Spectral Reconstruction with Deep Neural Networks

Lukas Kades

Cold Quantum Coffee

- Heidelberg University -

arXiv: 1905.04305, Lukas Kades, Jan M. Pawlowski, Alexander Rothkopf, Manuel Scherzer, Julian M. Urban, Sebastian J. Wetzel, Nicolas Wink, and Felix Ziegler

May 14, 2019

Outline

- Physical motivation - the inverse problem
- Existing methods
- Neural network based reconstruction
- Comparison
- Problems of reconstructions with neural networks
- Possible improvements
- Conclusion

Physical motivation

Real-time properties of strongly correlated quantum systems

- Time has to be analytically continued into the complex plane
- Explicit computations involve numerical steps

Propagator — Spectral function — Källen-Lehmann kernel

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

How to reconstruct the spectral function from noisy Euclidean propagator data to extract their physical structure?

The (inverse) problem

$$= K_{ij}$$

$$G_i := G(p_i) = \int_0^\infty \frac{\mathrm{d}\omega}{\pi} \frac{\omega\,\rho(\omega)}{\omega^2 + p_i^2} \simeq \sum_{j=1}^{N_\omega} K(p_i, \omega_j) \Delta\omega_j\, \rho_j \quad \Leftrightarrow \quad \rho_j = K_{ij}^{-1} G_i$$

Properties:

- Mostly very small eigenvalues - hard to invert numerically
- Ill-conditioned: A small error in the initial propagator data can result in large deviations in the reconstruction
- Suppression of additional structures for large frequencies

The (inverse) problem

$$= K_{ij}$$

$$G_i := G(p_i) = \int_0^\infty \frac{d\omega}{\pi} \frac{\omega\, \rho(\omega)}{\omega^2 + p_i^2} \simeq \sum_{j=1}^{N_\omega} K(p_i, \omega_j) \Delta\omega_j\, \rho_j \quad \Leftrightarrow \quad \rho_j = K_{ij}^{-1} G_i$$

Properties:

- Mostly very small eigenvalues - hard to invert numerically
- Ill-conditioned: A small error in the initial propagator data can result in large deviations in the reconstruction
- Suppression of additional structures for large frequencies

How to tackle such an inverse problem?

Specifying the problem

- Discretised noisy propagator points:

$$G = \{G_i + \sigma\eta_i\}_{i=1}^{N_p}$$

- Consisting of 1, 2 or 3 Breit-Wigners:

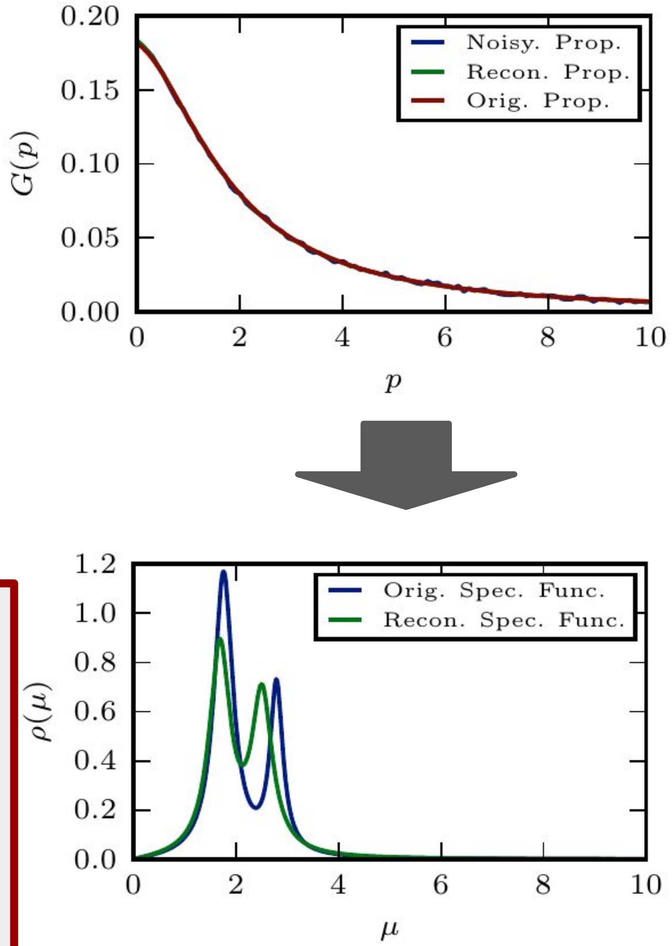
$$\rho^{(BW)}(\omega|A, M, \Gamma) = \frac{4A\Gamma\omega}{(M^2 + \Gamma^2 - \omega^2)^2 + 4\Gamma^2\omega^2}$$

