Supplementary Material for:
Neural Oscillators are Universal

A Another universality result for neural oscillators

The universal approximation Theorem 3.1 immediately implies another universal approximation results for neural oscillators, as explained next. We consider a continuous map $F:\mathbb{R}^p \to \mathbb{R}^q$; our goal is to show that F can be approximated to given accuracy ϵ by suitably defined neural oscillators. Fix a time interval $[0,T]$ for (an arbitrary choice) $T=2$. Let $K_0 \subset \mathbb{R}^p$ be a compact set. Given $\xi \in \mathbb{R}^p$, we associate with it a function $u_\xi(t) \in C_0([0,T];\mathbb{R}^p)$, by setting

$$u_\xi(t) := t\xi. \tag{A.1}$$

Clearly, the set $K := \{u_\xi \,|\, \xi \in K_0\}$ is compact in $C_0([0,T];\mathbb{R}^p)$. Furthermore, we can define an operator $\Phi : C_0([0,T];\mathbb{R}^p) \to C_0([0,T];\mathbb{R}^q)$, by

$$\Phi(u)(t) := \begin{cases} 0, & t \in [0,1), \\ (t-1)F(u(1)), & t \in [1,T]. \end{cases} \tag{A.2}$$

where $F:\mathbb{R}^p \to \mathbb{R}^q$ is the given continuous function that we wish to approximate. One readily checks that Φ defines a causal and continuous operator. Note, in particular, that

$$\Phi(u_\xi)(T) = (T-1)F(u_\xi(1)) = F(\xi),$$

is just the evaluation of F at ξ, for any $\xi \in K_0$.

Since neural oscillators can uniformly approximate the operator Φ for inputs $u_\xi \in K$, then as a consequence of Theorem 3.1 and (2.3), it follows that, for any $\epsilon > 0$ there exists $m \in \mathbb{N}$, matrices $W \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{m\times p}$ and $A \in \mathbb{R}^{q\times m}$, and bias vectors $b \in \mathbb{R}^m$, $c \in \mathbb{R}^q$, such that for any $\xi \in K_0$, the neural oscillator system,

$$\begin{cases} \ddot{y}_\xi(t) = \sigma\left(Wy_\xi(t) + tV\xi + b\right), & \text{(A.3)} \\ y_\xi(0) = \dot{y}_\xi(0) = 0, & \text{(A.4)} \\ z_\xi(t) = Ay_\xi(t) + c, & \text{(A.5)} \end{cases}$$

satisfies

$$|z_\xi(T) - F(\xi)| = |z_\xi(T) - \Phi(u_\xi)(T)| \le \sup_{t\in[0,T]} |z_\xi(t) - \Phi(u_\xi)(t)| \le \epsilon,$$

uniformly for all $\xi \in K_0$. Hence, neural oscillators can be used to approximate an arbitrary continuous function $F:\mathbb{R}^p \to \mathbb{R}^q$, uniformly over compact sets. Thus, neural oscillators also provide universal function approximation.

B Proof of Theorem 3.1

B.1 Proof of Lemma 3.4

Proof. We can rewrite $y(t) = \frac{1}{\omega}\int_0^t u(\tau)\sin(\omega(t-\tau))\,d\tau$. By direct differentiation, one readily verifies that $y(t)$ so defined, satisfies

$$\dot{y}(t) = \int_0^t u(\tau)\cos(\omega(t-\tau))\,d\tau + [u(\tau)\sin(\omega(t-\tau))]_{\tau=t} = \int_0^t u(\tau)\cos(\omega(t-\tau))\,d\tau,$$

in account of the fact that $\sin(0) = 0$. Differentiating once more, we find that

$$\begin{aligned} \ddot{y}(t) &= -\omega\int_0^t u(\tau)\sin(\omega(t-\tau))\,d\tau + [u(\tau)\cos(\omega(t-\tau))]_{\tau=t} \\ &= -\omega^2 y(t) + u(t). \end{aligned}$$

Thus $y(t)$ solves the ODE (2.6), with initial condition $y(0) = \dot{y}(0) = 0$. □

B.2 Proof of Fundamental Lemma 3.5

Reconstruction of a continuous signal from its sine transform. Let $[0, T] \subset \mathbb{R}$ be an interval. We recall that we define the windowed sine transform $\mathcal{L}_t u(\omega)$ of a function $u : [0, T] \to \mathbb{R}^p$, by

$$\mathcal{L}_t u(\omega) = \int_0^t u(t - \tau) \sin(\omega\tau)\, d\tau, \quad \omega \in \mathbb{R}.$$

In the following, we fix a compact set $K \subset C_0([0, T]; \mathbb{R}^p)$. Note that for any $u \in K$, we have $u(0) = 0$, and hence K can be identified with a subset of $C((-\infty, T]; \mathbb{R}^p)$, consisting of functions with $\mathrm{supp}(u) \subset [0, T]$. We consider the reconstruction of continuous functions $u \in K$. We will show that u can be approximately reconstructed from knowledge of $\mathcal{L}_t(\omega)$. More precisely, we provide a detailed proof of the following result:

***Lemma* B.1.** Let $K \subset C((-\infty, T]; \mathbb{R}^p)$ be compact, such that $\mathrm{supp}(u) \subset [0, T]$ for all $u \in K$. For any $\epsilon, \Delta t > 0$, there exists $N \in \mathbb{N}$, frequencies $\omega_1, \dots, \omega_N \in \mathbb{R} \setminus \{0\}$, phase-shifts $\vartheta_1, \dots, \vartheta_N \in \mathbb{R}$ and weights $\alpha_1, \dots, \alpha_N \in \mathbb{R}$, such that

$$\sup_{\tau \in [0, \Delta t]} \left| u(t - \tau) - \sum_{j=1}^N \alpha_j \mathcal{L}_t u(\omega_j) \sin(\omega_j \tau - \vartheta_j) \right| \leq \epsilon,$$

for all $u \in K$ and for all $t \in [0, T]$.

