442.022 - Nonlinear Signal Processing

Assignment 3

Author: Sophie Steger, Felix Stollberger
Date: Graz, August 12, 2021

Contents

1 Nonlinear map

Assignment 3:

Problem 1

Consider the nonlinear map

$$x_{n+1} = x_n e^{-r(1-x_n)} \tag{1}$$

and perform the following analyses.

Task a) (10 Points, ★) Identify all fixed points of the system and perform a local stability analysis.

- *Fixed points*:

$$x_{n+1} = x_n$$

$$\Longrightarrow \quad \cancel{x} = \cancel{x} e^{-r(1-x)} \qquad \Big| : x \, , \quad x \neq 0$$

$$1 = e^{-r(1-x)} \qquad \Big| \ln(\cdot)$$

$$0 = -r(1-x) \qquad \Big| : (-r) \, , \quad r \neq 0$$

$$0 = 1 - x$$

$$\boxed{x_1^* = 1} \qquad \boxed{x_2^* = 0}$$

- *Stability analysis*:

$$f'(x) := \frac{\partial f(x)}{\partial x} = \frac{\partial}{\partial x}\left(x e^{-r(1-x)}\right)$$

$$= e^{-r(1-x)} + x \cdot e^{-r(1-x)} \cdot r$$

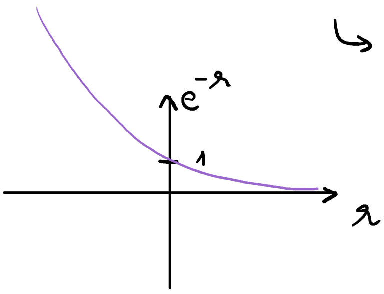
$$= e^{-r(1-x)}(1 + rx)$$

▷ $x_1^* = 1$: $f'(x_1^*) = (1 + r)$ ↪ $-1 < 1 + r < 1$

$$\begin{cases} \text{stable attractor}, & -2 < r < 0 \\ \text{unstable repellor}, & \text{else} \end{cases}$$

▷ $x_2^* = 0$: $f'(x_2^*) = e^{-r}$ ↪ $-1 < e^{-r} < 1$

$$\begin{cases} \text{stable attractor}, & r > 0 \\ \text{unstable}, & \text{else} \end{cases}$$

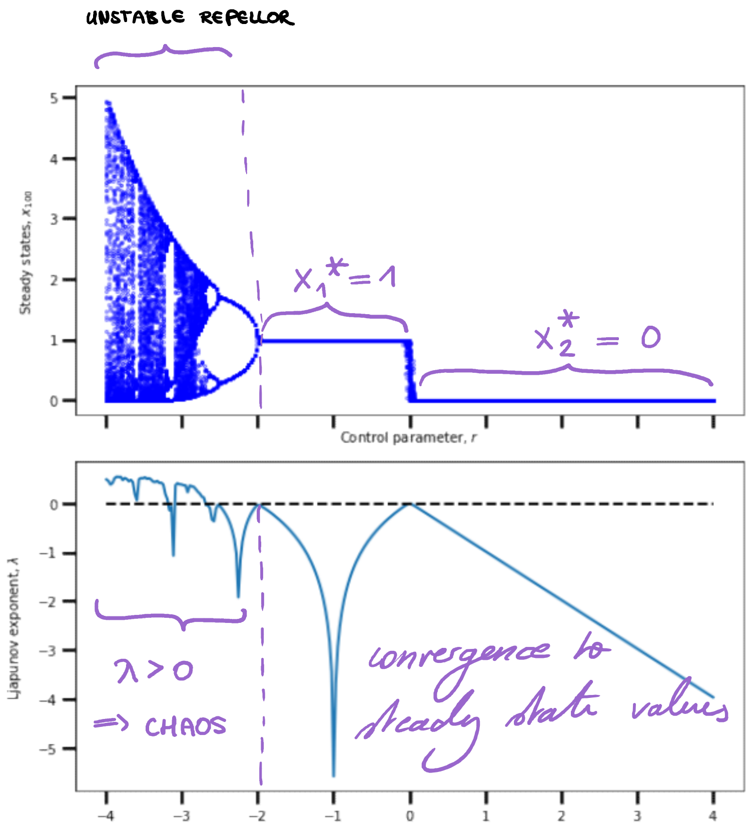


Task b) (5 Points, ★) Empirically find the range of initial points x_0 and control parameters r for which the system converges to a fixed point or exhibits chaotic behavior. For these interesting configurations, generate a plot that shows steady-state amplitudes (bifurcation diagram) and Ljapunov exponents as a function of the control parameter r. Explain what you can see in these plots.

$r \backslash x_0$	$(-\infty, 0)$	$(0, \infty)$
$(-\infty, -2)$	OVERFLOW	UNSTABLE REPELLOR
$(-2, 0)$	OVERFLOW	STABLE ATTRACTOR

$r \backslash x_0$	$(-\infty, 1)$	$(1, \infty)$
$(0, \infty)$	STABLE ATTRACTOR	OVERFLOW

here: $x_0 \sim \mathcal{U}(0,1)$



1.0.1 Task b: Discussion

The parameters x_0 and r for which the system converges or exhibits chaotic behavior were found empirically. The region of parameter pairs (x_0, r) are summarized on the previous page. Additionally, a bifurcation diagram and a plot of the Ljapunov exponents is generated with uniform distribution of initial states $x_0 \sim \mathcal{U}(0, 1)$ where all interesting regions are present and no overflow occurs.

