A Sampling theory in higher dimensions

The sampling theory — in its original form — is only applicable to one dimensional signals. However, it can be extended to higher dimensions in a straightforward manner. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function in $L^1(\mathbb{R}^n)$, which we think of as a higher mode signal. Let $I(\Omega_1, \ldots, \Omega_n)$ denote an n-dimensional rectangle about the origin with side lengths $\Omega_1, \ldots, \Omega_n$. Suppose that the Fourier transform $\widehat{f}$ vanishes identically outside of of $I(\Omega_1, \ldots, \Omega_n)$. Then

$$f(t_1, \ldots, t_n) = \sum_{m_1=-\infty}^{\infty} \cdots \sum_{m_n-\infty}^{\infty} f\left(\frac{m_1}{2\Omega_1}, \ldots, \frac{m_n}{2\Omega_n}\right) sinc\big(2\Omega_1\big(t_1 - \frac{n}{2\Omega_1}\big)\big) \cdots sinc\big(2\Omega_n\big(t_n - \frac{n}{2\Omega_n}\big)\big).$$

Thus we see that sampling f on the lattice defined by lengths $\left(\frac{1}{2\Omega_1}, \ldots, \frac{1}{2\Omega_n}\right)$ and taking shifted sinc functions of bandwidth $2\Omega_k$, for $1 \leq k \leq n$, we can reconstruct the function f as in the one dimensional case. Note that as in the case of the one-dimensional Nyquist-Shanon theorem, in order for perfect reconstruction one needs to sample at larger than twice the dominant frequency present in the signal. Therefore, in practise one would take the maximum of $\Omega = \max_i\{\Omega_i\}$ and sample at a frequency of 2Ω.

Curse of dimensionality. While the multidimensional Nyquist-Shanon sampling theorem provides a convenient theoretical framework in which to understand signal processing problems in higher dimensions. It does not come without problems. In practise, the multidimensional sampling theorem is extremely inefficient.

The main issue with sampling in higher dimensions is that there is an exponential increase in volumes of cubes (or rectangles/balls) associated with adding extra dimensions. To see this, imagine we had a signal $f : [0, 1] \to \mathbb{R}$ whose dominant frequency was 50-Hertz. Let us then suppose we wish to perform a reconstruction by using a sample rate of 100-Hertz. This means that we would need to sample exactly $10^2 = 100$ points from the unit interval $[0, 1]$ each spaced at a distance of 0.01. Now, imagine that we had a 10 mode signal $g : [0, 1]^{10} \to \mathbb{R}$ on the unit cube whose dominant frequency was also 50-Hertz. We wish to perform a 100-Hertz sample rate reconstruction of g as we did for f. Now we see a problem, in this instance we would need to sample $(10^2)^1 0 = 10^2 0$ points from the 10-dimensional cube. Thus when using a sampling distance of 0.01 we see that the 10-dimensional cube $[0, 1]^1 0$ is 10^{18}-times larger than the 1-dimensional cube $[0, 1]$. This exponential increase in the amount of sample points needed to reconstruct a high mode signal is referred to as the curse of dimensionality and is a mathematical consequence of the fact that volumes of many mathematical shapes grow exponentially with dimension. This makes the sampling theory of Nqyquist and Shanon some what unusable in practise for higher mode signals.

There have been other reconstruction techniques, most notable compressed sensing, that have shown far superior performance than classical sampling due to their ability to break the Nyquist limit and allow far fewer sampling points. However, such techniques have the added problem that they are memory intensive for high mode signals. As we show INRs offer a convenient middle ground that makes them perfectly suitable for signal reconstruction in higher mode signal settings.

B Proofs of results in section 3.2

B.1 Preliminaries

We will be using the basic theory of Hilbert spaces in L^2. Namely, the space of square integrable functions on $\mathbb{R}$ will be denoted by $L^2(\mathbb{R})$ and we recall that this is defined as the vector space of equivalence classes of measurable functions on $\mathbb{R}$ with the following inner product

$$\langle f, g \rangle_{L^2} = \int_{\mathbb{R}} f \cdot g. \tag{14}$$

We wil also need to make use of the Sobolev spaces of order r, denoted by $W_2^r(\mathbb{R})$. We define this space as the space of L^2-functions that have r weak derivatives that are also in $L^2(\mathbb{R})$.