Proof. **Step 0: (Equicontinuity)** We recall the following fact from topology. If $K \subset C((-\infty, T]; \mathbb{R}^p)$ is compact, then it is equicontinuous; i.e. there exists a continuous modulus of continuity $\phi : [0, \infty) \to [0, \infty)$ with $\phi(r) \to 0$ as $r \to 0$, such that

$$|u(t - \tau) - u(t)| \leq \phi(\tau), \quad \forall\, \tau \geq 0,\ t \in [0, T],\ \forall\, u \in K. \tag{B.1}$$

Step 1: (Connection to Fourier transform) Fix $t_0 \in [0, T]$ and $u \in K$ for the moment. Define $f(\tau) = u(t_0 - \tau)$. Note that $f \in C([0, \infty); \mathbb{R}^p)$, and f has compact support $\mathrm{supp}(f) \subset [0, T]$. We also note that, by (B.1), we have

$$|f(t + \tau) - f(t)| \leq \phi(\tau), \quad \forall\, \tau \geq 0,\ t \in [0, T].$$

We now consider the following odd extension of f to all of $\mathbb{R}$:

$$F(\tau) := \begin{cases} f(\tau), & \text{for } \tau \geq 0, \\ -f(-\tau), & \text{for } \tau \leq 0. \end{cases}$$

Since F is odd, the Fourier transform of F is given by

$$\widehat{F}(\omega) := \int_{-\infty}^{\infty} F(\tau) e^{-i\omega\tau}\, d\tau = i \int_{-\infty}^{\infty} F(\tau) \sin(\omega\tau)\, d\tau = 2i \int_0^T f(\tau) \sin(\omega\tau)\, d\tau = 2i \mathcal{L}_{t_0} u(\omega).$$

Let $\epsilon > 0$ be arbitrary. Our goal is to uniformly approximate $F(\tau)$ on the interval $[0, \Delta t]$. The main complication here is that F lacks regularity (is discontinuous), and hence the inverse Fourier transform of $\widehat{F}$ does not converge to F uniformly over this interval; instead, a more careful reconstruction based on mollification of F is needed. We provide the details below.

Step 2: (Mollification) We now fix a smooth, non-negative and compactly supported function $\rho : \mathbb{R} \to \mathbb{R}$, such that $\mathrm{supp}(\rho) \subset [0, 1]$, $\rho \geq 0$, $\int_{\mathbb{R}} \rho(t)\, dt = 1$, and we define a mollifier $\rho_\epsilon(t) := \frac{1}{\epsilon}\rho(t/\epsilon)$. In the following, we will assume throughout that $\epsilon \leq T$. We point out that $\mathrm{supp}(\rho_\epsilon) \subset [0, \epsilon]$, and hence, the mollification $F_\epsilon(t) = (F * \rho_\epsilon)(t)$ satisfies, for $t \geq 0$:

$$\begin{aligned} |F(t) - F_\epsilon(t)| &= \left| \int_0^\epsilon (F(t) - F(t + \tau)) \rho_\epsilon(\tau)\, d\tau \right| = \left| \int_0^\epsilon (f(t) - f(t + \tau)) \rho_\epsilon(\tau)\, d\tau \right| \\ &\leq \left\{ \sup_{\tau \in [0, \epsilon]} |f(t) - f(t + \tau)| \right\} \int_0^\epsilon \rho_\epsilon(\tau)\, d\tau \leq \phi(\epsilon). \end{aligned}$$

In particular, this shows that

$$\sup_{t\in[0,T]} |F(t) - F_\epsilon(t)| \le \phi(\epsilon),$$

can be made arbitrarily small, with an error that depends only on the modulus of continuity ϕ.

Step 3: (Fourier inverse) Let $\widehat{F}_\epsilon(\omega)$ denote the Fourier transform of F_ϵ. Since F_ϵ is smooth and compactly supported, it is well-known that we have the identity

$$F_\epsilon(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{F}_\epsilon(\omega) e^{-i\omega\tau}\, d\omega, \qquad \forall\, t \in \mathbb{R},$$

where $\omega \mapsto \widehat{F}_\epsilon(\omega)$ decays to zero very quickly (almost exponentially) as $|\omega| \to \infty$. In fact, since $F_\epsilon = F * \rho_\epsilon$ is a convolution, we have $\widehat{F}_\epsilon(\omega) = \widehat{F}(\omega)\widehat{\rho}_\epsilon(\omega)$, where $|\widehat{F}(\omega)| \le 2\|f\|_{L^\infty} T$ is uniformly bounded, and $\widehat{\rho}_\epsilon(\omega)$ decays quickly. In particular, this implies that there exists a $L = L(\epsilon, T) > 0$ *independent of* f, such that

$$\left| F_\epsilon(\tau) - \frac{1}{2\pi}\int_{-L}^{L} \widehat{F}(\omega)\widehat{\rho}_\epsilon(\omega) e^{-i\omega\tau}\, d\omega \right| \le 2T\|f\|_{L^\infty} \int_{|\omega|>L} |\widehat{\rho}_\epsilon(\omega)|\, d\omega \le \|f\|_{L^\infty}\epsilon, \qquad \forall\, \tau \in \mathbb{R}. \tag{B.2}$$