Objectives (the actual inverse problem):

- Case 1: Try to predict the underlying parameters:

$$\theta = \{A_i, M_i, \Gamma_i \,|\, 0 \leq i < N_{\mathrm{BW}}\}$$

- Case 2: Try to predict a discretised spectral function: $\rho = \{\rho_i\}_{i=1}^{N_\omega}$



Bayesian inference

What is that? -

Bayesian inference

What is that? -

- An optimization algorithm that uses Bayes’ theorem to deduce properties of an underlying posterior distribution.

(cf. Wikipedia: Statistical Inference)

Reminder: Bayes' Theorem

Given:

- Discretised propagator data: $G = \{G_i + \sigma\eta_i\}_{i=1}^{N_p}$
- Parameters of the Breit-Wigner functions: $\theta = \{A_i, M_i, \Gamma_i \,|\, 0 \leq i < N_{\mathrm{BW}}\}$

Prior probability

$$p(\theta|G) \propto p(\theta)p(G|\theta)$$

Posterior probability of θ given propagator data G

Probability of propagator data G given Breit-Wigner functions parameterised by θ

GrHMC method (Existing methods I) 1804.00945, A.K. Cyrol et al.

$$p(\theta|G) \propto p(\theta)p(G|\theta)$$

$$L = \| G - G(\theta_{sug}) \|_2^2$$

- Based on a hybrid Monte Carlo algorithm to map out the posterior distribution
- Enables the computation of expectation values: $\langle\theta\rangle, \langle\theta^2\rangle, \dots$

$$p(G|\theta) \propto \exp(-L)$$

$$A(\theta'|\theta_t) = \min\left(1, \frac{p(\theta'|G)g(\theta_t|\theta')}{p(\theta_t|G)g(\theta'|\theta_t)}\right)$$

Aims particularly at a prediction of the underlying parameters (Case 1)

BR method (Existing methods II) 1307.6106, Y. Burnier, A. Rothkopf

$$S_{\mathrm{BR}} = \int d\omega \left(1 - \frac{\rho(\omega)}{m(\omega)} + \log\left[\frac{\rho(\omega)}{m(\omega)}\right]\right)$$

$$p(\rho|G) \propto p(\rho)p(G|\rho)$$

$$p(G|\rho) \propto \exp\left(-L - \gamma(L - N_\tau)^2\right)$$

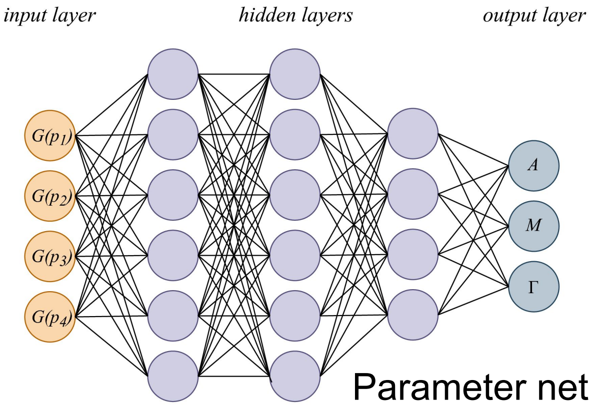
$$p(\rho) \propto \exp\left(S_{\mathrm{BR}}\right)$$

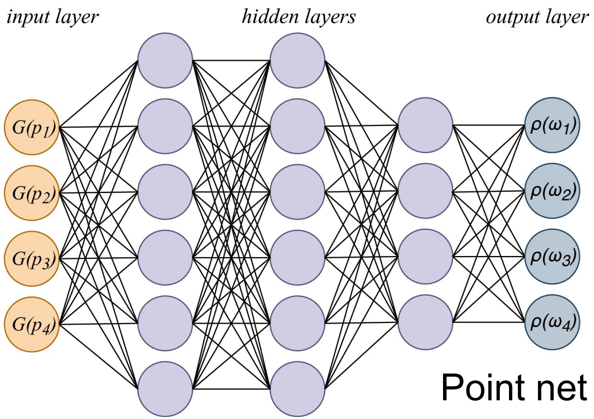
- Based on a gradient descent algorithm to find the maximum (Maximum A Posteriori - MAP)
- Incorporation of certain constraints (smoothness, scale invariance, etc.)

