Automatic Differentiation for Reconstructing Spectral Functions with Neural Networks

Speaker: **Lingxiao Wang** (FIAS)

with: **Shuzhe Shi** (McGill U. and SBU) and **Kai Zhou** (FIAS)

accepted* by** *NeurIPS 2021* ***"Machine Learning and Physical Science" workshop

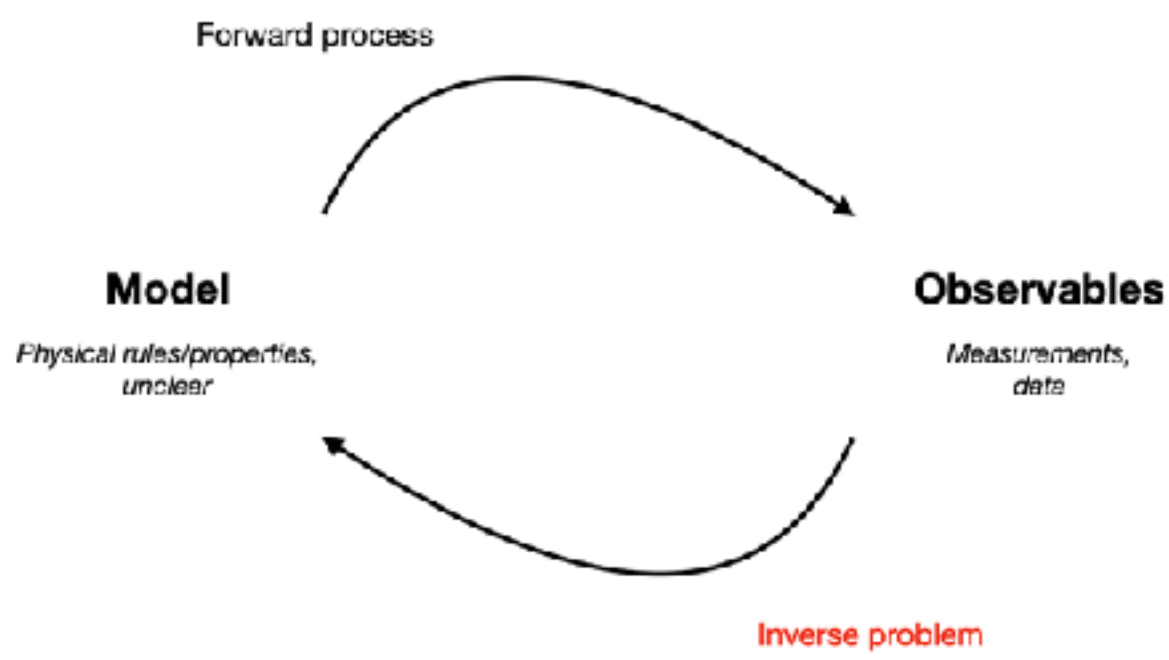
physics paper is preparing

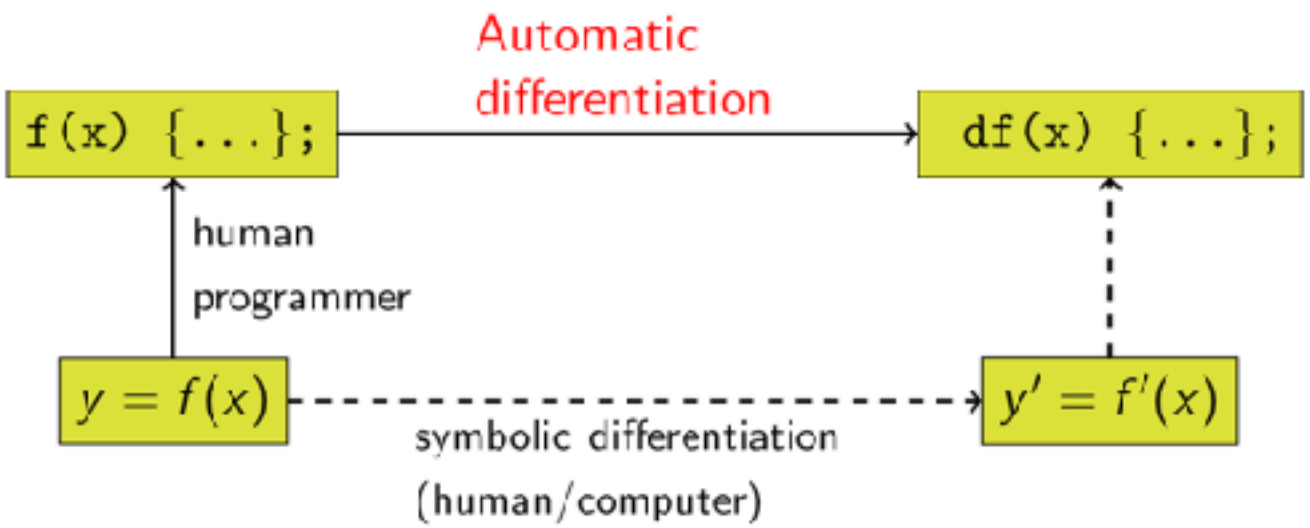
09 Nov. 2021, ***"Cold Quantum Coffee"*** seminar, ITP of Heidelberg University

Main Content

AD for Inverse Problem

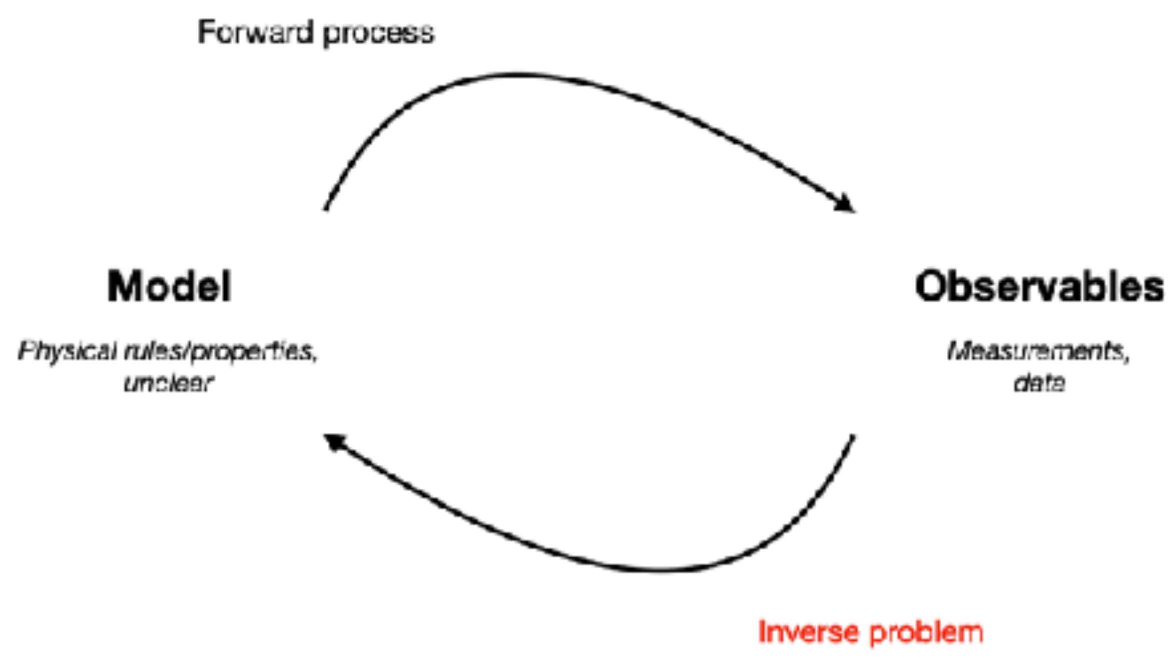
- Inverse Problem
  - Ill-posed problem
  - Reconstruct spectral functions
- AD for reconstruction
  - AD framework
    - Automatic Differentiation
    - Neural Network representations
  - Preliminary results
- Summary





Inverse problem

Inverse Problem



- **Mathematics**
- **IP in Physics**
  - QCD physics
  - Condensed matter physics
  - Optics
- **IP in Science and Technology**
  - Signal processing
  - Epidemiology
  - Material design
  - …

$$O = M(p) + N$$

- *O:* Observations/Outcomes
- *M:* Model/Function
- *p:* Parameters/Incomes
- *N:* Noise in real-world

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*.

