Ill posed problems and invertible neural networks

Seminar: Vielteilchentheorie Schwerpunkt Machine learning

Giorgio Lovato

07.02.2023

Content

Consider the following process:

$$\boldsymbol{f} : \mathbb{R}^{\mathrm{n}} \to \mathbb{R}^{\mathrm{m}}, \boldsymbol{x} \to \boldsymbol{y},$$

where the calculation of $\boldsymbol{y}$ for a given $\boldsymbol{x}$ is straight forward. What about the inverse function

$$\boldsymbol{g} = \boldsymbol{f}^{-1} : \mathbb{R}^{\mathrm{m}} \to \mathbb{R}^{\mathrm{n}}, \boldsymbol{y} \to \boldsymbol{x}?$$

In the general case it is not trivial to deduce $\boldsymbol{x}$ from $\boldsymbol{y}$ even if the forward process is known.

If for example the intrinsic dimension k of $\mathbf{y}$ is smaller than the nominal dimension n of $\mathbf{x}$

$\rightarrow$ Information is lost during the forward process

$\rightarrow$ Any $\mathbf{y}$ may be assigned a large number of different $\mathbf{x}$

$\Rightarrow$ The problem is *ill-posed*

If for example the intrinsic dimension k of y is smaller than the nominal dimension n of x

$\rightarrow$ Information is lost during the forward process

$\rightarrow$ Any y may be assigned a large number of different x

$\Rightarrow$ The problem is *ill-posed*

However this situation is actually quite common in physics:

- A set of quantities is measured/numerically obtained (y)
- From this the hidden variables of a model ($f(x)$) are to be inferred
- But the inverse calculation is expensive, difficult or not even known

An example from statistical mechanics

QMC calculations yield values of dynamic correlation functions at discrete imaginary times τ_{m}
These $G(\tau_{\mathrm{m}})$ (imaginary time Green function) relates to different spectral functions $A(\omega)$ via

$$G(\tau_{\mathrm{m}}) = \int_{-\infty}^{\infty} \mathcal{K}(\tau_{\mathrm{m}}, \omega) A(\omega) \mathrm{d}\omega,$$

which is called a Fredholm integral equation of the first kind
The kernel $\mathcal{K}$ is known and depends on the specific G, A

An example from statistical mechanics

QMC calculations yield values of dynamic correlation functions at discrete imaginary times τ_{m}

These $G(\tau_{\mathrm{m}})$ (imaginary time Green function) relates to different spectral functions $A(\omega)$ via

$$G(\tau_{\mathrm{m}}) = \int_{-\infty}^{\infty} \mathcal{K}(\tau_{\mathrm{m}}, \omega) A(\omega) \mathrm{d}\omega,$$

which is called a Fredholm integral equation of the first kind

The kernel $\mathcal{K}$ is known and depends on the specific G, A

Example:

$$\mathcal{K}(\tau_{\mathrm{m}}, \omega) = -\frac{\exp(-\tau_{\mathrm{m}}\omega)}{\exp(-\beta\omega) \pm 1}$$

While forward process $A(\omega) \rightarrow G(\tau_\mathrm{m})$ hard but doable, inverse is very complicated
In fact: naive solution by least squares fit yields

While forward process $A(\omega) \to G(\tau_\mathrm{m})$ hard but doable, inverse is very complicated
In fact: naive solution by least squares fit yields

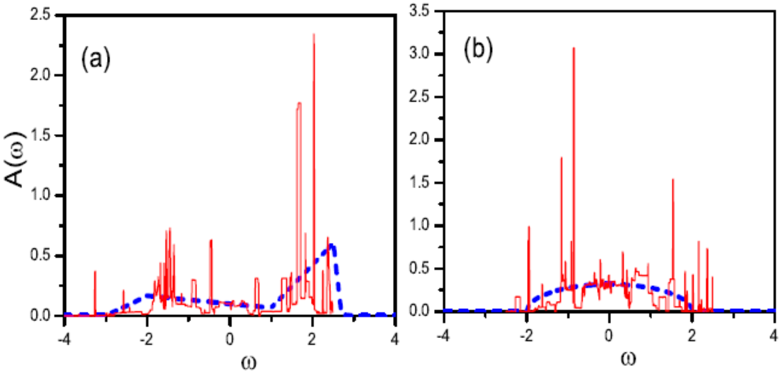

Figure: Red line: fit; Blue dashed line: actual spectrum (Andrey S. Mishchenko, 2012)

