CS230

Deep Learning for Physics Discovery
Winter 2021

Wai Tong Chung
wtchung@stanford.edu

Danyal Mohaddes Khorassani
danmohad@stanford.edu

Jacob Roy West
jrwest@stanford.edu

Abstract

A variational autoencoder is employed to gain insights into two physical systems: a particle settling under gravity in a viscous fluid, and a chemical reactor. The network architecture allows the interpretation of neuron activations in the latent layer in terms of physical parameters governing system behavior. The network takes as input a set of observations, a question regarding the state of nature under particular conditions, and an answer describing the state of nature under said conditions. The architecture was shown to yield accurate, interpretable results in both physical systems of interest, correctly discerning the parameters governing the physical systems and storing them in the latent neuron activations, and providing additional physical insight into the systems at hand.

1 Introduction

From Archimedes to Ibn-Khwarazmi to Newton, physicists have always relied upon a combination of observation of nature and deductive reasoning to discover the laws of nature. Physicists generally have access to a set of *observations* $\{o\}$, often from experimentation, that are indicative of a deterministic relationship among variables of interest. The objective of the physicist is then to determine the function f which represents the deterministic mapping between supposed physical conditions termed *questions* q to particular states of nature termed *answers* a, i.e. $f : q \to a$. Crucially, once f has been found, its *structure* is then analyzed, for it is the *interpretation* of this structure that yields insight into the inner workings of natural laws. That f be merely an accurate mapping from q to a is a necessary condition for meaningful physical analysis, but it is not sufficient.

Traditional neural network architectures used in deep learning are akin to "black-boxes", in that with a sufficiently large observational data set $\{o\}$ they can be trained to offer an accurate mapping function $f : q \to a$, but offer little insight. This is because the activations of hidden layers in a deep neural network, though containing the information necessary to achieve the aforementioned accurate mapping, represent a high-dimensional non-linear parameter space. This is in stark contrast to traditional physics-based models, where each model parameter represents specific physical mechanisms or their interaction. Consequently, physics-based models have a high degree of interpretability. Nevertheless, developments in interpretable deep learning have suggested a promising path towards automatic scientific discovery.

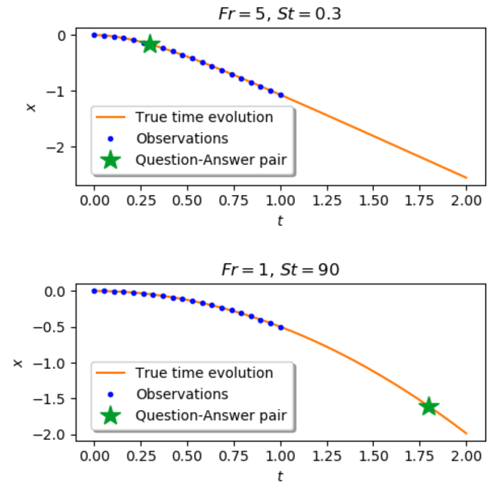


Figure 1: Example data for the particle motion problem.

A deep learning model called *Scinet*, which incorporates a variational autoencoder [1], was shown to automatically extract physical laws from several fundamental physics problems. When employed to predict planetary motion from raw geocentric planetary angles, the model's interpretable latent activations were shown to transform planetary angles from geocentric to heliocentric coordinates, thus demonstrating the network's ability to automatically recover necessary intermediate steps within complex calculations from unprocessed observations. To this end, we have applied the *SciNet* network architecture to extract interpretable scientific insight from canonical problems in fluid dynamics and combustion.

We will discuss the details of the network architecture in Sec. 4. Briefly, the inputs to the encoder network are a set of observations $|\{o\}| = N_o$, where each observation is a vector in $\mathcal{R}^n$ of observed variables $[v_1, v_2, ..., v_n]$. The *data* input to the decoder network is a question q, which is a variable which defines the condition under which the state of nature is to be inquired, with other inputs coming from other parts of the network. The output of the decoder network is the answer a, which is the state of nature under the conditions stipulated by the question q.

