

Bachelorarbeit  
zur Erlangung des akademischen Grades  
Bachelor of Science

Zachary Schellin  
376930

January 14, 2021

# Contents

<b>Contents</b>	<b>2</b>
<b>1 Introduction</b>	<b>3</b>
1.1 Objective of this thesis . . . . .	3
1.2 Thesis outline . . . . .	3
1.3 State of the art . . . . .	3
<b>2 The BGK Model</b>	<b>4</b>
2.1 Macroscopic features of Hydrodynamic and Rarefied Gas Flows in the SOD-Shock Tube . . . . .	5
<b>3 Deep learning</b>	<b>7</b>
3.1 Autoencoders . . . . .	9
3.1.1 Hyperparameters for the Fully Connected Autoencoder . . . . .	10
3.1.2 Hyperparameters for the Convolutional Autoencoder . . . . .	13
<b>4 Reduced Order Model</b>	<b>14</b>
4.1 Offline Phase . . . . .	15
4.1.1 Full order BGK model . . . . .	15
4.1.2 Reduced Basis by POD and Autoencoder . . . . .	15
4.2 Online Phase . . . . .	17
4.2.1 Reduced Order Model . . . . .	17
<b>5 Results</b>	<b>18</b>
5.1 Hydrodynamic Regime . . . . .	18
5.2 Rarefied Regime . . . . .	20
5.3 Discussion and Outlook . . . . .	20
<b>References</b>	<b>24</b>
5.4 Appendix A . . . . .	25

# 1 Introduction

The Bhatnagar-Gross-Krook equation (BGK) is a kinetic collision model of ionized and neutral gases valid for rarefied as well as other pressure regimes [1]. Generating data of such a flow field is essential for various industry and scientific applications[REF]. With the intention to reduce time and cost during the data generating process, experiments were substituted with computational fluid dynamics (CFD) computations. Consequently reduced-order models (ROMs) coupled to aforementioned computations were introduced to further the reduction of time and cost. The thriving field of artificial intelligence operates in model order reduction for data visualization/analysis since the 80's (Quelle?) and has now surfaced in fluid mechanics. This thesis will cover the use of artificial intelligence for model order reduction in fluid mechanics.

## 1.1 Objective of this thesis

Due to the non-linearity of transport problems in particular shock fronts, the construction of a robust ROM for those cases poses several challenges.

## 1.2 Thesis outline

- What is the BGK Model
- What is the SOD and BGK in SOD
- What is deep learning what are autoencoders
- hyperparameters for autoencoders
- What is a reduced order model
- offline phase
  - > my data in 1d bgk in sod shock tube FOM Data
  - > what is pod and Rb by POD
  - > RB by autoencoder
- online phase
  - > what is my ROM
  - > results - results by ROM for FC and for CONV
- variation of intrinsic variables
- Comparion to POD intrinsic variables number and quality

## 1.3 State of the art

State of the art model reduction of dynamical systems can be done via proper orthogonal decomposition (POD) which is an algorithm feeding on the idea of singular value decomposition (SVD)[2][3]. POD captures a low-rank representation on a linear manifold. So called POD modes, derived from SVD, describe the principle components of a problem which can be coupled within a Galerkin framework to produce an approximation of a lower dimension  $r$ .

$$f(x) \approx \tilde{f}(x) \quad \text{with } \tilde{f} = \sum_{k=1}^r a_k \psi_k(x) \quad \text{where } \psi_k \quad \text{are orthonormal functions. (1)}$$

Bernard et al. use POD-Galerkin with an additional population of their snapshot database via optimal transport for the proposed BGK equation, bisecting computational run time (cost) in conjunction with an approximation error of  $\sim 1\%$  in [4]. Artificial intelligence in the form of autoencoders replacing the POD within a Galerkin framework is evaluated against the POD performance by Kookjin et al. for advection-dominated problems[5] resulting in sub 0.1% errors. An additional time inter- and extrapolation is evaluated. Using machine learning/ deep learning for reduced order modeling in CFD is a novel approach although "the idea of autoencoders has been part of the historical landscape of neural networks for decades"[6, p.493]. Autoencoders, or more precisely learning internal representations by the delta rule (backpropagation) and the use of hidden units in a feed forward neural network architecture, premiered by Rumelhart et al. (1986) [7]. Through so called hierarchical training Ballard et al.(1987) introduce a strategy to train auto autoassociative networks (nowadays referred to as autoencoders), in a reasonable time promoting further development despite computational limitations [8]. The so called bottleneck of autoencoders yields a non-smooth and entangled representation thus being uninterpretable by practitioners[9] leading to developments in this field. Rifai et al. introduce the contractive autoencoder (CAE) for classification tasks (2011), with the aim to extract robust features which are insensitive to input variations orthogonal to the low-dimensional non-linear manifold by adding a penalty on the frobenius norm of the intrinsic variables with respect to the input, surpassing other classification algorithms [9]. Subsequent development emerges with the manifold tangent classifier (MTC) [10]. A local chart for each datapoint is obtained hence characterizing the manifold which in turn improves classification performance. On that basis a generative process for the CAE is developed. Through movements along the manifold with directions defined by the Jacobian of the bottleneck layer with respect to the input  $\vec{x}_m = JJ^T$ , sampling is realized [11].... Proper orthogonal decomposition (POD) and it's numerous variants like shifted-POD[?], POD-Galerkin[?], POD+I [?] to name only a few of them, try to solve this problem by.....

## 2 The BGK Model

This section covers the kinetic gas model, the BGK model, on which a model order reduction will be performed in the following. In addition the SOD-shocktube, on which the BGK model will be tested is discussed.

The BGK model was introduced by, and named after, physicists Prabhu L. Bhatnagar, Eugene P. Gross and Max Krook in 1954 [1]. It is an approximation of the standard Boltzmann equation. More precisely the r.h.s of the Boltzmann equation is approximated by the BGK operator [12]

$$\partial_t f + v \partial_x f = \frac{1}{\tau} (M_f - f). \quad (2)$$

It consists of the relaxation time  $\tau(x, t)$ , the Maxwellian distribution  $M_f$  and  $f(x, v, t)$  the probability of a gas particle having a microscopic velocity  $v$  in phase space  $(x, v, t)$ . The left side of the BGK model is a transport equation for  $f(x, v, t)$  with transport velocity  $v$ . In the Maxwellian distribution  $n(x, t)$  is the number of particles in space and time,  $R$  is the specific gas constant,  $T^0(x, t)$  is the temperature and  $u(x, t)$  is the macroscopic

velocity

$$M_f = \frac{n(x, t)}{\sqrt{2\pi R T^0(x, t)}} \exp\left(-\frac{(v - u(x, t))^2}{2R T^0(x, t)}\right). \quad (3)$$

Note that in this thesis the BGK model is discussed in one dimension. Hence the BGK model needs to be evaluated for three independent variables  $x$ ,  $v$  and  $t$  as seen above. Now what makes the BGK model especially attractive for model order reduction? The fruitfulness of performing a model order reduction on the BGK model becomes clearer when looking at its space and velocity discretization

$$\partial_t f_{j,k} = -(v_k)_1 D_x f|_{j,k}(t) + \frac{1}{\tau} (M_{f,j,k}(t) - f_{j,k}(t)). \quad (4)$$

Here a uniform grid is considered with  $x_j = j\Delta x$ ,  $j \in \mathbb{Z}$ ,  $v_k = k\Delta v$ ,  $k \in \mathbb{Z}$  and  $t^n = n\Delta t$ ,  $n \in \mathbb{N}$  on which  $f_{j,k} = f(x_j, v_k, t)$  and  $M_{f,j,k} = M_f(x_j, v_k, t)$  are evaluated at point  $(x_j, v_k)$  in a time instance  $t$ . For brevity  $D_x f|_{j,k}$  is the discrete space derivative at  $(x_j, v_k)$  [12]. Now the PDE in eq. (2) is broken down into a system of ODE's in time, for which every ODE is a linear advection equation with constant scalar speed  $v_k$  and a source term. To continue let's consider  $K$  to be the number of gridpoints in velocity and  $J$  to be the number of grid points in space. Then a total of  $KJ$  first order differential equations need to be evaluated in 1D. Obviously in three dimensions the system of ODE's inflates up to  $K^3 N^3$  first order differential equations eq. (4). This in turn drives the evaluation of the BGK model at the edge of intractability for dense meshes in 3D and all together stimulate the want for a reduced order model.

In order to evaluate  $M_f$  and therefore eq. (4) the macroscopic velocity  $u(x, t)$  is required. Other macroscopic quantities of the gas flow, namely the density  $\rho$ , the impulse  $\rho u$  and the energy  $E$  are the expected values or moments of  $f$  in velocity space and can be calculated with

$$\rho(x, t) = \int f \, dv, \quad (5)$$

$$\rho(x, t)u(x, t) = \int vf \, dv, \quad (6)$$

$$E(x, t) = \int \frac{1}{2}v^2 f \, dv. \quad (7)$$

As an illustrative example of the linkage between

## 2.1 Macroscopic features of Hydrodynamic and Rarefied Gas Flows in the SOD-Shock Tube

- intrinsic code variables for hydro and rare add  
pls

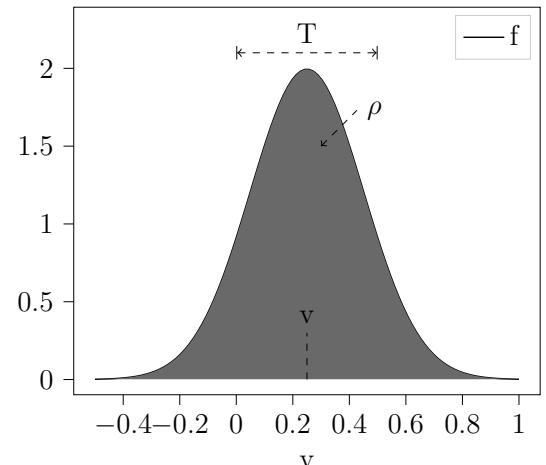


Figure 1: Illustration of the linkage between the macroscopic quantities of the gas flow and  $f$  the distribution function

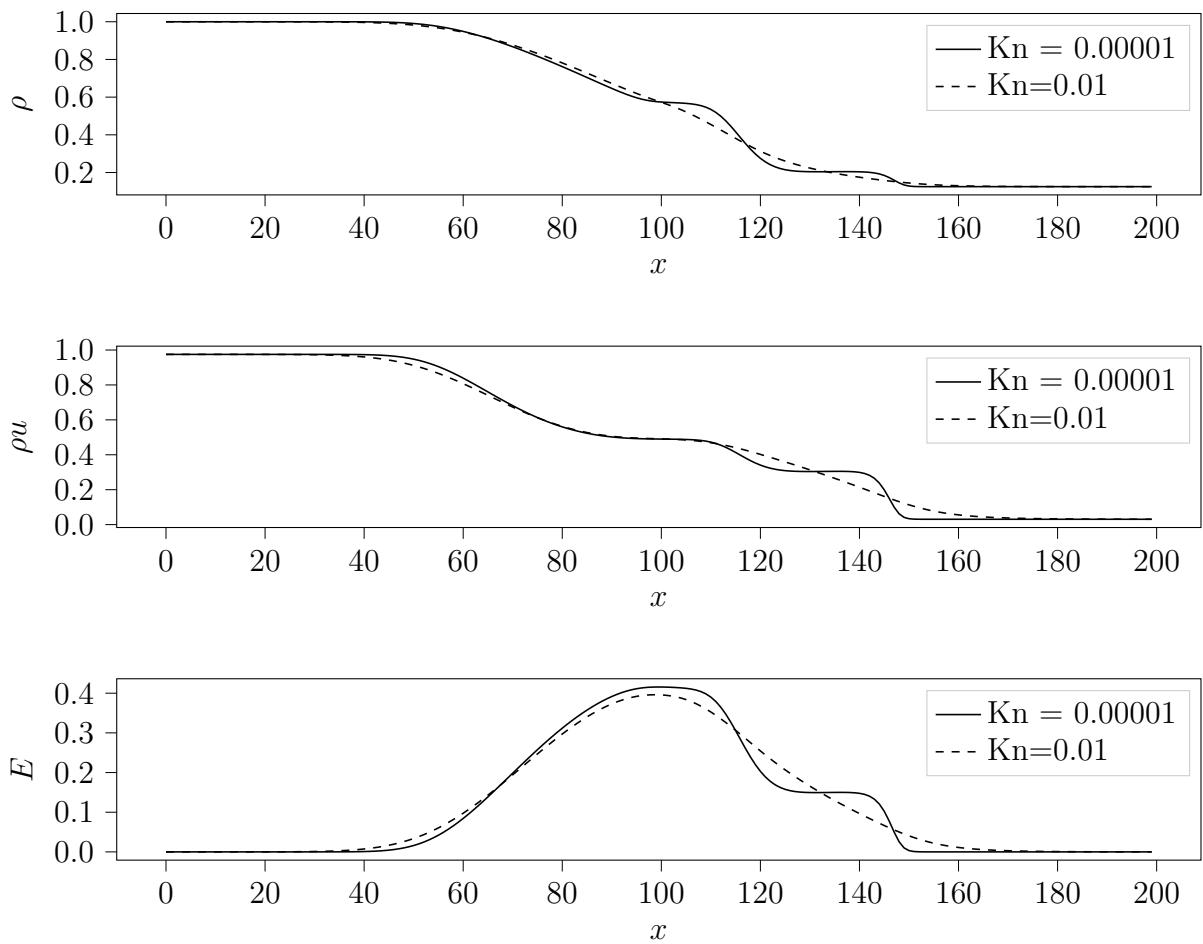


Figure 2: Macroscopic quantities  $\rho$ ,  $\rho u$  and  $E$  in the SOD schock tube at the last time stamp. The quantites are displayed for the hydrodynamic regime in - - lines and for the rarefied regime in - lines.