Of course there exists a multitude of methods (see Andrey S. Mishchenko (2012) for overview):

- Tikhonov-Phillips regularization method
- Maximum entropy method
- Stochastic optimization method

However in the following we will discuss

Invertable Neural Networks (INN)

Invertible neural networks

Idea

- Construct a NN which architecture allows for trivial inversion
- Train the INN on the known and simple forward process as well as the inverse
- Record lost information in the forward process via a latent variable

$\rightarrow$Ardizzone et al. (2018)

What is known

measurement: $y \in \mathbb{R}^{m}$, system state: $x \in \mathbb{R}^{n}$,

theoretical model: $y = f(x)$

What is known

measurement: $\boldsymbol{y} \in \mathbb{R}^{\mathrm{m}}$, system state: $\boldsymbol{x} \in \mathbb{R}^{\mathrm{n}}$,
theoretical model: $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$

As the model $\boldsymbol{f}(\boldsymbol{x})$ imposes an information loss, we seek the conditional probability distribution $p(\boldsymbol{x}|\boldsymbol{y})$

A mathematical derivation is not feasible

Invertible neural networks

Problem specification

What is known

measurement: $\boldsymbol{y} \in \mathbb{R}^{\mathrm{m}}$, system state: $\boldsymbol{x} \in \mathbb{R}^{\mathrm{n}}$,
theoretical model: $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$

As the model $\boldsymbol{f}(\boldsymbol{x})$ imposes an information loss, we seek the conditional probability distribution $p(\boldsymbol{x}|\boldsymbol{y})$

A mathematical derivation is not feasible

What is done

The actual probability density is instead approximated by $q(\boldsymbol{x}|\boldsymbol{y})$ using a neural network trained on an arbitrary amount of $(\boldsymbol{x}, \boldsymbol{f}(\boldsymbol{x}))$ with a suitable prior for $\boldsymbol{x}$

Invertible neural networks

Problem specification

Training

Given x the network computes y^*

Given y the network computes $q(x|y)$

Both results are then evaluated and the network is updated correspondingly

Invertible neural networks

Problem specification

Training

Given $\boldsymbol{x}$ the network computes $\boldsymbol{y}^*$
Given $\boldsymbol{y}$ the network computes $q(\boldsymbol{x}|\boldsymbol{y})$
Both results are then evaluated and the network is updated correspondingly

Latent variable

While training the forward process:
Latent variable $\boldsymbol{z}$ is randomly drawn from normal distribution to encode lost information
This ensures that the inverse process is a deterministic function $g(\boldsymbol{y}, \boldsymbol{z}) = \boldsymbol{x}$
Thus all hidden parameters can be recovered

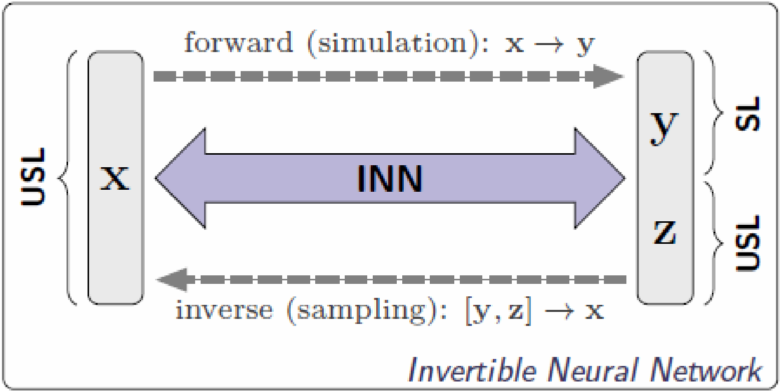

Figure: Schematic visualization (Ardizzone et al., 2018)