For a control parameter value $r < -2$, we know from the initial stability analysis that the system behaves like an unstable repellor. In the beginning, the bifurcation diagram shows two different steady state values. In this region, the Ljapunov exponents are negative, indicating non-chaotic behavior. For r further decreasing, the Ljapunov exponent exceeds zero - i.e. the system is divergent. In between the chaotic regime, there are some islands of stability where the Ljapunov exponent is negative. Interestingly, for a certain value of control parameter r, the system jumps between three steady state values. We refer to the main theorem of the paper [2] implying chaotic behavior around this point, which is also visible from the Ljapunov plot.
For the value range $-2 < r < 0$, the system is a stable attractor towards the first fixed point $x_1^* = 1$. Similarly, for values $r > 0$ the system converges to the second fixed point $x_1^* = 0$. Decreasing Ljapunov exponents indicate a faster convergence to the fixed points.

2 Duffing equation

Problem 2

Consider the Duffing equation

$$\ddot{x} + \delta\dot{x} - \alpha x + x^3 = \beta\cos(\omega t) \tag{2}$$

where $\alpha, \beta, \delta, \omega \in \mathbb{R}_+$, and perform the following analyses.

Task a) (10 Points) Transform the given system description into a system of first order ordinary differential equations.

$$\left.\begin{aligned} x_1 &= x \\ x_2 &= \dot{x} \\ x_3 &= \omega t \end{aligned}\right\} \qquad \begin{aligned} &(\mathrm{I}) & \dot{x}_1 &= x_2 \\ &(\mathrm{II}) & \dot{x}_2 &= -\delta x_2 + \alpha x_1 - x_1^3 + \beta\cos(\omega t) \\ &(\mathrm{III}) & \dot{x}_3 &= \omega \end{aligned}$$

$$\underline{x} = [x_1 \quad x_2 \quad x_3]^T \quad \triangleright \quad \dot{\underline{x}} = F(\underline{x}) = \begin{bmatrix} x_2 \\ \alpha x_1 - x_1^3 - \delta x_2 + \beta\cos(\omega t) \\ \omega \end{bmatrix}$$

Task b) (20 Points) For $\beta = 0$ (autonomous case) identify all fixed points of the system and perform a local stability analysis.

$$\boxed{\ddot{x} + \delta\dot{x} - \alpha x + x^3 = 0}$$

$$\underline{x} = [x_1 \quad x_2]^T \qquad F(\underline{x}) = \begin{bmatrix} x_2 \\ \alpha x_1 - x_1^3 - \delta x_2 \end{bmatrix}$$

- Fixed points: $F(\underline{x}) \overset{!}{=} \underline{0}$

$$\begin{bmatrix} x_2 \\ \alpha x_1 - x_1^3 - \delta x_2 \end{bmatrix} \overset{!}{=} \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Longrightarrow \quad x_2^* = 0$$

$$\alpha x_1 - x_1^3 - \delta \cdot 0 \overset{!}{=} 0$$

$$x_1(\alpha - x_1^2) = 0 \qquad \begin{aligned} &\rightarrow x_{1,1}^* = 0 \\ &\rightarrow x_{1,2}^* = +\sqrt{\alpha} \\ &\rightarrow x_{1,3}^* = -\sqrt{\alpha} \end{aligned}$$

$$\boxed{\begin{aligned} &\bullet\ \underline{x}_1^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ &\bullet\ \underline{x}_2^* = \begin{bmatrix} \sqrt{\alpha} \\ 0 \end{bmatrix} \\ &\bullet\ \underline{x}_3^* = \begin{bmatrix} -\sqrt{\alpha} \\ 0 \end{bmatrix} \end{aligned}}$$

- Stability Analysis: Compute eigenvalues of Jacobian matrix $J_{F(\underline{x})}\Big|_{\underline{x}=\underline{x}^*}$

$$J_{F(\underline{x})} = \begin{bmatrix} \partial f_1/\partial x_1 & \partial f_1/\partial x_2 \\ \partial f_2/\partial x_1 & \partial f_2/\partial x_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \alpha - 3x_1^2 & -\delta \end{bmatrix}$$

EIGENVALUES: $\left| J_{F(\underline{x}^*)} - \lambda I \right| = \left| \begin{bmatrix} -\lambda & 1 \\ \alpha - 3x_1^2 & -\delta-\lambda \end{bmatrix} \right|$

① $x_{1,1}^* = 0$: $\left| J_{F(\underline{x}_1^*)} - \lambda I \right| = \left| \begin{bmatrix} -\lambda & 1 \\ \alpha & -\delta-\lambda \end{bmatrix} \right|$

$$= \lambda(\lambda+\delta) - \alpha$$
$$= \lambda^2 + \lambda\delta - \alpha \overset{!}{=} 0$$