### 3 Deep learning

In this section deep learning with the focus on fully connected autoencoders and convolutional autoencoders will be introduced. Starting off with addressing important terminology whilst presenting the concept of autoencoders and continuing with the introduction of the fully connected and convolutional forward-pass, this section closes with ADAM [?] an update to the backpropagation algorithm as well as important training methods.

The term deep learning situated around the much broader field of artificial intelligence stems from the use of deep feed forward networks also called multi layer perceptrons (MLP) or feed forward neural networks. Deep recurrent neural networks (RNN) are also used in this field but won't be covered in this thesis. In contrast to RNNs, information flows forward through these networks, which explains the name feed forward. In the following I will use the abbreviation MLP when talking about the aforementioned algorithm. Network refers to the typical composition of many different functions.

The task of any MLP is to approximate a function  $f^*(\mathbf{x}; \Theta) \approx f(\mathbf{x})$  through learning the values of the parameters  $\Theta$ . As mentioned before  $f^*$  is a composition of functions eg. eq. (8).

$$f^*(\mathbf{x}; \Theta) = f^3(f^2(f^1(\mathbf{x}; \Theta^1), \Theta^2), \Theta^3) \quad (8)$$

In eq. (8) each function  $f^i$  is called layer. In this example  $f^1$  is called the input layer,  $f^3$  the output layer and  $f^2$  a hidden layer. Hence a layer is a vector-to-vector function. In this context a unit describes the corresponding vector-to-scalar functions of one layer. The width of a layer is referred to as the dimension of the vector valued input. The depth of the network describes the number of composed functions. In autoencoders the dimensions of input and output layer are identical.

Autoencoders are a special kind of neural network, that have a central hidden layer, that outputs a code  $c$  which should contain useful features of the input  $x$  while usually being of a lower dimension. Hereafter the code will be addressed as intrinsic variables highlighting the property of containing useful features of the input  $x$ . Autoencoders can be split in two parts, the encoder  $h(x) = c$  which compresses the input and outputs the intrinsic variables  $c$  and the decoder  $g(h) = \hat{x}$  which reconstructs the input from the intrinsic variables to output  $\hat{x}$ . The goal of autoencoders conflicts with the training objective. The former is to produce a code that describes the intrinsic features of the input, while the latter is to minimize the difference between  $x$  and  $\hat{x}$ . Therefore autoencoders need to be restrained from learning the identity function perfectly which in turn should drive the model to choose which instance to copy.

Obviously deep learning emphasizes the focus on the depth of a model. This is because linear models with just one layer can only approximate linear functions. Adding a non-linear activation function to the output of the proposed model wouldn't be sufficient in modeling any nonlinear behavior of the function. However the universal approximation theorem [13] states that MLPs with at least one hidden layer and any non-linear activation function can approximate any function given that enough hidden layers can be provided. In conclusion, MLPs are universal approximators [6].

There are several types of layers, that can be used in MLPs. For this thesis it is sufficient to treat so called *fully connected layers* and *convolutional layers*. Fully connected layers

are called `Linear`[?] in PyTorch[?] because they compute a linear transformation of the input eq. (9), where  $x$  is the input vector,  $A$  is the weight matrix and  $b$  is a bias vector. This is the forward pass of a linear layer. The learnable parameters  $\Theta$  are in this case the values in  $A$  and  $b$ . For a linear layer which takes a vector of size  $i$  as input and outputs a vector of size  $o$  there are  $l = i \times o + o$  learnable parameters.

$$y = xA^T + b \quad (9)$$

Continuing with the forward pass in convolutional layers, which are called `Conv2D`[?] in Pytorch. Convolutional layers are usually applied when the input data has a known grid-like topology [6]. While for fully connected layers, the input size is fixed, convolutional layers can be applied to inputs of various sizes. Furthermore, they are sparse by construction and share parameters making them equivariant [6]. Even though the name implies the use of the convolutional operation in eq. (10), PyTorch and many other neural network libraries instead use the cross correlation, which is an operation closely related to a convolution[6][?]. The convolution in 10 operates on two functions  $x$  and  $w$ , where the latter is a weighting function of the former which makes  $s$  the weighted average of the input  $x$ . In eq. (11)  $x$  is represented by  $I(m, n)$ , and  $w$  by  $K(m, n)$  respectively. Note that while eq. (10) is scalar valued and eq. (11) is two dimensional, only for illustrative reasons.

$$s(t) = (x * w)(t) = \int x(a)w(t - a) da \quad (10)$$

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n)K(i - m, j - n) \quad (11)$$

One drawback in implementing the discrete convolution eq. (11) is that there can be invalid values for  $m$  and  $n$ . This can be solved by exploiting the commutative property of the convolution, which results for the discrete convolution in a flipped kernel relative to the input, eq. (12). Without the need of flipping the kernel which is not always possible, i.e. exploiting the commutative property, cross correlation is adopted eq. (13).

$$S(i, j) = (K * I)(i, j) = \sum_m \sum_n I(i - m, j - n)K(m, n) \quad (12)$$

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i + m, j + n)K(m, n) \quad (13)$$

The given equations illustrate the movement of the kernel over a two dimensional input, but leaving out two basic accompanying designs. First is the so called stride of the kernel, which results in an increased down sampling with increased stride size. Considering strides of the kernel along one dimension of the input results in a shrinkage of that dimension by

$$o = \frac{i - k}{s} + 1. \quad (14)$$

Here  $o$ ,  $i$ ,  $s$  and  $k$  are the output, input, stride and kernel size of one dimension respectively [14]. Second are the so called channels, which allow convolutional layers to extract a different feature for every channel from the same input. For a two dimensional input like

images this could be different features for the same location on the image, like edges and color in the RGB color space [6]. Adding strides and kernel to eq. (13) yields

$$S_{i,j,k} = c(\mathbf{K}, \mathbf{I}, s)_{i,j,k} \sum_{l,m,n} [\mathbf{I}_{l,(j-1)\times s+m, (k-1)\times s+n} \mathbf{K}_{i,l,m,n}] . \quad (15)$$

The kernel  $\mathbf{K}$  is a four dimensional tensor with  $i$  indexing into the output channels of  $S$ ,  $l$  indexing into the input channels of  $I$ ,  $m$  and  $n$  indexing into the rows and columns. The input  $\mathbf{I}$  and output  $S$  are three dimensional tensors with  $j$  and  $k$  indexing into the rows and columns. Note that in PyTorch eq. (15) is a so called valid cross correlation (valid convolution) [?] meaning the kernel will only move over input units for which all  $m$  and  $n$  are inside the rows and columns of the input. Zero padding the input can prevent the kernel from omitting corners in that case.

For autoencoders the necessity to perform a transposition of the applied layers arises, which is straight forward for fully connected layers. Convolutional layers with strides greater than unity on the other hand the transposition needs the kernel to be padded with zeros to realize an upsampling of the input data [14]. Note that the padding of the kernel with zeros is only used to illustrate how the upsampling works. In eq. (16) taken from [6] multiplications with zero are omitted. The size of one dimension during upsampling can be calculated with the transposition of eq. (14).

$$t(\mathbf{K}, \mathbf{H}, s)_{i,j,k} = \sum_{\substack{l,m \\ s,t \\ (l-1)\times s+m=j}} \sum_{\substack{n,p \\ s,t \\ (n-1)\times s+p=k}} \sum_q \mathbf{K}_{q,i,m,p} \mathbf{H}_{q,l,n} \quad (16)$$

Forward-propagation is then referred to as the compositional evaluation of each layer. For the example in eq. (8) the outputs of each layer would then be:  $f^1(\mathbf{x}, \Theta^1) = \mathbf{p}$ ,  $f^2(\mathbf{p}, \Theta^2) = \mathbf{q}$  and  $f^3(\mathbf{q}, \Theta^3) = \hat{\mathbf{y}}$ .

Subsequently the cost function  $J(\Theta)$ , which will be discussed later on, can be computed. Afterwards back-propagation returns the gradients w.r.t. the layer parameters  $\Theta$  to finally compute updated parameters  $\Theta$ . The name back-propagation refers to the use of the chain rule of calculus to obtain the gradients of each layer. Assuming again the example in eq. (8) back-propagation would be equations eq. (19) to eq. (17):

$$\frac{\partial J}{\partial \Theta^3} = \frac{\partial J}{\partial \hat{y}_i} \frac{\partial \hat{y}}{\partial \Theta^3} \quad (17)$$

$$\frac{\partial J}{\partial \Theta^2} = \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial q} \frac{\partial q}{\partial \Theta^2} \quad (18)$$

$$\frac{\partial J}{\partial \Theta^1} = \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial p} \frac{\partial p}{\partial \Theta^1} \quad (19)$$

While the term back-propagation is solely used for the method to compute the gradients for each layer in a backward fashion, meaning from the last layer to the first, the update of the parameters is done in an optimization step. -L2-error as performance metrics

### 3.1 Autoencoders

Autoencoders have many hyperparameters determining their capability for compression and subsequent reconstruction. These parameters include : *depth*, *width of layers*, *activation functions*, *batch-size*, *learning rate*, *number of filters*, *stride width*, *kernel size*.

Their finding is discussed in this section, for both Hand R.

For the fully connected autoencoder the order of determining the hyperparameters is as follows: Depth → Hidden Units → Batch Size → Activation Functions. For the convolutional autoencoder the order of determining the hyperparameters is as follows: Depth → Channels → Batch Size → Activation Functions. Both models are evaluated on the course of the training. Figures showing the training process for all experiments are provided in the appendix. After training the L2-Error?? is evaluated on the unshuffled, complete dataset as described in section 4.1.1.

Important to mention is that the difficulty of finding the right set of hyperparameters for MLPs led to an intensive search prior to the creation of this contribution. The difficulty lies in the fact that there is little systematic knowledge about how the hyperparameters interact in the model and answer to a variety of input data. Goodfellow et al. point out that with a combination of intuition, certain methods and first of all experience practitioners find hyperparameters that work well [6]. As for this thesis a list of the prior search is provided in section 5.4. This list does not claim to cover the complete search and can by no means give enough insight to enable reproducability by another person. For this reason an insight into the methods employed to find the set of hyperparameters used in this contribution are described in the following.

### 3.1.1 Hyperparameters for the Fully Connected Autoencoder

To start with a working model a guess is needed to be made about the initial design of the architecture. Here the hyperparameters already found are used as a starting point and shown in Table 1.

To start with the number of layers or depth as described in section 3 is determined, as

Mini batch size	Intrinsic dimensions	Epochs	Learing rate	activation code/rest
16	3	2000	1e-4	Tanh/LeakyReLU

Table 1: Initial hyperparameter selection

they set a consequential part of the representational capacity of the model and therefore can initiate over- and underfitting at an early stage of the parameter search. ?? and ?? in ?? show the training and validation error for five designs shown in table 2.

For the hydrodynamic regime a number of layers greater than four results in a slight

Number of layers	Reduction per layer
10	40 → 40 → 20 → 10 → 5 → 3
8	40 → 40 → 20 → 10 → 3
6	40 → 40 → 20 → 3
4	40 → 40 → 3
2	40 → 3

Table 2: Initial hyperparameter selection

overfitting at an early stage of training (at around 100 epochs). Below four layers an underfitting can be observed. Hence yielding the conclusion, that four layers result in the

best performing net at this early stage. Overfitting occurs only after the 1000th epoch and is less than with the other three nets that show overfitting. Note, that contrary to expected results, the train error is lower than the test error. The random shuffling during preprocessing might be taken into account here. Nonetheless solely overfitting is defined by differences between train and test loss. In Addition four layers and lower show a stable training in relation to the rest.

The analysis of the number of layers for the rarefied regime is not showing overfitting as obvious as before. The network with two layers underfits in contrast to the other nets whereas nets with more than two layers reach a similar train - and test loss. Networks with more than 4 layers, again show an unstable training compared to the net with four layers. Train and test loss show a diverging behaviour after around the 100th epoch.

In conclusion the net with four layers performs best for both the hydrodynamic and the rarefied regime.

Training duration is raised from 2000 epochs to 5000, as training did now converge completely as seen in ?? and fig. 14.

In the following the width of the hidden layer will be analysed. There are two available hidden layers for the autoencoder. But as the architecture of the decoder mirrors that of the encoder only one parameter needs to be varied. For both the hydrodynamic and the rarefied regime five experiments are conducted, varying the hidden units from ten to fifty. Results for  $\Pi_h$  and  $\Pi_r$  are shown in table 3. For Hand Rforty hidden units per-

Hidden units	10	20	30	40	50
Error for $\Pi_h$	0.0054	0.0027	0.0036	0.0017	0.0032
Shrinkage factor for H	0.3	0.1	0.015	0.075	0.06
Error for $\Pi_r$	0.0078	0.0027	0.0022	0.0015	0.0025
Shrinkage factor for R	0.5	0.25	0.016	0.0125	0.01

Table 3: L2-Error for different number of hidden units for  $\Pi_h$  and  $\Pi_r$ .

forms best with an error around  $1e^{-3}$  and a shrinkage factor of 0.075 for H and 0.0125 for R. The combination of representational capacity set by the number of hidden units and shrinkage factor has an optimum here. When decreasing the number of hidden units, the representational capacity of the model also decreases. At the same time the shrinkage factor increases, therefore less information needs to be compressed in one step. The error increases with decreasing the number of hidden units from forty. This can be explained by the lower representational capacity. When increasing the number of hidden units to fifty the error also increases though, the representational capacity grows. Here the decreased shrinkage factor could be responsible for that.

Next the mini-batch size is analysed. Results are displayed in table 4 for H and in table 5 for R. Additionally section 5.4 shows the training for both input data H and R. For both input data experiments are first conducted with mini-batch sizes in the range of : [2, 4, 8, 16, 32].

