# Machine learning for quantum field theories with a sign problem

Andrei Alexandru

Paulo Bedaque, Henry Lamm, Scott Lawrence

Phys. Rev. D96 (2017), no. 9 094505, [arXiv:1709.01971]

AI for Nuclear Physics
Jefferson Laboratory, 2020

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

# The plan

- Motivation
- Sign problem, complexification and contour deformation
- Generalized thimble method
- The learnifold: neural network parametrization
- Case study: Thirring model in 1+1 dimensions
- Summary and outlook

# Motivation

- Physical models of interest require non-perturbative calculations that have a sign problem:
  - QCD at finite baryon density (RHIC, neutron star structure, etc)
  - Real time dynamics for strongly coupled QFT
  - Strongly correlated electrons (Hubbard model, etc.)
- Complex path methods are likely to work for a large class of problems.

# QFT on the lattice

- The partition function is expressed as a path integral
- The fields are sampled on a grid; differential operators are replaced by finite difference ones

$$Z = \int \mathcal{D}\phi \, e^{-S[\phi]} \to Z_{\mathrm{latt}} = \int_{\mathbb{R}^N} \prod_i \mathrm{d}\phi_i \, e^{-S[\phi]}$$

$$S_{\mathrm{latt}} = \sum_n \left[ \tilde{m}\phi_n^2 + \sum_\alpha \kappa_\alpha \phi_n \phi_{n+\hat{\alpha}} + \tilde{\lambda}\phi_n^4 \right]$$

- The partition function is a many-dimensional integral over **real** variables
- The integrand has **no singularity** for both bosonic and fermionic theories

# Monte-Carlo sampling

- QFT correlators are statistical averages

$$\langle O \rangle = \frac{1}{Z} \int \mathcal{D}\phi \, e^{-S(\phi)} O(\phi)$$

- Estimate using importance sampling

$$\langle O \rangle \approx \frac{1}{N} \sum_{i=1}^{N} O(\phi_i) \qquad \{\phi_1, \ldots, \phi_N\} \text{ with } P(\phi) = \frac{1}{Z} e^{-S(\phi)}$$

- Stochastic errors decrease with sample size

$$\sigma_{\langle O \rangle} = \sqrt{\frac{\langle O^2 \rangle - \langle O \rangle^2}{N}} \propto \frac{1}{\sqrt{N}}$$

# Sign problem

- When the partition function is not real direct Monte-Carlo sampling is not possible

- The usual workaround involves *reweighting*

$$Z_0 = \int \mathcal{D}\phi \left| e^{-S(\phi)} \right| = \int \mathcal{D}\phi \, e^{-S_R(\phi)}$$

$$\langle O(\phi) \rangle_Z = \frac{\left\langle O(\phi) e^{-iS_I(\phi)} \right\rangle_{Z_0}}{\left\langle e^{-iS_I(\phi)} \right\rangle_{Z_0}}$$

- Sampling is done based on $S_R$; $S_I$ is introduced in the observable: $\{\phi_1, \ldots, \phi_N\}$ with $P(\phi) \propto e^{-S_R(\phi)}$

$$\langle O \rangle \approx \frac{1}{N} \sum_{i=1}^{N} O(\phi_i) e^{-iS_I(\phi)} \Bigg/ \frac{1}{N} \sum_{i=1}^{N} e^{-iS_I(\phi)}$$

# Sign problem

- A *sign problem* appears when the phase average is nearly zero (or zero): $e^{-iS_I(\phi_1)} + \ldots + e^{-iS_I(\phi_N)} \ll N$