- $$\lambda_1 = \frac{-\delta \pm \sqrt{\delta^2 + 4\alpha}}{2}$$

$$\lambda_1^+ = \frac{-\delta + \sqrt{\delta^2+4\alpha}}{2} > 0$$

$$\lambda_1^- = \frac{-\delta - \sqrt{\delta^2+4\alpha}}{2} < 0$$

$\boxed{\text{SADDLE POINT}}$

② $x_{1,2}^* = \sqrt{\alpha}$

③ $x_{1,3}^* = -\sqrt{\alpha}$

$\left| J_{F(\underline{x}_2^*)} - \lambda I \right| = \left| \begin{bmatrix} -\lambda & 1 \\ \alpha - 3\alpha & -\delta-\lambda \end{bmatrix} \right|$

$$= \lambda(\lambda+\delta) + 2\alpha$$
$$= \lambda^2 + \lambda\delta + 2\alpha$$

- $$\lambda_{2/3} = \frac{-\delta \pm \sqrt{\delta^2 - 8\alpha}}{2}$$

- $\lambda_{2/3}^+ = \dfrac{-\delta + \sqrt{\delta^2 - 8\alpha}}{2} < 0$ $\boxed{\text{ATTRACTOR}}$
- $\lambda_{2/3}^- = \dfrac{-\delta - \sqrt{\delta^2 - 8\alpha}}{2}$

$\delta^2 < 8\alpha \Rightarrow \sqrt{\delta^2 - 8\alpha} = j \cdot \sqrt{8\alpha - \delta^2}$

$\Rightarrow \operatorname{Re}\{\lambda_{2,3}^-\} < 0$ $\boxed{\text{STABLE SPIRAL}}$

$\delta^2 > 8\alpha \Rightarrow \operatorname{Re}\{\lambda_{2,3}\} < 0$ $\boxed{\text{ATTRACTOR}}$

$\delta^2 = 8\alpha \Rightarrow \lambda_{2,3}^+ = \lambda_{2,3}^- < 0$

$\boxed{\text{DEGENERATE NODE}}$

SPECIAL CASES :

for $\alpha = 0$, $\beta \neq 0$

$\Longrightarrow \lambda^{+}_{1/2/3} = 0$ NO JUDGEMENT POSSIBLE

$\lambda^{-}_{1/2/3} < 0$

for $\beta = 0$, $\alpha \neq 0$

$\lambda_1^{+} > 0$, $\lambda_1^{-} < 0$ SADDLE POINT

$\mathrm{Re}\{\lambda^{+/-}_{2/3}\} = 0$ NO JUDGEMENT POSSIBLE

$\alpha = 0, \beta = 0 \Longrightarrow \lambda^{+/-}_{1/2/3} = 0$ NO JUDGEMENT POSSIBLE

Task c) (20 Points) For $\omega = 1$ and the following sets of parameters characterize the systems behavior.

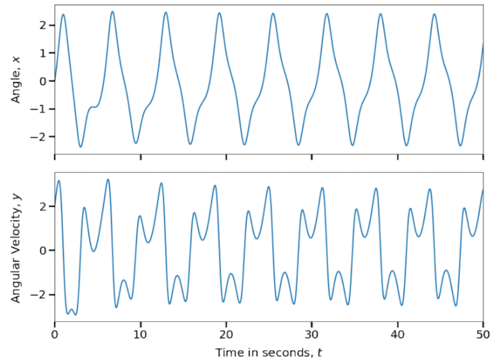
(i) $\alpha = 0$, $\beta = 5$ and $\delta = 0.7$.

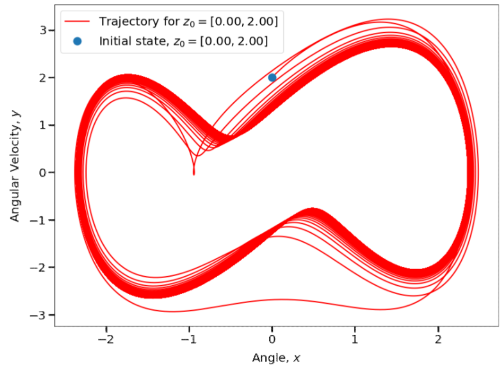
(ii) $\alpha = 0$, $\beta = 20$ and $\delta = 0.17$.

(iii) $\alpha = 1$, $\beta = 0.4$ and $\delta = 0.25$.

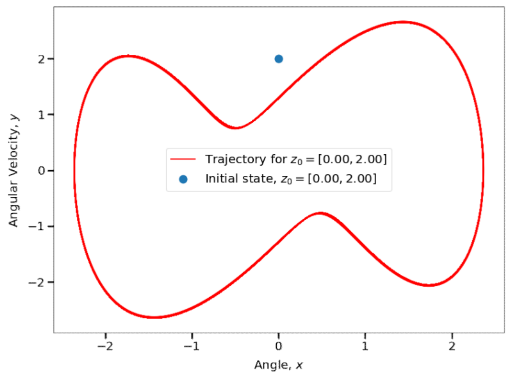
For each case provide an illustrative plot of a trajectory in time domain and in the phase plane and explain your plots.