Proof of prop. 3.3. The sinc function is in L^2 and further since its Fourier transform is the rectangular function it is easy to see using the fact that the Fourier transform is an isometry of L^2 that it must satisfy the Partition of unity conditon.

The Gaussian $e^{x^2/2s^2}$ is also in L^2, however this function does not satisfy the partition of unity condition. The reason being that by the Poisson summatic formula it suffice to show that the translates over the integers of the Fourier transform sums to 1. However, the Fourier transform of a Gaussian is a Gaussian, therefore outside the rectangle $[-\pi/\beta, \pi/\beta]$ the exponential decay cannot contribute anything to the sum in the partition of unity condition. Hence the sum could not sum to 1. It follows the translates of the Gaussian can only form a weak Riesz basis.

In general, the Fourier transform of a wavelet is localized in phase and frequency, hence as in the case of the Gaussian above, they will be in L^2 and form a weak Riesz basis but in general they might not form a Riesz basis. Conditions have been given for a wavelet to form a Riesz basis, see Sun & Zhou (2002), though this is outside the scope of this work.

In order to form a Riesz basis $ReLU$ would have to be in $L^2(\mathbb{R})$, which it is not. On the other hand, given $x \in \mathbb{R}$ we have that

$$\sum_{k\in\mathbb{Z}} ReLU(x+k) = \sum_{k\geq -x, k\in\mathbb{Z}} ReLU(x+k) = \sum_{k\geq -x, k\in\mathbb{Z}} (x+k) = \infty$$

showing that there is no way $ReLU$ could satisfy the partition of unity condition.

A similar proof shows that translates of sine cannot form a Riesz/weak Riesz basis. □

B.1.1 RESULTS ON THE ERROR KERNEL AND PUC CONDITION

We recall from sec. 3.3 that the understanding of the sampling properties of the shifted basis functions F_k comes down to analysing the error kernel $E_{\widetilde{F},F}$. The reason being was that the average error $\overline{\epsilon}_s(T)^2$ is a good predictor of the true error $\epsilon_s(T)^2$.

We sketch a proof showing that the vanishing of the error kernel in the limit $T \to 0$ for a suitable test function $\widetilde{F}$ is equivalent to F satisfying the partition of unity condition. We will do this under two assumptions:

A1. The Fourier transform of F is continuous at 0.

A2. The Fourier transform of $\widetilde{F}$ is continuous at 0.

A3. The sampled signal s we wish to reconstruct is contained in W_2^r for some $r > \frac{1}{2}$. This assumption is needed so that the quantity ϵ_{corr} goes to zero as $T \to 0$.

We remark that an explicit construction of $\widetilde{F}$ will be given after the proof as during the course of the proof we will see what conditions we need to impose for the construction of $\widetilde{F}$ from F.

From the definition of the approximation operator, equation 3, we have that

$$\lim_{T\to 0} ||f - A_T(f)||^2_{L^2} = \lim_{T\to 0} \int_{-\infty}^{\infty} E_{\widetilde{F},F}(T\omega)|\hat{s}(\omega)|^2 \frac{d\omega}{2\omega} \tag{15}$$

where we remind the reader that the error kernel $E_{\widetilde{F},F}$ is given by equation equation 6. We now observe that if $\widetilde{F}$ is a function such that $\hat{\widetilde{F}}$ is bounded and F satisfies the first Riesz condition, condition 1 from defn. 3.1, then by definition it follows that $E_{\widetilde{F},F}$ is bounded. Therefore in the above integral we can apply the dominated convergence theorem and compute

$$\lim_{T\to 0} ||s - A_T(s)||^2_{L^2} = \int_{-\infty}^{\infty} \lim_{T\to 0} E_{\widetilde{F},F}(T\omega)|\hat{s}(\omega)|^2 \frac{d\omega}{2\omega} \tag{16}$$

$$= E_{\widetilde{F},F}(0) \int_{-\infty}^{\infty} |\hat{s}(\omega)|^2 \frac{d\omega}{2\omega} \tag{17}$$

$$= E_{\widetilde{F},F}(0)||s||^2 \tag{18}$$

where to get the second equality we have used assumptions A1 and A2 above and to get the third equality we have used the fact that the Fourier transform is an isometry from L^2 to L^2.