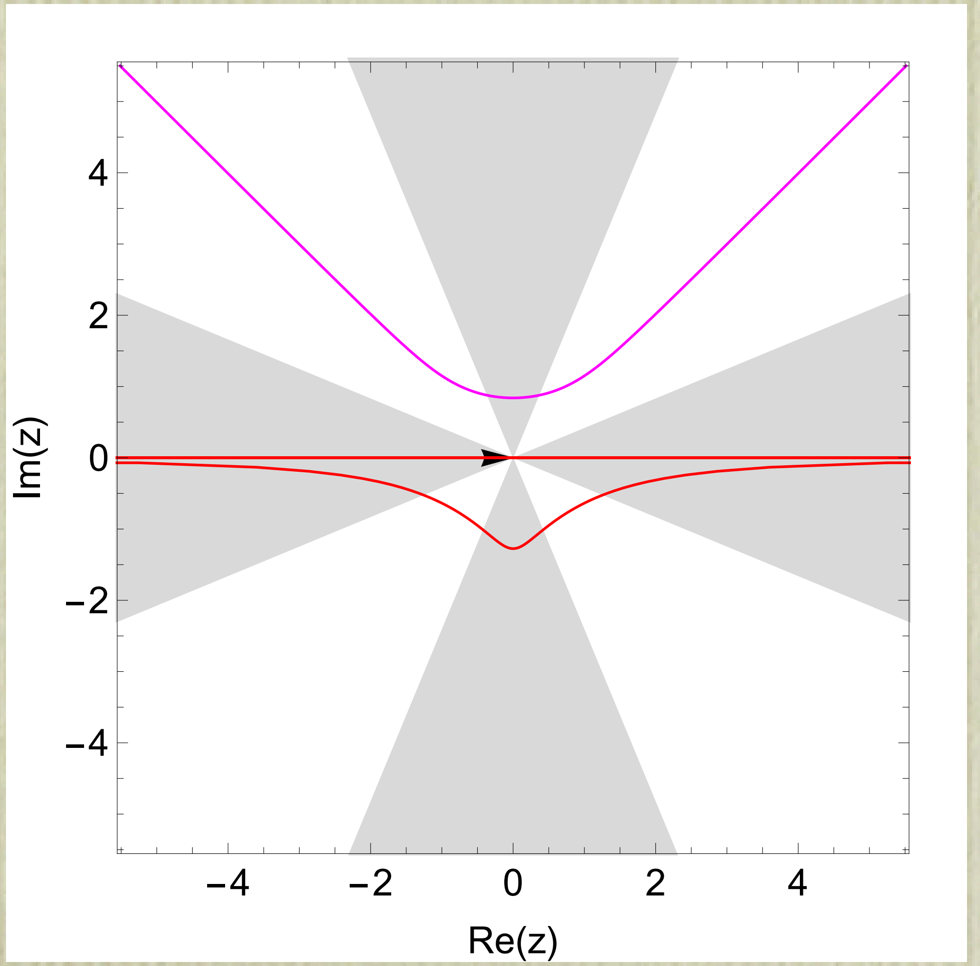
- The cost of the calculation is inversely proportional to the phase average: $N \propto \left\langle e^{-iS_I(\phi)} \right\rangle^{-2}$

- For example in QCD

$$\left\langle e^{-iS_I} \right\rangle_{Z_0} = \frac{Z}{Z_0} = e^{-\beta V (f_{\text{baryon}} - f_{\text{isospin}})} \to 0 \text{ as } V \to \infty$$