Aims particularly at a prediction of a discretised spectral function (Case 2)

Neural network based reconstruction

$$L = \| G - G(\theta_{sug}) \|_2^2$$





$$L = \| \rho - \rho(\theta_{sug}) \|_2^2$$

$$L = \| \theta - \theta_{sug} \|_2^2$$

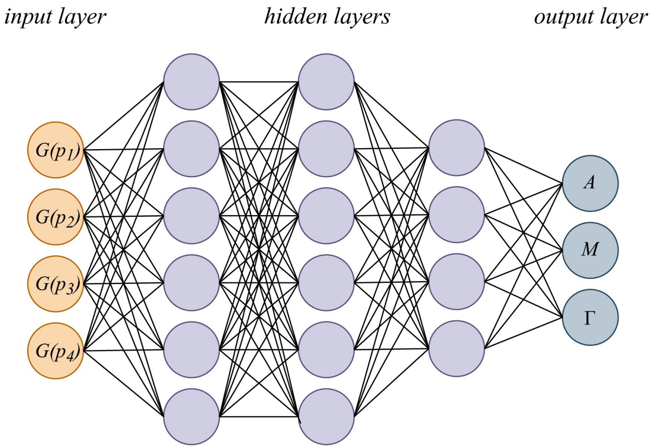
New!

- Based on a feed-forward network architecture
- A definition of a large set of loss functions is possible

Aims at a correct prediction for both cases - a discretised spectral function or the underlying parameters

Training procedure

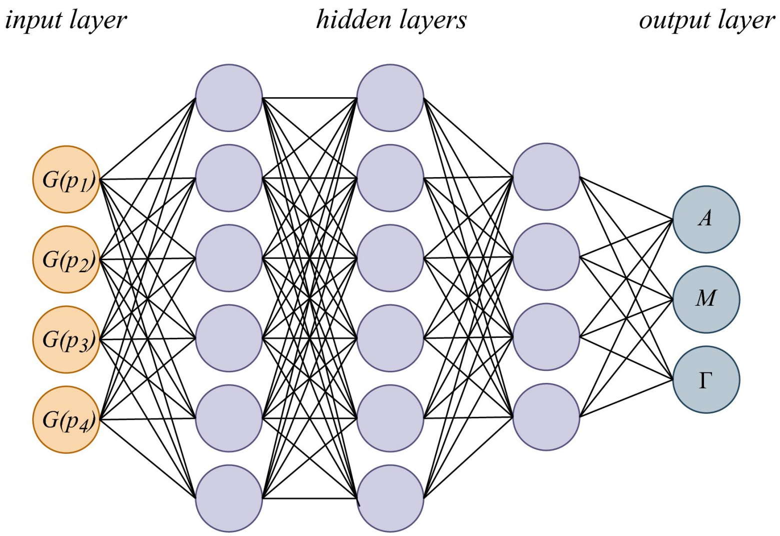
1. Generate training data:
$$\{(\theta, G)\,|\,\theta = \{A_i, M_i, \Gamma_i \sim \mathrm{Uni}\,|\,0 \leq i < N_{\mathrm{BW}}\} \wedge G \sim G(\theta) + \sigma\eta\}$$
2. Forward pass: $\theta_{sug} = \mathrm{net}(G)$
3. Compute the loss: $L = \| \theta - \theta_{sug} \|_2^2$
4. Backward pass (Backpropagation): Adapt network parameters for better prediction
5. Repeat until convergence



The <u>inverse integral transformation is parametrised</u> by the hidden variables of the neural network.

Potential advantages of neural networks

- Parametrisation of the inverse integral transformation



Potential advantages of neural networks

- Parametrisation of the inverse integral transformation
- Optimisation/Training based directly on arbitrary representations of the spectral function - much larger set of possible loss functions

$$L = \| \rho - \rho(\theta_{sug}) \|_2^2$$

$$L = \| G - G(\theta_{sug}) \|_2^2$$

$$L = \| \theta - \theta_{sug} \|_2^2$$

Potential advantages of neural networks

- Parametrisation of the inverse integral transformation
- Optimisation/Training based directly on arbitrary representations of the spectral function - much larger set of possible loss functions
- Provides implicit regularisation by training data or explicit, by additional regularisation terms in the loss function

Potential advantages of neural networks

- Parametrisation of the inverse integral transformation
- Optimisation/Training based directly on arbitrary representations of the spectral function - much larger set of possible loss functions
- Provides implicit regularisation by training data or explicitly, by additional regularisation terms in the loss function
- Computationally much cheaper (after training)
- More direct access to try-and-error scenarios for the exploration of more appropriate loss functions, etc.