We thus see that the statement $\lim_{T\to 0} ||s - Q_T(s)||_{L^2}^2 = 0$ is equivalent to $E_{\widetilde{F},F}(0) = 0$. From equation 6 this is equivalent to

$$E_{\widetilde{F},F}(0) = |1 - \hat{\widetilde{F}}(0)\hat{F}(0)|^2 + |\hat{\widetilde{F}}(0)|^2 \sum_{k\neq 0} |\hat{F}(2\pi k)|^2 = 0. \tag{19}$$

We see that $E_{\widetilde{F},F}(0)$ is a sum of positive terms and hence will vanish if and only if all the terms in the summands vanish. Looking at the first summand we see that we need $\hat{\widetilde{F}}(0)\hat{F}(0) = 1$, which can hold if and only if both factors are not zero. We normalise the function F so that $\hat{F}(0) = \int F(x)dx = 1$. Thus the conditions that need to be satisfied are

$$\hat{\widetilde{F}}(0) = 1 \text{ and } \sum_{k\neq 0} |\hat{F}(2\pi k)|^2 = 0. \tag{20}$$

We can rewrite the second condition in equation 20 as

$$\hat{F}(2\pi k) = \delta_k \tag{21}$$

where δ denotes the Dirac delta distribution. From this viewpoint we then immediately have that the second condition can be written in the form

$$\sum_k F(x+k) = 1 \tag{22}$$

which is precisely the partition of unity condition.

The function $\widetilde{F}$ is easy to choose. Let $\mathcal{S}$ denote Schwartz space of Schwartz functions in L^2. It is well known that this space is dense in $L^2(\mathbb{R})$ and that the Fourier transform maps $\mathcal{S}$ onto itself. Therefore, in the Fourier domain let $\widetilde{\mathcal{S}}$ denote the set of Schwartz functions f such that $\hat{f}(0) \neq 0$. Note that $\widetilde{\mathcal{S}}$ is dense in L^2 and elements in $\widetilde{\mathcal{S}}$ are continuous at the origin. In order to define $\widetilde{F}$ we simply take any element $f \in \widetilde{\mathcal{S}}$ and let $\widetilde{F} = \frac{1}{\hat{f}(0)} f$. In fact, if we denote the space $\overline{\mathcal{S}}$ to consist of those Schwartz functions f whose Fourier transform satisfies $\hat{f}(0) = 1$, then it is easy to see that $\overline{\mathcal{S}}$ is dense in L^2. Thus the space $\overline{\mathcal{S}}$ can be used as a test space for $\widetilde{F}$ and is the defining test space for the approximation operator A_T.

B.1.2 Proofs of main results in section 3.3

Proof of theorem 3.4. We first note that by condition 1 in defn. 3.1. The space $V(F)$ is a subspace of $L^2(\mathbb{R})$. Therefore, the space $V(F)$ with the induced L^2-norm forms a well-defined normed vector space.

Since $g \in V(F)$ we can write $g = \sum_{k=-\infty}^{\infty} a(k)F(x-k)$ in L^2. Therefore, there exists a $k(\epsilon)$ such that

$$\left|\left| g - \sum_{k=-k(\epsilon)}^{k(\epsilon)} a(k)F(x-k) \right|\right|_{L^2} < \epsilon. \tag{23}$$

We can then define a 2-layer neural network f with $n(\epsilon) = 2k(\epsilon)$ neurons as follows: Let the weights in the first layer be the constant vector $[1, \cdots, 1]^T$ and the associated bias to be the vector $[-k(\epsilon), -k(\epsilon)+1, \ldots, k(\epsilon)]^T$. Let the weights associated to the second layer be the vector $[a(-k(\epsilon)), a(-k(\epsilon)+1), \cdots, a(k(\epsilon))]$ and the associated bias be 0. These weights and biases will make up the parameters for the neural network f and in the hidden layer we take F as the non-linearity.