Step 4: (Quadrature) Next, we observe that, since F and ρ_ϵ are compactly supported, their Fourier transform $\omega \mapsto \widehat{F}(\omega)\widehat{\rho}_\epsilon(\omega)e^{-i\omega\tau}$ is smooth; in fact, for $|\tau| \le T$, the Lipschitz constant of this mapping can be explicitly estimated by noting that

$$\begin{aligned}\frac{\partial}{\partial\omega}\left[\widehat{F}(\omega)\widehat{\rho}_\epsilon(\omega)e^{-i\omega\tau}\right] &= \frac{\partial}{\partial\omega}\int_{\mathrm{supp}(F_\epsilon)} (F*\rho_\epsilon)(t) e^{i\omega(t-\tau)}\, dt \\ &= \int_{\mathrm{supp}(F_\epsilon)} i(t-\tau)(F*\rho_\epsilon)(t) e^{i\omega(t-\tau)}\, dt.\end{aligned}$$

We next take absolute values, and note that any t in the support of F_ϵ obeys the bound $|t| \le T + \epsilon \le 2T$, while $|\tau| \le T$ by assumption; it follows that

$$\mathrm{Lip}\left(\omega \mapsto \widehat{F}(\omega)\widehat{\rho}_\epsilon(\omega)e^{-i\omega\tau}\right) \le (2T+T)\|F\|_{L^\infty}\,\|\rho_\epsilon\|_{L^1} = 3T\|F\|_{L^\infty}, \quad \forall\, \tau \in [0,T].$$

It thus follows from basic results on quadrature that for an equidistant choice of frequencies $\omega_1 < \cdots < \omega_N$, with spacing $\Delta\omega = 2L/(N-1)$, we have

$$\left| \frac{1}{2\pi}\int_{-L}^{L} \widehat{F}(\omega)\widehat{\rho}_\epsilon(\omega)e^{-i\omega\tau}\, d\omega - \frac{\Delta\omega}{2\pi}\sum_{j=1}^{N} \widehat{F}(\omega_j)\widehat{\rho}_\epsilon(\omega_j)e^{-i\omega_j\tau} \right| \le \frac{CL^2\, 3T\|F\|_{L^\infty}}{N}, \quad \forall\, \tau \in [0,T],$$

for an absolute constant $C > 0$, independent of F, T and N. By choosing N to be even, we can ensure that $\omega_j \ne 0$ for all j. In particular, recalling that $L = L(T,\epsilon)$ depends only on ϵ and T, and choosing $N = N(T,\epsilon)$ sufficiently large, we can combine the above estimate with (B.2) to ensure that

$$\left| F_\epsilon(\tau) - \frac{\Delta\omega}{2\pi}\sum_{j=1}^{N} \widehat{F}(\omega_j)\widehat{\rho}_\epsilon(\omega_j)e^{-i\omega_j\tau} \right| \le 2\|f\|_{L^\infty}\epsilon, \quad \forall\, \tau \in [0,T],$$

where we have taken into account that $\|F\|_{L^\infty} = \|f\|_{L^\infty}$.

Step 5: (Conclusion) To conclude the proof, we recall that $\widehat{F}(\omega) = 2i\mathcal{L}_{t_0}u(\omega)$ can be expressed in terms of the sine transform $\mathcal{L}_t u$ of the function u which was fixed at the beginning of Step 1. Recall also that $f(\tau) = u(t_0 - \tau)$, so that $\|f\|_{L^\infty} = \|u\|_{L^\infty}$. Hence, we can write the real part of $\frac{\Delta\omega}{2\pi}\widehat{F}(\omega_j)\widehat{\rho}_\epsilon(\omega_j)e^{-i\omega_j\tau} = \frac{\Delta\omega}{2\pi}2i\mathcal{L}_{t_0}u(\omega_j)\widehat{\rho}_\epsilon(\omega_j)e^{-i\omega_j\tau}$, in the form $\alpha_j\mathcal{L}_{t_0}(\omega_j)\sin(\omega_j\tau - \vartheta_j)$ for coefficients $\alpha_j \in \mathbb{R}$ and $\theta_j \in \mathbb{R}$ which depend only on $\Delta\omega$ and $\widehat{\rho}_\epsilon(\omega_j)$, but are independent of u. In particular, it follows that

$$\begin{aligned}\sup_{\tau\in[0,\Delta t]}\left|u(t_0-\tau)-\sum_{j=1}^{N}\alpha_j\mathcal{L}_{t_0}u(\omega_j)\sin(\omega_j\tau-\vartheta_j)\right| &= \sup_{t\in[0,\Delta t]}\left|F(\tau)-\mathrm{Re}\left(\frac{\Delta\omega}{2\pi}\sum_{j=1}^{N}\widehat{F}(\omega_j)\widehat{\rho}_\epsilon(\omega_j)e^{-i\omega_j\tau}\right)\right| \\ &\le \sup_{\tau\in[0,\Delta t]}\left|F(\tau)-\frac{\Delta\omega}{2\pi}\sum_{j=1}^{N}\widehat{F}(\omega_j)\widehat{\rho}_\epsilon(\omega_j)e^{-i\omega_j\tau}\right| \\ &\le \sup_{\tau\in[0,\Delta t]}|F(\tau)-F_\epsilon(\tau)| \\ &\quad + \sup_{\tau\in[0,\Delta t]}\left|F_\epsilon(\tau)-\frac{\Delta\omega}{2\pi}\sum_{j=1}^{N}\widehat{F}(\omega_j)\widehat{\rho}_\epsilon(\omega_j)e^{-i\omega_j\tau}\right|.\end{aligned}$$