- In QCD the calculation cost increases exponentially with the volume

# Contour deformation



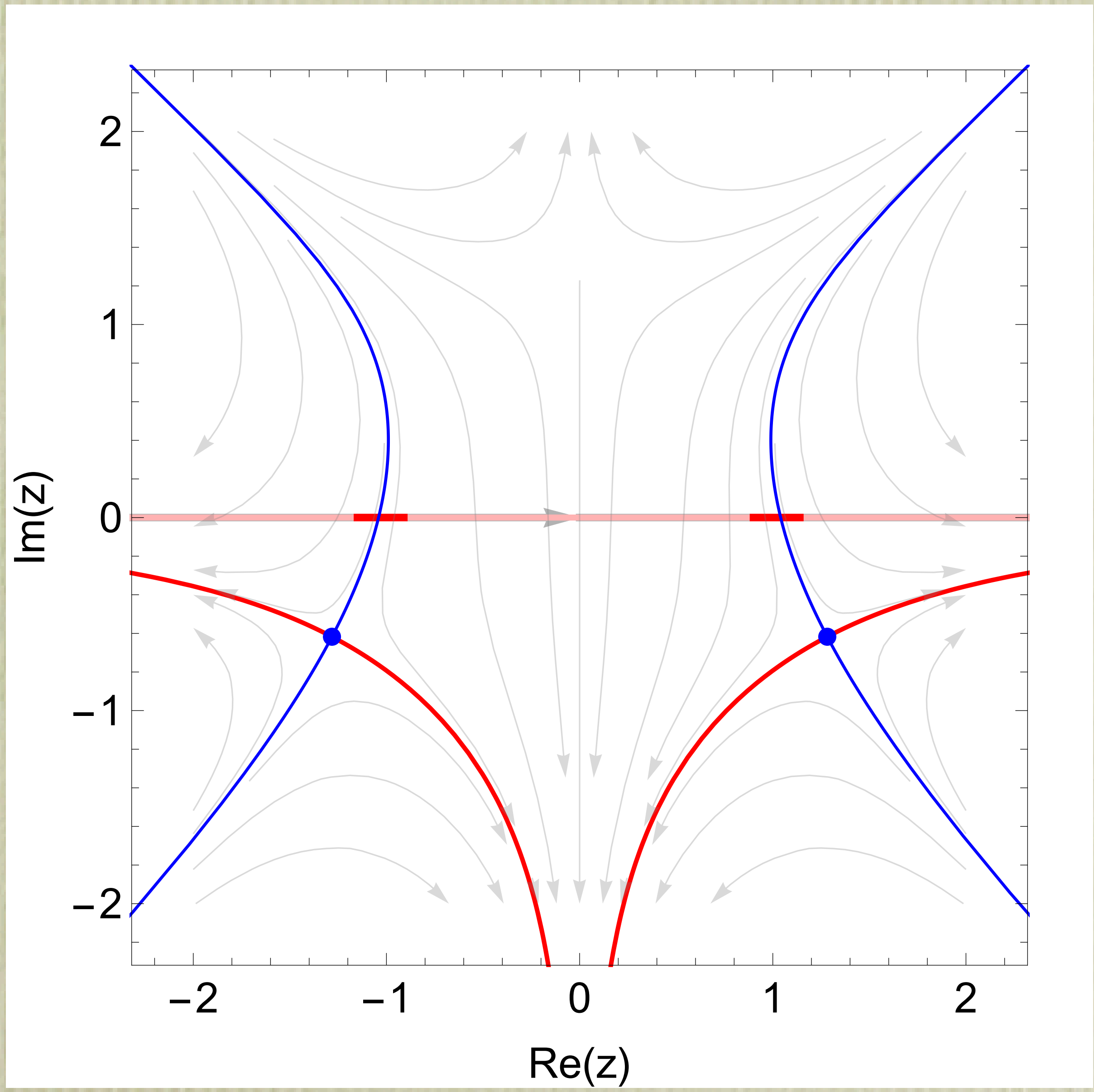

$$Z = \int_{-\infty}^{\infty} \mathrm{d}x \, e^{-S(x)}$$

$$S(x) = x^4 - x^2 + 10ix$$

$$Z = \int_{\mathcal{C}} \mathrm{d}z \, e^{-S(z)}$$

- Generalized Cauchy's theorem
- Deformation in the field variable space (lattice geometry unchanged)