Applying equation 23 we obtain that

$$||f(\theta) - g||_L^2 < \epsilon. \tag{24}$$

□

Proof of prop. 3.9. We start by proving the proposition for the case that $s \in W_2^1$. We then note that in this case by thm. 3.7, there is a $\Omega > 0$ sufficiently small such that $\epsilon_{corr} < \frac{\epsilon}{2}$. Furthermore, by lemma 3.8 we have that average approximation error $\overline{\epsilon}(\Omega) < \frac{\epsilon}{2}$ for Ω sufficiently small. Therefore, by taking $f_\Omega = A_\Omega(s) \in V_\Omega(F)$ the proposition follows for signal in W_2^1.

For the general case of signals $s \in L^2$, we proceed as follows. We use the fact that $C_c^\infty(\mathbb{R})$ is dense in L^2. This implies we can find an $f \in C_c^\infty(\mathbb{R})$ such that $||f - s||_{L^2} < \frac{\epsilon}{2}$. Furthermore, since $s \in C_c^\infty$ it lies in W_2^1. By the above we have that there exists $\Omega > 0$ such that $||f - A_\Omega(f)||_{L^2} < \frac{\epsilon}{2}$. The result then follows by an application of the triangle inequality. □

Proof of thm. 3.10. *Proof.* By prop. 3.9 there exists an $\Omega > 0$ sufficiently small and an $f_\Omega \in V_\Omega(F)$ such that

$$||s - f_\Omega||_{L^2} < \frac{\epsilon}{2}. \tag{25}$$

As f_Ω lies in $V_\Omega(F)$ we can write $f_\Omega = \sum_{k=-\infty}^{\infty} a_\Omega(k) F(\frac{1}{\Omega}(x - \Omega k))$. This implies there exists $k(\epsilon) > 0$ such that

$$\left|\left| f_\Omega - \sum_{k=-k(\epsilon)}^{k(\epsilon)} a_\Omega(k) F\left(\frac{1}{\Omega}(x - \Omega k)\right)\right|\right| < \frac{\epsilon}{2}. \tag{26}$$

We define a neural network $\mathcal{N}$ with $n(\epsilon) = 2k(\epsilon)$ neurons in its hidden layer as follows. The weights in the first layer will be the constant vector $[1, \ldots, 1]^T$ and the associated bias will be the vector $[-\Omega k(\epsilon), -\Omega k(\epsilon) + 1, \ldots, \Omega k(\epsilon)]^T$. The weights associated to the second layer will be $[a(-k(\epsilon)), \ldots, a(k(\epsilon))]$ and the bias for this layer will be 0. These weights and biases will make up the parameters θ for the neural network. In the hidden layer we take as activation the function F_Ω. With these parameters and activation function, equation 26 implies that

$$||\mathcal{N}(\theta) - f_\Omega||_{L^2} < \frac{\epsilon}{2}. \tag{27}$$

An application of the triangle inequality then proves the theorem. □

□

C On Taken's embedding theorem

Taken's embedding theorem is a delay embedding theorem giving conditions under which the strange attractor of a dynamical system can be reconstructed from a sequence of observations of the phase space of that dynamical system.

The theorem constructs an embedding vector for each point in time

$$x(t_i) = [x(t_i), x_(t_i + n\Delta t), \ldots, x(t_i + (d-1)n\Delta t)]$$

Where d is the embedding dimension and n is a fixed value. The theorem then states that in order to reconstruct the dynamics in phase space for any n the following condition must be met

$$d \geq 2D + l$$

where D is the box counting dimension of the strange attractor of the dynamical system which can be thought of as the theoretical dimension of phase space for which the trajectories of the system do not overlap.