By Steps 1 and 3, the first term on the right-hand side is bounded by $\le \phi(\epsilon)$, while the second one is bounded by $\le 2\sup_{u\in K}\|u\|_{L^\infty}\epsilon \le C\epsilon$, where $C = C(K) < \infty$ depends only on the compact set $K \subset C([0,T];\mathbb{R}^p)$. Hence, we have

$$\sup_{\tau\in[0,\Delta t]}\left|u(t_0-\tau)-\sum_{j=1}^{N}\alpha_j\mathcal{L}_{t_0}u(\omega_j)\sin(\omega_j\tau-\vartheta_j)\right| \le \phi(\epsilon)+C\epsilon.$$

In this estimate, the function $u\in K$ and $t_0\in[0,T]$ were arbitrary, and the modulus of continuity ϕ as well as the constant C on the right-hand side depend only on the set K. it thus follows that for this choice of α_j, ω_j and ϑ_j, we have

$$\sup_{u\in K}\sup_{t\in[0,T]}\sup_{\tau\in[0,\Delta t]}\left|u(t-\tau)-\sum_{j=1}^{N}\alpha_j\mathcal{L}_{t}u(\omega_j)\sin(\omega_j\tau-\vartheta_j)\right| \le \phi(\epsilon)+C\epsilon.$$

Since $\epsilon > 0$ was arbitrary, the right-hand side can be made arbitrarily small. The claim then readily follows.

□

The next step in the proof of the fundamental Lemma 3.5 needs the following preliminary result in functional analysis,

***Lemma* B.2.** Let $\mathcal{X},\mathcal{Y}$ be Banach spaces, and let $K\subset\mathcal{X}$ be a compact subset. Assume that $\Phi:\mathcal{X}\to\mathcal{Y}$ is continous. Then for any $\epsilon>0$, there exists a $\delta>0$, such that if $\|u-u^K\|_{\mathcal{X}}\le\delta$ with $u\in\mathcal{X}$, $u^K\in K$, then $\|\Phi(u)-\Phi(u^K)\|_{\mathcal{Y}}\le\epsilon$.

Proof. Suppose not. Then there exists $\epsilon_0>0$ and a sequence u_j, u_j^K, ($j\in\mathbb{N}$), such that $\|u_j - u_j^K\|_{\mathcal{X}} \le j^{-1}$, while $\|\Phi(u_j)-\Phi(u_j^K)\|_{\mathcal{Y}}\ge\epsilon_0$. By the compactness of K, we can extract a subsequence $j_k\to\infty$, such that $u_{j_k}^K\to u^K$ converges to some $u^K\in K$. By assumption on u_j, this implies that

$$\|u_{j_k}-u^K\|_{\mathcal{X}}\le\|u_{j_k}-u_{j_k}^K\|_{\mathcal{X}}+\|u_{j_k}^K-u^K\|_{\mathcal{X}}\overset{(k\to\infty)}{\longrightarrow}0,$$

which, by the assumed continuity of Φ, leads to the contradiction that $0<\epsilon_0\le\|\Phi(u_{j_k})-\Phi(u^K)\|_{\mathcal{Y}}\to 0$, as $k\to\infty$. □

Proof of Lemma 3.5. Now, we can prove the fundamental Lemma in the following,

Proof. Let $\epsilon>0$ be given. We can identify $K\subset C_0([0,T];\mathbb{R}^p)$ with a compact subset of $C((-\infty,T];\mathbb{R}^p)$, by extending all $u\in K$ by zero for negative times, i.e. we set $u(t)=0$ for $t<0$. Applying Lemma B.2, with $\mathcal{X}=C_0([0,T];\mathbb{R}^p)$ and $\mathcal{Y}=C_0([0,T];\mathbb{R}^q)$, we can find a $\delta>0$, such that for any $u\in C_0([0,T];\mathbb{R}^p)$ and $u^K\in K$, we have

$$\|u-u^K\|_{L^\infty}\le\delta \quad\Rightarrow\quad \|\Phi(u)-\Phi(u^K)\|_{L^\infty}\le\epsilon. \tag{B.3}$$

By the inverse sine transform Lemma B.1, there exist $N \in \mathbb{N}$, frequencies $\omega_1, \dots, \omega_N \neq 0$, phase-shifts $\vartheta_1, \dots, \vartheta_N$ and coefficients $\alpha_1, \dots, \alpha_N$, such that for any $u \in K$ and $t \in [0, T]$:

$$\sup_{\tau \in [0,T]} \left| u(t-\tau) - \sum_{j=1}^{N} \alpha_j \, \mathcal{L}_t u(\omega_j) \sin(\omega_j \tau - \vartheta_j) \right| \leq \delta.$$

Given $\mathcal{L}_t u(\omega_1), \dots, \mathcal{L}_t u(\omega_N)$, we can thus define a reconstruction mapping $\mathcal{R} : \mathbb{R}^N \times [0, T] \to C([0, T]; \mathbb{R}^p)$ by

$$\mathcal{R}(\beta_1, \dots, \beta_N; t)(\tau) := \sum_{j=1}^{N} \alpha_j \beta_j \sin(\omega_j (t-\tau) - \vartheta_j).$$

Then, for $\tau \in [0, t]$, we have

$$|u(\tau) - \mathcal{R}(\mathcal{L}_t u(\omega_1), \dots, \mathcal{L}_t u(\omega_N); t)(\tau)| \leq \delta.$$