Inverse Problem

Examples

- **Ill-posed problem**
  - A small error in the initial mearuments can result in large deviations of reconstructions
  - Noise and discontinuities of observed data
- **Medical physics**
  - From x-ray intensity to tissue image
- **Image processing**
  - Deblurring
- **Numerical differentiation**

Hadamard (1865–1963): A problem is called well-posed,
if **a solution exists, the solution is unique,**
and the solution depends continuously on the given data.

Computerized tomography (CT)

$\overset{F}{\longmapsto}$

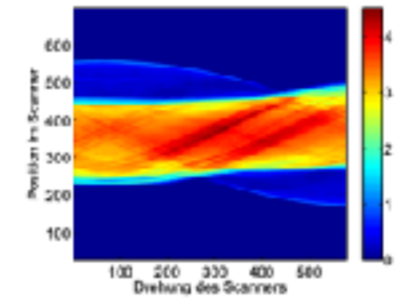


Image — Measurements

Direct problem: Simulate/predict the measurements
(from knowledge of the interior density distribution)
Given x calculate $F(x) = y$*!*

Inverse problem: Reconstruct/image the interior distribution
(from taking x-ray measurements)
Given y solve $F(x) = y$*!*

Image deblurring

$\overset{F}{\longmapsto}$

x
True image

$y = F(x)$
Blurred image

Direct problem: Simulate/predict the blurred image
(from knowledge of the true image)
Given x calculate $F(x) = y$*!*

Inverse problem: Reconstruct/image the true image
(from the blurred image)
Given y solve $F(x) = y$*!*

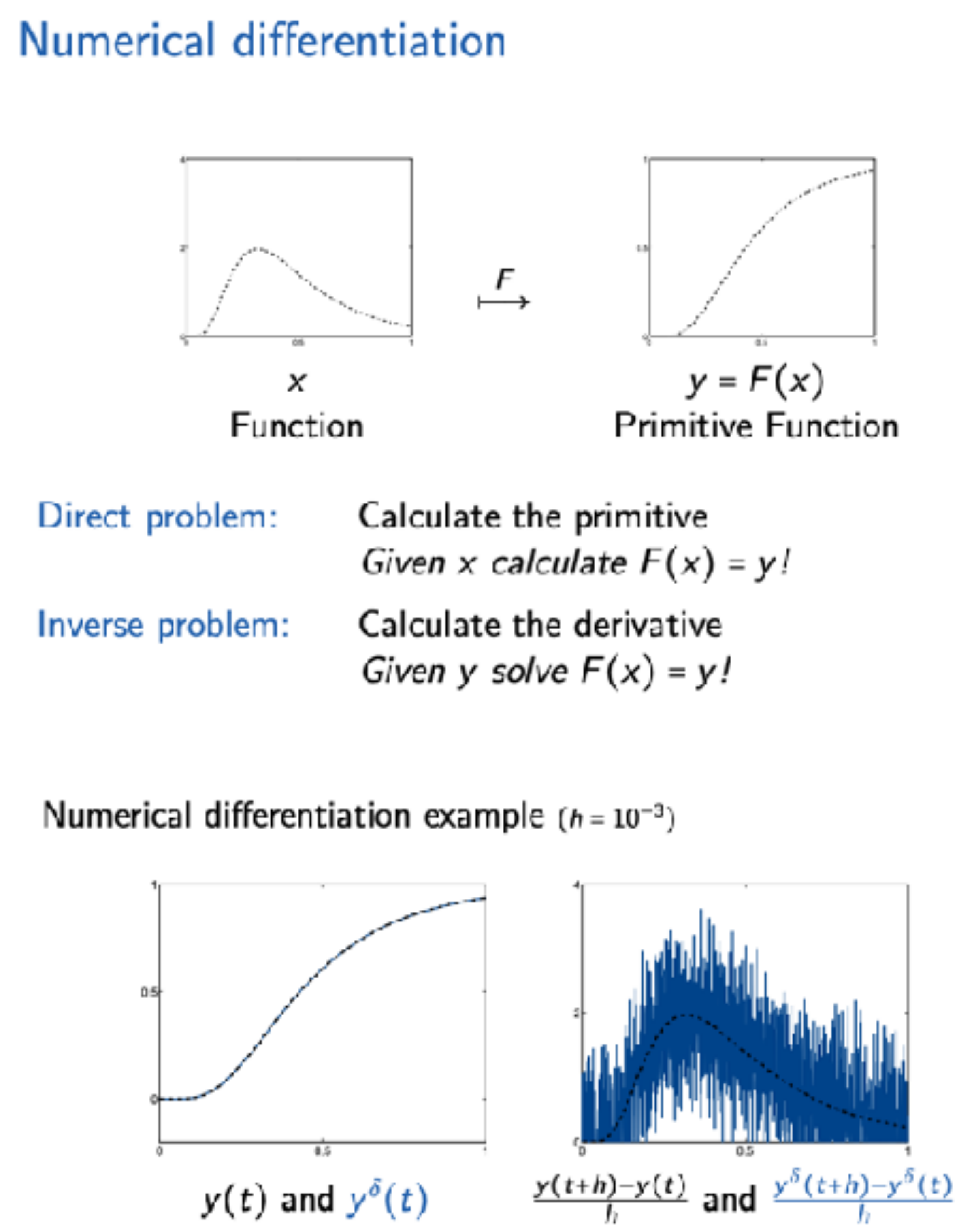
B. Harrach (2015). *Introduction to Inverse Problems*

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*.

Inverse Problem

Examples

- **Ill-posed problem**
    - A small error in the initial mearuments can result in large deviations in reconstructions
    - Noise and discontinuities of observed data
- **Numerical differentiation**
    - *F* is a integral operation
    - step length $h = 10^{-3}$
    - rebuilt error $\delta \to 0, y^{\delta} \to \hat{y}$
- Optics, Condensed matter, QCD Physics,…



B. Harrach (2015). *Introduction to Inverse Problems*

Hadamard (1865–1963): A problem is called well-posed,
if **a solution exists, the solution is unique,**
and the solution depends continuously on the given data.

Inverse Problem

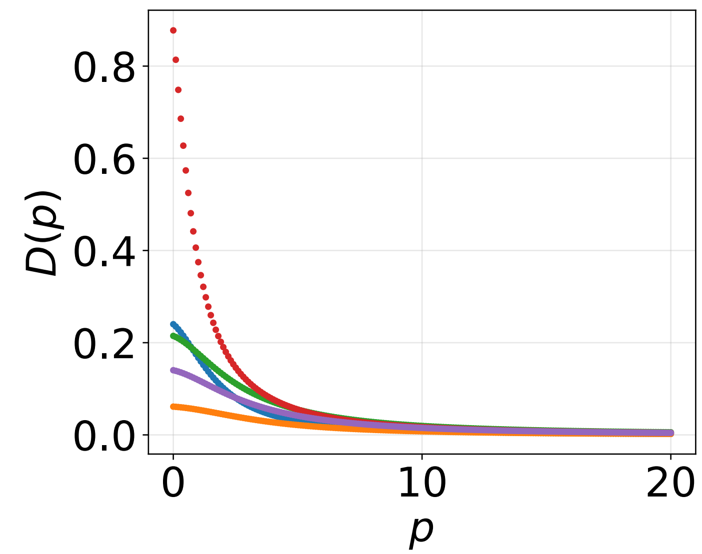
Rebuild spectral functions

- **Real-time properties of strongly correlated quantum systems**
  - Time has to be analytically continued into the complex plane
  - Reconstruct the spectral function from **noisy Euclidean propagator** data (e.g.,Lattice QCD) to extract their physical structures
- Methods
  - **Classical methods**
    - Truncated Singular Value Decomposition (TSVD)
    - Tikhonov regularization, …

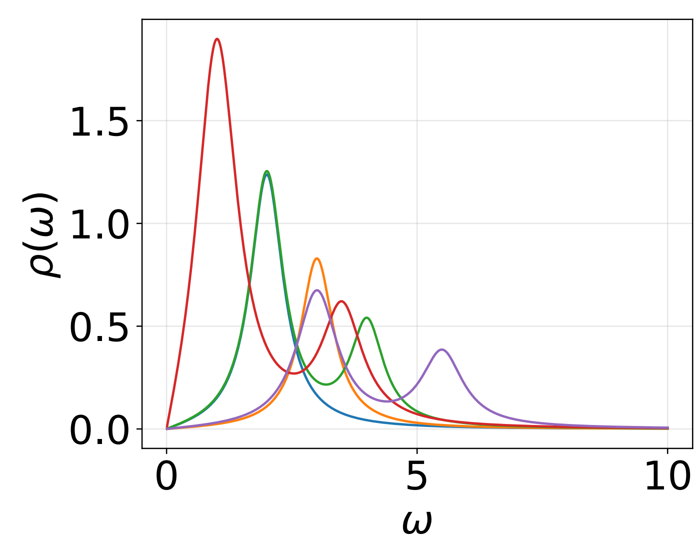
    H. W. Engl and C. W. Groetsch, editors , *Inverse and Ill-Posed Problems* (Academic Press, Boston, 1987).
  - **Baysian methods: Maximum Entropy Method (MEM)**

    M. Jarrell and J. E. Gubernatis, *Bayesian Inference and the Analytic Continuation of Imaginary-Time Quantum Monte Carlo Data*, Physics Reports **269**, 133 (1996).