Drawbacks of the theorem: The theorem does not provide conditions as to what the best n is and in practise when D is not known it does not provide conditions for the embedding dimension d. The quantity $n\Delta t$ is the amount of time delay that is being applied. Extremely short time delays cause the values in the embedding vector to almost be the same, and extremely large time delays cause the value to be uncorrelated random variables. The following papers show how one can find the time delay in practise Kim et al. (1999); Small (2005). Furthermore, in practise estimating the embedding dimension is often done by a false nearest neighbours algorithm Kennel et al. (1992).

Thus in practise time delay embeddings for the reconstruction of dynamics can require the need to carry further experiments to find the best time delay length and embedding dimension.

D RELATION TO UNIVERSAL APPROXIMATION

Thms. 3.4 and 3.10 can be interpreted as universal approximation theorems for signals in $L^2(\mathbb{R})$. The classic universal approximation theorems are generally for functions on bounded domains . In 92' W. A Light extended those results on bounded domains to a universal approximation for continuous function on $\mathbb{R}^n$ by sigmoid activated networks Light (1992). His result can also be made to hold for sinc activated networks, and since the space of continuous functions is dense in $L^2(\mathbb{R})$ his proof easily extends to give a universal approximation result for sinc activated 2 layer networks for signals in $L^2(\mathbb{R})$. Thus thm. 3.10 can be seen as giving a different proof of W.A. Light's result.

Although it seems like such results have been known through classical methods, we would like to emphasize that the importance of thm. 3.10 comes in how it relates to sampling theory. Given a signal $s \in L^2(\mathbb{R})$ that is bandlimited, the Nyquist-Shannon sampling theorem. This classical theorem, as denoted in equation **??**, allows signal reconstruction using shifted sinc functions while explicitly specifying the coefficients of these shifted sinc functions. These coefficients correspond to samples of the signal, represented as $s(n/2\Omega)$. In cases where the signal is not bandlimited, prop. 3.9 still enables signal reconstruction via shifted sinc functions, albeit without a closed formula for the coefficients involved. This is precisely where thm. 3.10 demonstrates its significance. The theorem reveals that the shifted sinc functions constituting the approximation can be encoded using a two-layer sinc-activated neural network. Notably, this implies that the coefficients can be learned as part of the neural network's weights, rendering such a sinc-activated network exceptionally suited for signal reconstruction in the L^2 space. In fact, thm. 3.10 shows that one does not need to restrict to sinc functions and that any activation that forms a Riesz basis will be optimal.

E DYNAMICAL EQUATIONS

Lorentz System: For the Lorenz system we take the parameters, $\sigma = 10$, $\rho = 28$ and $\beta = \frac{8}{3}$. The equations defining the system are:

$$\frac{dx}{dt} = \sigma(-x + y) \tag{28}$$

$$\frac{dy}{dt} = -xz + \rho x - y \tag{29}$$

$$\frac{dz}{dt} = -xy - \beta z \tag{30}$$

Van der Pol Oscillator: For the Van der Pol oscillator we take the parameter, $\mu = 1$. The equations defining the system are:

$$\frac{dx}{dt} = \mu(x - \frac{1}{3}x^3 - y) \tag{31}$$

$$\frac{dy}{dt} = \frac{1}{\mu}x \tag{32}$$

Chen System: For the Chen system we take the parameters, $\alpha = 5$, $\beta = -10$ and $\delta = -0.38$. The equations defining the system are:

$$\frac{dx}{dt} = \alpha x - yz \tag{33}$$

$$\frac{dy}{dt} = \beta y + xz \tag{34}$$

$$\frac{dz}{dt} = \delta z + \frac{xy}{3} \tag{35}$$

Rössler System: For the Rössler system we take the parameters, $a = 0.2$, $b = 0.2$ and $c = 5.7$. The equations defining the system are:

$$\frac{dx}{dt} = -(y + z) \tag{36}$$

$$\frac{dy}{dt} = x + ay \tag{37}$$

$$\frac{dz}{dt} = b + z(x - c) \tag{38}$$

Generalized Rank 14 Lorentz System: For the following system we take parameters $a = \frac{1}{\sqrt{2}}$, $R = 6.75r$ and $r = 45.92$. The equations defining the system are:

$$\frac{d\psi_{11}}{dt} = -a\left(\frac{7}{3}\psi_{13}\psi_{22} + \frac{17}{6}\psi_{13}\psi_{24} + \frac{1}{3}\psi_{31}\psi_{22} + \frac{9}{2}\psi_{33}\psi_{24}\right) - \sigma\frac{3}{2}\psi_{11} + \sigma a\frac{2}{3}\theta_{11} \tag{39}$$

$$\frac{d\psi_{13}}{dt} = a\left(-\frac{9}{19}\psi_{11}\psi_{22} + \frac{33}{38}\psi_{11}\psi_{24} + \frac{2}{19}\psi_{31}\psi_{22} - \frac{125}{38}\psi_{31}\psi_{24}\right) - \sigma\frac{19}{2}\psi_{13} + \sigma a\frac{2}{19}\theta_{13} \tag{40}$$

$$\frac{d\psi_{22}}{dt} = a\left(\frac{4}{3}\psi_{11}\psi_{13} - \frac{2}{3}\psi_{11}\psi_{31} - \frac{4}{3}\psi_{13}\psi_{31}\right) - 6\sigma\psi_{22} + \frac{1}{3}\sigma a\theta_{22} \tag{41}$$

$$\frac{d\psi_{31}}{dt} = a\left(\frac{9}{11}\psi_{11}\psi_{22} + \frac{14}{11}\psi_{13}\psi_{22} + \frac{85}{22}\psi_{13}\psi_{24}\right) - \frac{11}{2}\sigma\psi_{31} + \frac{6}{11}\sigma a\theta_{31} \tag{42}$$

$$\frac{d\psi_{33}}{dt} = a\left(\frac{11}{6}\psi_{11}\psi_{24}\right) - \frac{27}{2}\sigma\psi_{33} + \frac{2}{9}\sigma a\theta_{33} \tag{43}$$

$$\frac{d\psi_{24}}{dt} = a\left(-\frac{2}{9}\psi_{11}\psi_{13} - \psi_{11}\psi_{33} + \frac{5}{9}\psi_{13}\psi_{31}\right) - 18\sigma\psi_{24} + \frac{1}{9}\sigma a\theta_{24} \tag{44}$$

$$\begin{aligned}\frac{d\theta_{11}}{dt} = a\Big(&\psi_{11}\theta_{02} + \psi_{13}\theta_{22} - \frac{1}{2}\psi_{13}\theta_{24} - \psi_{13}\theta_{02} + 2\psi_{13}\theta_{04} + \psi_{22}\theta_{13} + \psi_{22}\theta_{31} + \psi_{31}\theta_{22}\\ &+ \frac{3}{2}\psi_{33}\theta_{24} - \frac{1}{2}\psi_{24}\theta_{13} + \frac{3}{2}\psi_{24}\theta_{33}\Big) + Ra\psi_{11} - \frac{3}{2}\theta_{11}\end{aligned} \tag{45}$$

$$\begin{aligned}\frac{d\theta_{13}}{dt} = a\Big(&-\psi_{11}\theta_{22} + \frac{1}{2}\psi_{11}\theta_{24} - \psi_{11}\theta_{02} + 2\psi_{11}\theta_{04} - \psi_{22}\theta_{11} - 2\psi_{31}\theta_{22}\\ &+ \frac{5}{2}\psi_{31}\theta_{24} + \frac{1}{2}\psi_{24}\theta_{11} + \frac{5}{2}\psi_{24}\theta_{31}\Big) + Ra\psi_{13} - \frac{19}{2}\theta_{13}\end{aligned} \tag{46}$$

$$\frac{d\theta_{22}}{dt} = a\Big(\psi_{11}\theta_{13} - \psi_{11}\theta_{31} - \psi_{13}\theta_{11} + 2\psi_{13}\theta_{31} + 4\psi_{22}\theta_{04} - \psi_{33}\theta_{11} + 2\psi_{24}\theta_{02}\Big) + 2Ra\psi_{22} - 6\theta_{22} \tag{47}$$

