ as well as the integral $\int_{\tau=0}^{\tau=t}\dfrac{e^{\lambda_2(\tau-t)} - e^{\lambda_1(\tau-t)}}{\lambda_1-\lambda_2}\mathrm{d}\tau$, which always has a closed-form expression depending λ_1 and λ_2. The absolute error at any t is then bounded by the sum of the products. In particular, Eq. 57 is the special case where we do not divide $[0, t]$ into subintervals ($n = 1$), which is discussed in Section A.2.

In the special case where $\max_{1\le i\le n}(t_i - t_{i-1}) \to 0$, there is

$$\int_{\tau=0}^{\tau=t} |r(\tau)| \frac{e^{\lambda_2(\tau-t)} - e^{\lambda_1(\tau-t)}}{\lambda_1 - \lambda_2} \mathrm{d}\tau = \sum_{i=1}^{n} M(t_{i-1}, t_i) \int_{\tau=t_{i-1}}^{\tau=t_i} \frac{e^{\lambda_2(\tau-t)} - e^{\lambda_1(\tau-t)}}{\lambda_1 - \lambda_2} \mathrm{d}\tau. \tag{58}$$

E.2 System of ODEs

Consider a Jordan chain of length 3 and eigenvalue $(\lambda + i\omega)$.

$$\begin{aligned} u_1'(t) + (\lambda + i\omega)u_1(t) + u_2(t) &= f_1(t) \\ u_2'(t) + (\lambda + i\omega)u_2(t) + u_3(t) &= f_2(t) \\ u_3'(t) + (\lambda + i\omega)u_3(t) &= f_3(t) \end{aligned}$$

The approximated solution given by the neural network incurs residuals $r_1(t), r_2(t), r_3(t)$, namely,

$$u_1'(t) + (\lambda + i\omega)u_1(t) + u_2(t) = f_1(t) + r_1(t) \tag{59}$$
$$u_2'(t) + (\lambda + i\omega)u_2(t) + u_3(t) = f_2(t) + r_2(t) \tag{60}$$
$$u_3'(t) + (\lambda + i\omega)u_3(t) = f_3(t) + r_3(t) \tag{61}$$

Eq. 61 implies that

$$|u_3 - u_3^*| \le e^{-\lambda t} \int_{\tau=0}^{\tau=t} e^{\lambda\tau} |r_3(\tau)| \mathrm{d}\tau$$

By triangle inequality, Eq. 60 becomes

$$\begin{aligned} |u_2' + \lambda u_2 + u_3^*| &\le |u_2' + \lambda u_2 + u_3| + |u_3 - u_3^*| \le |r_2(t)| + e^{-\lambda t} \int_{\tau=0}^{\tau=t} e^{\lambda t} |r_3(\tau)| \mathrm{d}\tau \\ |u_2 - u_2^*| &\le e^{-\lambda t} \int_{\tau=0}^{\tau=t} e^{\lambda\tau} |r_2(\tau)| \mathrm{d}\tau + e^{-\lambda t} \int_{s=0}^{s=t} \left(\int_{\tau=0}^{\tau=s} e^{\lambda\tau} |r_3(\tau)| \mathrm{d}\tau \right) \mathrm{d}s \\ &= e^{-\lambda t} \int_{\tau=0}^{\tau=t} e^{\lambda\tau} |r_2(\tau)| \mathrm{d}\tau + e^{-\lambda t} \int_{\tau=0}^{\tau=t} \left(\int_{s=\tau}^{s=t} e^{\lambda\tau} |r_3(\tau)| \mathrm{d}s \right) \mathrm{d}\tau \\ &= e^{-\lambda t} \int_{\tau=0}^{\tau=t} e^{\lambda\tau} |r_2(\tau)| \mathrm{d}\tau + e^{-\lambda t} \int_{\tau=0}^{\tau=t} (t-\tau) e^{\lambda\tau} |r_3(\tau)| \mathrm{d}\tau \end{aligned}$$