    M. Asakawa, Y. Nakahara, and T. Hatsuda, *Maximum Entropy Analysis of the Spectral Functions in Lattice QCD*, Progress in Particle and Nuclear Physics **46**, 459 (2001).
  - Supervised learning inverse mapping
  - Gausian process, Radial Basis Functions(RBF), sVAE



$$D(p) \equiv \int_0^\infty K(p,\omega)\rho(\omega)\mathrm{d}\omega$$



Inverse Problem

Rebuild spectral functions

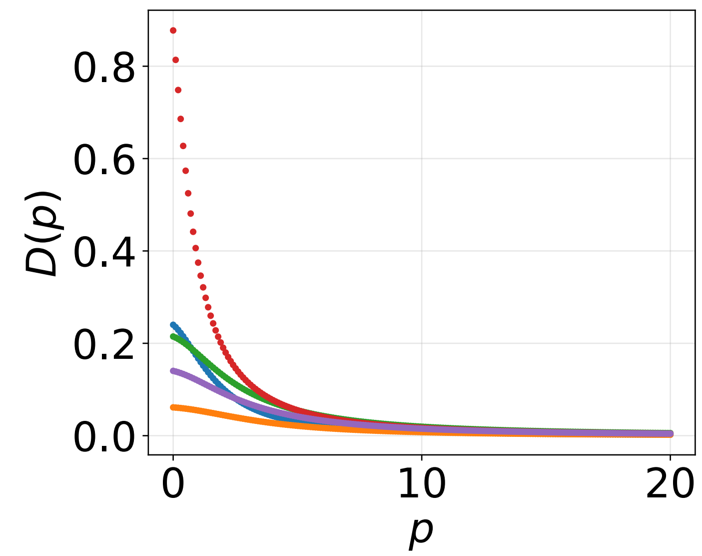
- **Real-time properties of strongly correlated quantum systems**
  - Time has to be analytically continued into the complex plane
  - Reconstruct the spectral function from **noisy Euclidean propagator** data (e.g.,Lattice QCD) to extract their physical structures
- Methods
  - Classical methods
  - Baysian methods: Maximum Entropy Method (MEM)
  - **Supervised Learning** inverse mapping

    L. Kades, J. M. Pawlowski, A. Rothkopf, M. Scherzer, J. M. Urban, S. J. Wetzel, N. Wink, and F. P. G. Ziegler, ***Spectral Reconstruction with Deep Neural Networks***, Phys. Rev. D **102**, 096001 (2020).

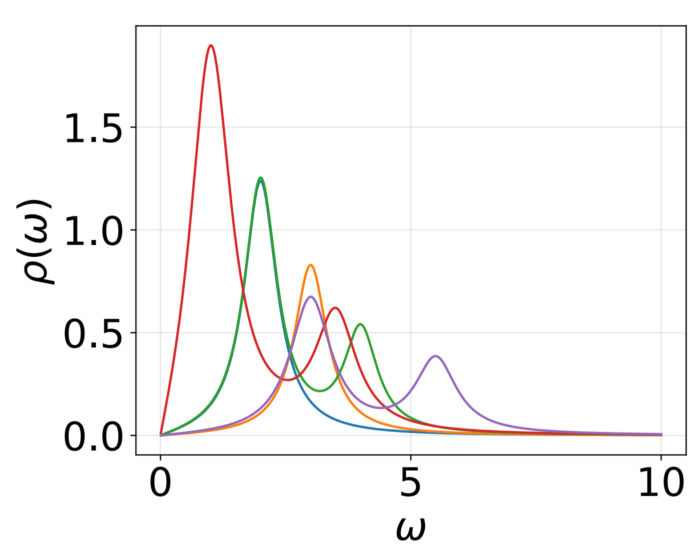
    R. Fournier, L. Wang, O. V. Yazyev, and Q. Wu, ***Artificial Neural Network Approach to the Analytic Continuation Problem***, Phys. Rev. Lett. **124**, 056401 (2020).

    H. Yoon, J.-H. Sim, and M. J. Han, ***Analytic Continuation via Domain Knowledge Free Machine Learning***, Phys. Rev. B **98**, 245101 (2018).
  - **Gausian process, Radial Basis Functions(RBF), sVAE(Variational AutoEncoder)**

    ArXiv:2107.13464, ArXiv:2106.08168, ArXiv:2110.13521



$$D(p) \equiv \int_0^\infty K(p,\omega)\rho(\omega)\mathrm{d}\omega$$



Inverse Problem

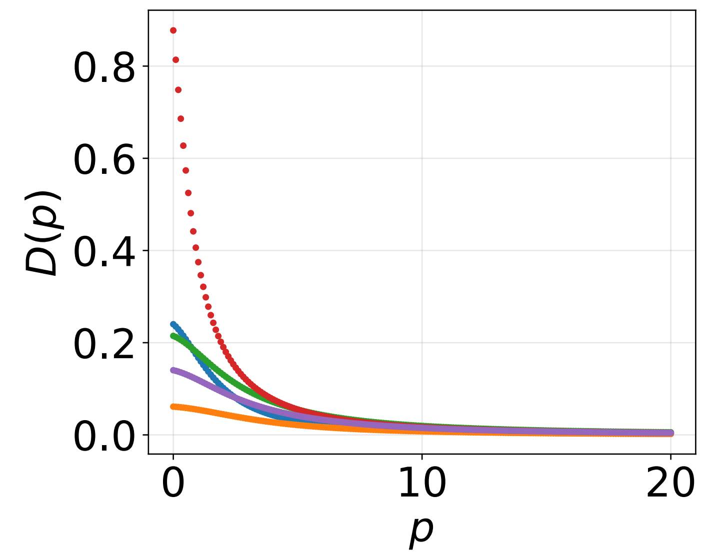
Mock data

- **Real-time properties of strongly correlated quantum systems**
  - Reconstruct the spectral function from **noisy Euclidean propagator** data (e.g.,Lattice QCD) to extract their physical structures
- Mock Data
  - *Kallen—Lehmann(KL)* representation

$$D(p) = \int_0^\infty \frac{\omega\,\rho(\omega)}{\omega^2 + p^2}\frac{\mathrm{d}\omega}{\pi}$$

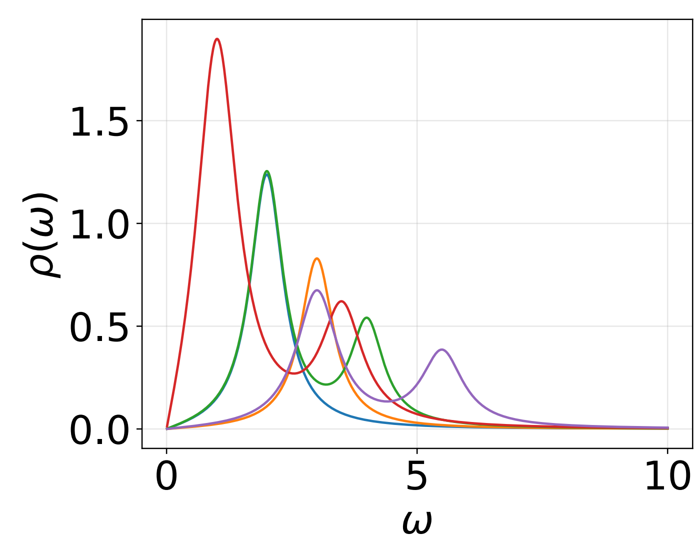
  - Breit-Wigner peaks $\rho^{(\mathrm{BW})}(\omega) = \frac{4A\Gamma\omega}{\left(M^2 + \Gamma^2 - \omega^2\right)^2 + 4\Gamma^2\omega^2}$