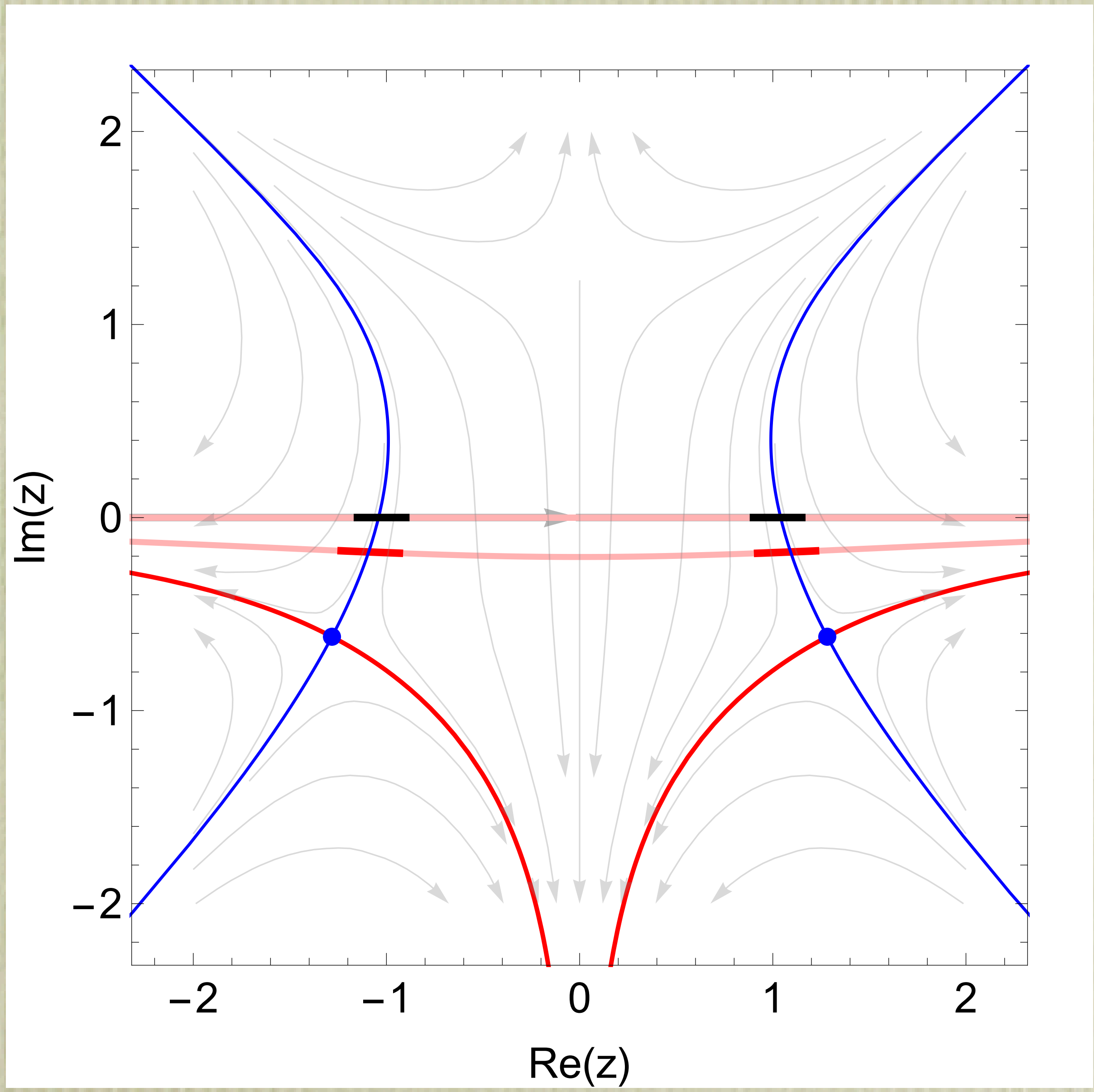
# Holomorphic gradient flow and Lefschetz thimbles