Apply the same procedure for Eq. 59, there is

$$\begin{aligned} &|u_1' + \lambda u_1 + u_2^*| \le |u_1' + \lambda u_1 + u_2| + |u_2 - u_2^*| \\ &\qquad \le |r_1(t)| + e^{-\lambda t} \int_{\tau=0}^{\tau=t} e^{\lambda\tau} |r_2(\tau)| \mathrm{d}\tau + e^{-\lambda t} \int_{\tau=0}^{\tau=t} (t-\tau) e^{\lambda\tau} |r_3(\tau)| \mathrm{d}\tau \\ &|u_1 - u_1^*| \le e^{-\lambda t} \int_{\tau=0}^{\tau=t} e^{\lambda\tau} |r_1(\tau)| \mathrm{d}\tau + e^{-\lambda t} \int_{s=0}^{s=\tau} \left(\int_{\tau=0}^{\tau=s} e^{\lambda\tau} |r_2(\tau)| \mathrm{d}\tau \right) \mathrm{d}s \\ &\qquad + e^{-\lambda t} \int_{s=0}^{s=\tau} \left(\int_{\tau=0}^{\tau=s} (s-\tau) e^{\lambda\tau} |r_3(\tau)| \mathrm{d}\tau \right) \mathrm{d}s \\ &\qquad = e^{-\lambda t} \int_{\tau=0}^{\tau=t} e^{\lambda\tau} |r_1(\tau)| \mathrm{d}\tau + e^{-\lambda t} \int_{\tau=0}^{\tau=t} \left(\int_{s=\tau}^{s=t} e^{\lambda\tau} |r_2(\tau)| \mathrm{d}s \right) \mathrm{d}\tau \\ &\qquad + e^{-\lambda t} \int_{\tau=0}^{\tau=t} \left(\int_{s=\tau}^{s=t} (s-\tau) e^{\lambda\tau} |r_3(\tau)| \mathrm{d}s \right) \mathrm{d}\tau \\ &\qquad = e^{-\lambda t} \int_{\tau=0}^{\tau=t} e^{\lambda\tau} |r_1(\tau)| \mathrm{d}\tau + e^{-\lambda t} \int_{\tau=0}^{\tau=t} (t-\tau) e^{\lambda\tau} |r_2(\tau)| \mathrm{d}\tau + e^{-\lambda t} \int_{\tau=0}^{\tau=t} \frac{(t-\tau)^2}{2} e^{\lambda\tau} |r_3(\tau)| \mathrm{d}\tau \\ &\qquad = \int_{\tau=0}^{\tau=t} e^{\lambda(\tau-t)} |r_1(\tau)| \mathrm{d}\tau + \int_{\tau=0}^{\tau=t} (t-\tau) e^{\lambda(\tau-t)} |r_2(\tau)| \mathrm{d}\tau + \int_{\tau=0}^{\tau=t} \frac{(t-\tau)^2}{2} e^{\lambda(\tau-t)} |r_3(\tau)| \mathrm{d}\tau \end{aligned}$$

Note that, with $M_k(a,b) = \max\limits_{a\le\tau\le b} |r_k(\tau)|$ ($k = 1, 2, 3$) and $0 = t_0 \le t_1 \le \cdots \le t_n = t$,

$$0 \le \int_0^t \frac{(t-\tau)^k}{k!} e^{\lambda(\tau-t)} |r(\tau)| \mathrm{d}\tau \le \sum_{i=1}^{n} M_k(t_{i-1}, t_i) \int_{t_{i-1}}^{t_i} \frac{(t-\tau)^k}{k!} e^{\lambda(\tau-t)} \mathrm{d}\tau \tag{62}$$

$$\le M_k(0,t) \int_0^t \frac{(t-\tau)^k}{k!} e^{\lambda(\tau-t)} \mathrm{d}\tau \tag{63}$$

Again, Eq. 62 shows one can evaluate the absolute error bound by dividing n subintervals. For each interval, one evaluates the maximum residual as well as the integral (which has a closed-form expression). Eq. 63 is the special case as discussed in Section A.3, where subintervals are not used ($n = 1$).