The results of the L2-Error show best performance between a mini-batch size of 8 and 16 for H. Likewise the results of the L2-Error for R show best performance between mini-batch sizes of 4 and 16. Therefore additional experiments are conducted. Three additional experiments for H with mini-batch sizes of: [10, 12, 24]. Two additional experiments for

Rwith mini-batch sizes of: [6, 10]. The training reveals that smaller batch sizes lead to oscillating traing and test errors. This can be overcome by a smaller learning rate, but also leads to increased computational time as more epochs are needed to achieve a comparable training and test loss at the last epoch. A mini-batch size of 8 is bordering the zone where the oscillations of training and test loss subside. At the same time a mini-batch size of 8 to 10 indicate also a growth of the L2-Error. In conclusion a mini-batch size of 8 is chosen to continue to working with for Hand R. The osccliations which make the training instable can be battled with a lower learning rate as soon as training starts to tremble.

Together with the mini-batch sizes, activations are examined. Hence experiments with

Batch Size	2	4	8	10	12	14	16	32
Error for $\Pi_h$	0.0011	0.0011	0.0011	0.0029	0.0018	0.0018	0.0013	0.0030

Table 4: L2-Error for different mini-batch sizes for  $\Pi_h$ .

Batch Size	2	4	6	8	10	16	32
Error for $\Pi_r$	0.0014	0.0010	0.0012	0.0012	0.0014	0.0017	0.0017

Table 5: L2-Error for different mini-batch sizes for  $\Pi_r$ .

different activation functions are performed with the innitally used mini-batch size of 16. Besides the mini-batch size, epochs are as well taken from the inital selection of hyperparameters as 2000 epochs. Unlike the previous two examinations the selection of activation functions shows early in training if a acceptable level of perfomance can be achieved. This is a personal observation made throughout working on this thesis. First five activations are applied to all the four layers. Second a combination of activation functions is studied, where the intrinsic variables is activated with a different function than the remaining layers. Results can be observed in table 6 and table 7. Both input

Activation function	ReLU	ELU	Tanh	SiLU	LeakyReLU
Error	0.0028	0.0019	0.0036	0.002	0.0039
Activation function	ELU/Tanh	LeakyReLU/Tanh	ELU/SiLU		
Error	0.0019	0.0017	0.0019		

Table 6: L2-Error different activation functions and combinations for H.

Activation function	ReLU	ELU	Tanh	SiLU	LeakyReLU
Error	0.0048	0.0029	0.0037	0.0134	0.0.0034
Activation function	ELU/Tanh	LeakyReLU/Tanh	ELU/SiLU		
Error	0.0.0032	0.0030	0.0036		

Table 7: L2-Error different activation functions and combinations for R.

data lead to different results this time. For that reason Hand Rare studied separately.

Starting with Hand the values of the L2-Error, it can be observed that, while ELU and SiLU stand out with the lowest value of the L2-Error, when applying the same activation function to all layers, all three coupled activations convince in the combination case. Applying different activations to the layers increases the representational capacity of the model since the model can feed on a greater variety of functions as explained in the previous subsection. The need for a great representational capacity reduces for  $H$ . The reasons are the linearity of that case as described in [1]. Hence two activations (ELU and SiLU) from a similar family of functions perform well on  $H$  without the need of a second activation. Still adding another activation render as good results concerning the L2-Error. When observing the train -and test loss in fig. 17 one can notice that overfitting clusters the activations into two groups. While they all achieve a similar MSE-Loss for train -and test data only ELU and Tanh are not showing overfitting. The combination of two activations functions yields an abundance of overfitting only for the combination of LeakyReLU with Tanh. In addition LeakyReLU with Tanh reaches together with ELU in combination with SiLU the lowest train -and test loss. The slight overfitting observed with the combination ELU and SiLU give rise to choose for the final model LeakyReLU with Tanh.

To continue with Rthe L2-Error is again observed in table 7, ELU stands out for the single activation. ELU with Tanh and LeakyReLU with Tanh meet the lowest L2-error for the combination of activations. Next observing the training for the three activations in fig. 18 one finds that all three show little to no overfitting, while the combination of LeakyReLU and Tanh deviates from that slightly. From this point no clean decision can be made which activation or combination to take. Therefore another run is conducted for the three. Results are shown in table 8. Only LeakyReLU with Tanh and ELU alone convince with a lower L2-loss. The train -and test loss does not give any hint which of the two remaining activations/ combination of activations to take concerning overfitting or a profound gradient for the train -and test loss. Hence a third run is performed taking the parameters already produced for the two models and training them for another 2000 epochs. This warm-starting of the models finally gives an answer on which one to take. The results in table 8 show that LeakyReLU with Tanh, just as for  $H$ , perform well on  $R$  with the given model. In fig. 19 the train and test loss during training are provided for the second and third run.

Activation function	ELU	ELU/Tanh	LeakyReLU/TanH
Error 2nd run	0.0034	0.0029	0.0025
Error 3rd run		0.0024	0.0019

Table 8: L2-Error different activation functions and combinations for  $H$ .

### 3.1.2 Hyperparameters for the Convolutional Autoencoder

The convolutional autoencoder architecture 1.1 is a composition of six convolutional and three fully connected layers, eq. (20).

$$y_p = f_C^9(f_C^8(f_C^7(f_C^6(f_F^6(f_F^5(f_F^4(f_F^3(f_C^3(f_C^2(f_C^1(y_0))))))))))) \quad (20)$$

## 4 Reduced Order Model

In this section model order reduction (ROM) will be introduced and two algorithms for obtaining a reduced basis (RB) are discussed. The proper orthogonal decomposition (POD) and autoencoders. In addition a reduced order model (ROM) based on the method of characteristics[15] is evaluated.

Model order reduction is a technique used for reducing the computational cost, which is computational resources as memory and computation power. Partial differential equations (PDEs) once discretized become a system of high dimensional ordinary differential equations (ODEs) as shown in section 2. Here model order reduction exploits the idea that every high dimensional dynamical-state space  $f(\mathbf{x}, \mu) \in \mathcal{D}$  can be described by a state-space or manifold of lower dimension  $\tilde{f}(\mu) \in \mathcal{E}$

$$f \approx \tilde{f} \quad \text{with} \quad \mathcal{D} \ll \mathcal{E}. \quad (21)$$

Reduced order modeling is partitioned into two successive phases called the *offline* - and the *online phase*. During the offline phase data or *snapshots* of a dynamical-system is generated through experiments or simulations of the full order model (FOM). The so called *snapshots*  $U = u(t_1), \dots, u(t_n)$  are created once, each representing one moment in time of the dynamical system. Next a mapping  $g$  is constructed such that  $\tilde{u} = g(u)$ , for which  $u(t_i) \approx \tilde{u}(t_i)$ . During the online phase the reduced order model is evaluated and the error is estimated by eg.  $\|u(t) - \tilde{u}(t)\|$ . Therefore the online phase may be described as stage of independence from the full order model. To continue the definition of the *intrinsic solution manifold dimensionality* by *Carlberg et al.* [5] is needed: Assuming the initial value problem

$$\mathbf{r}^n(\mathbf{x}^n; \mu) = 0, \quad n = 1, \dots, N_t \quad (22)$$

has a unique solution for each parameter instance  $\mu \in \mathcal{D}$ , the intrinsic dimensionality of the solution manifold  $\{\mathbf{x}(t, \mu) | t \in [0, T], \mu \in \mathcal{D}\}$  is (at most)  $p^* = n_\mu + 1$ , as a mapping  $(t, \mu) \mapsto \mathbf{x}$  is unique in this case. This provides a practical lower bound on the dimension of a nonlinear trial manifold for exactly representing the dynamical-system state. In summary the intrinsic solution manifold dimensionality, in the following referred to as the number of intrinsic variables, are in number as much as there are parameters that can describe the whole dynamical-system state plus one, the time  $t$ . As for the full order BGK equation section 2 the intrinsic variables could be three for the macroscopic velocity  $U(\mathbf{x}, t)$ , three for the microscopic velocities  $\xi$  of the collision operator  $Q(f, f)$ , three for  $T, \rho$  and  $\nu$  plus one equaling to a total of ten intrinsic variables for the 3D case. In the following subsection the available data for this thesis is introduced along with a proposal for the number of intrinsic variables.

The BGK equation is valid for gases in the hydrodynamic regime as well as for rarefied gases. As described in section 2 depending on the equilibrium of the BGK equation, the solution transitions between a pure Boltzmann - and a Maxwellian distribution. The former is known to be well approximated by linear methods as the SVD, while the latter poses problems due to the non-linear behavior.

## 4.1 Offline Phase

### 4.1.1 Full order BGK model

The number of intrinsic variables can therefore be determined for the 1D case of the BGK equation. The parameters  $\mu$  describing the gas flow in the hydrodynamic regime has therefore one macroscopic velocity  $U(x)$ , and the DIE ORIGINAL DTATENSTRUJTUR BESCHREIBEN For the autencoder using fully connected layers, the input vectors  $y_o \in \mathbb{R}^{n_s}$  of size  $n_{\text{input}} = n_\xi = 40$  are arranged in the sampling matrix  $S_{AE} \in \mathbb{R}^{n_s \times n_\xi}$  as seen in eq. (23) resulting in  $n_S = 5000$  available samples. Note that the POD uses the same matrix transposed  $S_{AE}^T$  as input. In the following hydrodynamic regime will reffer for the input data for knudsen number 10e-4 and rarefied regime will refer to the input data for knudsen numbers 10e2. - insert unshuffled set pls

$$S_{AE} = \begin{bmatrix} f(\xi_1, t_1, x_1) & \cdots & f(\xi_n, t_1, x_1) \\ f(\xi_1, t_1, x_2) & \cdots & f(\xi_n, t_1, x_2) \\ \vdots & \vdots & \vdots \\ f(\xi_1, t_1, x_n) & \cdots & f(\xi_n, t_1, x_n) \\ f(\xi_1, t_2, x_1) & \cdots & f(\xi_n, t_2, x_1) \\ \vdots & \vdots & \vdots \\ f(\xi_1, t_n, x_n) & \cdots & f(\xi_n, t_n, x_n) \end{bmatrix} \quad S_{Conv} = \begin{bmatrix} n_{Filters} & f(\xi_1, \mathbf{t}, \mathbf{x}) \\ n_{Filters} & f(\xi_2, \mathbf{t}, \mathbf{x}) \\ \vdots & \\ n_{Filters} & f(\xi_n, \mathbf{t}, \mathbf{x}) \end{bmatrix} \quad (24)$$

$$(23)$$

Convolutional autoencoders use a different sampling matrix  $S_{Conv}$  due to their two dimensional capability resulting in  $n_S = 40$  available samples eq. (24).  $N_{Filters}$  varies over the succeeding layers, growing with the shrinkage of  $(\mathbf{t}, \mathbf{x})$ . - In the following the data of the BGK full order model with a knudens number of  $Kn = 0.00001$ , describing a gasflow in the hyrodynamic regime will be reffere to as  $\Pi_h$ . Equally the data of the BGK full order model with a knudsen number of  $Kn = 0.01$  will be refferd to as  $\Pi_r$ .

### 4.1.2 Reduced Basis by POD and Autoencoder

The singular value decomposition of the input  $X$  [REF to Section 1] gives the optimal low-rank approximation  $\tilde{X}$  of  $X$  eq. (25)[Eckard-Young].

$$[\text{!htbp}] \quad \underset{\tilde{X}, s.t. \text{rank}(\tilde{X})=r}{\text{argmin}} \quad \|X - \tilde{X}\|_F = \tilde{U} \tilde{\Sigma} \tilde{V}^* \quad (25)$$

Intrinsic variables	3	4	5	6	7	8	9	10
Error	0.0327	0.0153	0.0087	0.0046	0.0021	0.0014	0.0005	0.0003

Table 9: L2-Error for different numbers of intrinsic variables for the rarefied gas flow. Calculations with two intrinsic variables are also performed, but not shown here because two intrinsic variables is considered trivial.

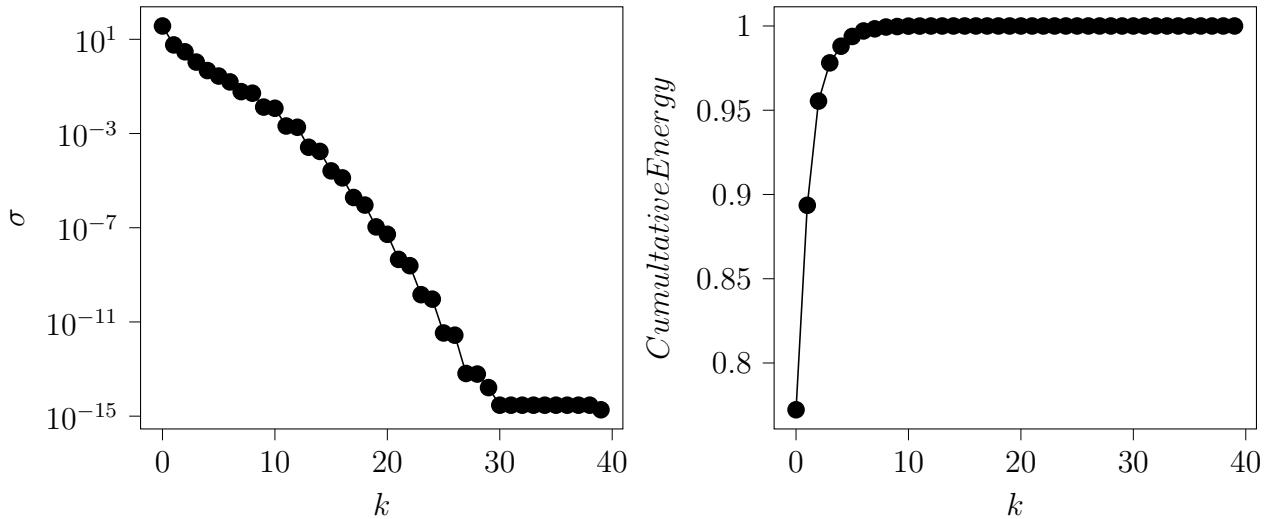


Figure 3: Singular values  $\sigma$  over  $k$  number of singular values left and *cumulative energy* over  $k$  right for the gas flow in the rarefied regime.