Comparison to existing methods

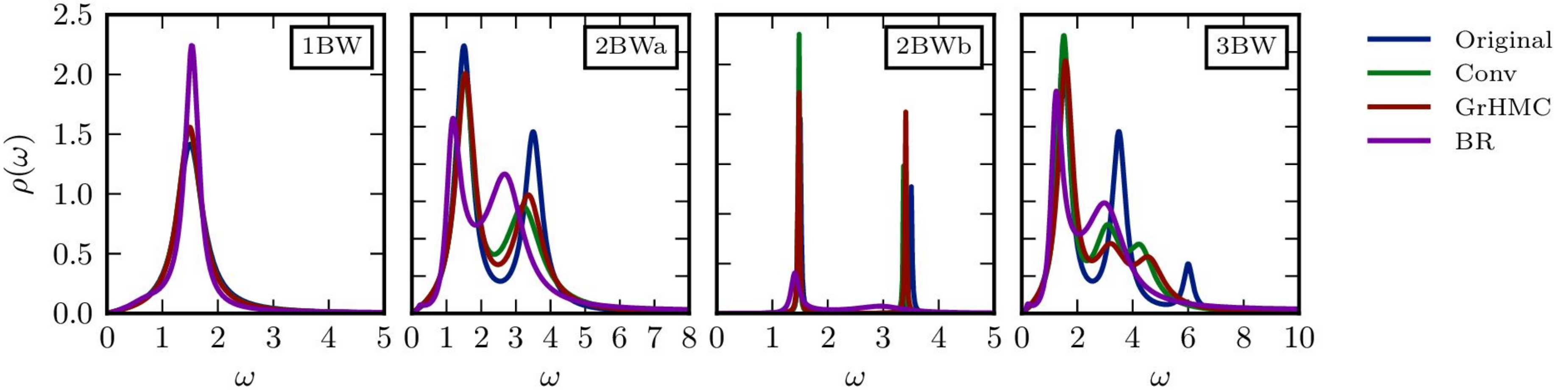
Neural network approach:

- Implicit Bayesian approach
- Optimum is learned a priori by a parametrisation by the neural network
- Based on arbitrary loss functions

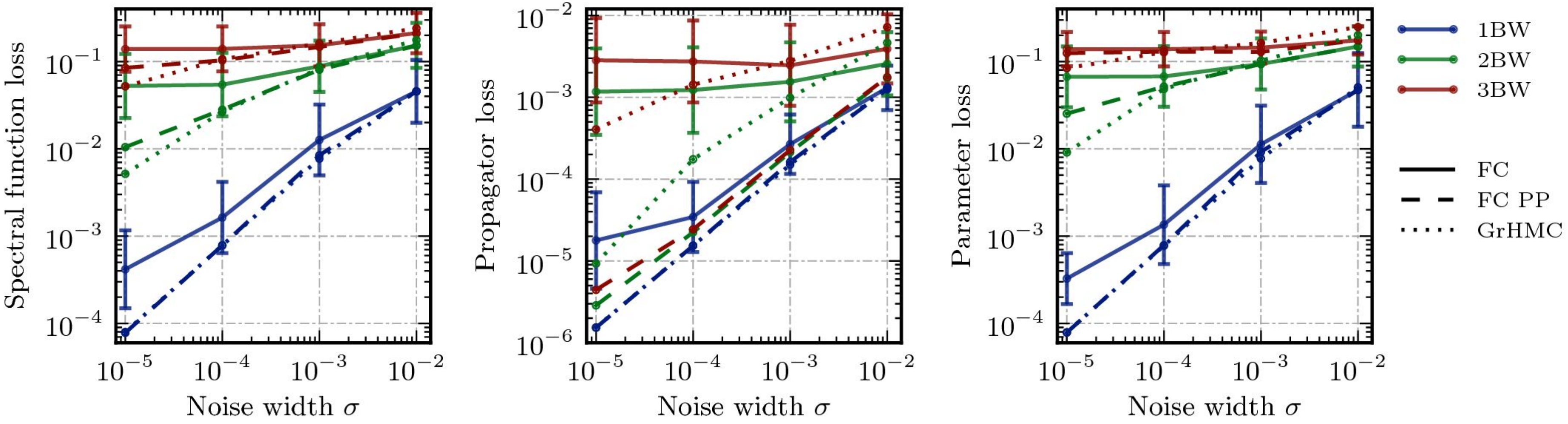
Existing methods:

- Explicit Bayesian approach
- Iterative optimization algorithm
- Restricted to propagator loss