We can now uniquely define $\Psi : \mathbb{R}^N \times [0, T^2/4] \to C_0([0, T]; \mathbb{R}^p)$, by the identity

$$\Psi(\mathcal{L}_t u(\omega_1), \dots, \mathcal{L}_t u(\omega_N); t^2/4) = \Phi\left(\mathcal{R}(\mathcal{L}_t u(\omega_1), \dots, \mathcal{L}_t u(\omega_N); t)\right).$$

Using the short-hand notation $\mathcal{R}_t u = \mathcal{R}(\mathcal{L}_t u(\omega_1), \dots, \mathcal{L}_t u(\omega_N); t)$, we have $\sup_{\tau \in [0,t]} |u(\tau) - \mathcal{R}_t u(\tau)| \leq \delta$, for all $t \in [0, T]$. By (B.3), this implies that

$$\left| \Phi(u)(t) - \Psi(\mathcal{L}_t u(\omega_1), \dots, \mathcal{L}_t u(\omega_N); t^2/4) \right| = |\Phi(u)(t) - \Phi(\mathcal{R}_t u)(t)| \leq \epsilon.$$

□

B.3 Proof of Lemma 3.6

Proof. Let $\omega \neq 0$ be given. For a (small) parameter $s > 0$, we consider

$$\ddot{y}_s = \frac{1}{s} \sigma(-s\omega^2 y_s + su), \quad y_s(0) = \dot{y}_s(0) = 0.$$

Let Y be the solution of

$$\ddot{Y} = -\omega^2 Y + u, \quad Y(0) = \dot{Y}(0) = 0.$$

Then we have, on account of $\sigma(0) = 0$ and $\sigma'(0) = 1$,

$$\begin{aligned} s^{-1}\sigma(-s\omega^2 Y + su) - [-\omega^2 Y + u] &= \frac{\sigma(-s\omega^2 Y + su) - \sigma(0)}{s} - \sigma'(0)[-\omega^2 Y + u] \\ &= \frac{1}{s} \int_0^s \frac{\partial}{\partial \zeta} \left[\sigma(-\zeta\omega^2 Y + \zeta u) \right] d\zeta - \sigma'(0)[-\omega^2 Y + u] \\ &= \frac{1}{s} \left(\int_0^s \left[\sigma'(-\zeta\omega^2 Y + \zeta u) - \sigma'(0) \right] d\zeta \right) \left[-\omega^2 Y + u \right]. \end{aligned}$$

It follows from Lemma 3.4 that for any input $u \in K$, with $\sup_{u \in K} \|u\|_{L^\infty} =: B < \infty$, we have a uniform bound $\|Y\|_{L^\infty} \leq BT/\omega$, hence we can estimate

$$| -\omega^2 Y + u| \leq B(\omega T + 1),$$

uniformly for all such u. In particular, it follows that

$$\left| s^{-1}\sigma(-s\omega^2 Y + su) - [-\omega^2 Y + u] \right| \leq B(T\omega + 1) \sup_{|x| \leq sB(T\omega+1)} |\sigma'(x) - \sigma'(0)|.$$

Clearly, for any $\delta > 0$, we can choose $s \in (0, 1]$ sufficiently small, such that the right hand-side is bounded by δ, i.e. with this choice of s,

$$\left| s^{-1}\sigma(-s\omega^2 Y(t) + su(t)) - [-\omega^2 Y(t) + u(t)] \right| \leq \delta, \quad \forall\, t \in [0, T],$$

holds for any choice of $u \in K$. We will fix this choice of s in the following, and write $g(y, u) := s^{-1}\sigma(-s\omega^2 y + su)$. We note that g is Lipschitz continuous in y, for all $|y| \leq BT/\omega$ and $|u| \leq B$, with $\mathrm{Lip}_y(g) \leq \omega^2 \sup_{|\xi| \leq B(\omega T+1)} |\sigma'(\xi)|$.

To summarize, we have shown that Y solves

$$\ddot{Y} = g(Y, u) + f, \qquad Y(0) = \dot{Y}(0) = 0,$$

where $\|f\|_{L^\infty} \le \delta$. By definition, y_s solves

$$\ddot{y}_s = g(y_s, u), \qquad y_s(0) = \dot{y}_s(0) = 0.$$

It follows from this that

$$\begin{aligned} |y_s(t) - Y(t)| &\le \int_0^t \int_0^\tau \{|g(y_s(\theta), u(\theta)) - g(Y(\theta), u(\theta))| + |f(\theta)|\} \, d\theta \, d\tau \\ &\le \int_0^t \int_0^\tau \left\{ \mathrm{Lip}_y(g) |y_s(\theta) - Y(\theta)| + \delta \right\} \, d\theta \, d\tau \\ &\le T\omega^2 \sup_{|\xi| \le B(\omega T + 1)} |\sigma'(\xi)| \int_0^t |y_s(\tau) - Y(\tau)| \, d\tau + T^2 \delta. \end{aligned}$$

Recalling that $Y(t) = \mathcal{L}_t u(\omega)$, then by Gronwall's inequality, the last estimate implies that

$$\sup_{t \in [0,T]} |y_s(t) - \mathcal{L}_t u(\omega)| = \sup_{t \in [0,T]} |y_s - Y| \le C\delta,$$

for a constant $C = C(T, \omega, \sup_{|\xi| \le B(\omega T+1)} |\sigma'(\xi)|) > 0$, depending only on T, ω, B and σ'. Since $\delta > 0$ was arbitrary, we can ensure that $C\delta \le \epsilon$. Thus, we have shown that a suitably rescaled nonlinear oscillator approximates the harmonic oscillator to any desired degree of accuracy, and uniformly for all $u \in K$.