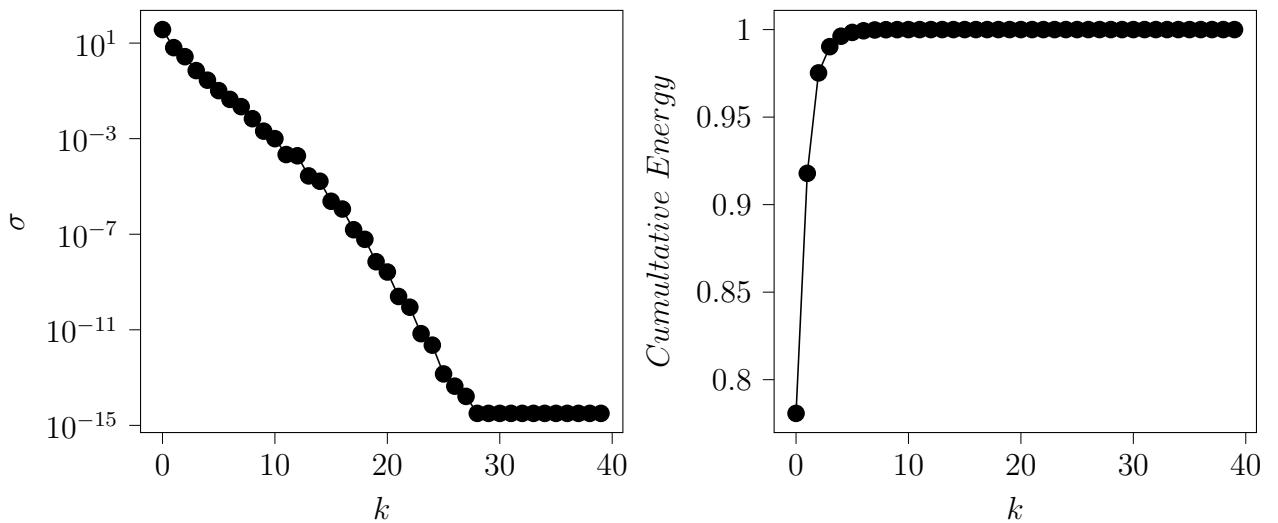


Figure 4: Singular values  $\sigma$  over  $k$  number of singular values left and *cumulative energy* over  $k$  right for the gas flow in the hydrodynamic regime.

Intrinsic variables	3	4	5	6	7	8	9	10
Error	0.0205	0.0081	0.0030	0.0013	0.0006	0.0002	$6.2e^{-5}$	$2.7e^{-5}$

Table 10: L2-Error for different numbers of intrinsic variables for the hydrodynamic gas flow. Calculations with two intrinsic variables are also performed, but not shown here because two intrinsic variables is considered trivial.

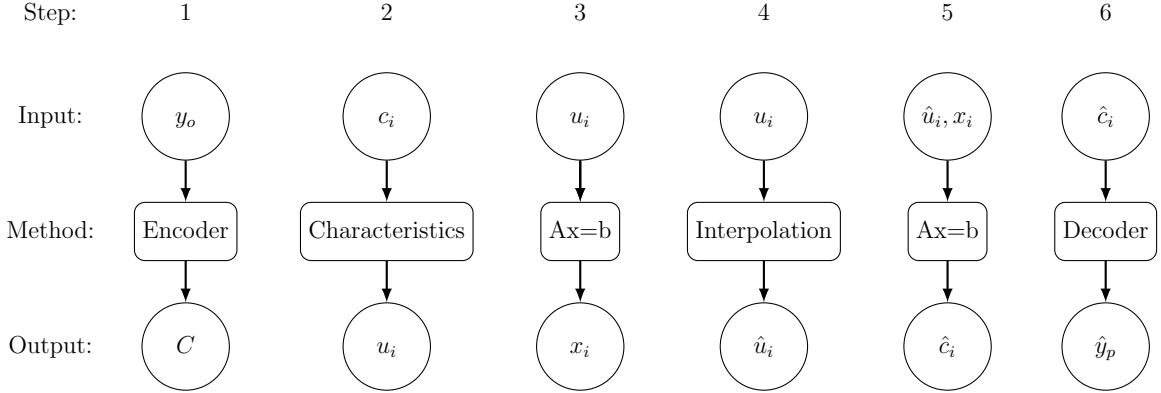


Figure 5: This figure shows the steps for obtaining a reduced order model (ROM). Decoder and Encoder need to be used after training. In step one  $y_0$  is the original input data,  $C$  is the Code. In step two  $c_i$  is the  $i$ -th intrinsic variable and  $u_i$  the corresponding characteristic. The eigenvalue problem in step 3 outputs  $x_i$  the eigenvector of  $A$ , a diagonal matrix composed of  $u_i$  and  $b$  is the corresponding  $i$ -th intrinsic variable  $c_i$ . In step 4  $\hat{u}_i$  is the interpolated vector to  $u_i$ . Step 5 solves the linear equation for the diagonal matrix  $A$  composed of  $\hat{u}_i$  times the eigenvector  $x_i$  of the eigenvalueproblem in step 3. The output is  $\hat{c}_i$  the  $i$ -th intrinsic variable corresponding to  $\hat{u}_i$  the  $i$ -th interpolated characteristic.

## 4.2 Online Phase

loremipsu

### 4.2.1 Reduced Order Model

The compression of the input data  $y_0$  yields a code  $C \in \mathbb{R}^{ix5000}$ , composed of the intrinsic variables  $c_i$ . The index  $i$  corresponds to the  $i$ -th intrinsic variable whereas their number is given by the input data. Each of them describes the transport of a discontinuity as seen in fig. 8. Hence the exploitability of the code in terms of constructing a ROM is not provided. On that account the method of characteristics [16] provides a means to bypass this shortage. It is necessary for  $c_i(x, t)$  to satisfy the conservative condition eq. (26) and the transport equation eq. (27).

$$\frac{d}{dt} \int c_i dx = \frac{d}{dt} f_i = const. \quad (26) \qquad \qquad \frac{\partial}{\partial t} c_i + \frac{\partial}{\partial x} f_i = 0 \quad (27)$$

The characteristics  $u_i$  describe the constant transport velocities for each variable  $c_i$  calculated using eq. (28). Subsequently enabling the usage of a simple polynomial interpolation of any degree. Furthermore a linear mapping  $A_i x_i = c_i$  can be applied for the reconstruction of interpolated code variables  $\hat{c}_i$ . Figure 5 depicts this approach in detail.

Questions concerning the capacity of this ROM, e.g. how many samples  $\hat{n}_t$  are needed to reconstruct  $n_t$  timestamps, are analysed in section 5.

$$u_i = \frac{f_i(c_i^-) - f_i(c_i^+)}{c_i^- - c_i^+} \quad (28)$$

## 5 Results

### 5.1 Hydrodynamic Regime

In search for a reduced model of the BGK equation, a first reduction and analysis of the provided data in the hydrodynamic regime is conducted. The error over the  $L_2$ -Norm derived from eq. (31) assigns a value to each reduction algorithm enabling a evaluation. Furthermore the conservation quantities given in ?? to ?? of the prediction are analysed

Algorithm	$L_2$
SVD	0.03
Fully Connected Autoencoder	0.002
Convolutional Autoencoder	0.02

Table 11: L2-Error over Batch-Size

over the time average of each quantity. This normalization is given in eq. (30), where  $\hat{\sigma}$  represents the given quantity. With more than 99% of the total cumulative energy  $S_N$  of the first five singular values calculated with eq. (29) the SVD provides an upper bound to the number of intrinsic features the autoencoder should extract. Figure 6 shows the singular values (left) and the cumulative energy (right).

$$S_N = \sum_{k=1}^N a_k \quad \text{with a sequence} \quad \{a_k\}_{k=1}^n \quad (29)$$

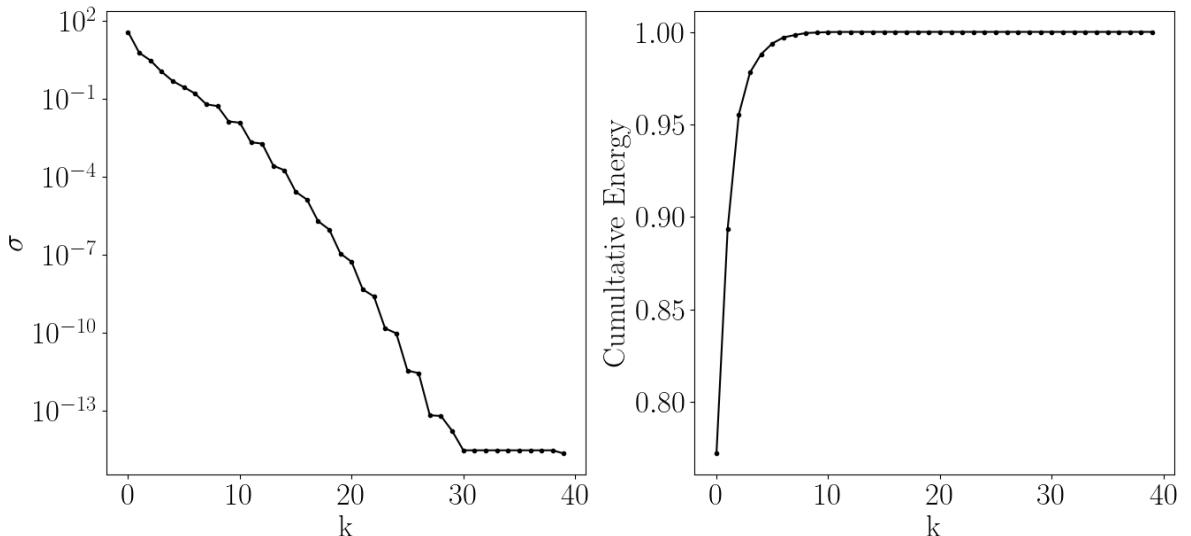


Figure 6: Singular Values (left) and cumulative energy (right) over the number of singular values

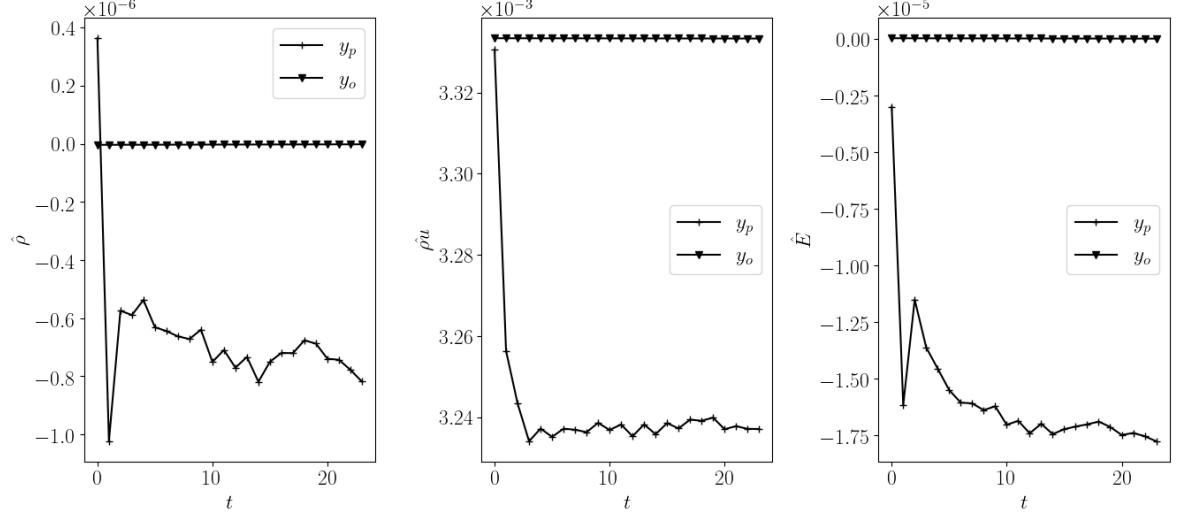
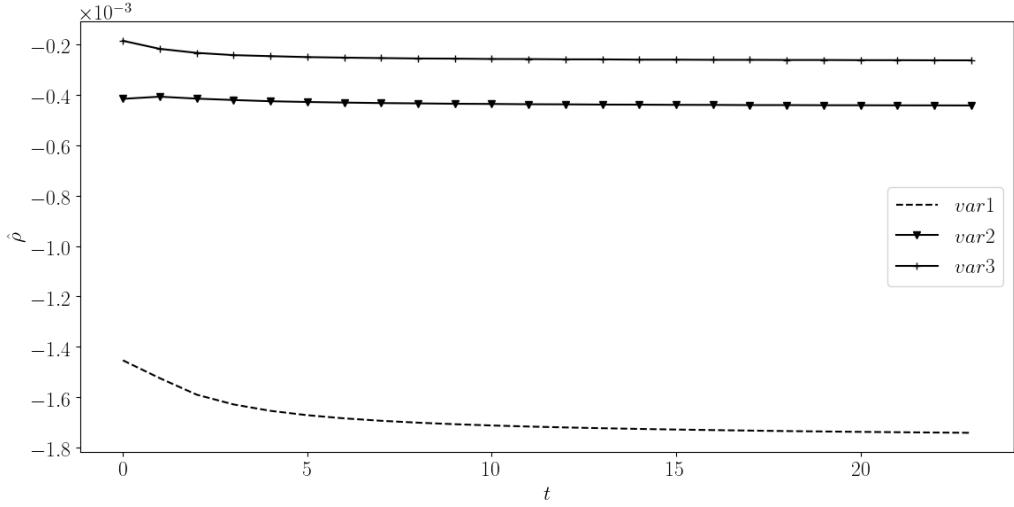


Figure 7: Normalized conservative quantities  $\hat{\rho}$ ,  $\hat{\rho}\hat{u}$  and  $\hat{E}$  as in eq. (30) for  $y_o$  and  $y_p$ .



$$\hat{\sigma} = \frac{\frac{d}{dt} \int \sigma dx}{\bar{\sigma}} = 0 \quad \text{and} \quad \bar{\sigma} = \frac{\iint \sigma dt dx}{\Delta t} \quad (30) \quad L_2 = \frac{\|y_o - y_p\|}{\|y_o\|} \quad (31)$$

Given, that the autoencoder is able to achieve a reconstruction that is equal or below the threshold of the  $L_2$ -Norm form existing methods like SVD ?? and that conservation is preserved, a reduced order model can not be derived as described in ???. Solely the the reconstruction can be validated. In addition the code needs to be a conservative system hence fulfilling ?? to ??.