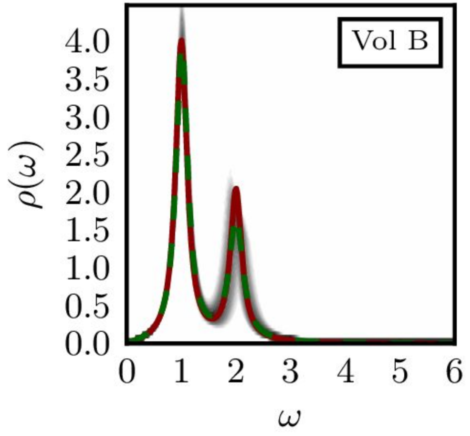
Numerical results I



Numerical results II

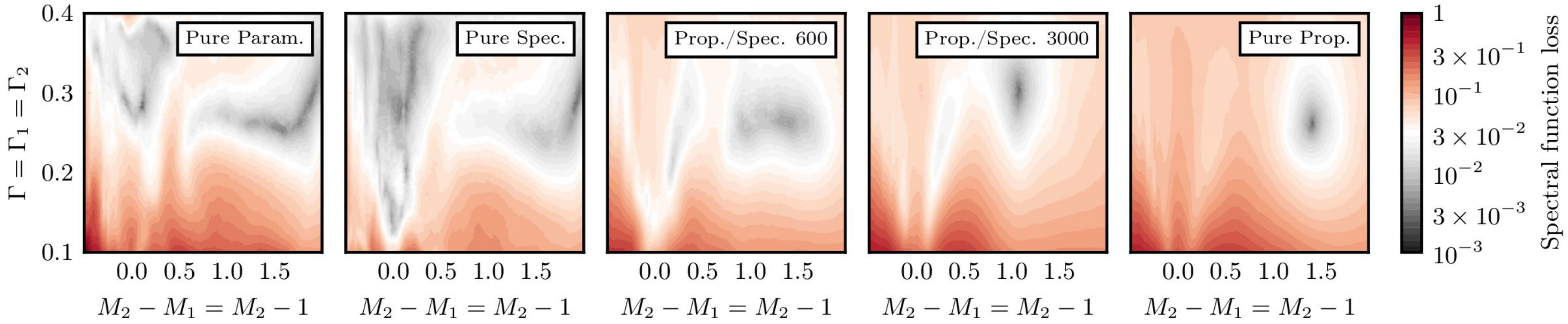


Problems of neural networks



Expressive power too small for large parameter spaces:

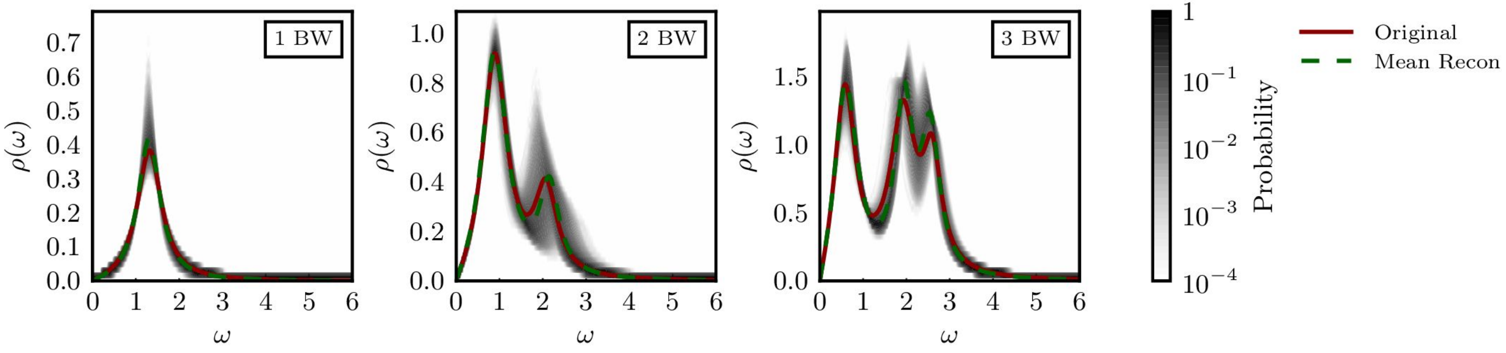
- Set of inverse transformations is too large
- Systematic errors due to a varying severity of the inverse problem



How to obtain **reliable reconstructions**?