(i)

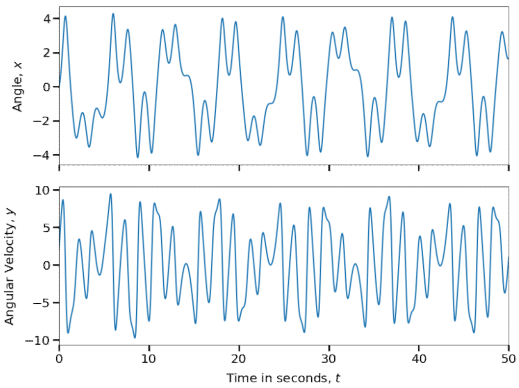


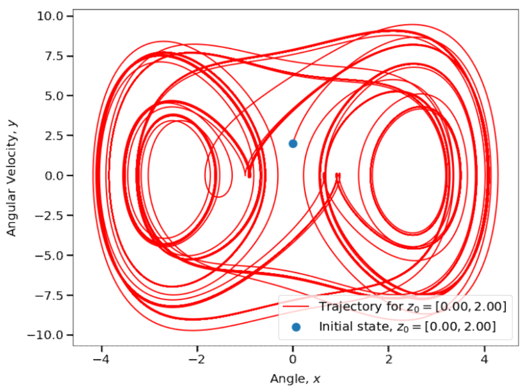


Starting at sample 20000

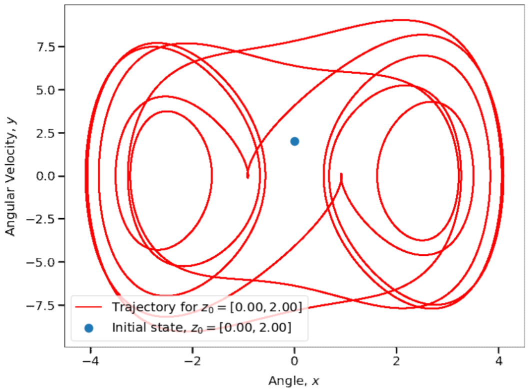


(ii)

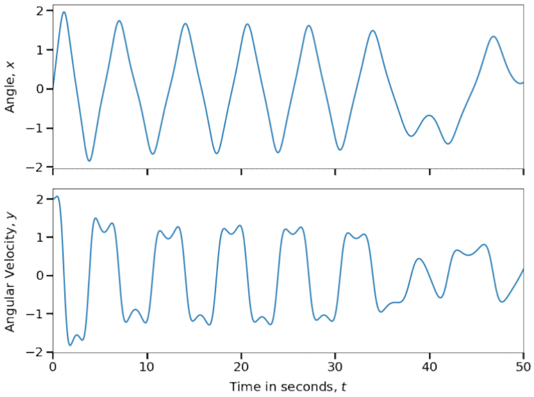


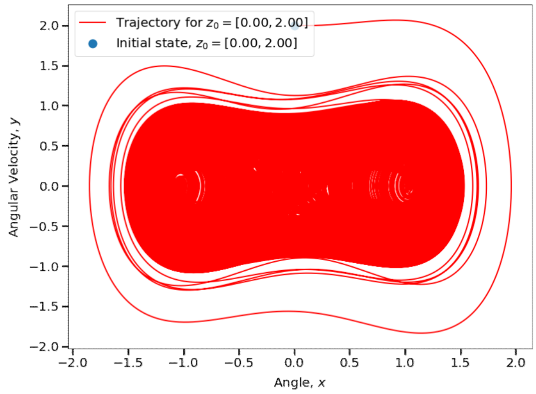


Starting at sample 20000

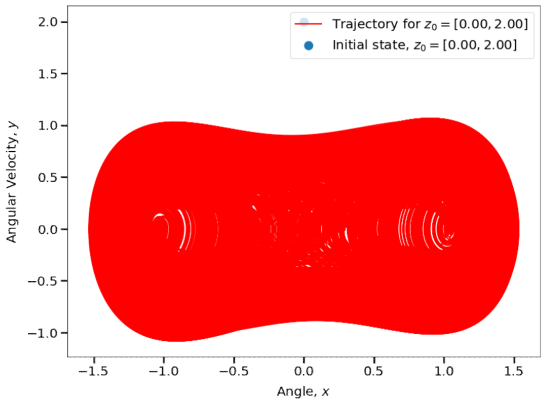


(iii)





Starting at sample 20000



2.0.1 Task c: Discussion

In the first case ($\alpha = 0, \beta = 5, \delta = 0.7$), both the amplitude and phase plane plot show periodic behavior. The periodicity is even more obvious when discarding the first 20.000 samples of the trajectory. Thus, the initial transient behavior of the system is ignored.
For the second case ($\alpha = 0, \beta = 20, \delta = 0.17$), again the first 20.000 samples are discarded. In the phase space we can observe that a certain trajectory is repeated over and over again indicating periodic behavior.
The last case ($\alpha = 1, \beta = 0.4, \delta = 0.25$) displays some interesting behavior of the system. At first, the time plot of the angle and angular velocity seem periodic. However, the phase space clearly does not show any periodicity. As no points in the phase space are visited again periodically, a certain region of the state space is filled by the trajectory.

Task d) (15 Points, ★) For the three system configurations above, generate a Poincaré section of your example trajectories. To do so, sample the trajectories periodically, e.g. at points $t = n \cdot T_s$, $n \in \mathbb{N}$, and plot your sampled points in the phase plane. Describe what you see.