$$\begin{aligned}\frac{d\theta_{31}}{dt} = a\Big(&\psi_{11}\theta_{22} - 2\psi_{13}\theta_{22} + \frac{5}{2}\psi_{13}\theta_{24} - \psi_{22}\theta_{11} + 2\psi_{22}\theta_{13} + 4\psi_{31}\theta_{02} - 4\psi_{33}\theta_{02}\\ &+ 8\psi_{33}\theta_{04} - \frac{5}{2}\psi_{24}\theta_{13}\Big) + 3Ra\psi_{31} - \frac{11}{2}\theta_{31}\end{aligned} \tag{48}$$

$$\frac{d\theta_{33}}{dt} = a\Big(\frac{3}{2}\psi_{11}\theta_{24} - 4\psi_{31}\theta_{02} + 8\psi_{31}\theta_{04} - \frac{3}{2}\psi_{24}\theta_{11}\Big) + 3Ra\psi_{33} - \frac{27}{2}\theta_{33} \tag{49}$$

$$\begin{aligned}\frac{d\theta_{24}}{dt} = a\Big(&\frac{1}{2}\psi_{11}\theta_{13} - \frac{3}{2}\psi_{11}\theta_{33} + \frac{1}{2}\psi_{13}\theta_{11} - \frac{5}{2}\psi_{13}\theta_{31} - 2\psi_{22}\theta_{02}\\ &- \frac{5}{2}\psi_{31}\theta_{13} - \frac{3}{2}\psi_{33}\theta_{11}\Big) + 2Ra\psi_{24} - 18\theta_{24}\end{aligned} \tag{50}$$

$$\begin{aligned}\frac{d\theta_{02}}{dt} = a\Big(&-\frac{1}{2}\psi_{11}\theta_{11} + \frac{1}{2}\psi_{11}\theta_{11} + \frac{1}{2}\psi_{11}\theta_{13} + \frac{1}{2}\psi_{13}\theta_{11} + \psi_{22}\theta_{24}\\ &- \frac{3}{2}\psi_{31}\theta_{31} + \frac{3}{2}\psi_{31}\theta_{33} + \frac{3}{2}\psi_{33}\theta_{31} + \psi_{24}\theta_{24}\Big) - 4\theta_{02}\end{aligned} \tag{51}$$

$$\frac{d\theta_{04}}{dt} = -a\Big(\psi_{11}\theta_{13} + \psi_{13}\theta_{11} + 2\psi_{22}\theta_{22} + 4\psi_{31}\theta_{33} + 4\psi_{33}\theta_{31}\Big) - 16\theta_{04} \tag{52}$$

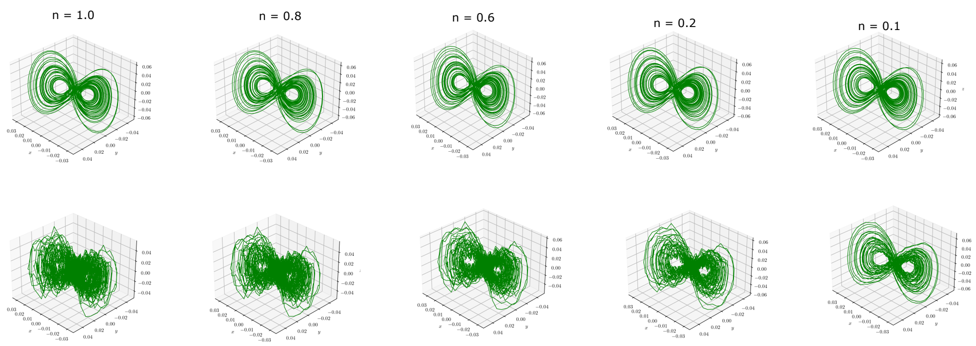

Figure 6: Robust recovery of dynamical systems from partial observations (Lorenz system). *Top row:* coordinate network. *Bottom row:* classical method.