$$Z = \int_{-\infty}^{\infty} \mathrm{d}x \, e^{-S(x)}$$

$$\boxed{\frac{dz}{d\tau} = \overline{\frac{\partial S}{\partial z}}}$$

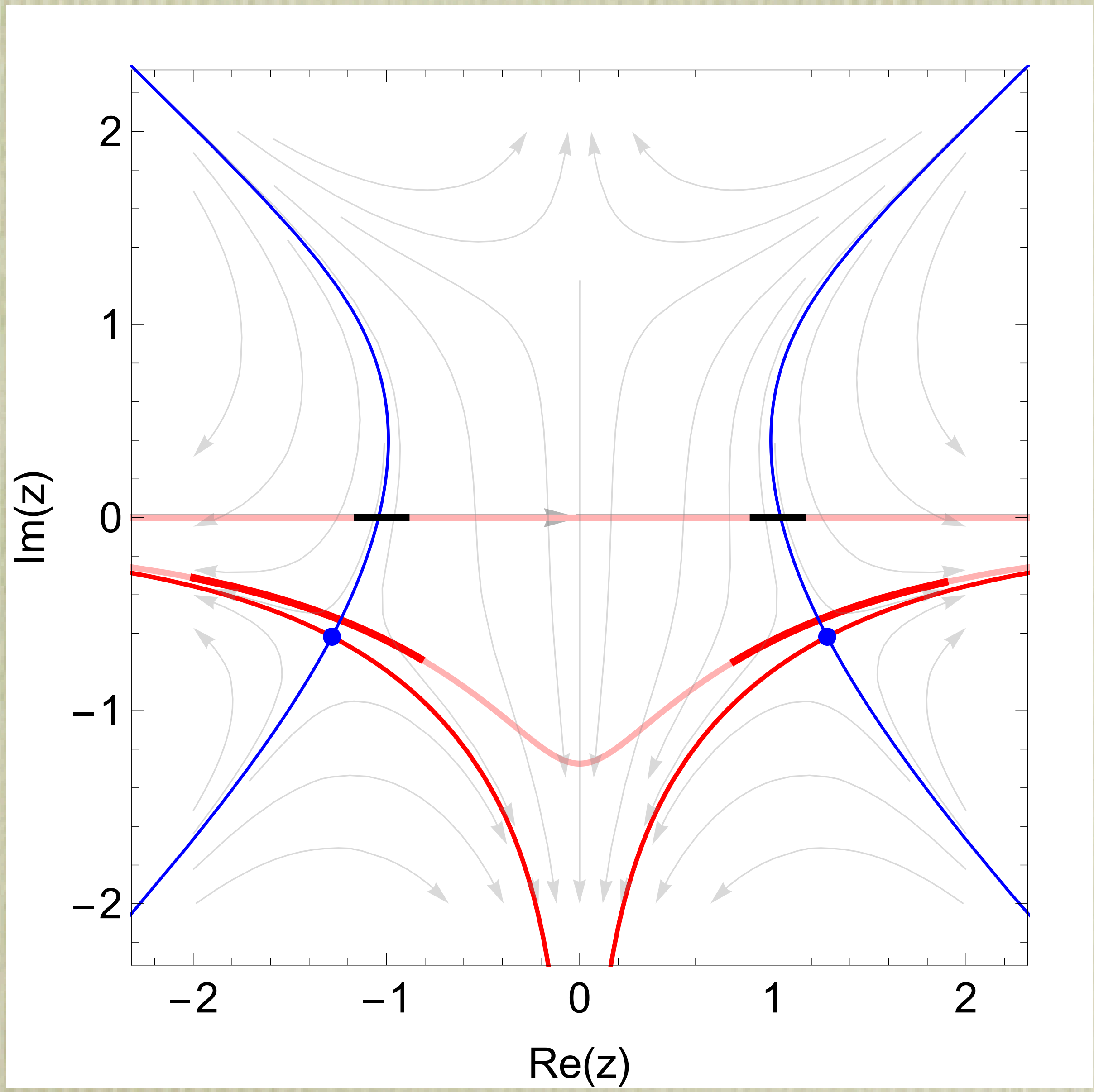
# Holomorphic gradient flow and Lefschetz thimbles