A parametrization obtained directly from one-loop perturbative quantum field theory:
A: amplitude, Γ: width, M: mass



+ Gaussian Noise

$$D(p) \equiv \int_0^\infty K(p, \omega)\rho(\omega)\mathrm{d}\omega$$



Automatic Differentiation for Reconstruction

Framework

Automatic differentiation

- **Automatic differentiation (AD)**
  - It refers to a general way of taking a program which **computes a value**, and **automatically constructing a procedure for computing derivatives of that value.**
  - Example

How we compute the derivatives of logistic least squares regression in a net

ω weights, b bias, $\sigma(z)$ activation function
x input, y output, t target, $\mathcal{L}$ loss function

Chain rule: $h'(x) = f'(g(x))g'(x).$

Computing the loss:

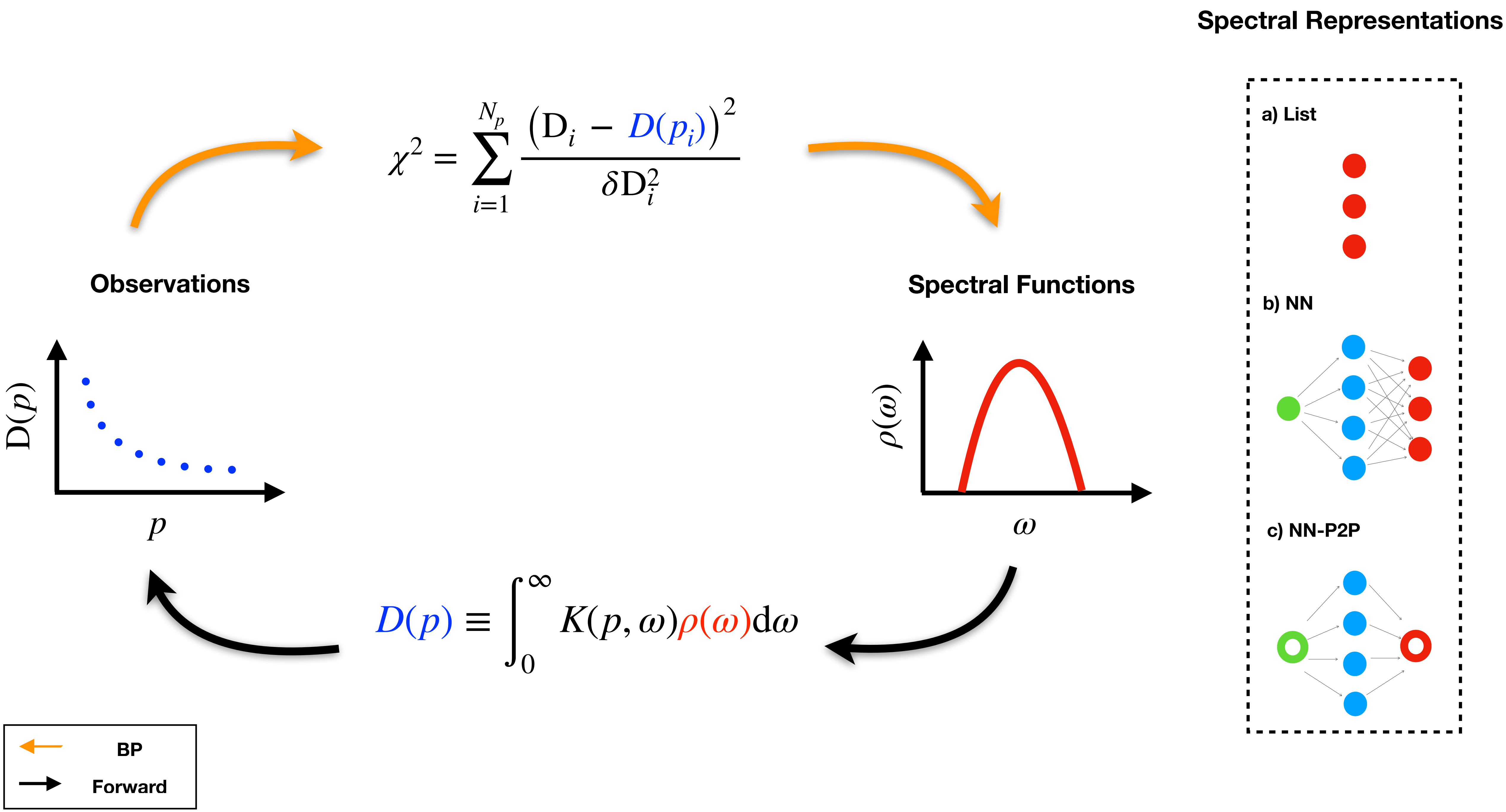
$$z = wx + b$$
$$y = \sigma(z)$$
$$\mathcal{L} = \frac{1}{2}(y - t)^2$$

Computing the derivatives:

$$\overline{\mathcal{L}} = 1$$
$$\bar{y} = y - t$$
$$\bar{z} = \bar{y}\sigma'(z)$$
$$\bar{w} = \bar{z}x$$
$$\bar{b} = \bar{z}$$

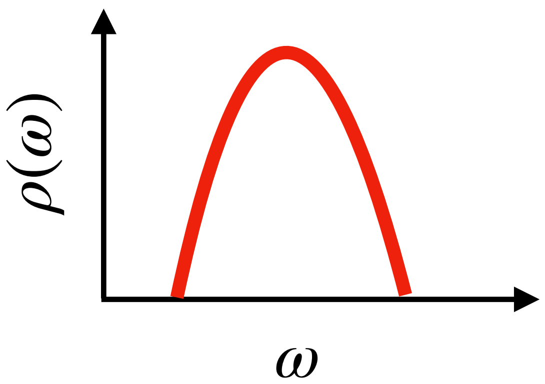
Framework

Automatic differentiation



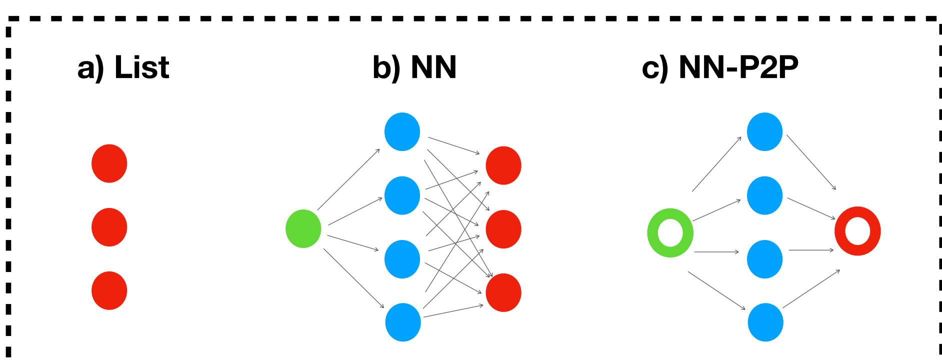
Framework

Automatic differentiation



$$D(p) \equiv \int_0^\infty K(p,\omega)\rho(\omega)\mathrm{d}\omega$$

$$L = \sum_{i=1}^{N_p} \frac{\left(\mathrm{D}_i - D(p_i)\right)^2}{\delta D_i^2}$$



$$\frac{\partial \Delta}{\partial \theta}$$

Gradient-based Optimization

Regularization

$$\Delta = \frac{\partial L}{\partial D(p)} K(p,\omega)$$

a) **List :** $(\rho_1, \rho_2, \cdots, \rho_{N_\omega})$

b) **NN :** $(\rho_1, \rho_2, \cdots, \rho_{N_\omega})$

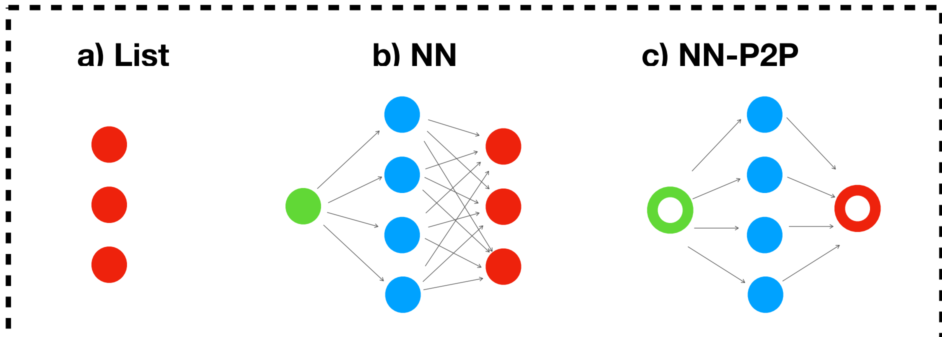
c) **NN-P2P :** $\rho_i(\omega_i)$

BP

Forward

Framework

Back-propagation



Gradient-based Optimization

Regularization

$$\Delta = \frac{\partial L}{\partial D(p)} K(p, \omega)$$

a) List : $(\rho_1, \rho_2, \cdots, \rho_{N_\omega})$

Differentiable variables : ρ_i

Adam, L2 ($\lambda = 10^{-3} \rightarrow 0$), Smoothness ($\lambda_s = 10^{-4} \rightarrow 0$)