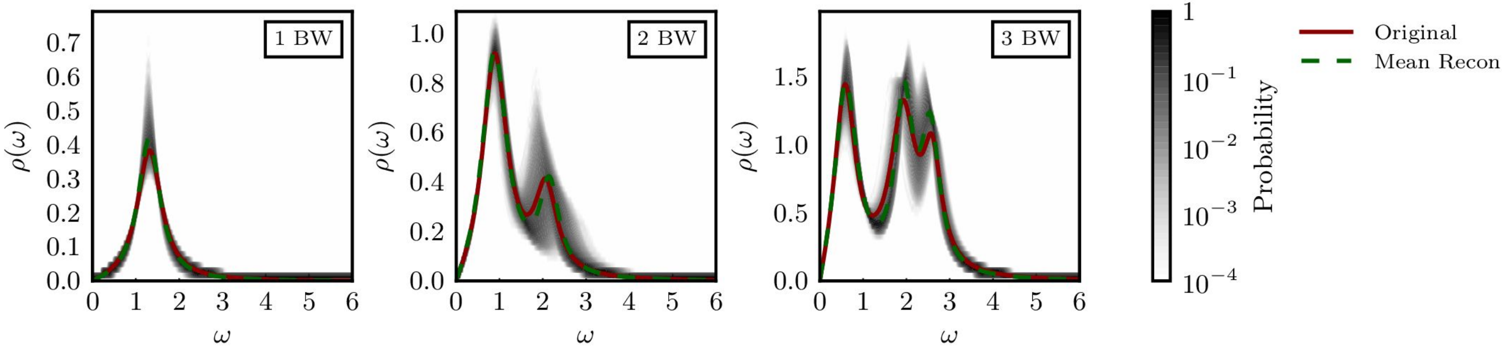
What is meant by reliable reconstructions?

- Locality of proposed solutions in parameter space (aims at a reduction of the strength of the ill-conditioned problem)



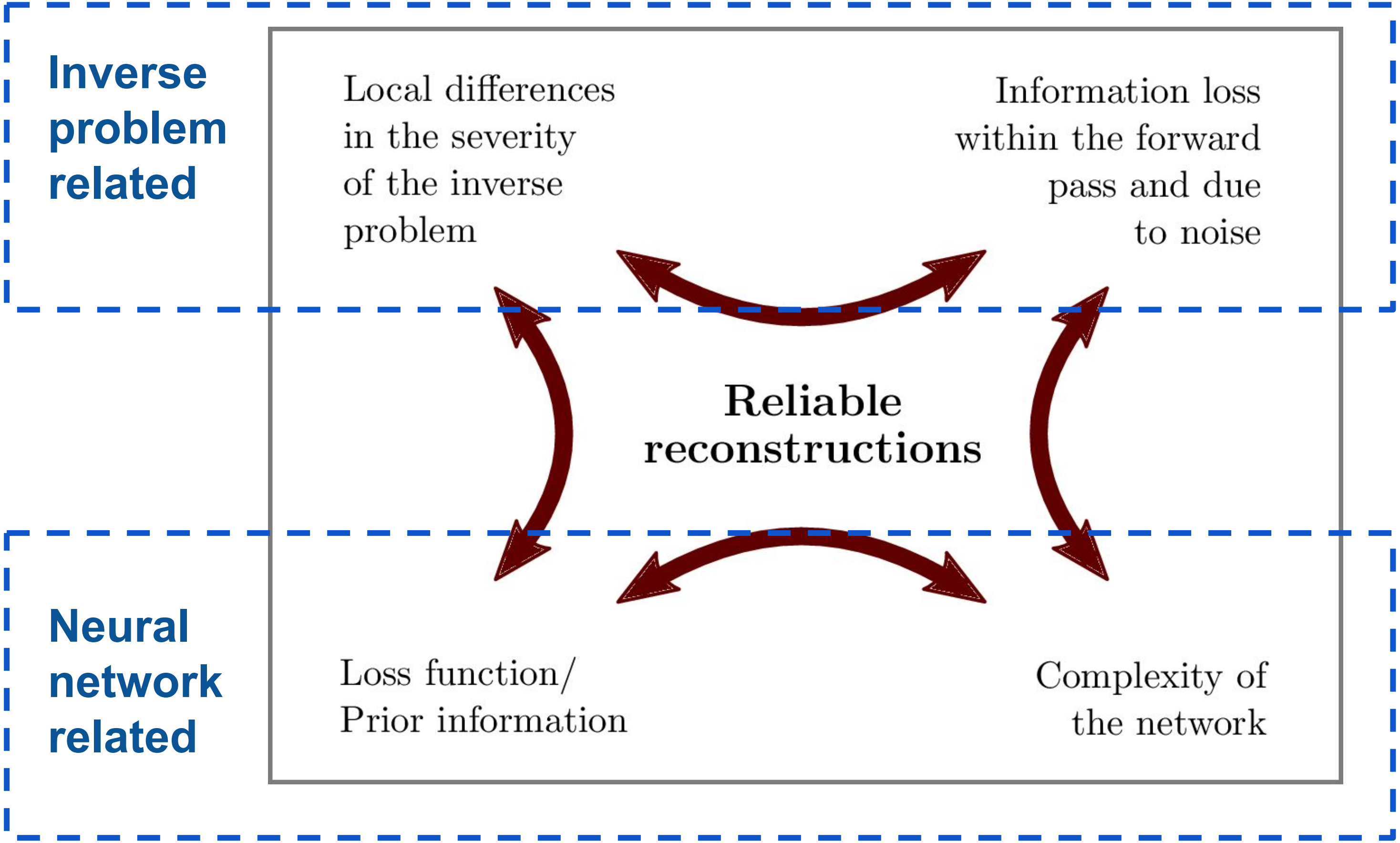
What is meant by reliable reconstructions?