2 Related work

There has been growing interest in the use of what can be broadly considered "data-driven" techniques for the advancement of the field of physics, as opposed to the traditional "pen-and-paper" methodology. A conservative approach (conservative in the sense of similarity with traditional physical analysis) was taken by Floryan and Graham [2]. The authors employ a spectral decomposition technique and combined it

CS230: Deep Learning, Winter 2018, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

Encoder	Decoder	# Training samples	# Dev samples	# Test samples
[500,100]	[100,100]	200,000	10,000	10,000

Table 1: *Scinet* architecture and dataset breakdown used for the particle motion sub-problem.

Encoder	Decoder	# Training samples	# Dev samples	# Test samples
[500,100]	[100,100]	1500	250	250

Table 2: *Scinet* architecture and dataset breakdown used for the chemistry sub-problem.

with a data-driven approach to extract the persistent features of physical systems across large ranges of time and scale. A more radical approach was taken by Wu and Tegmark [3], who aimed to develop an "AI physicist". They combine a number of ideas from both artificial intelligence and classical physics to develop an "agent" whose purpose is to learn new theories of physics and attempt to unify them when presented with data in an unsupervised setting. The authors successfully demonstrate their model by applying their agent to accurately predict particle motion in domains governed by a harmonic potential field, a gravitational field and an electromagnetic field. An approach that lies between the two aforementioned extremes is that proposed by Iten et al. [1], upon which this project is based. The authors refrain from creating an agent-based automatic physics discovery system, but instead use an encoder-decoder architecture where a latent neuron layer is designed to capture physical insights in a fashion inspired by the process physicists employ in traditional analysis. The authors apply their methodology to various physical systems and demonstrate that their architecture is well-suited to capturing key insights into system behavior and stores them in the activations of the latent neurons, such as the number of parameters necessary to accurately describe system dynamics, and even coordinate transformations.

3 Dataset and Features

3.1 Particle Motion

The motion of a particle in a viscous fluid is typically described by ODEs relating the forces on the particle to its velocity [4]. These forces are parametrized by properties of the fluid and particle, and its velocity, which may not be constant in time. A very simple model for particle motion is described by two constant parameters, the Stokes number, St which describes the drag force, and the Froude number Fr, which describes the gravitational force. Given a particle initially at rest at position $x = 0$, its future position at time t is given by

$$x(t) = \mathrm{St}^2\mathrm{Fr}[1 - \frac{1}{\mathrm{St}}t - \exp(-t/\mathrm{St})] \tag{1}$$

Fr is sampled uniformly from the range $[0, 5]$ and St is sampled logarithmically from the range $[0.01, 100]$ to synthetically generate training data for this problem. The observations are a set of 50 $x(t)$, uniformly spaced in the range $t = [0, 1]$, evaluating equation 1. The "question" is a time randomly chosen in the range $t_q = [0, 2]$, and the "answer" is the position $x(t_q)$. This forces the decoder to extrapolate to a time where it does not have direct observations. Example observations, questions, and answers for two different parameter combinations are shown in Figure 1.

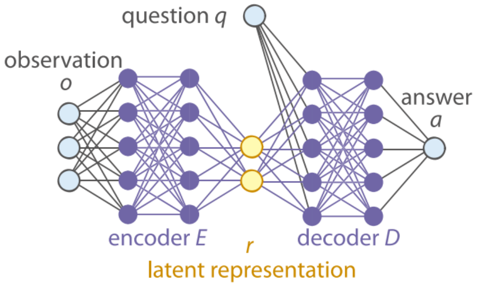


Figure 2: Architecture of *Scinet* (taken from [1], Figure 1b)

This problem setup is very similar to the pendulum problem studied in [1], and so the same architecture is used. The architecture details and dataset size are given in Table 1.

3.2 Chemistry

We apply the chemistry package Cantera [5] to calculate the formation of CO_2 and water from methane and oxygen with a single step chemistry in an isothermal zero-dimensional reactor:

$$\mathrm{CH_4} + 2\mathrm{O_2} \longrightarrow \mathrm{CO_2} + 2\mathrm{H_2O} \tag{2}$$

$$\frac{d[\mathrm{CO_2}]}{dt} = k_f(T)[\mathrm{CH_4}][\mathrm{O_2}]^2 \tag{3}$$

where the forward rate constant k_f is a function of temperature and the brackets $[\phi]$ refer to the concentration of scalar ϕ. 2000 samples of data are generated from range $T = [1000, 3000]$ K.