b) NN : $(\rho_1, \rho_2, \cdots, \rho_{N_\omega})$

Differentiable variables : Network weights $\{\theta\}$

Adam, L2 ($\lambda = 10^{-3} \rightarrow 0$), Smoothness ($\lambda_s = 10^{-4} \rightarrow 0$)

c) NN-P2P : $\rho_i(\omega_i)$

Differentiable variables : Network weights $\{\theta\}$

Adam, L2 ($\lambda = 10^{-6} \rightarrow 0$)

Gradient-based Optimization

Adam : $\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t$

Regularization

L2 : $\lambda ||\theta||_2^2$

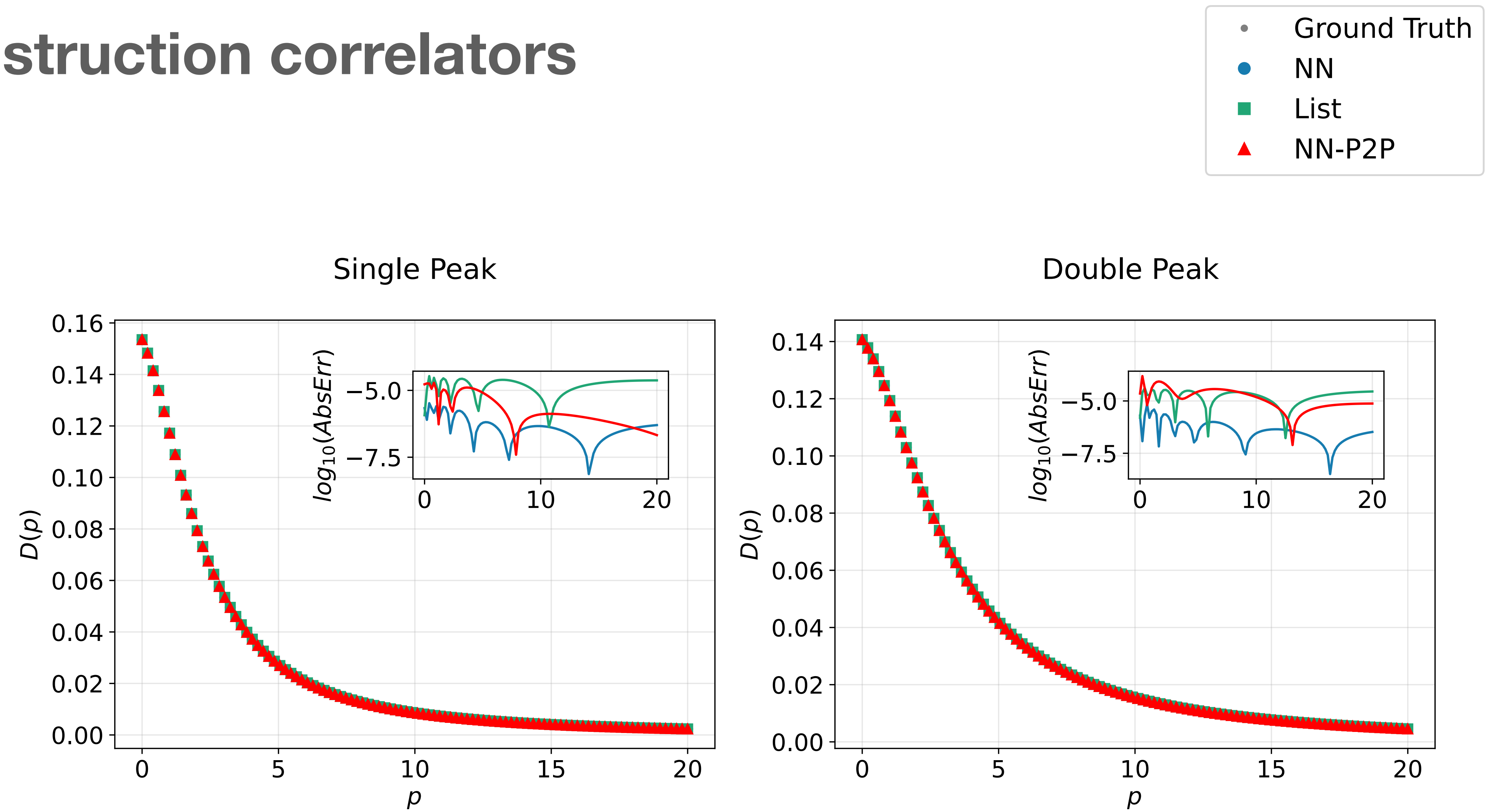
Smoothness: $\lambda_s \sum_i^{N_\omega} (\rho_i - \rho_{i-1})^2$

Physical Prior

Positive-defined condition: Softplus $log(1 + e^x)$

Preliminary results

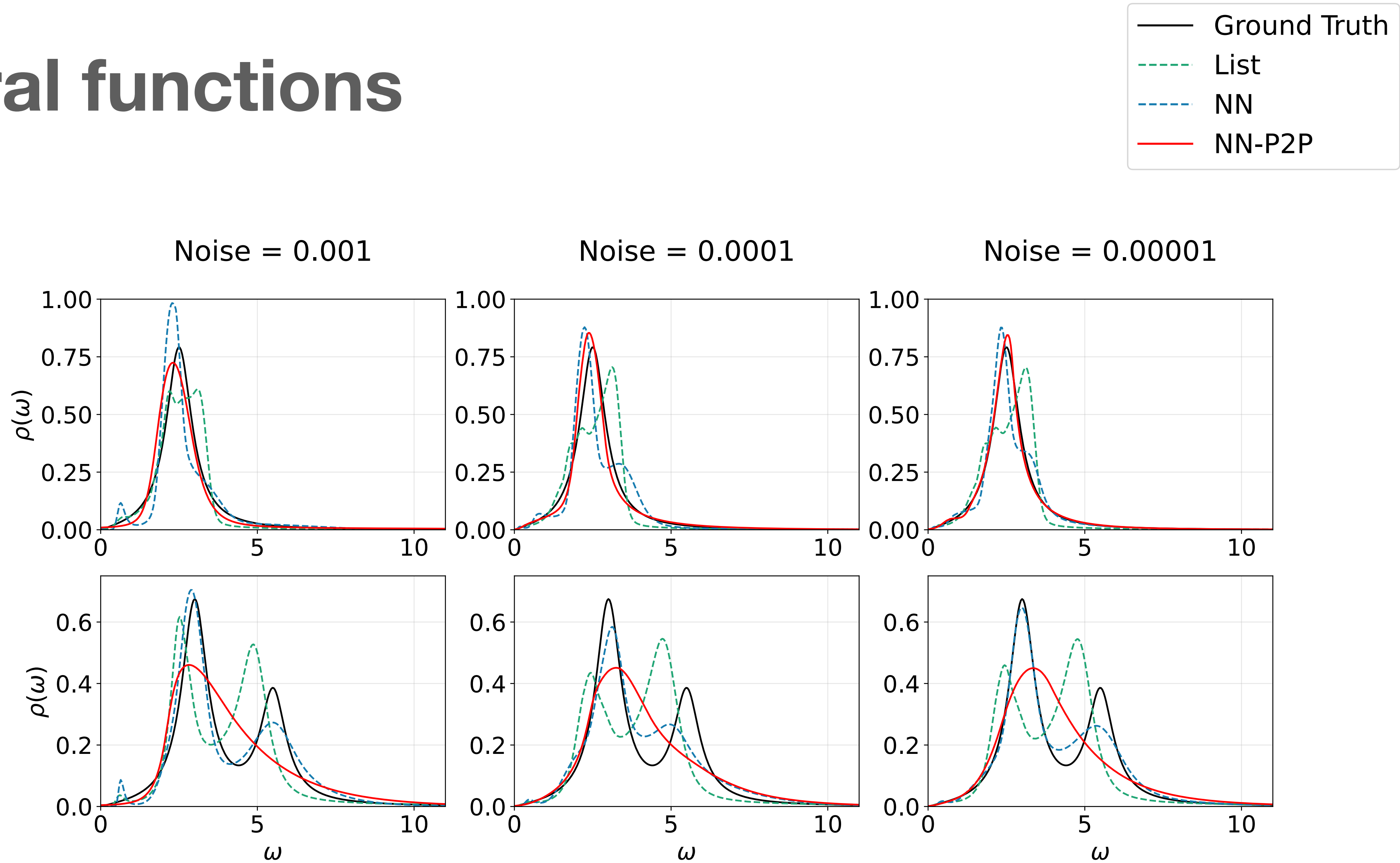
Reconstruction correlators



- Two samples : spectral functions with single or double Breit-Wigner peaks
- **Reconstruction absolute error** $|\mathrm{D}_i - D(p_i)| < 10^{-5}$ as the same magnitude of noise