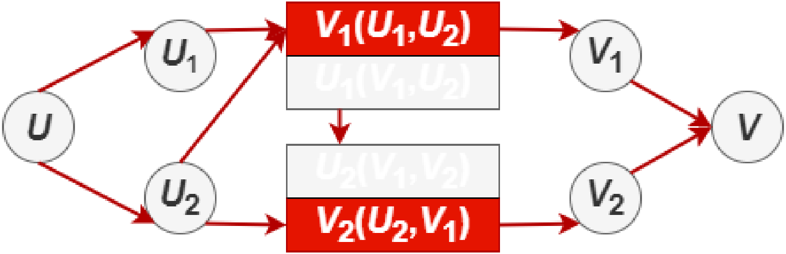

Figure: Forward process

$$
\begin{aligned}
\boldsymbol{V}_1 &= \boldsymbol{U}_1 \odot \exp[\boldsymbol{s}_2(\boldsymbol{U}_2)] + \boldsymbol{t}_2(\boldsymbol{U}_2) \\
\boldsymbol{V}_2 &= \boldsymbol{U}_2 \odot \exp[\boldsymbol{s}_1(\boldsymbol{V}_1)] + \boldsymbol{t}_1(\boldsymbol{V}_1)
\end{aligned}
$$

Invertible neural networks

Architecture

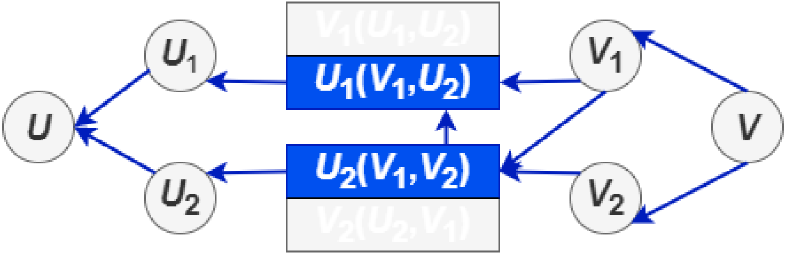

Figure: Inverse process

$$
\begin{aligned}
\boldsymbol{U}_2 &= [\boldsymbol{V}_2 - \boldsymbol{t}_1(\boldsymbol{V}_1)] \odot \exp[-\boldsymbol{s}_1(\boldsymbol{V}_1)] \\
\boldsymbol{U}_1 &= [\boldsymbol{V}_1 - \boldsymbol{t}_2(\boldsymbol{U}_2)] \odot \exp[-\boldsymbol{s}_2(\boldsymbol{U}_2)]
\end{aligned}
$$

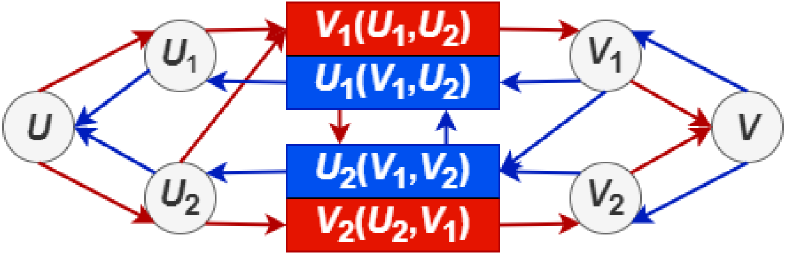

Figure: Both processes

$$V_1 = U_1 \odot \exp[s_2(U_2)] + t_2(U_2)$$
$$V_2 = U_2 \odot \exp[s_1(V_1)] + t_1(V_1)$$
$$U_2 = [V_2 - t_1(V_1)] \odot \exp[-s_1(V_1)]$$
$$U_1 = [V_1 - t_2(U_2)] \odot \exp[-s_2(U_2)]$$