The "observation" data consists of CO_2 mass fraction Y_{CO2} as a function of time before time $t = [0, 12.5]$ μs, with a timestep of 0.5 μs. The "answer" consists of Y_{CO2} at $t = [0, 25]$ μs, and the "question" consists of 3 neurons corresponding to time, Y_{CH4}, and Y_{O2}. The training details for *Scinet* are provided in Table 2.

4 Methods

We employ the *Scinet* architecture in this study, shown in Figure 2. *Scinet* trains an encoder on observation data o, which condenses the input to a much smaller latent representation. Then, instead of reconstructing the input directly, the latent representation and question q are fed into a decoder network to provide an answer a. Users provide triples of (o, q, a) as training data. This structure mimics the scientific task of observing experiments, creating a model, and assessing the accuracy of the model predictions.

Training phase	Batch size	Learning rate	β	# Epochs
1	512	10^{-3}	10^{-3}	1,000
2	1024	10^{-4}	10^{-3}	500
3	1024	10^{-5}	10^{-3}	500

Table 3: Hyperparameters used to train *Scinet* on the particle motion sub-problem.

Since the goal of this network is a model which is both descriptive and interpretable, there are two parts to the *Scinet* loss function, which is based on the β-VAE cost function [6]:

$$Cost = \frac{1}{N_{batch}} \sum_{i=1}^{N_{batch}} \left(||\mathbf{a}_i - \hat{\mathbf{a}}_i||^2 - \frac{\beta}{2} \left(\sum_{j=1}^{N_{latent}} \left(\log(\sigma_{ij}^2) - \mu_{ij}^2 - \sigma_{ij}^2 - 1 \right) \right) \right) \tag{4}$$

where the first term represents the distance between the correct and predicted answers, **a** and **â**. The second term approximates the KL divergence under certain assumptions detailed in the supplemental material of [1]. μ_{ij} and σ_{ij} represent the mean and standard deviation of the latent activation distribution for latent neuron j on example i. β is a hyperparameter which controls the relative importance of the two objectives. Individual terms are summed over N_{batch} examples in a batch. Depending on the physical problem addressed, *Scinet* uses different hyperparameters, but generally speaking, it uses 2 fully connected layers each in the encoder and decoder. The Adam optimizer is used, with $\mathcal{O}(100,000)$ examples, 5% of which are used in the dev set.

5 Results and Discussion

5.1 Particle Motion

The training hyperparameters are largely adapted from the "pendulum" problem in [1], but also using multiple phases as in their "heliocentric" problem. The parameters used during training for all of the following results are shown in Table 3. Decreasing the learning rate and increasing the batch size through multiple phases of training was helpful in reducing the large oscillations in the loss function, which can be seen in Fig. 3. The difference between training and dev losses is consistently about 3 orders of magnitude smaller than the losses themselves, which indicates that overfitting is not a problem.

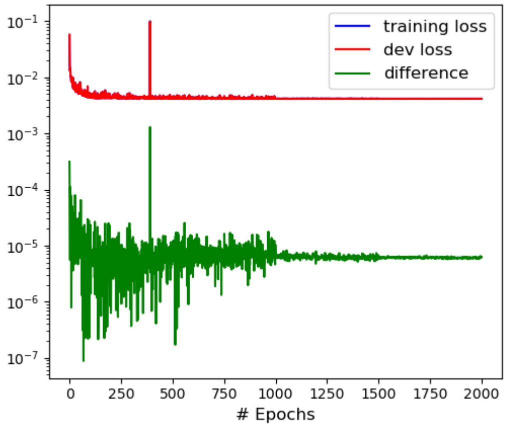


Figure 3: Training and dev losses during training of *Scinet* with 2 latent neurons, and the absolute value of their difference.