Preliminary results

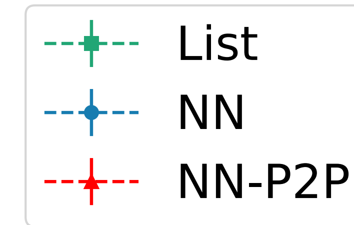
Spectral functions



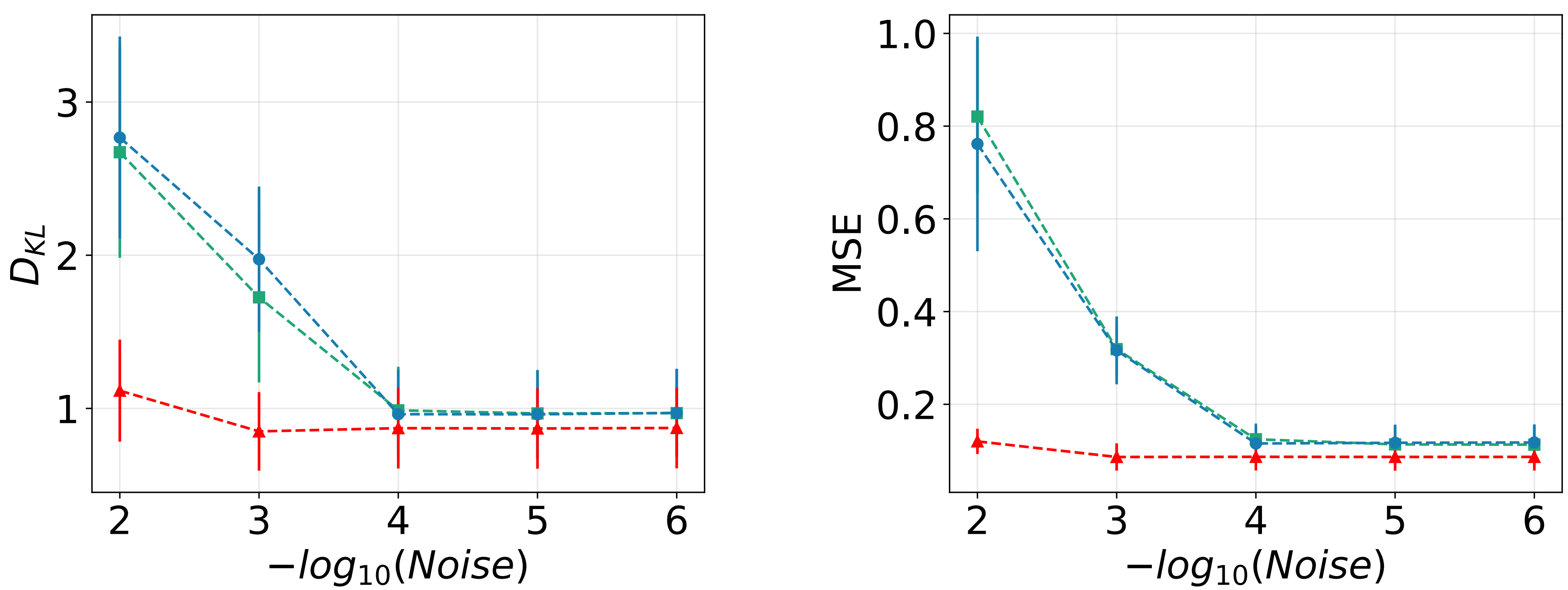
- **Reconstruction performance** will be better with noise decreasing
- **NN-P2P** gets the best consistency in single peak case
- NN and List can represent a more diverse spectrum in double peak case

Preliminary results

Quantified Evaluation



$$D_{KL}(p\|q) = \int \mathrm{d}\mathbf{x}\, p(\mathbf{x}) \log \frac{p(\mathbf{x})}{q(\mathbf{x})}$$

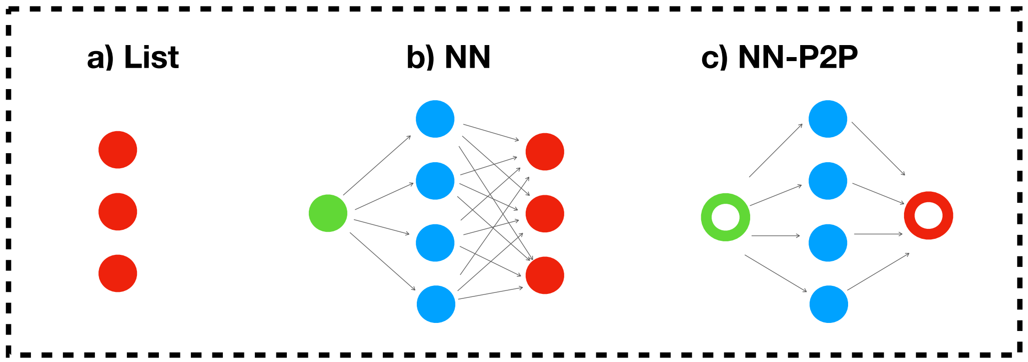


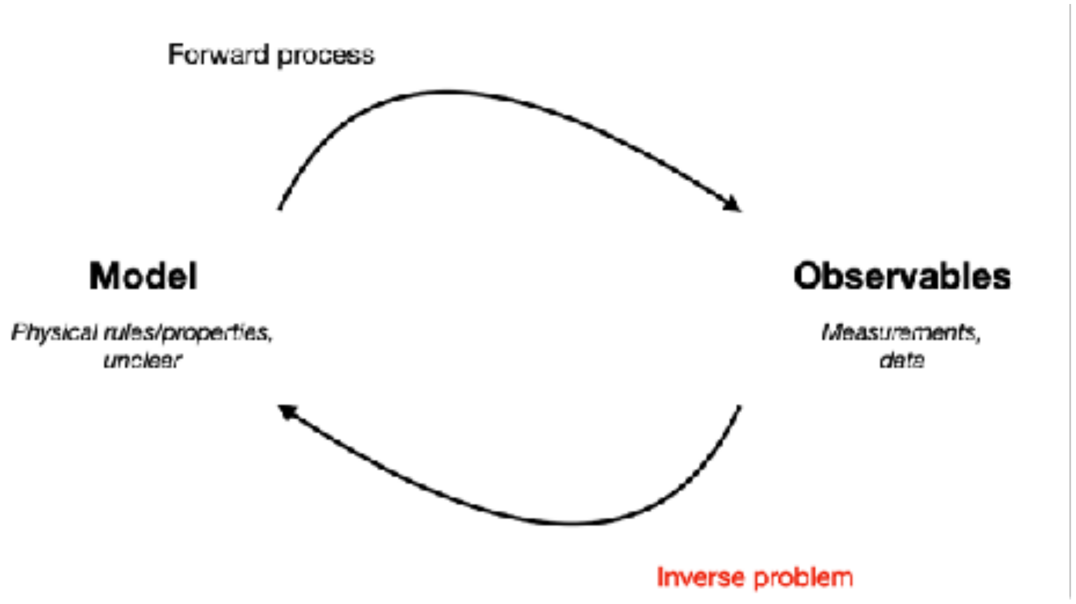
- 100 spectra with mixed single and double peaks
- K-L divergence and MSE decreasing with noise decreasing
- **NN-P2P** is better than others, because its intrinsic continuity of one-to-one mapping