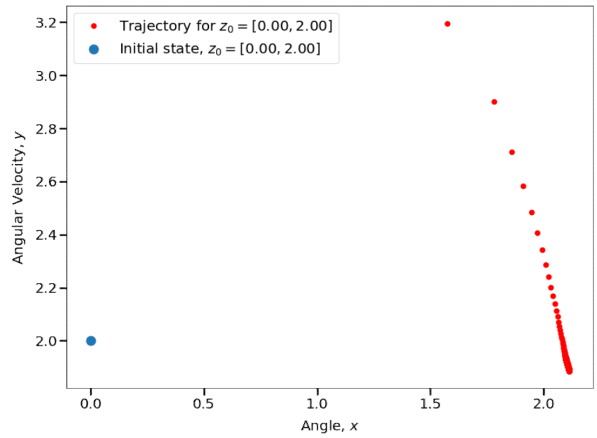
POINCARÉ MAP

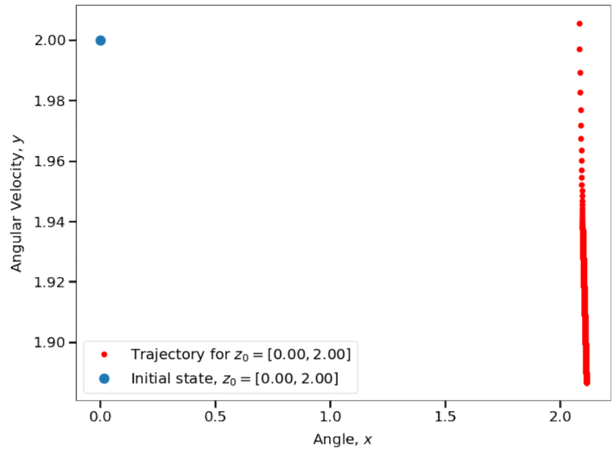
↳ we sample at: $t = n \cdot T_s \overset{!}{=} k \cdot 2\pi$

with $k \in \mathbb{Z}$

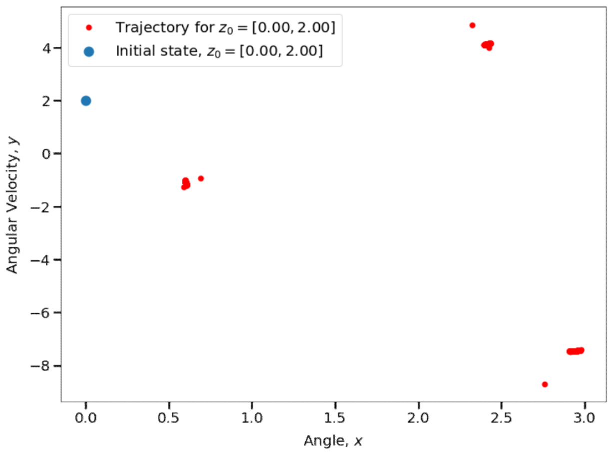
starting at sample 20000

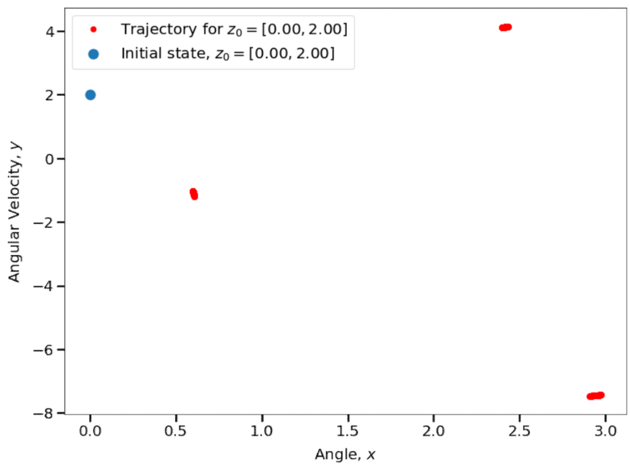
(i)



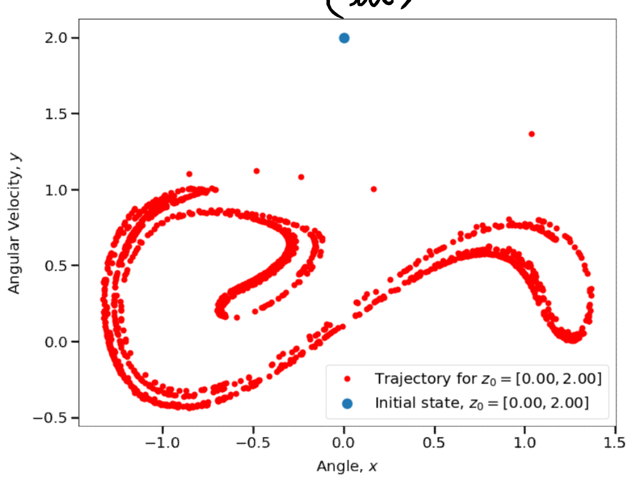


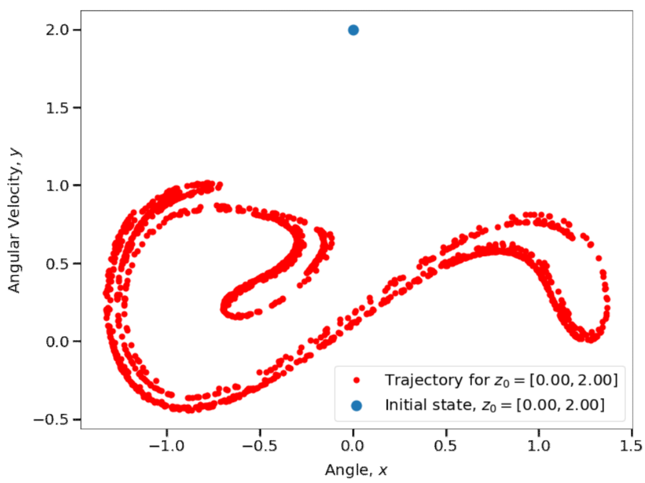
(ii)





(iii)





2.0.2 Task d: Discussion

A Poincaré section is a projection of the phase space trajectory onto tranversal hyperplane with a reduced dimensionality. In our case, the required periodic sampling of the trajectory creates a Poincaré section perpendicular to the time axis in the phase space.