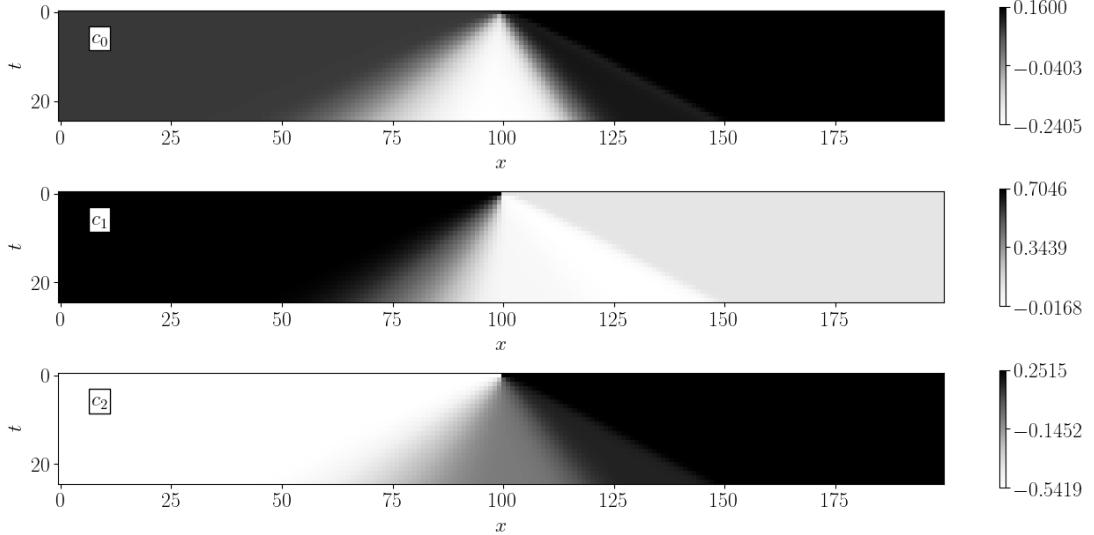


Figure 8: Code variable  $c_1, c_2$  and  $c_3$  over space  $x$  and time  $t$  of the fully connected autoencoder

## 5.2 Rarefied Regime

The number of intrinsic variables defining the rarefied regime is unknown. Therefore experiments varying the number of intrinsic variables from two to ten are performed. Table 12 shows the results of two runs. From seven up to ten intrinsic variables onward the L2-Error reaches values around  $1.2e^{-3}$ . From five and six intrinsic variables a shift happens to a value for the L2-Error around  $2.3e^{-3}$ . A lowering of the number of intrinsic variables up to two increases the L2-Error further. A comparison between results obtained from the SVD and the autoencoder, show that both algorithms perform nearly the same with eight intrinsic variables with an L2-Error of  $1.4e^{-3}$ . With more than eight intrinsic variables

Intrinsic variables	3	4	5	6	7	8	9	10
Error 1st run	0.0048	0.0025	0.0026	0.0027	0.0013	0.0014	0.0013	0.0009
Error 2nd run	0.0038	0.0024	0.0016	0.0015	0.0011	0.0010	0.0012	0.0010

Table 12: L2-Error for different numbers of intrinsic variables for the rarefied gas flow. Experiments with two intrinsic variables are also performed, but not shown here because two intrinsic variables is considered trivial.

## 5.3 Discussion and Outlook

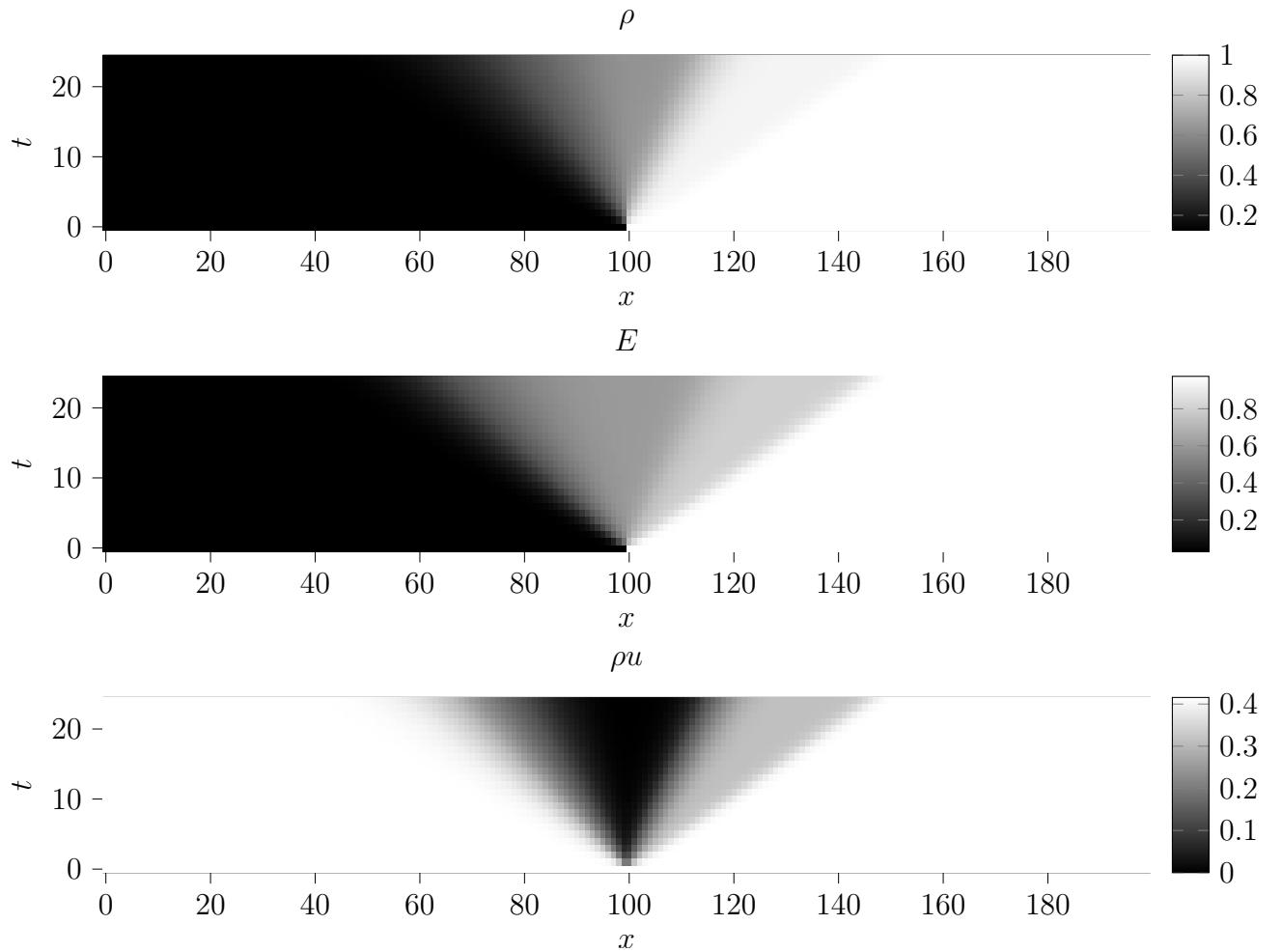


Figure 9: Macroscopic quantities  $\rho$ ,  $E$  and  $\rho u$  of the original data.

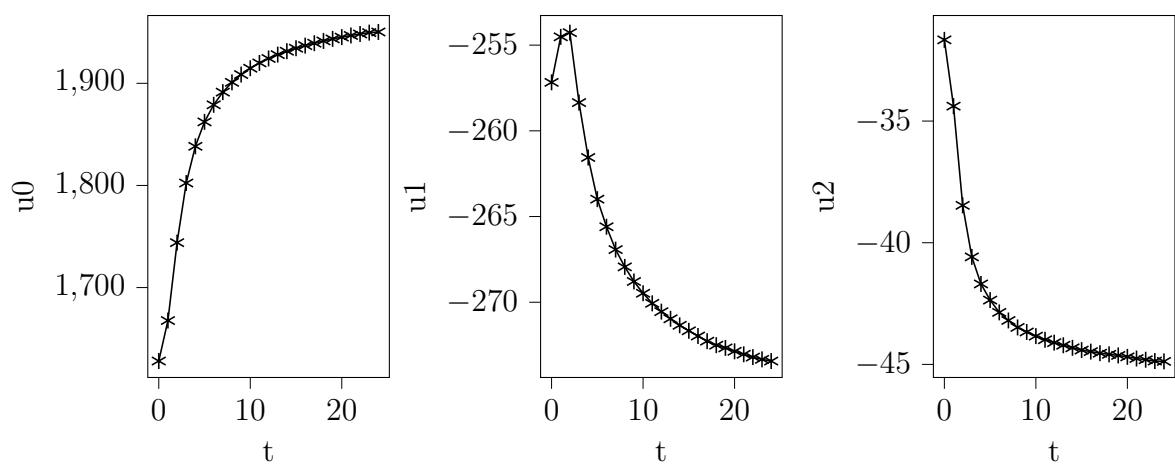


Figure 10: Characteristic velocities  $u_0$ ,  $u_1$ ,  $u_2$  of the code variables  $\text{var}0$ ,  $\text{var}1$ ,  $\text{var}2$  respectively calculated as described in Section 4.2.1.

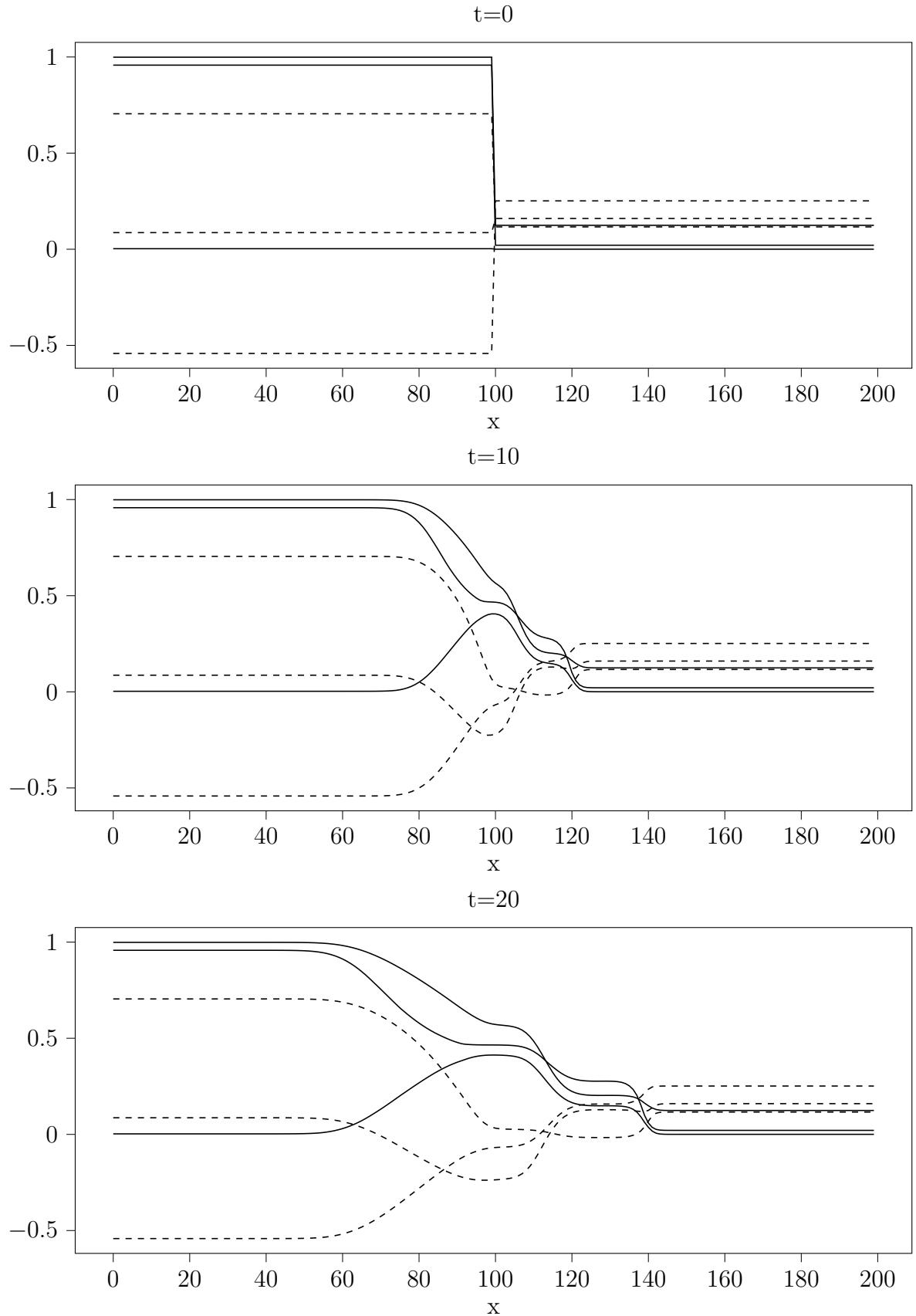


Figure 11: Code variables  $\text{var}0$ ,  $\text{var}1$ ,  $\text{var}3$  (dashed lines ---) and macroscopic quantities  $\rho$ ,  $E$ ,  $\rho u$  (full lines —) for three timestamps  $t_0$ ,  $t_{10}$ ,  $t_{20}$ .