Summary

Summary
and Outlooks

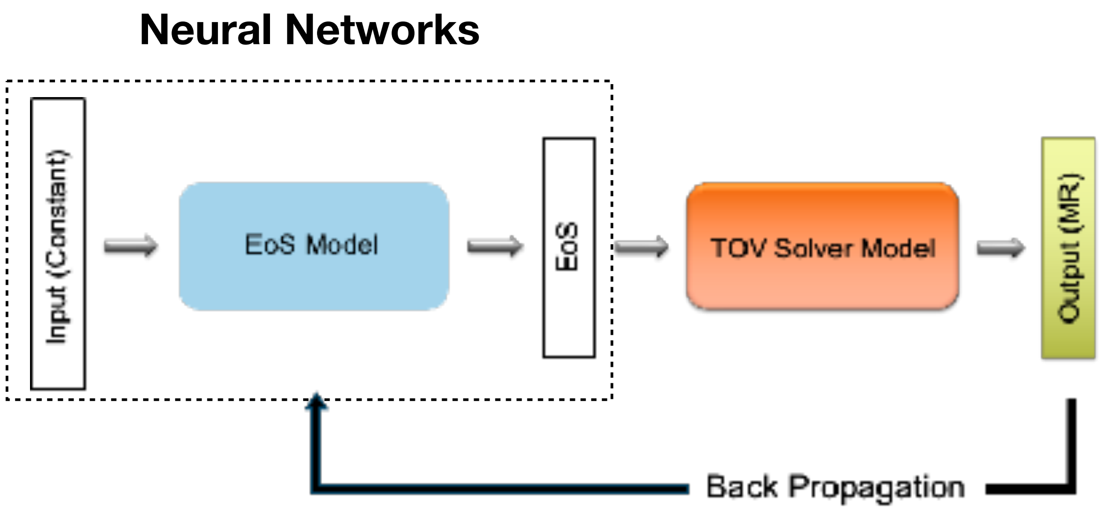
- **Take-home messages**
  - AD can solve inverse problem using error information unsupervisedly
  - Neural network representations can help us to embed physical regularization into reconstructing
- **Future works**
  - Optimization algorithm
  - Maximum Entropy Method (MEM)
- Related works
  - Neutron Star, *in preparing*
  - Bottonium potential, *arXiv:2105.07862*

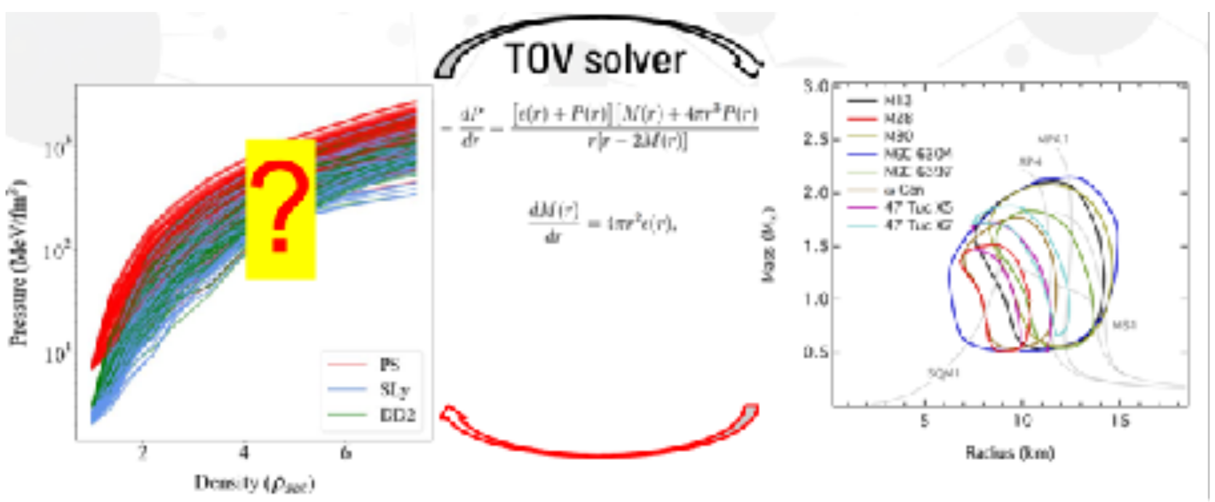




Summary
and Outlooks

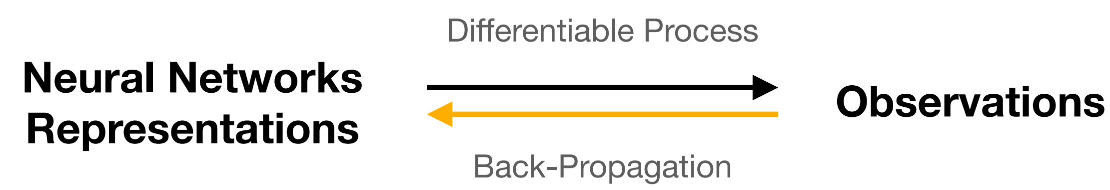
- **Take-home messages**
  - AD can solve inverse problem using error information unsupervisedly
  - Neural network representations can help us to embed physical regularization into reconstructing
- **Future works**
  - Optimization algorithm
  - Maximum Entropy Method (MEM)
- Related works
  - **Neutron Star,** ***in preparing***
  - Bottomonium potential, *arXiv:2105.07862*