The functions $s_{1,2}$ and $t_{1,2}$

- Arbitrarily complex functions
- Encode the *Neural Network* part

The functions $s_{1,2}$ and $t_{1,2}$

- Arbitrarily complex functions
- Encode the *Neural Network* part

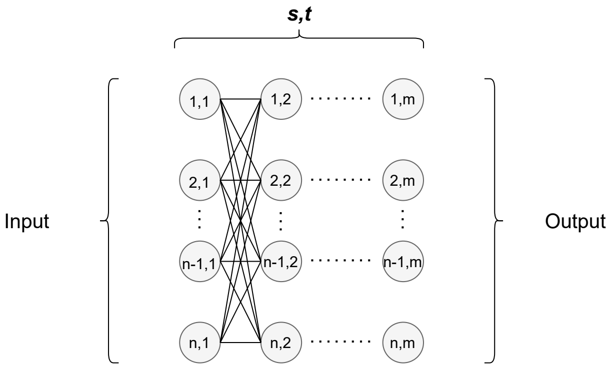

Figure: Representation of s, t

Recap

- Network is made up of interconnected modules
- Each module guarantees invertibility
- Within those modules: *classical NN*
- Fully connected NN with width n and depth m
- Training changes parameters in $\boldsymbol{s}$, $\boldsymbol{t}$

Gaussian Mixture Model (GMM)

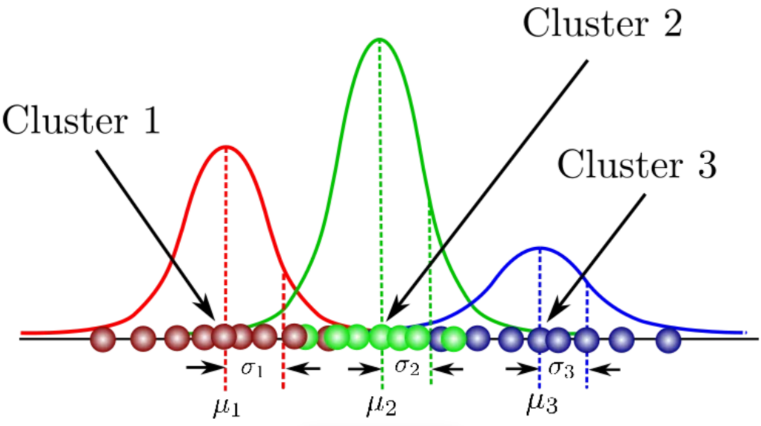

Figure: Representation of a GMM with 3 components [Carrasco (2019)]

Gaussian Mixture Model (GMM)

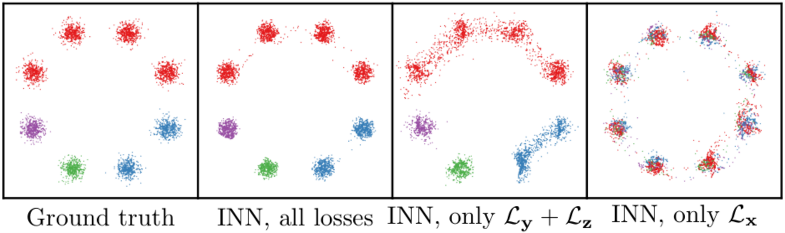

Figure: Different INN trained on GMM with 8 components and 4 labels [Ardizzone et al. (2018)]

Oxygen saturation of living tissue

The reflective spectrum of tissue depends on:

- oxygen saturation s_{O_2}
- blood volume fraction ν_{hb}
- scattering magnitude a_{mie}
- anisotropy g
- tissue layer thickness d

$s_{O_2} \rightarrow$ **Tumor detection**

Oxygen saturation of living tissue

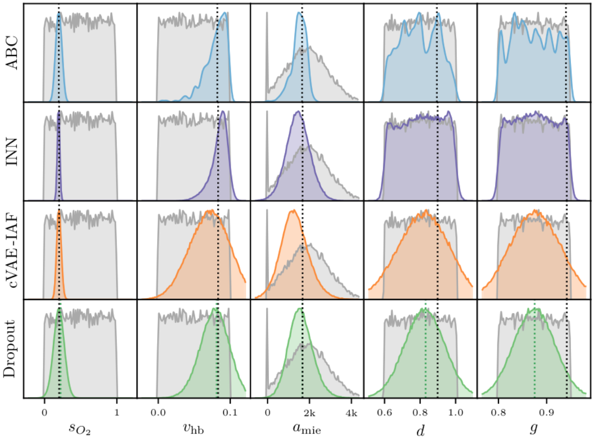

Figure: Results of INN compared to competitors [Ardizzone et al. (2018)]