To finish the proof, we observe that y solves

$$\ddot{y} = \sigma(-\omega^2 y + su), \qquad y(0) = \dot{y}(0) = 0,$$

if, and only if, $y_s = y/s$ solves

$$\ddot{y}_s = s^{-1}\sigma(-s\omega^2 y_s + su), \qquad y_s(0) = \dot{y}_s(0) = 0.$$

Hence, with $W = -\omega^2$, $V = s$, $b = 0$ and $A = s^{-1}$, we have

$$\sup_{t \in [0,T]} |Ay(t) - \mathcal{L}_t u(\omega)| = \sup_{t \in [0,T]} |y_s(t) - \mathcal{L}_t u(\omega)| \le \epsilon.$$

This concludes the proof. □

B.4 Proof of Lemma 3.7

Proof. Let $\epsilon, \Delta t$ be given. By the sine transform reconstruction Lemma B.1, there exists $N \in \mathbb{N}$, frequencies $\omega_1, \dots, \omega_N$, weights $\alpha_1, \dots, \alpha_N$ and phase-shifts $\vartheta_1, \dots, \vartheta_N$, such that

$$\sup_{\tau \in [0, \Delta t]} \left| u(t - \tau) - \sum_{j=1}^N \alpha_j \mathcal{L}_t u(\omega_j) \sin(\omega_j \tau - \vartheta_j) \right| \le \frac{\epsilon}{2}, \quad \forall\, t \in [0, T], \ \forall\, u \in K, \tag{B.4}$$

where any $u \in K$ is extended by zero to negative times. It follows from Lemma 3.6, that there exists a coupled oscillator network,

$$\ddot{y} = \sigma(w \odot y + Vu + b), \qquad y(0) = \dot{y}(0) = 0,$$

with dimension $m = pN$, and $w \in \mathbb{R}^m$, $V \in \mathbb{R}^{m \times p}$, and a linear output layer $y \mapsto \widetilde{A}y$, $\widetilde{A} \in \mathbb{R}^{m \times m}$, such that $[\widetilde{A}y(t)]_j \approx \mathcal{L}_t u(\omega_j)$ for $j = 1, \dots, N$; more precisely, such that

$$\sup_{t \in [0,T]} \sum_{j=1}^N |\alpha_j| \left| \mathcal{L}_t u(\omega_j) - [\widetilde{A}y]_j(t) \right| \le \frac{\epsilon}{2}, \quad \forall\, u \in K. \tag{B.5}$$

Composing with another linear layer $B : \mathbb{R}^m \simeq \mathbb{R}^{p \times N} \to \mathbb{R}^p$, which maps $\boldsymbol{\beta} = [\beta_1, \dots, \beta_N]$ to

$$B\boldsymbol{\beta} := \sum_{j=1}^N \alpha_j \beta_j \sin(\omega_j \Delta t - \vartheta_j) \in \mathbb{R}^p,$$

we define $A := B\widetilde{A}$, and observe that from (B.4) and (B.5):

$$\begin{aligned}\sup_{t\in[0,T]} |u(t-\Delta t) - Ay(t)| &\le \sup_{t\in[0,T]} \left| u(t-\Delta t) - \sum_{j=1}^N \alpha_j \mathcal{L}_t u(\omega_j) \sin(\omega_j \Delta t - \vartheta_j) \right| \\ &\quad + \sup_{t\in[0,T]} \sum_{j=1}^N |\alpha_j| \left| \mathcal{L}_t u(\omega_j) - [\widetilde{A}y]_j(t) \right| |\sin(\omega_j \Delta t - \vartheta_j)| \\ &\le \epsilon. \end{aligned}$$

□

B.5 Proof of Lemma 3.8

Proof. Fix Σ, Λ, γ as in the statement of the lemma. Our goal is to approximate $u \mapsto \Sigma\sigma(\Lambda u + \gamma)$.

Step 1: (nonlinear layer) We consider a first layer for a hidden state $y = [y_1, y_2]^T \in \mathbb{R}^{p+p}$, given by

$$\begin{Bmatrix} \ddot{y}_1(t) = \sigma(\Lambda u(t) + \gamma) \\ \ddot{y}_2(t) = \sigma(\gamma) \end{Bmatrix}, \quad y(0) = \dot{y}(0) = 0.$$

This layer evidently does not approximate $\sigma(\Lambda u(t) + \gamma)$; however, it does encode this value in the second derivative of the hidden variable y_1. The main objective of the following analysis is to approximately compute $\ddot{y}_1(t)$ through a suitably defined additional layer.

Step 2: (Second-derivative layer) To obtain an approximation of $\sigma(\Lambda u(t) + \gamma)$, we first note that the solution operator

$$\mathcal{S} : u(t) \mapsto \eta(t), \quad \text{where } \ddot{\eta}(t) = \sigma(\Lambda u(t) + \gamma) - \sigma(\gamma), \quad \eta(0) = \dot{\eta}(0) = 0,$$

defines a continuous mapping $\mathcal{S} : C_0([0,T];\mathbb{R}^p) \to C_0^2([0,T];\mathbb{R}^p)$, with $\eta(0) = \dot{\eta}(0) = \ddot{\eta}(0) = 0$. Note that η is very closely related to y_1. The fact that $\ddot{\eta} = 0$ is important to us, because it allows us to *smoothly* extend η to negative times by setting $\eta(t) := 0$ for $t < 0$ (which would not be true for $y_1(t)$). The resulting extension defines a compactly supported function $\eta : (-\infty, 0] \to \mathbb{R}^p$, with $\eta \in C^2((-\infty, T];\mathbb{R}^p)$. Furthermore, by continuity of the operator $\mathcal{S}$, the image $\mathcal{S}(K)$ of the compact set K under $\mathcal{S}$ is compact in $C^2((-\infty, T];\mathbb{R}^p)$. From this, it follows that for small $\Delta t > 0$, the second-order backward finite difference formula converges,