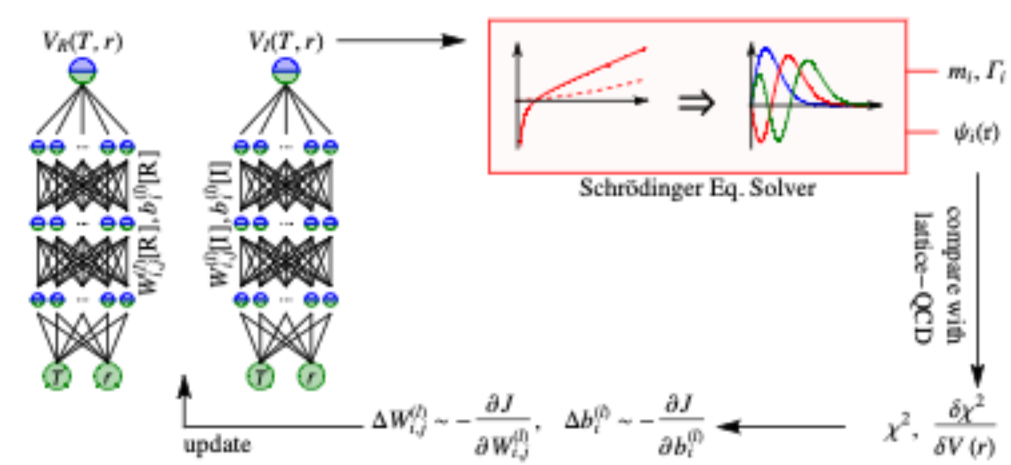


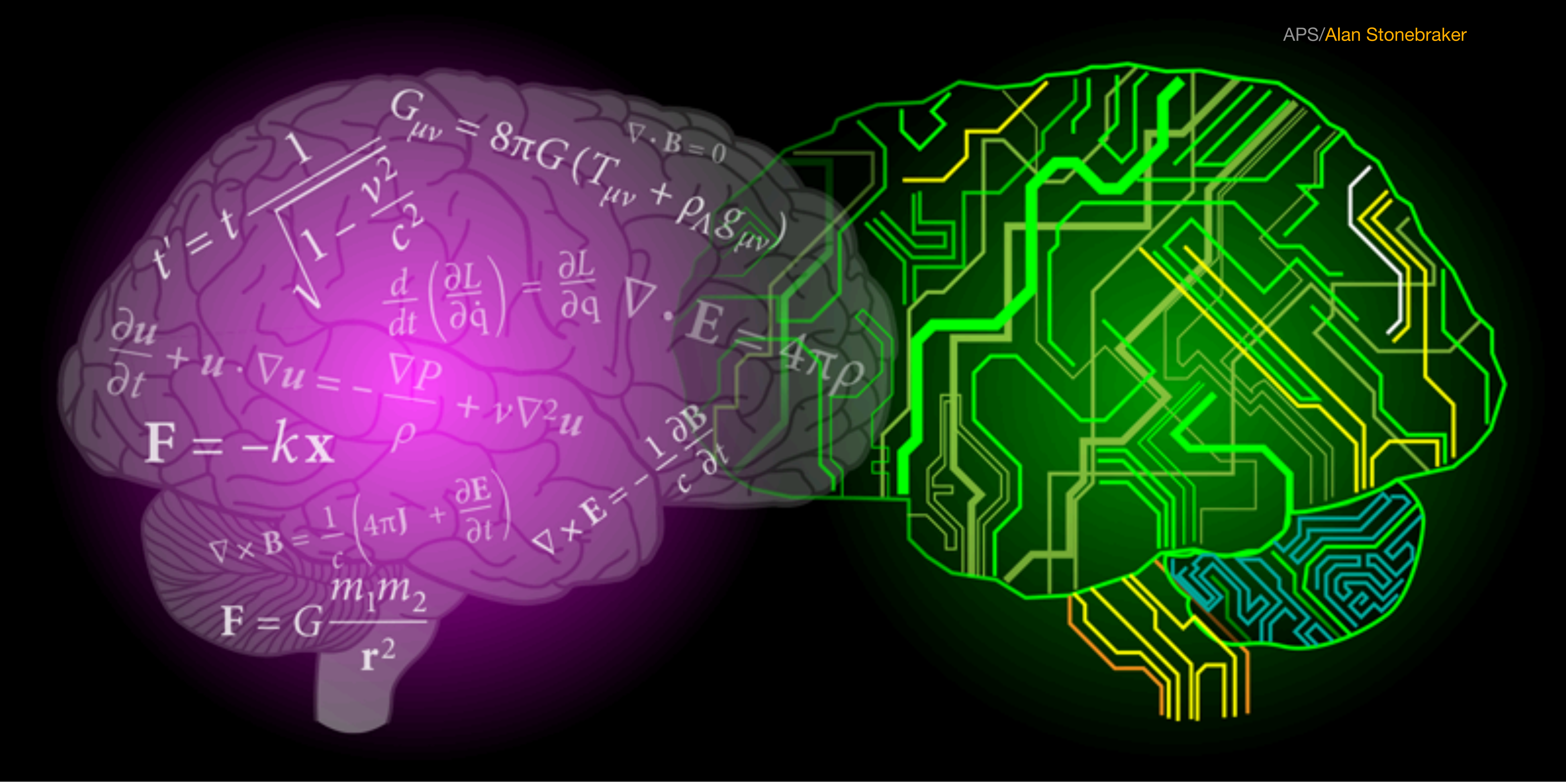


Summary
and Outlooks

- **Take-home messages**
    - AD can solve inverse problem using error information unsupervisedly
    - Neural network representations can help us to embed physical regularization into reconstructing
- **Future works**
    - Optimization algorithm
    - Maximum Entropy Method (MEM)
- Related works
    - Neutron Star, *in preparing*
    - **Bottomonium potential,** ***arXiv:2105.07862***







Future

AD in Physics, opportunities and challenges

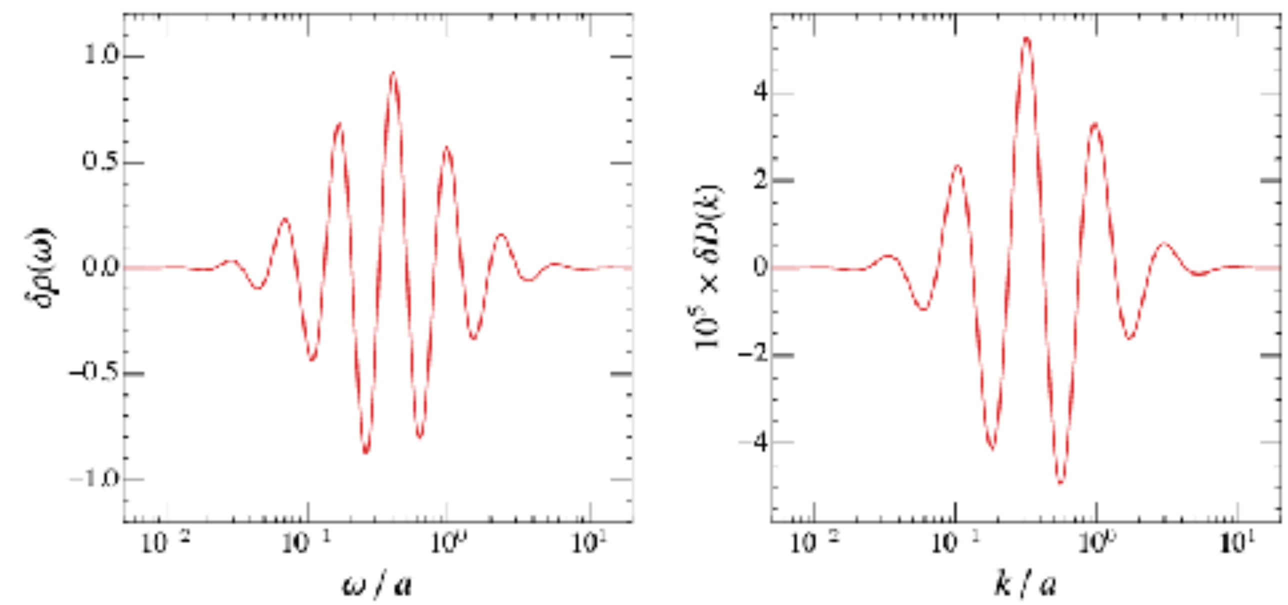
Backups

- **noise cases null-mode**

$$\|\rho_1 - \rho_2\|^2 = \sum_{i=\pm} \int_0^\infty \left(b_{i,s}^{[\rho_1]} - b_{i,s}^{[\rho_2]} \right)^2 \mathrm{d}s \,,$$

$$\|D_1 - D_2\|^2 = \sum_{i=\pm} \int_0^\infty \left(b_{i,s}^{[D_1]} - b_{i,s}^{[D_2]} \right)^2 \mathrm{d}s = \sum_{i=\pm} \int_0^\infty \lambda_s^2 \left(b_{i,s}^{[\rho_1]} - b_{i,s}^{[\rho_2]} \right)^2 \mathrm{d}s \,,$$

- **It can be suppressed by** regularizations and neural network representations



$f(s)$	$\delta\rho = \int f(s)\psi_{\pm,s}(x)\mathrm{d}s$
$(2\pi\sigma^2)^{-\frac{1}{2}} \exp[-\frac{(s-s_0)^2}{2\sigma^2}]$	$e^{-\frac{\sigma^2 \ln^2(x/a)}{2}} \psi_{\pm,s_0}(x)$
$(2\pi\sigma^2)^{-\frac{1}{2}} \exp[-\frac{(s-s_0)^2}{2\sigma^2}](\frac{s-s_0}{\sigma})^n$	$e^{-\frac{\sigma^2 \ln^2(x/a)}{2}} \mathrm{H}_n(\sigma \ln(x/a)/\sqrt{2})\,(\frac{\partial_{s_0}}{\sqrt{2}\ln x})^n \psi_{\pm,s_0}(x)$
$(2\pi\sigma^2)^{-\frac{1}{2}} \exp[-\frac{(s-s_0)^2}{2\sigma^2}]\, e^{\frac{\sigma^2 k^2}{2}} \cos[k(s-s_0)]$	$e^{-\frac{\sigma^2 \ln^2(x/a)}{2}} \cosh(\sigma^2 k \ln(x/a))\, \psi_{\pm,s_0}(x)$
$(2\pi\sigma^2)^{-\frac{1}{2}} \exp[-\frac{(s-s_0)^2}{2\sigma^2}]\, e^{\frac{\sigma^2 k^2}{2}} \sin[k(s-s_0)]$	$\mp e^{-\frac{\sigma^2 \ln^2(x/a)}{2}} \sinh(\sigma^2 k \ln(x/a))\, \psi_{\pm,s_0}(x)$

TABLE I: Noise terms with analytical form. Here, H_n are the Hermite polynomials.

Backups

- **Classical MEM**
  - Explicit physical prior
  - Classical optimizors and regularizations
  - Unique solution and uncertainty estimation
- **Supervised Learning**
  - Implicit physical prior (hidden in data)
  - Modern optimizors and regularizations
  - Unique solution? and uncertainty estimation (with Bayesian NN)
- **Our method (NN+AD)**
  - Explicit physical prior (rigorous backward process and physical restrictions)
  - Modern optimizors and regularizations
  - Unique solution and uncertainty estimation (with MEM)