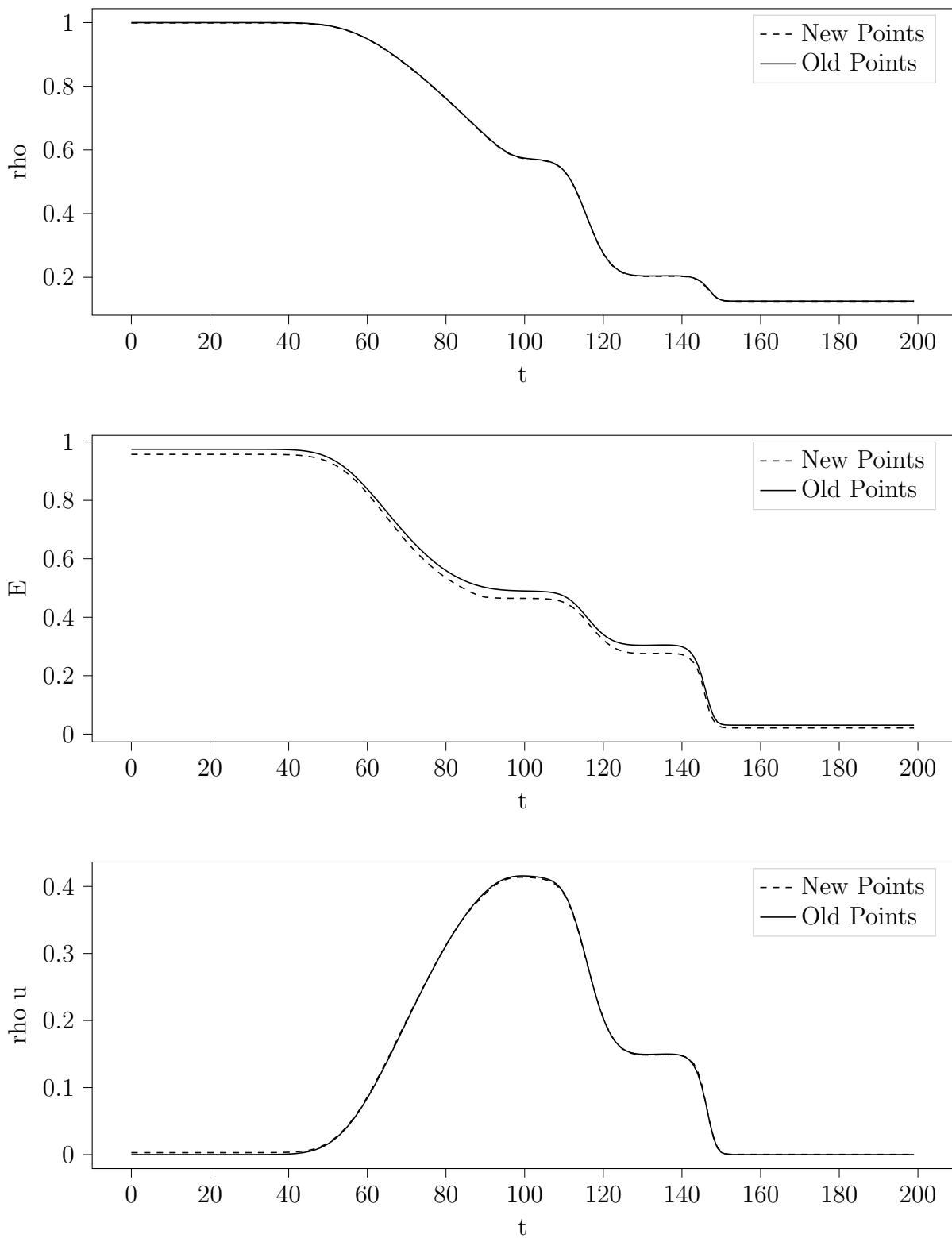


Figure 12: Resulting macroscopic quantities  $\rho$ ,  $E$ ,  $\rho u$  after interpolating in time using the interpolation method described in 4.2.1. The interpolated quantity lies at timestamp  $t=24.5$ , while the original quantity lies at timestamp  $t=25$ .

## References

- [1] Bhatnagar, Gross, and Krook. A model for collision processes in gases. 1954.
- [2] Thomas Franz. *Reduced-order modeling of steady transonic flows via manifold learning*. 2016.
- [3] Steve L. Brunton and J. Nathan Kutz. *Data driven science and engineering*. 2019.
- [4] Florian Bernard, Angelo Iollo, and Sébastien Riffaud. Reduced-order model for the bgk equation based on pod and optimal transport. 2018.
- [5] Kookjin Lee and Kevin T. Carlberg. Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders. 2019.
- [6] Ian J. Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, Cambridge, MA, USA, 2016. <http://www.deeplearningbook.org>.
- [7] D.E. Rumelhart, G.E. Hinton, and R.J. Williams. Learning internal representations by error propagation. 1986.
- [8] Dana H. Ballard. Modular learning in neural networks. 1987.
- [9] Salah Rifai, Pascal Vincent, Xavier Muller, Xavier Glorot, and Yoshua Bengio. Contractive auto-encoders: Explicit invariance during feature extraction. 2011.
- [10] Salah Rifai, Yann N Dauphin, Pascal Vincent, Yoshua Bengio, and Xavier Muller. *The Manifold Tangent Classifier*. Curran Associates, Inc., 2011.
- [11] Salah Rifai, Yoshua Bengio, Yann Dauphin, and Pascal Vincent. A generative process for sampling contractive auto-encoders. 2012.
- [12] Gabriella Puppo. Kinetic models of bgk type and their numerical integration. 2019.
- [13] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, January 1989.
- [14] Vincent Dumoulin and Francesco Visin. A guide to convolution arithmetic for deep learning. 2018.
- [15] Julius Reiss. Skript zu cfd 1. 2018.
- [16] Hervé Le Dret and Brigitte Lucquin. *Partial Differential Equations: Modeling, Analysis and Numerical Approximation* -. Birkhäuser, Basel, 2016.

## 5.4 Appendix A

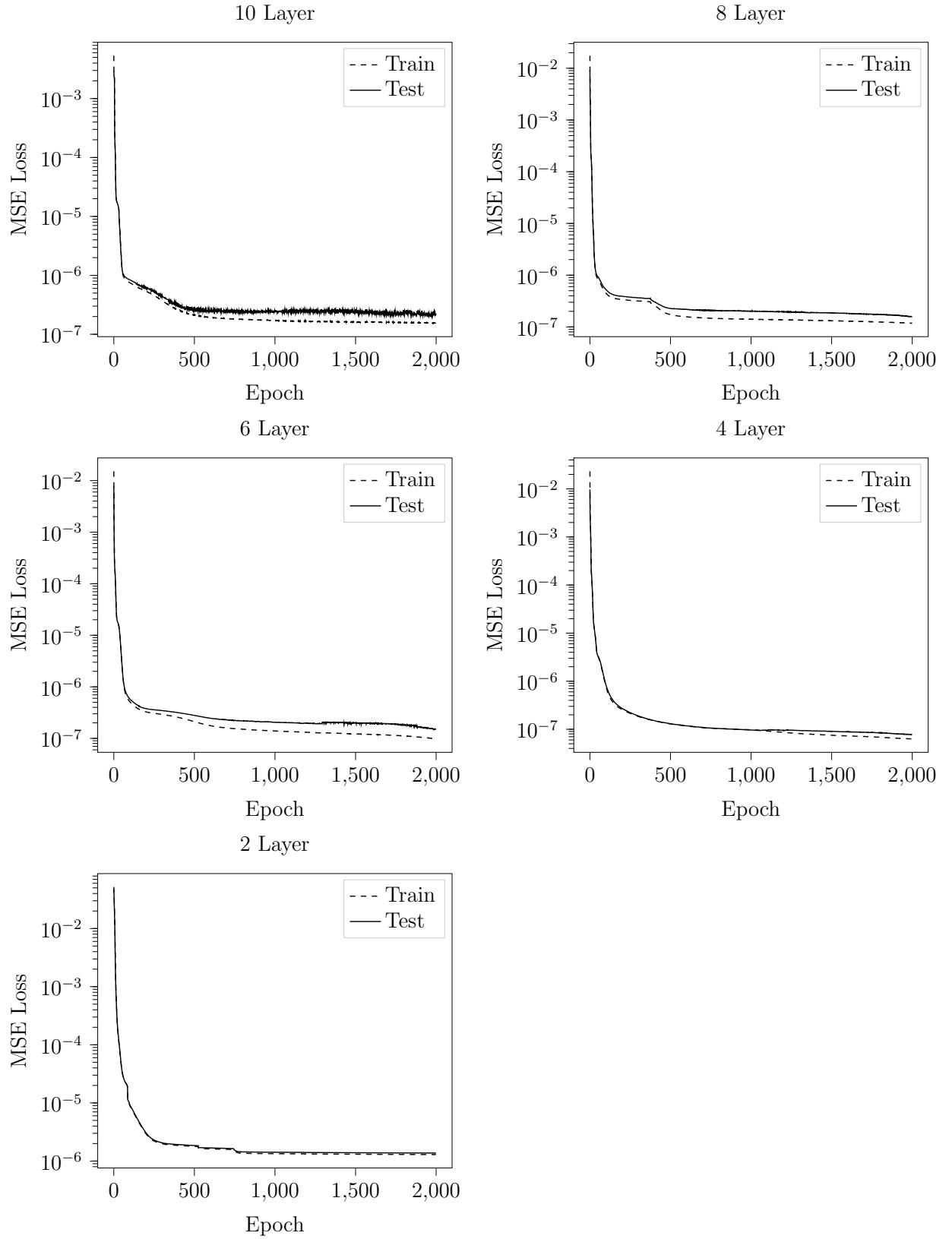


Figure 13: Experiments with five different depth on  $\mathbf{H}$ . Training and test loss are shown over 2000 epochs.

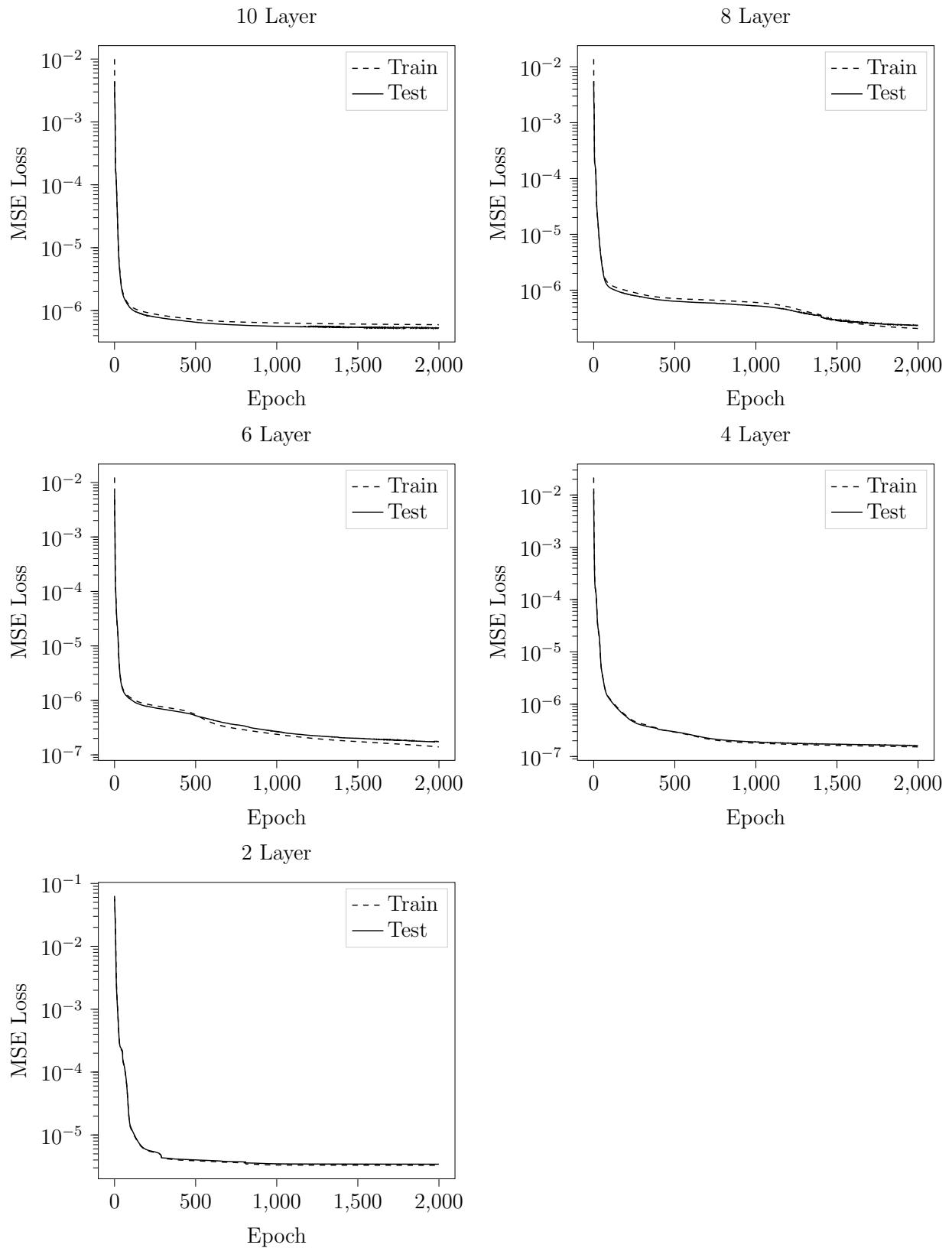


Figure 14: Experiments with five different depth on R. Training and test loss are shown over 2000 epochs.