$$\sup_{t\in[0,T]} \left| \frac{\eta(t) - 2\eta(t-\Delta t) + \eta(t - 2\Delta t)}{\Delta t^2} - \ddot{\eta}(t) \right| = o_{\Delta t \to 0}(1), \quad \forall \eta = \mathcal{S}(u),\, u \in K,$$

where the bound on the right-hand side is uniform in $u \in K$, due to equicontinuity of $\{\ddot{\eta} \,|\, \eta = \mathcal{S}(u),\, u \in K\}$. In particular, the second derivative of η can be approximated through *linear combinations of time-delays of* η. We can now choose $\Delta t > 0$ sufficiently small so that

$$\sup_{t\in[0,T]} \left| \frac{\eta(t) - 2\eta(t-\Delta t) + \eta(t - 2\Delta t)}{\Delta t^2} - \ddot{\eta}(t) \right| \le \frac{\epsilon}{2\|\Sigma\|}, \quad \forall y = \mathcal{S}(u),\, u \in K,$$

where $\|\Sigma\|$ denotes the operator norm of the matrix Σ. By Lemma 3.7, applied to the input set $\widetilde{K} = \mathcal{S}(K) \subset C_0([0,T];\mathbb{R}^p)$, there exists a coupled oscillator

$$\ddot{z}(t) = \sigma(w \odot z(t) + V\eta(t) + b), \quad z(0) = \dot{z}(0) = 0, \tag{B.6}$$

and a linear output layer $z \mapsto \widetilde{A}z$, such that

$$\sup_{t\in[0,T]} \left| [\eta(t) - 2\eta(t-\Delta t) + \eta(t-2\Delta t)] - \widetilde{A}z(t) \right| \le \frac{\epsilon \Delta t^2}{2\|\Sigma\|}, \quad \forall \eta = \mathcal{S}(u),\, u \in K.$$

Indeed, Lemma 3.7 shows that time-delays of any given input signal can be approximated with any desired accuracy, and $\eta(t) - 2\eta(t - \Delta) - \eta(t - 2\Delta)$ is simply a linear combination of time-delays of the input signal η in (B.6).

To connect $\eta(t)$ back to the $y(t) = [y_1(t), y_2(t)]^T$ constructed in Step 1, we note that

$$\ddot{\eta} = \sigma(Au(t) + b) - \sigma(b) = \ddot{y}_1 - \ddot{y}_2,$$

and hence, taking into account the initial values, we must have $\eta \equiv y_1 - y_2$ by ODE uniqueness. In particular, upon defining a matrix $\widetilde{V}$ such that $\widetilde{V}y := Vy_1 - Vy_2 \equiv V\eta$, we can equivalently write (B.6) in the form,

$$\ddot{z}(t) = \sigma(w \odot z(t) + \widetilde{V}y(t) + b), \quad z(0) = \dot{z}(0) = 0. \tag{B.7}$$

Step 3: (Conclusion)

Composing the layers from Step 1 and 2, we obtain a coupled oscillator

$$\ddot{y}^\ell = \sigma(w^\ell \odot y^\ell + V^\ell y^{\ell-1} + b^\ell), \quad (\ell = 1, 2),$$

initialized at rest, with $y^1 = y$, $y^2 = z$, such that for $A := \Sigma\widetilde{A}$ and $c := \Sigma\sigma(\gamma)$, we obtain

$$\begin{aligned}
\sup_{t\in[0,T]} \left|\left[Ay^2(t) + c\right] - \Sigma\sigma(\Lambda u(t) + \gamma)\right| &\le \|\Sigma\| \sup_{t\in[0,T]} \left|\widetilde{A}z(t) - \left[\sigma(\Lambda u(t) + \gamma) - \sigma(\gamma)\right]\right| \\
&= \|\Sigma\| \sup_{t\in[0,T]} \left|\widetilde{A}z(t) - \ddot{\eta}(t)\right| \\
&\le \|\Sigma\| \sup_{t\in[0,T]} \left|\widetilde{A}z(t) - \frac{\eta(t) - 2\eta(t-\Delta t) + \eta(t-2\Delta t)}{\Delta t^2}\right| \\
&\quad + \|\Sigma\| \sup_{t\in[0,T]} \left|\frac{\eta(t) - 2\eta(t-\Delta t) + \eta(t-2\Delta t)}{\Delta t^2} - \ddot{\eta}(t)\right| \\
&\le \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\end{aligned}$$

This concludes the proof. □

B.6 Proof of Theorem 3.1

Proof. **Step 1:** By the Fundamental Lemma 3.5, there exist N, a continuous mapping Ψ, and frequencies $\omega_1, \dots, \omega_N$, such that

$$|\Phi(u)(t) - \Psi(\mathcal{L}_t u(\omega_1), \dots, \mathcal{L}_t u(\omega_N); t^2/4)| \le \epsilon,$$

for all $u \in K$, and $t \in [0, T]$. Let M be a constant such that

$$|\mathcal{L}_t u(\omega_1)|, \dots, |\mathcal{L}_t u(\omega_N)|, \frac{t^2}{4} \le M,$$

for all $u \in K$ and $t \in [0, T]$. By the universal approximation theorem for ordinary neural networks, there exist weight matrices Σ, Λ and bias γ, such that

$$|\Psi(\beta_1, \dots, \beta_N; t^2/4) - \Sigma\sigma(\Lambda\boldsymbol{\beta} + \gamma)| \le \epsilon, \quad \boldsymbol{\beta} := [\beta_1, \dots, \beta_N; t^2/4]^T,$$

holds for all $t \in [0, T]$, $|\beta_1|, \dots, |\beta_N| \le M$.