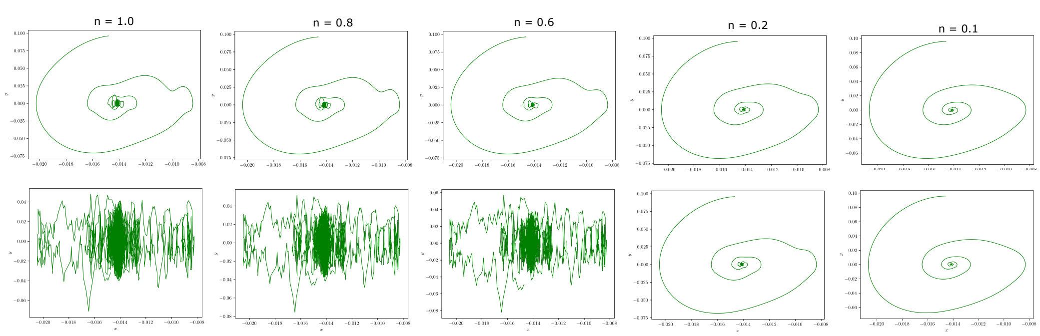

Figure 7: Robust recovery of dynamical systems from partial observations (Duffing system). *Top row:* coordinate network. *Bottom row:* classical method.

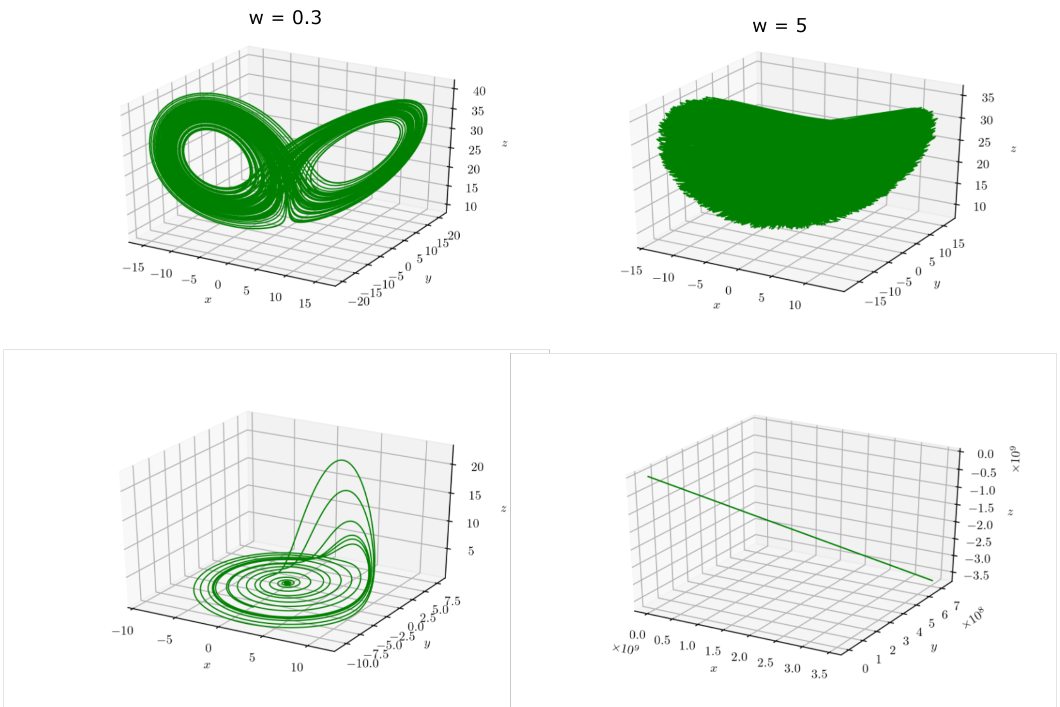

Figure 8: The top row and the bottom row depicts the SINDy reconstructions obtained for the Lorenz system and the Rossler system, respectively, using coordinate networks. As ω is increased in the sinc function, the coordinate network allows more higher frequencies to be captured, resulting in noisy reconstructions.