Star cluster evolution

- Simulation of a star cluster
- At each time step calculation of spectra y
- As well as:
  - Ionizing Luminosity
  - Ionizing Emission Rate
  - Cloud Density
  - Expansion Velocity
  - Age of cluster

$\Rightarrow$ **Effect on surrounding gas cloud**

Star cluster evolution

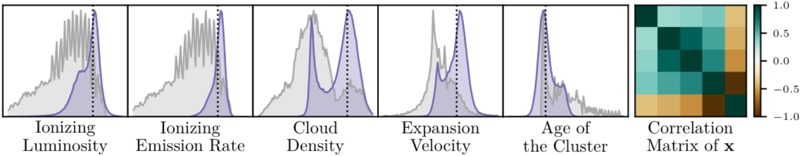

Figure: Parameters of a specific time step in a simulation [Ardizzone et al. (2018)]

Conclusion

- In science *ill posed problems* are faced
- They are difficult to solve with high ambiguity
- Specific architectures allow for invertibility in NN
- This can be used to train a NN on the forward and inverse process simultaneously
- Combined with a latent variable, multimodalities, correlations and unrecoverable variables can be identified

$\Rightarrow$ As a next step, apply this procedure to analytic continuation

References

Published as a conference paper at ICLR 2019

Analyzing Inverse Problems with Invertible Neural Networks

Lynton Ardizzone[1], **Jakob Kruse**[1], **Sebastian Wirkert**[2],
Daniel Rahner[3], **Eric W. Pellegrini**[3], **Ralf S. Klessen**[3],
Lena Maier-Hein[2], **Carsten Rother**[1], **Ullrich Köthe**[1]
[1]Visual Learning Lab Heidelberg, [2]German Cancer Research Center (DKFZ),
[3]Zentrum für Astronomie der Universität Heidelberg (ZAH)
[1]`lynton.ardizzone@iwr.uni-heidelberg.de`,
[2]`s.wirkert@dkfz-heidelberg.de`, [3]`daniel.rahner@uni-heidelberg.de`

Abstract

For many applications, in particular in natural science, the task is to determine hidden system parameters from a set of measurements. Often, the forward process from parameter- to measurement-space is well-defined, whereas the inverse problem is ambiguous: multiple parameter sets can result in the same measurement. To fully characterize this ambiguity, the full posterior parameter *distribution*, conditioned on an observed measurement, has to be determined. We argue that a particular class of neural networks is well suited for this task – so-called Invertible Neural Networks (INNs). Unlike classical neural networks, which attempt to solve the ambiguous inverse problem directly, INNs focus on learning the forward process, using additional latent output variables to capture the information otherwise lost. Due to invertibility, a model of the corresponding inverse process is learned implicitly. Given a specific measurement and the distribution of the latent variables, the inverse pass of the INN provides the full posterior over parameter space. We prove theoretically and verify experimentally, on artificial data and real-world problems from medicine and astrophysics, that INNs are a powerful analysis tool to find multi-modalities in parameter space, uncover parameter correlations, and identify unrecoverable parameters.

References

Andrey S. Mishchenko, Erik Koch Frithjof Anders Mark Jarrell, Eva Pavarini (2012), *Stochastic Optimization Method for Analytic Continuation*. Verlag des Forschungszentrum Jülich.

Ardizzone, Lynton, Jakob Kruse, Sebastian Wirkert, Daniel Rahner, Eric W. Pellegrini, Ralf S. Klessen, Lena Maier-Hein, Carsten Rother, and Ullrich Köthe (2018), "Analyzing inverse problems with invertible neural networks." URL `https://arxiv.org/abs/1808.04730`.

Carrasco, Oscar Contreras (2019), "Gaussian mixture models explained from intuition to implementation." URL `https://towardsdatascience.com/gaussian-mixture-models-explained-6986aaf5a95`.