A key goal of *Scinet* is to correctly determine the minimum number of parameters needed to fully describe a dataset, where each parameter is represented by the activation of a different latent neuron. To assess the network's performance, we used the same procedure to train networks with $1-4$ latent neurons. The components of the loss function (reconstruction loss and KL divergence loss) are plotted against the number of latent neurons in Figure 4. This shows a substantial improvement in reconstruction loss going from 1 to 2 neurons, and then no added benefit. The KL divergence loss increases slightly when using more than 1 neuron, but this is somewhat expected as there are more latent neurons to sum over. The fact that the KL loss does not increase beyond 2 neurons indicates that the additional neurons are not active.

When the latent space is too small, the network is forced to learn a representation which is too compact to represent all the characteristics of the data. When the latent space is too large, *Scinet* should set those unecssary activations to zero. This behavior is demonstrated in Figure 5, which shows the latent activations on test data for networks with $1-3$ latent neurons, plotted against the input parameters. A few observations can be made:

- With more than 1 neuron, the ordering of the latent activations is probably a result of random initialization. Nothing was done to ensure similarity among the first neurons.
- The contours of the first activation are very similar across cases. However, with the 2 neuron network, the activation appears to be multiplied by -1. This is probably also a result of random initialization, and it indicates that the learned parameters could be simple transformations of the true parameters.
- The fact that the 1 neuron network has learned essentially the same representation as the first neurons of the other networks suggests that *Scinet* could be used to identify useful reduced models for physical phenomena, which have a more limited range of applicability.

Figure 5 can only be made when the input parameters are known. However, it suggests that for networks with several active latent neurons, the extracted feature can be understood by dropping out different neurons and assessing the kinds of errors the network makes. This may help suggest to the scientist the kinds of patterns the network has been able to find.

5.2 Chemistry

Training *Scinet* for the chemistry problem adopted a similar approach to the particle problem in section Section 5.1, with hyperparameters shown in Table 4. Since only 2000 datapoints are used for train, dev, and test sets, a smaller batch size has been adopted. Training is found to have converged around 200 epochs as shown in Fig. 6b. The difference between training and dev losses is consistently about 2 orders of magnitude smaller than the losses themselves, which indicates that overfitting is not a problem. Additionally, the predictions from *Scinet* are found to be in good agreement with the correponding true CO_2 mass fractions as shown in Fig. 6a.

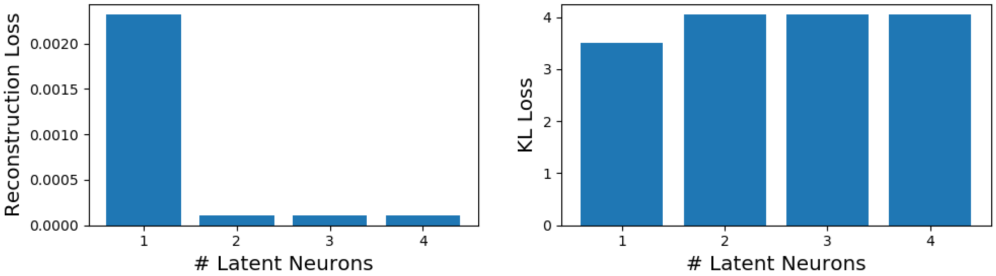


Figure 4: Components of the loss function evaluated on test dataset as a function of the number of latent neurons in the particle problem.