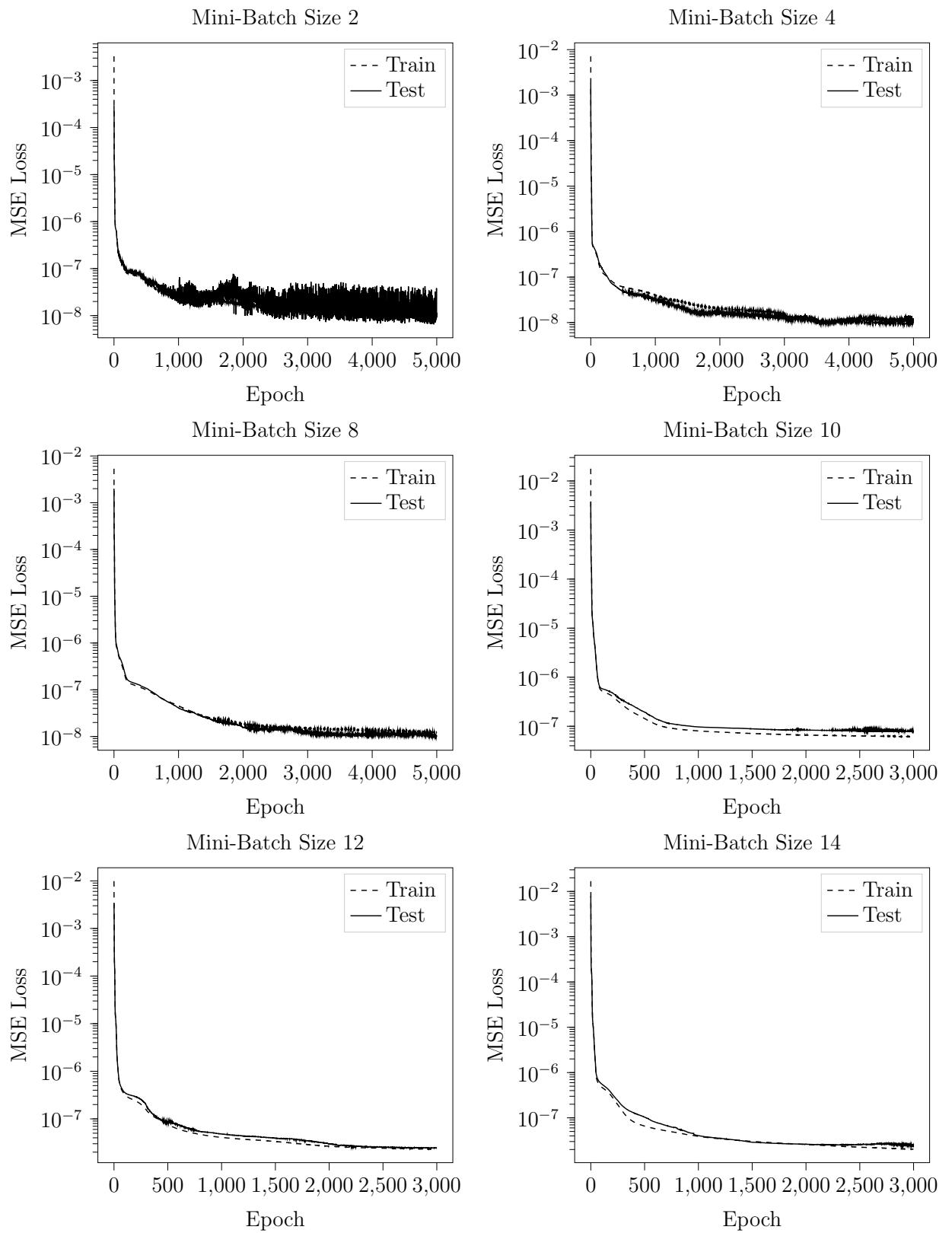


Figure 15: Training with different mini-batch sizes for  $\mathbf{H}$ . Train -and test loss is shown over 5000 epochs.

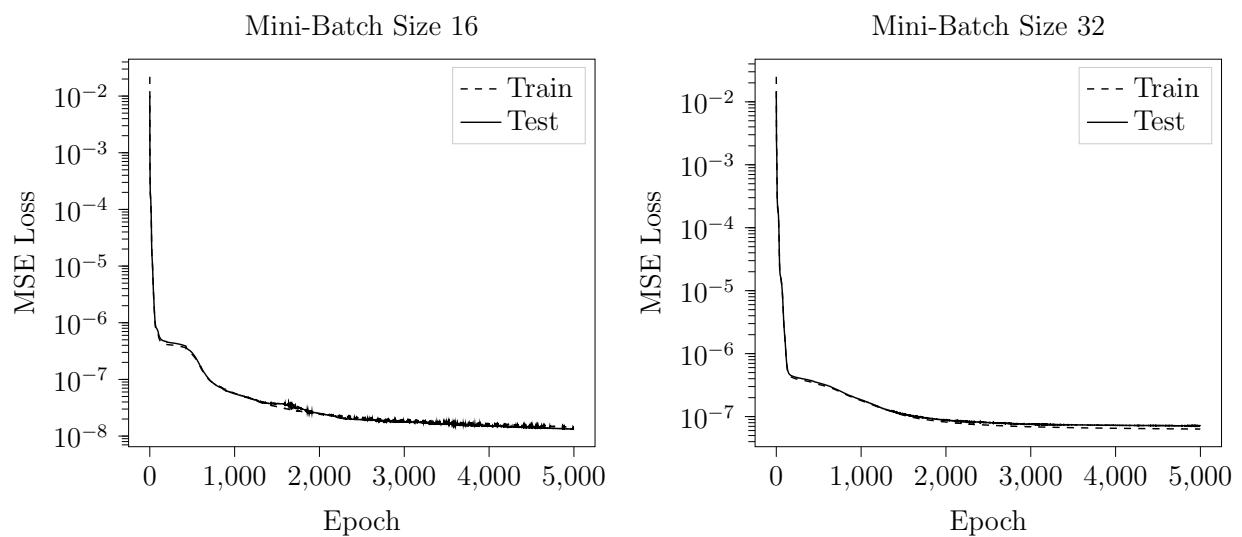


Figure 15: Training with different mini-batch sizes for  $H$ . Train -and test loss is shown over 5000 epochs.

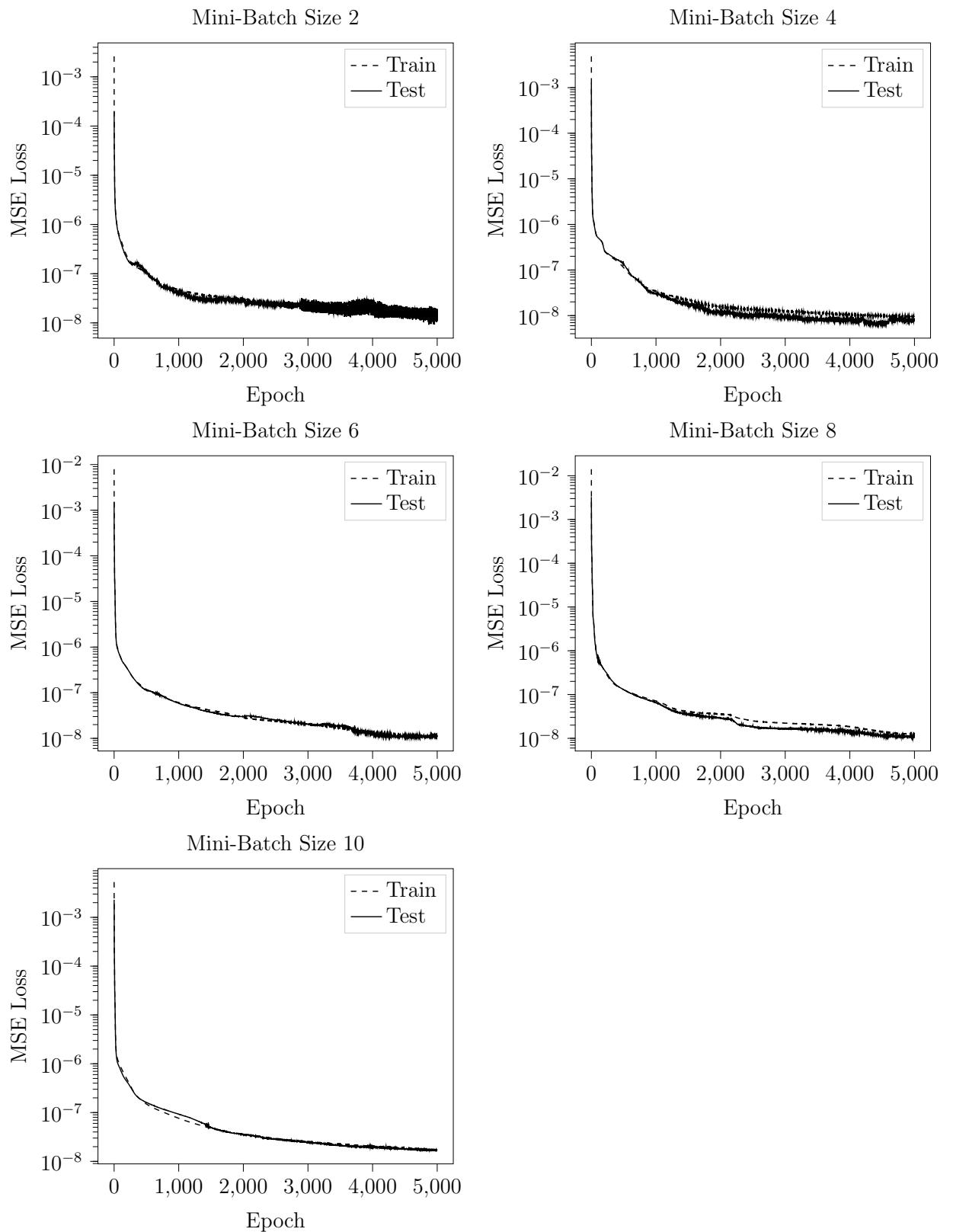


Figure 16: Training with different mini-batch sizes for R. Train -and test loss is shown over 5000 epochs.

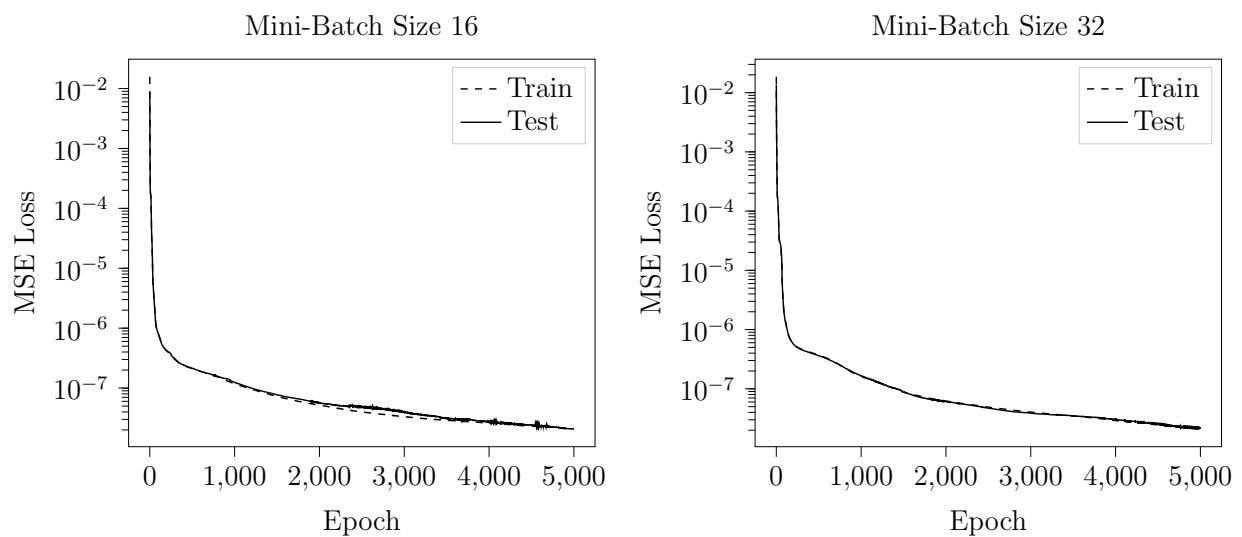


Figure 16: Training with different mini-batch sizes for R. Train -and test loss is shown over 5000 epochs.