As already discussed in the previous task, the transient behaviour of the system could be also observed in the Poincaré section, as it is directly derived from the trajectory of the system. Therefore, we have again included plots with-, and without the first 20.000 samples, eliminating most of the transient effects.

Starting with the first parameter set ($\alpha = 0, \beta = 5, \delta = 0.7$), a convergence to a single point could be observed in the Poincaré section. This is typical for periodic systems, and verifies again that the first system is indeed periodic. Nevertheless, the convergence happens only asymptotically, and causes the similar appearance of both figures with-, and without the first 20.000 samples. Note that the scaling of the axis changes by a factor of ten between the first plot (with transient parts), and the second figure.

The second parameter set causes also a similar behavior, now with a convergence to three points in the Poincaré section. The transient effects are not as observable as in the previous case because the scaling of the axes is inherently coarser due to their distribution in the Poincaré section.

The last case delivers definitely the most interesting result, as a whole new trajectory emerges on the Poincaré section. This is a clear indicator for chaotic behaviour, as the oscillator changes between different states in the phase space when observing it at equally spaced points in time. In combination with the phase space trajectory from the previous task, this confirms again that the Duffing oscillator is chaotic for the given parameter set ($\alpha = 1, \beta = 0.4, \delta = 0.25$).

3 WaveNet

3.1 Problem 3

This problem dealt with vocal synthesis using the WaveNet architecture. Therefore, we had to record ourself for several seconds while producing one of the five vowels ('a', 'e', 'i', 'o', 'u'). Those recordings served as training data for the WaveNet.

3.1.1 Implementing the residual block

The given repository contained already a basic implementation of the WaveNet, with the exception of the residual blocks, which had to be implemented by us.

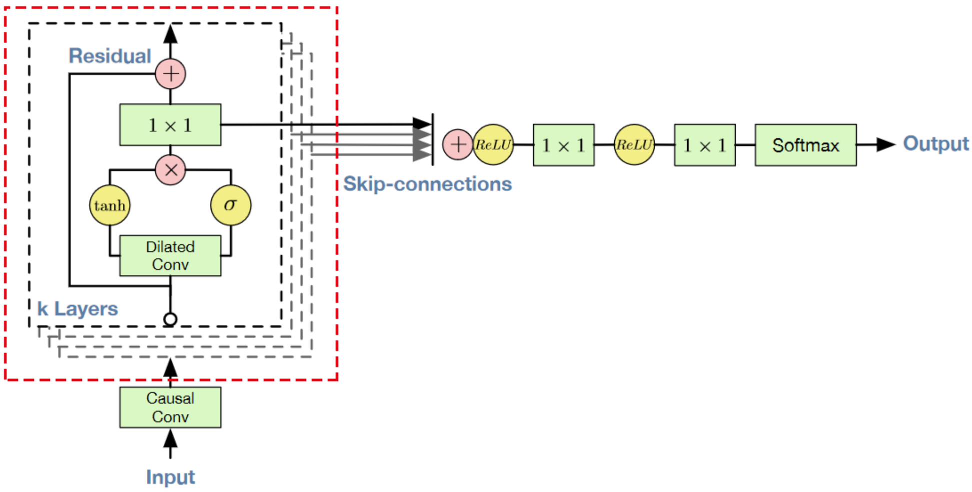


Figure 3.1: Suggested architecture for the residual block (red rectangle) by van den Oord et al. [3].

The residual block is a crucial part of the WaveNet architecture, as described by van den Oord et al. [3]. Figure 3.1 shows a a basic implementation, as given by the author, which served us as a reference for our own model.

The core of a residual block is a dilated convolution, where the dilation factor is fixed, but doubles in each hidden layer in the network. Those dilated convolutions allow to construct a network with a much lager field of view compared to normal convolutions, while maintaining similar complexity. The basic idea behind this dilation is to skip input variables with a certain step, which allows to apply the filter over an area lager than its convolution kernel [3]. The WaveNet features a stack of those dilated convolutions, where the output of each layer serves as an input for the next layer, increasing the receptive field to several thousand samples. In the considered speech generation application, this equals to about 0.25 to 0,5 s audio, which is sufficient to gain enough information of the waveform for successful prediction-, and generation.

The computation of the residual in the block is further based on a parallel evaluation of the output of the dilated convolution using a tanh-, as well as sigmoidal activation function, followed by multiplication of both, and a 1×1 convolution. This 1×1 convolution is actually equivalent to applying a fully connected layer on the result of the multiplication.