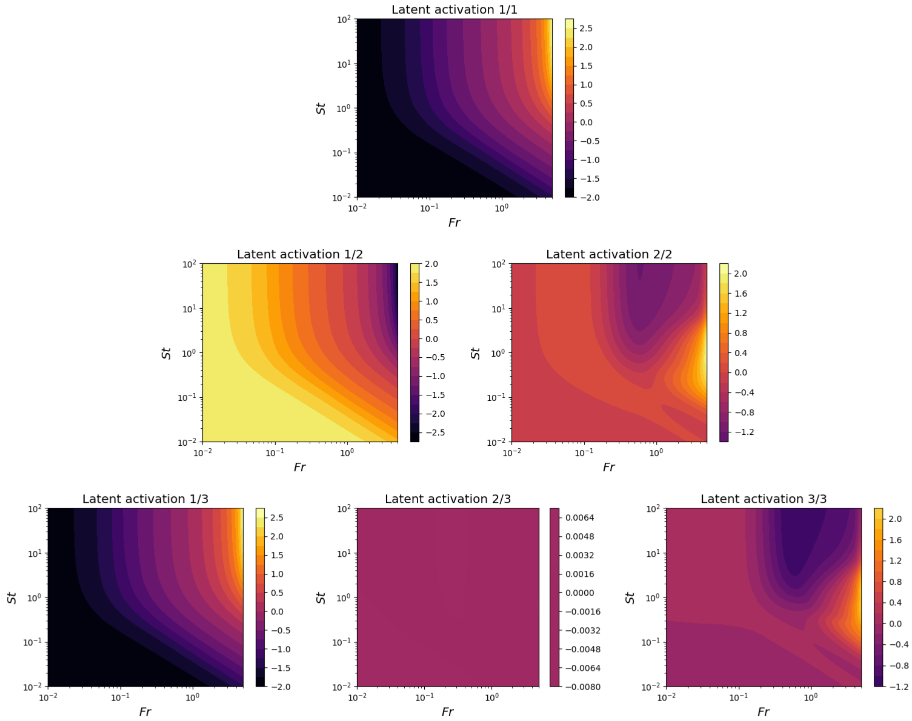
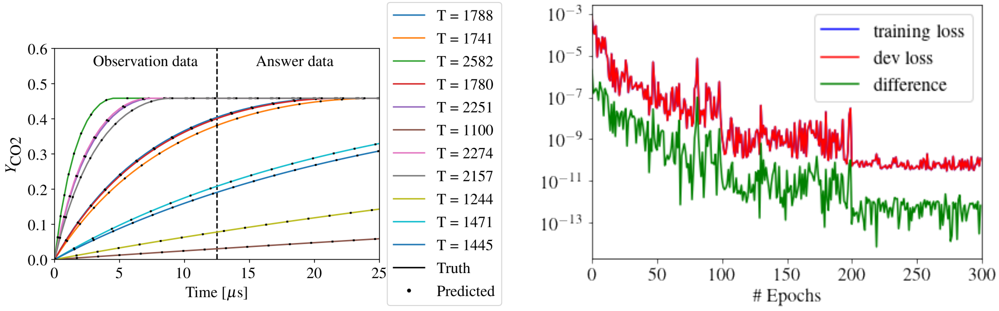


Figure 5: Comparison of latent space activations as a function of the governing parameters and the number of latent neurons in the particle problem.



(a)

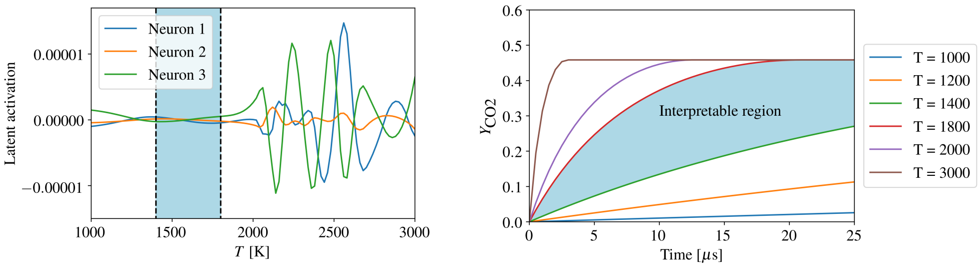
(b)

Figure 6: (a) Predictions and (b) losses from the chemistry problem.

Training phase	Batch size	Learning rate	β	# Epochs
1	32	10^{-3}	10^{-3}	100
2	32	10^{-4}	10^{-3}	100
3	32	10^{-5}	10^{-3}	100

Table 4: Hyperparameters used to train *Scinet* on the chemistry sub-problem.

Figure 7a shows the latent space activations for the range of temperatures explored in the chemistry problem. At first glance, no discernible relationship is observed between the latent activations and temperature. However, the behavior latent space can be divided into 3 regions. At $T < 1400$ K, the formation of CO_2 is incomplete for the timeframe investigated, and can be described as a linear relationship. Thus, the latent space will not fully describe the reaction involved. At $T > 1800$ K, the autoencoder is asked to predict CO_2 formation after the reaction is completed, where the CO_2 mass fraction is constant. Thus, the latent space does not provide meaningful information on the reaction. In essence, meaningful information can only be extracted when $T = [1400, 1800]$ K, which is highlighted in blue in Fig. 7. This is shown in Fig. 8 where all three latent space activations possess a strong linear relationship with the forward rate constant of the reaction k_f, which is the sole parameter in Eq. (3), in the interpretable region.