- Locality of proposed solutions in parameter space (aims at a reduction of the strength of the ill-conditioned problem)
- Homogeneous distribution of losses in parameter space



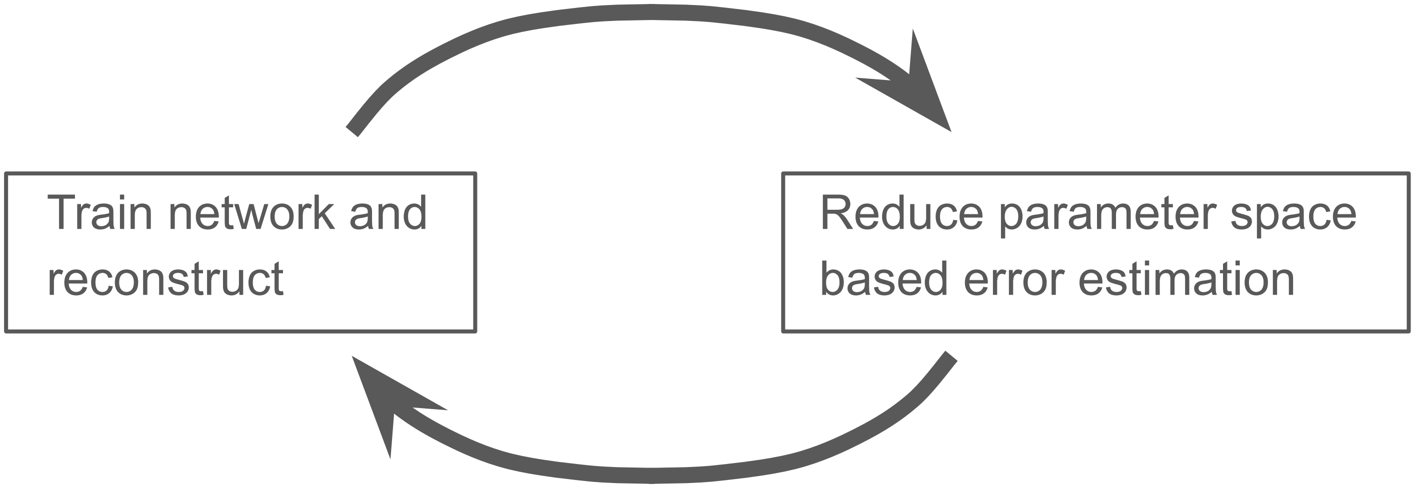
➢ Spectral reconstructions with a reliable error estimation

Factors for reliable reconstructions



Iterative procedure

➢ Reliable reconstructions allow an iterative procedure implemented by a successive reduction of the parameter space