The implementation of this residual block was therefore relatively straightforward, and is here shown on the forward-pass of the block.

```
pad_length = round(self.dilation*(self.kernel_size -1))

x_pad = F.pad(x,(pad_length , 0))

z = torch.tanh(self.feat_conv(x_pad)) * torch.sigmoid(self.gate_conv(x_pad))

if fill_buffer:
    self.conv_buffer = z[:, :, -self.buffer_size:]
    assert self.conv_buffer.shape == (1, self.num_channels , self.buffer_size)

z = self.mix_conv(z)
skip = z
residual= z + x
```

Listing 3.1: Forward pass of the residual block

Proper padding of the input signal was necessary to ensure that the output of the dilated convolution had the same length as the bypassed original input signal, which was added to the output of the 1×1 convolution for generating the residual.

Based on this architecture, our WaveNet was trained on the recorded audio signals for 850 epochs. The optimal hyperparameter set was found by manual optimization, but was limited by the training time of around 10 min on a hexacore i7 architecture. The final WaveNet had 13 residual blocks per WaveNet cell, and a kernel size of 64 in the residual convolutions. The dilation factor was set to two, and the number of hidden neurons in the 1×1 convolution layer to 128. With this configuration, a minimal training loss of $6 \cdot 10^{-4}$ was archived, for training on an approximately 6 s long sequence of the vowel 'O', spoken by a male speaker.The receptive field covered 16384 samples (approx. 1s) in this configuration. Figure 3.2 shows a prediction of the WaveNet of the waveform of an 'O', compared to the real, recorded sequence. Excellent accordance between both wave forms could be observed, indicating good modeling performance on the trained vowel.

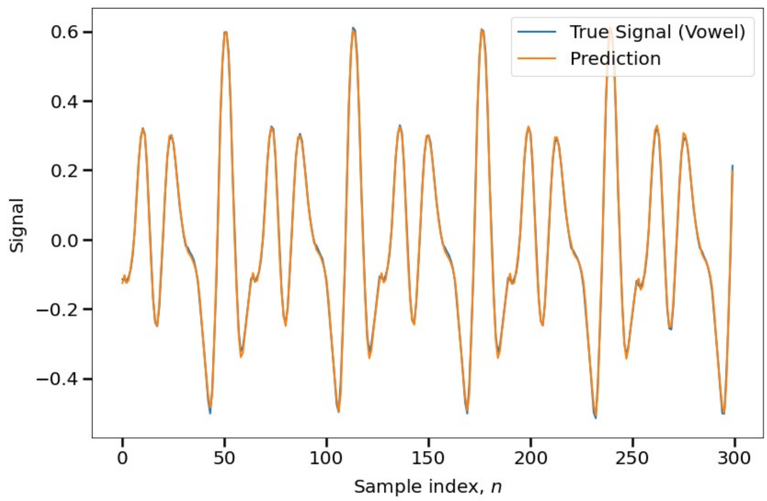
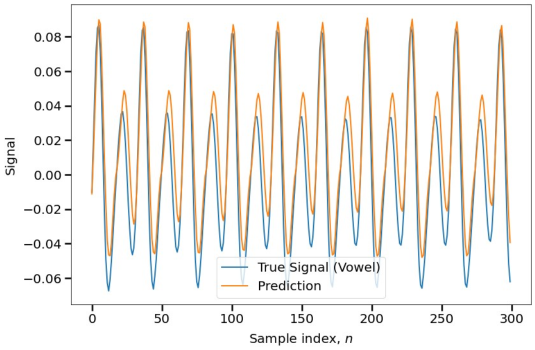


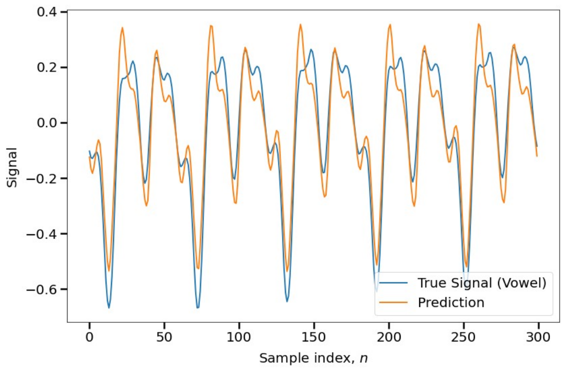
Figure 3.2: Comparison of the predicted, and the real waveform of an 'O'

The model could be also used to predict the wave form of vowels of a different speaker which where not used for training. The following Figure 3.3(a) shows for example a prediction of an

'O' of a female speaker, while the model was trained on a male voice. Additionally, different vowels could be predicted, which is shown in Figure 3.3(a) where a 'U' is predicted based on the WaveNet, trained on an 'O'. Although the accuracy is clearly limited compared to the original case, the general waveform is still conserved in both examples.



(a) Prediction of a female voice



(b) Prediction of a different vowel

Figure 3.3: Prediction of different sequences from the training sequence

3.1.2 Generating synthetic vowels

The generation of the synthetic vowel was generally very similar to the prediction done in the previous task, with the main difference that the samples where now sequentially read from the convolution buffer, which is initialized with just ones.

The generated vowels differed drastically in quality, ranging from nearly inaudible chaotic noises to clearly recognizable vowels. WaveNets with a kernel size of 64 outperformed networks with smaller kernel sizes, longer training increased also the quality of the generated sequence. Our final version was able to produce a clear, steady vowel, but with sill slightly varying amplitude. Nevertheless, this quality was maintained for generated sequences with lengths up to 20.000 samples, corresponding to around 2,3 s of raw audio. Figure 3.4 shows a comparison of our generated vowel and the recorded sequence, used for training. A slight amplitude deviation could be observed, but the much more relevant waveform was correctly covered by the WaveNet, and sounded like the actual vowel.