(a) Latent activation vs temperature.

(b) CO2 mass fraction vs temperature.

Figure 7: Interpretable regions of the dataset in the chemistry problem.

6 Conclusion/Future Work

A variational autoencoder network architecture called *SciNet* was applied to two problems in physics: that of a particle settling in a viscous fluid under gravity, and chemical reactor. It was sought to use the network to glean physical insights regarding the nature of the problems considered through the analysis of the latent space in the trained network. In the former problem, the network was shown to correctly identify the number of parameters necessary to fully describe the system, in that the addition of latent neurons beyond the two required did not result in improved accuracy. Furthermore, upon consideration of the latent neuron activations, it was clear that the network set the activation of the additional neurons to zero, allowing the clear interpretation of the activations as the correct parametrization of the physical system. In the latter problem, the network allowed the segmentation of the parameter space into *interpretable* and *non-interpretable* regions, which followed directly from the system dynamics. In particular, the interpretable region corresponded to a regime where the latent neuron activations varied near-linearly with the known system parameters, showing that the network discovered the key variables controlling system behavior. In the non-interpretable region, the latent neuron activations did not yield information regarding system parameters, which we interpreted to mean that system behavior was physically agnostic to the known system parameters in this regime.

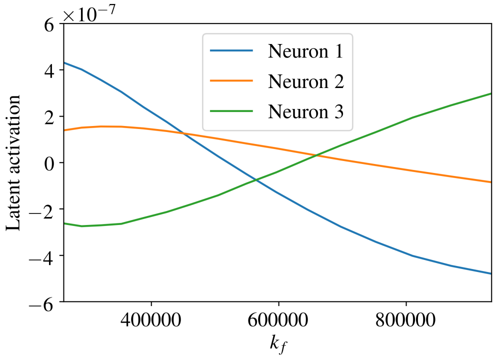


Figure 8: Latent activation vs forward rate constant in the interpretable region.

Future work would focus primarily on the interpretation of results when the governing parameters are not known beforehand. In particular, it is not immediately clear how to interpret the latent neuron activations when one does not know what physical parameters one seeks to interpret. Prior knowledge of the physical system could potentially be applied to divine the correct interpretation of the activations, but a more methodical and data-driven approach should be developed. Another important avenue for future work is the analysis of stochastic systems. Presently, our analysis has focused on purely deterministic systems, and the ability to extract and interpret system parameters in stochastic problems would be beneficial in a variety of subfields in physics.

7 Contributions

WTC developed the chemistry *SciNet* application. DMK developed a thermodynamics *SciNet* application (discussed in the proposal and milestone) and contributed to the dataset generation for the chemistry application. JRW developed the particle motion *SciNet* application. All authors contributed to analysis and report writing.

8 References

[1] R. Iten, T. Metger, H. Wilming, L. del Rio, and R. Renner, "Discovering physical concepts with neural networks," *Phys. Rev. Lett.*, vol. 124, p. 010508, Jan 2020.

[2] D. Floryan and M. D. Graham, "Discovering multiscale and self-similar structure with data-driven wavelets," *Proceedings of the National Academy of Sciences*, vol. 118, no. 1, 2021.

[3] T. Wu and M. Tegmark, "Toward an ai physicist for unsupervised learning.(2018)," *arXiv preprint arXiv:1810.10525*, 1810.

[4] M. R. Maxey and J. J. Riley, "Equation of motion for a small rigid sphere in a nonuniform flow," *The Physics of Fluids*, vol. 26, no. 4, pp. 883–889, 1983.

[5] D. G. Goodwin, R. L. Speth, H. K. Moffat, and B. W. Weber, "Cantera: An object-oriented software toolkit for chemical kinetics, thermodynamics, and transport processes," 2018. https://www.cantera.org.

[6] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, "beta-vae: Learning basic visual concepts with a constrained variational framework," 2016.