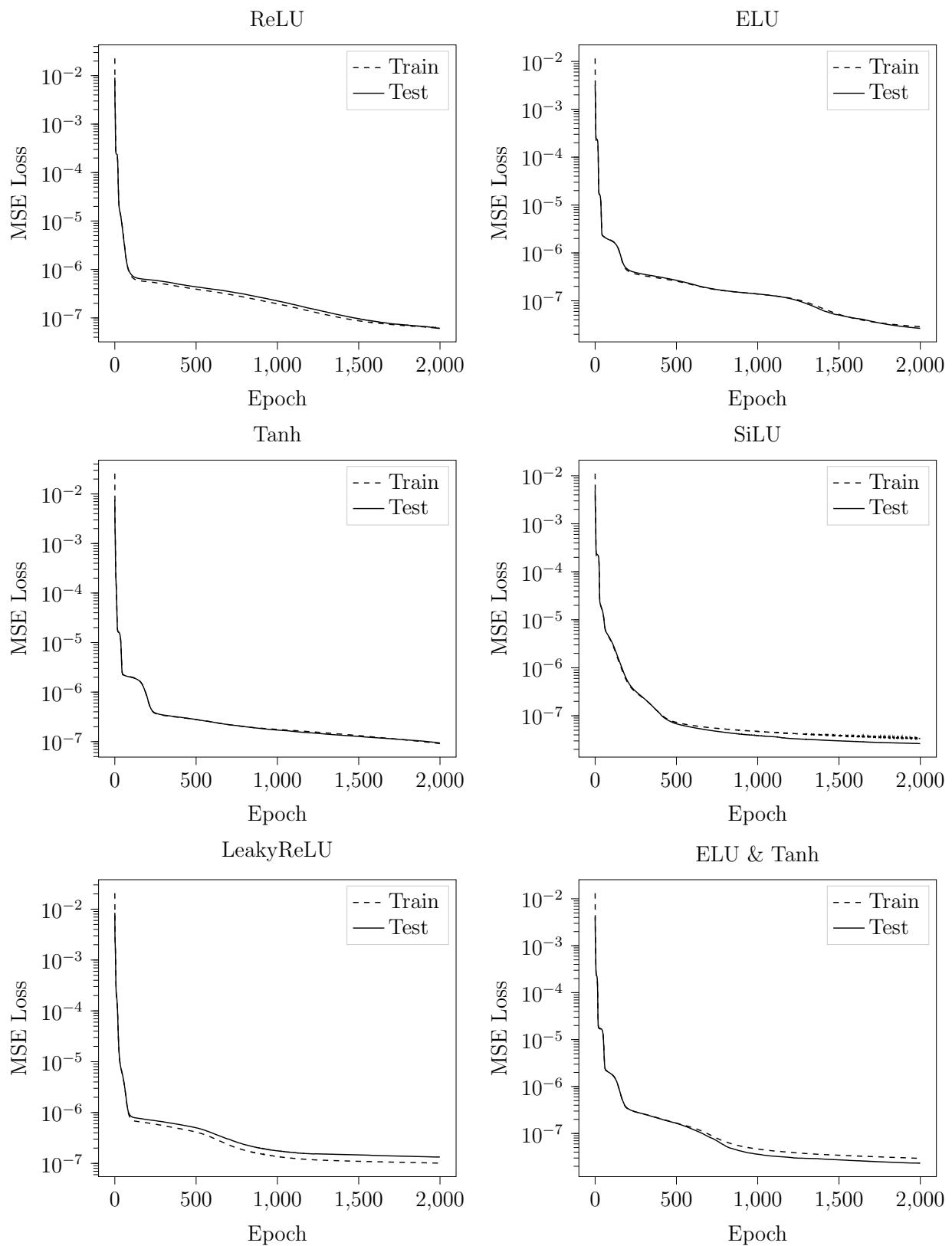


Figure 17: balblabla

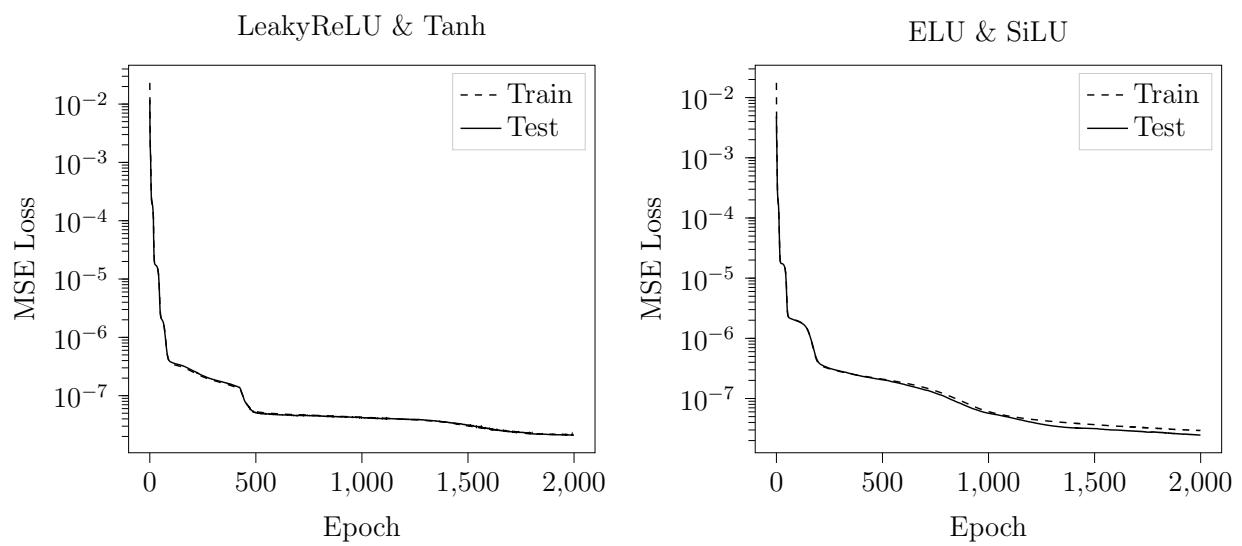


Figure 17

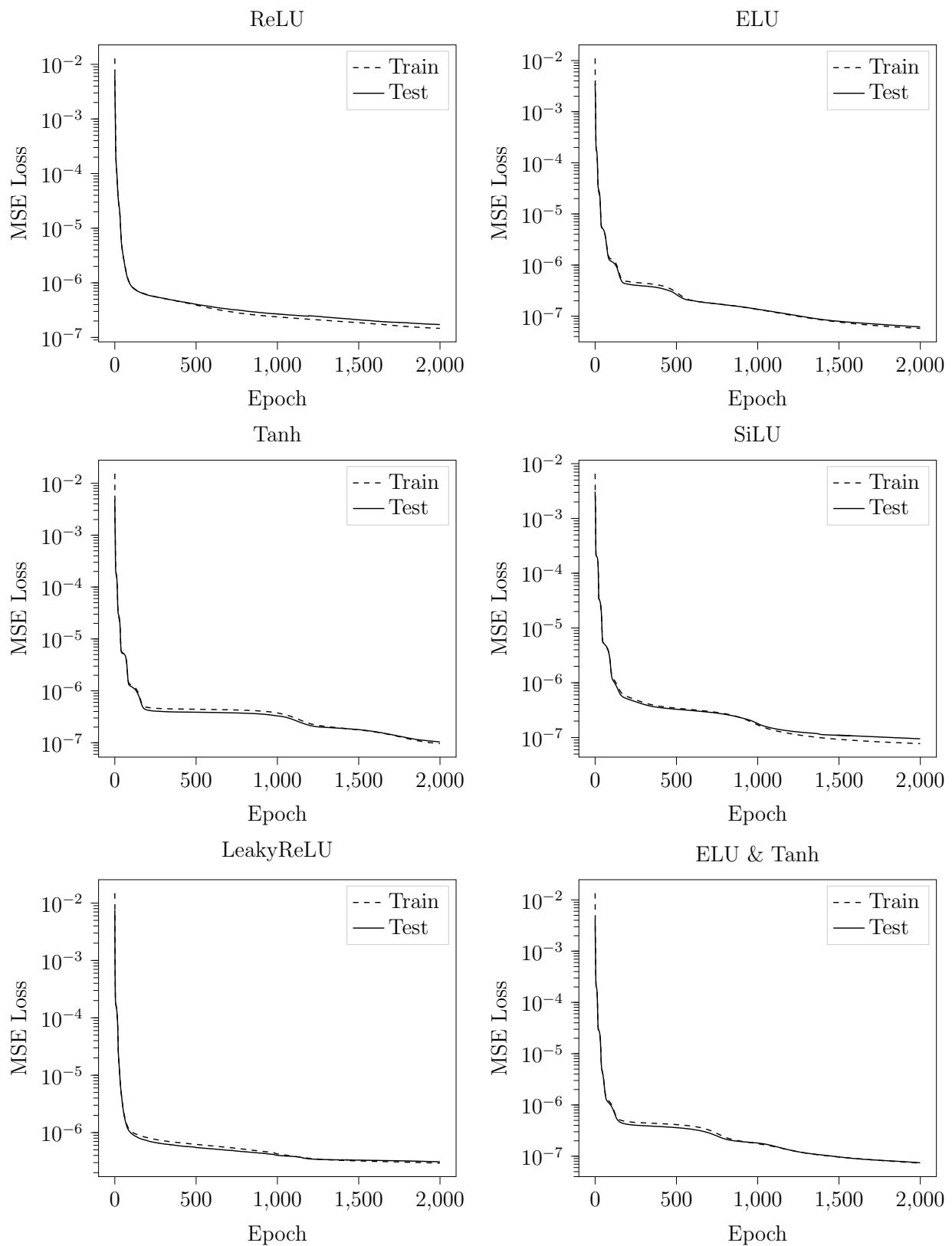


Figure 18: blablabla

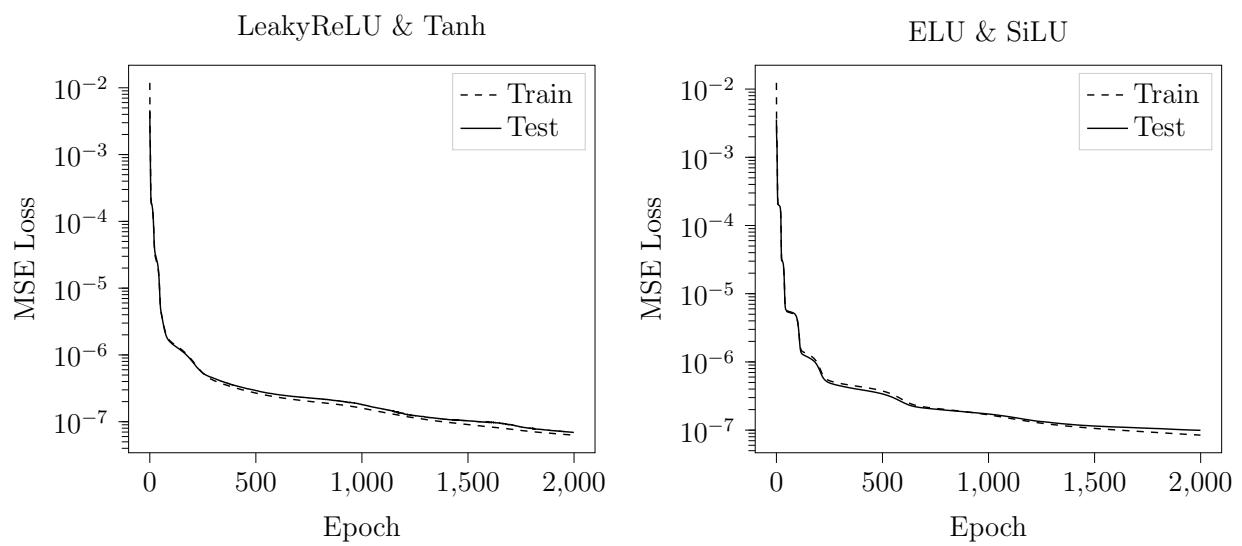


Figure 18: blablabla

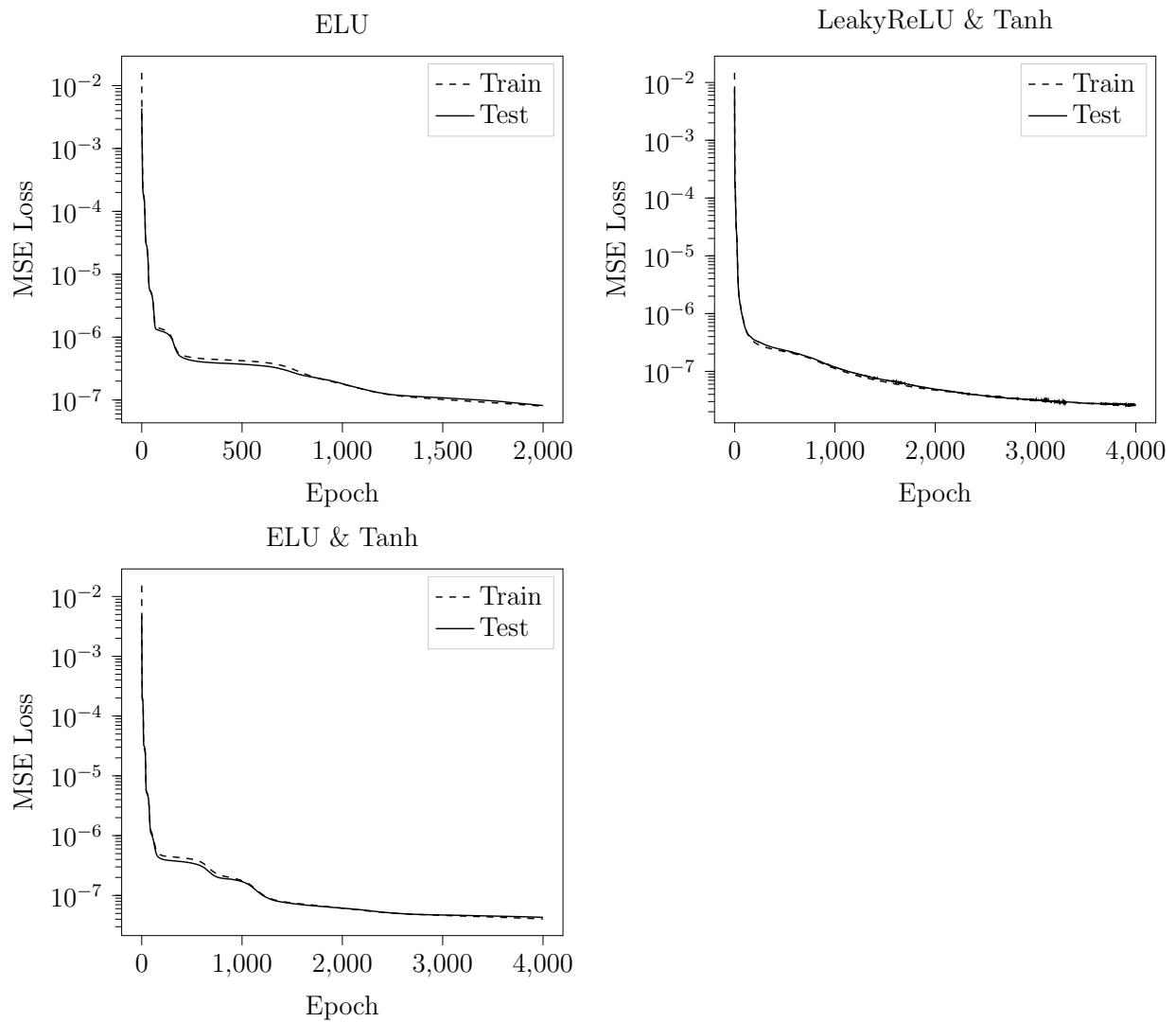


Figure 19: blablabla