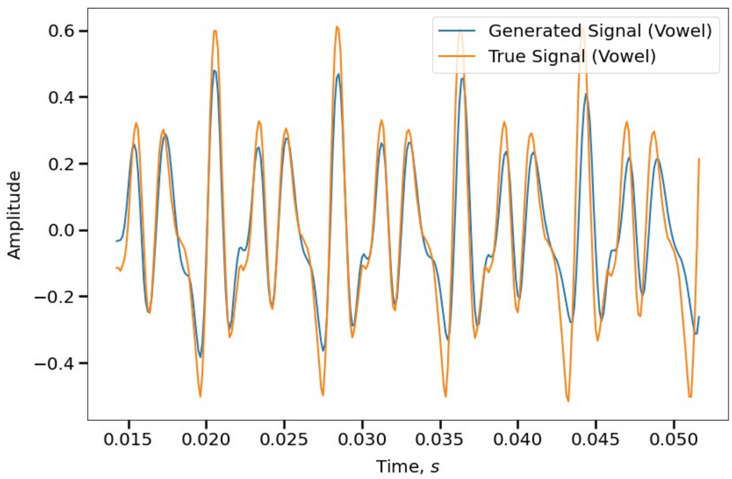


Figure 3.4: Generated waveform of a synthetic 'O'

Figure 3.5 shows also a detail of the generated waveform 19.000 samples (resp. 2,25 s) after the initialisation. Little to no deviations from the original waveform could be observed, indicating a sufficient long-term stability to generate several seconds of audio. [1] A sample sequence of this waveform (repeated three times) can be found under the following Link.

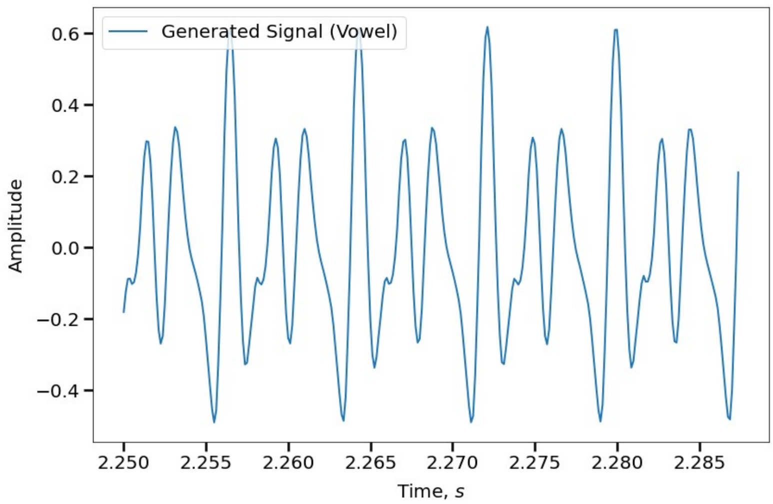


Figure 3.5: Generated waveform of a synthetic 'O' after two seconds, indicating stability of the WaveNet

The output of the WaveNet is a chaotic signal, which is actually the biggest advantage compared to conventional speech synthesizers, using periodic wave forms. The slight variations between each period ensure a more natural sounding speech, which makes the WaveNet one of the best currently available speech synthesizer. [1]

3.1.3 Vision for the future of speech synthesis

Siri, Alexa, Google assistant, ... not only speech recognition plays an important role but also speech synthesizers are a part of our everyday life. There is no need to look at the phone when using navigation while driving - a computer generated voice is describing the route in detail. Speech synthesis is not just a fun gimmick, it can immensely facilitate tasks. We think that for us humans it is not only important to simply understand the content of a spoken sentence, but also the way a sentence is spoken - intonation, timbre, and emotions are almost equally as important (e.g. compare waveforms of wavenets with vocoders[2]). An important application of speech synthesizers are text-to-speech systems which make it possible for people who suffer from low vision to use smart phones, computers or even read text from print (e.g. labels of medication). Modern speech synthesizers like wavenets are able to produce more natural sounding words instead of simply concatenating phonemes. In theory, it is possible to train a network with own voice data to generate individual sounding voice. Also, we could observe that a recent update of the app TikTok included a new text-to-speech feature. This allows creators to add text description to videos which is additionally read out loud by a speech synthesizer, making the platform more inclusive for deaf and visually impaired people. Overall, for the English language the feature works surprisingly well and is used abundantly[3]. Thus, we can observe that speech synthesizers are important, many applications already exist and it is an important part of everyday life for many of us already today - and for some people an essential tool to manage life more independently.

1 Longer sequences where not tested due to the long generation time (approx. 5 minutes for 2 s)
2 Parametric speech synthesizers used to reduce bandwidth when transmitting voice signals
3 Although - as expected - it can produce quite funny sounding sentences for other languages, e.g. German

List of Figures

Bibliography

[1] Deepmind (2021). Wavenet: A generative model for raw audio. Available at `https://deepmind.com/blog/article/wavenet-generative-model-raw-audio` (2021/08/12).

[2] Li, T.-Y. and Yorke, J. A. (1975). Period three implies chaos. *The American Mathematical Monthly*, 82(10):985–992.

[3] van den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., and Kavukcuoglu, K. (2016). WaveNet: A Generative Model for Raw Audio. pages 1–15.