Step 2: Fix $\epsilon_1 \le 1$ sufficiently small, such that also $\|\Sigma\|\|\Lambda\|\mathrm{Lip}(\sigma)\epsilon_1 \le \epsilon$, where $\mathrm{Lip}(\sigma) := \sup_{|\xi|\le\|\Lambda\|M+|\gamma|+1} |\sigma'(\xi)|$ denotes an upper bound on the Lipschitz constant of the activation function over the relevant range of input values. It follows from Lemma 3.6, that there exists an oscillator network,

$$\ddot{y}^1 = \sigma(w^1 \odot y^1 + V^1 u + b^1), \quad y^1(0) = \dot{y}^1(0) = 0, \tag{B.8}$$

of depth 1, such that

$$\sup_{t\in[0,T]} |[\mathcal{L}_t u(\omega_1), \dots, \mathcal{L}_t u(\omega_N); t^2/4]^T - A^1 y^1(t)| \le \epsilon_1,$$

for all $u \in K$.

Step 3: Finally, by Lemma 3.8, there exists an oscillator network,

$$\ddot{y}^2 = \sigma(w^2 \odot y^2 + V^2 y^1 + b^1),$$

of depth 2, such that

$$\sup_{t\in[0,T]} |A^2 y^2(t) - \Sigma\sigma(\Lambda A^1 y^1(t) + \gamma)| \le \epsilon,$$

holds for all y^1 belonging to the compact set $K_1 := \mathcal{S}(K) \subset C_0([0,T];\mathbb{R}^{N+1})$, where $\mathcal{S}$ denotes the solution operator of (B.8).

Step 4: Thus, we have for any $u \in K$, and with short-hand $\mathcal{L}_t u(\boldsymbol{\omega}) := (\mathcal{L}_t u(\omega_1), \dots, \mathcal{L}_t u(\omega_N))$,

$$\begin{aligned}\left|\Phi(u)(t) - A^2 y^2(t)\right| \le &\left|\Phi(u)(t) - \Psi(\mathcal{L}_t u(\boldsymbol{\omega}); t^2/4)\right| \\ &+ \left|\Psi(\mathcal{L}_t u(\boldsymbol{\omega}); t^2/4) - \Sigma\sigma(\Lambda[\mathcal{L}_t u(\boldsymbol{\omega}); t^2/4] + \gamma)\right| \\ &+ \left|\Sigma\sigma(\Lambda[\mathcal{L}_t u(\boldsymbol{\omega}); t^2/4] + \gamma) - \Sigma\sigma(\Lambda A^1 y^1(t) + \gamma)\right| \\ &+ \left|\Sigma\sigma(\Lambda A_1 y_1(t) + \gamma) - A^2 y^2(t)\right|.\end{aligned}$$

By step 1, we can estimate

$$\left|\Phi(u)(t) - \Psi(\mathcal{L}_t u(\boldsymbol{\omega}); t^2/4)\right| \le \epsilon, \quad \forall\, t \in [0,T],\ u \in K.$$

By the choice of Σ, Λ, γ, we have

$$\left|\Psi(\mathcal{L}_t u(\boldsymbol{\omega}); t^2/4) - \Sigma\sigma(\Lambda[\mathcal{L}_t u(\boldsymbol{\omega}); t^2/4] + \gamma)\right| \le \epsilon, \quad \forall\, t \in [0,T],\ u \in K.$$

By construction of y^1 in Step 2, we have

$$\begin{aligned}&\left|\Sigma\sigma(\Lambda[\mathcal{L}_t u(\boldsymbol{\omega}); t^2/4] + \gamma) - \Sigma\sigma(\Lambda A_1 y_1(t) + \gamma)\right| \\ &\qquad \le \|\Sigma\| \mathrm{Lip}(\sigma) \|\Lambda\| \left|[\mathcal{L}_t u(\boldsymbol{\omega}); t^2/4] - A^1 y^1(t)\right| \\ &\qquad \le \|\Sigma\| \mathrm{Lip}(\sigma) \|\Lambda\|\, \epsilon_1 \\ &\qquad \le \epsilon,\end{aligned}$$

for all $t \in [0,T]$ and $u \in K$. By construction of y^2 in Step 3, we have

$$\left|\Sigma\sigma(\Lambda A^1 y^1(t) + \gamma) - A^2 y^2(t)\right| \le \epsilon, \quad \forall\, t \in [0,T],\ u \in K.$$

Thus, we conclude that

$$|\Phi(u)(t) - A^2 y^2(t)| \le 4\epsilon,$$

for all $t \in [0,T]$ and $u \in K$. Since $\epsilon > 0$ was arbitrary, we conclude that for any causal and continuous operator $\Phi : C_0([0,T];\mathbb{R}^p) \to C_0([0,T];\mathbb{R}^q)$, compact set $K \subset C_0([0,T];\mathbb{R}^p)$ and $\epsilon > 0$, there exists a coupled oscillator of depth 3, which uniformly approximates Φ to accuracy ϵ for all $u \in K$. This completes the proof. □