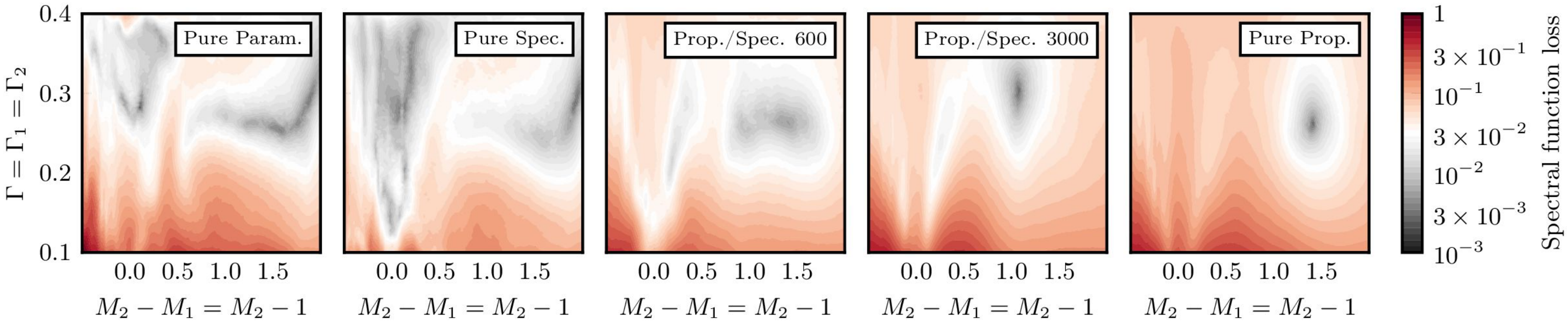
How to obtain **reliable reconstructions**?

Future work I - Training data and learning loss functions

- Search for algorithms to artificially manipulate the loss landscape
- Discover more appropriate loss functions for existing methods

➢ Reduction of the strength of the ill-conditioned problem

1707.02198, Santos et al.
1810.12081, Wu et al.

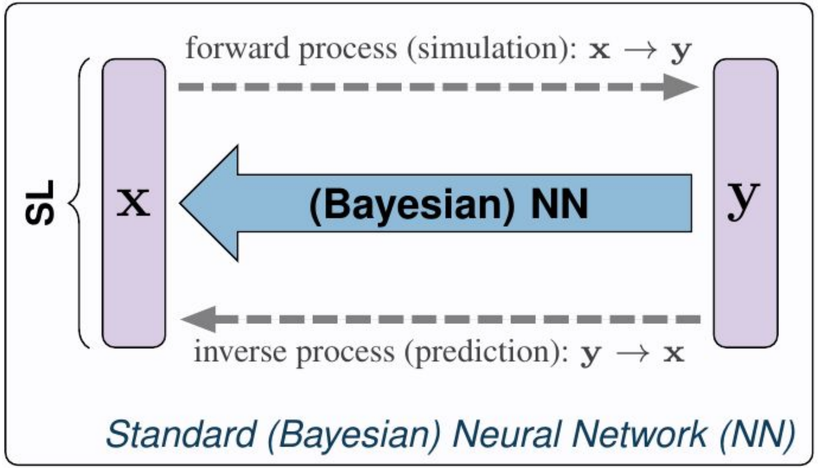


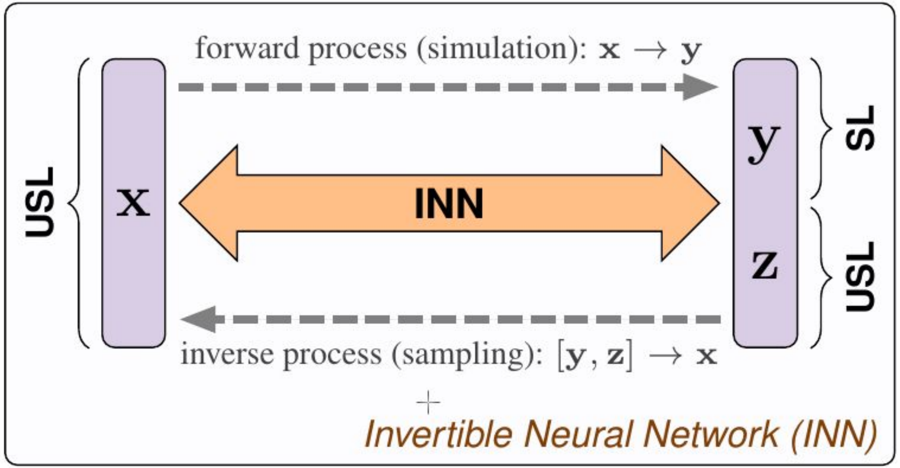
➢ Results in locality of solutions and a homogeneous loss distribution

Future work II - Invertible neural networks

1808.04730, Ardizzone et al.

- Particular network architecture that is trained in both directions - <u>invertible</u>
- Allows Bayesian Inference by sampling





➢ Enables a reliable error estimation

Conclusion

- Recapitulation of the inverse problem of spectral reconstruction
- Introduction of a reconstruction scheme based on deep neural networks
- Analysed problems regarding reconstructions with neural networks
- Proposed solutions for this problems for future work

Further future work

- Gaussian processes
- Application on physical data