$$Z = \int_{-\infty}^{\infty} \mathrm{d}x \, e^{-S(x)}$$

$$\boxed{\frac{dz}{d\tau} = \overline{\frac{\partial S}{\partial z}}}$$

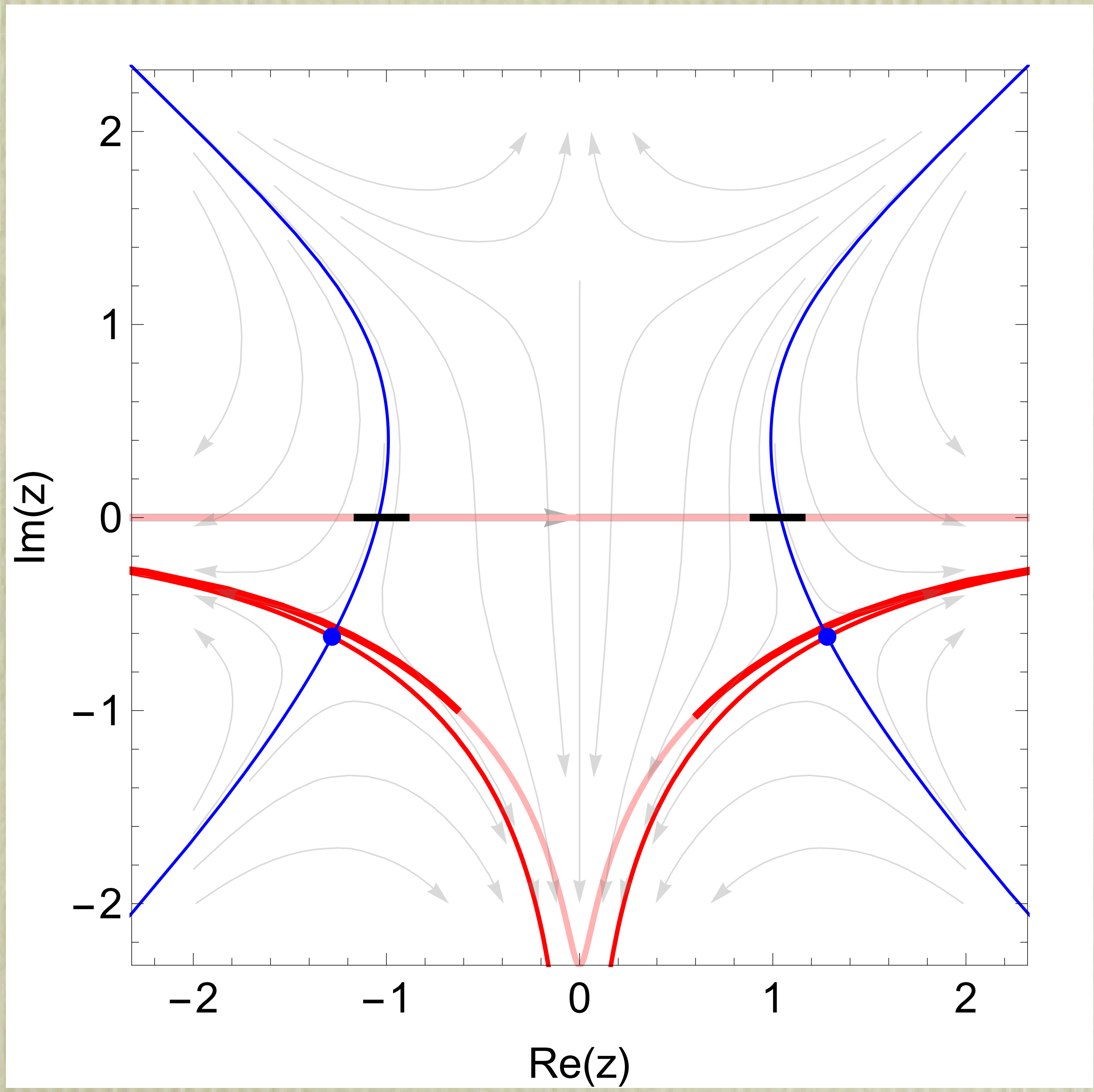
# Holomorphic gradient flow and Lefschetz thimbles




$$Z = \int_{-\infty}^{\infty} \mathrm{d}x \, e^{-S(x)}$$

$$\boxed{\frac{dz}{d\tau} = \overline{\frac{\partial S}{\partial z}}}$$

# Holomorphic gradient flow and Lefschetz thimbles



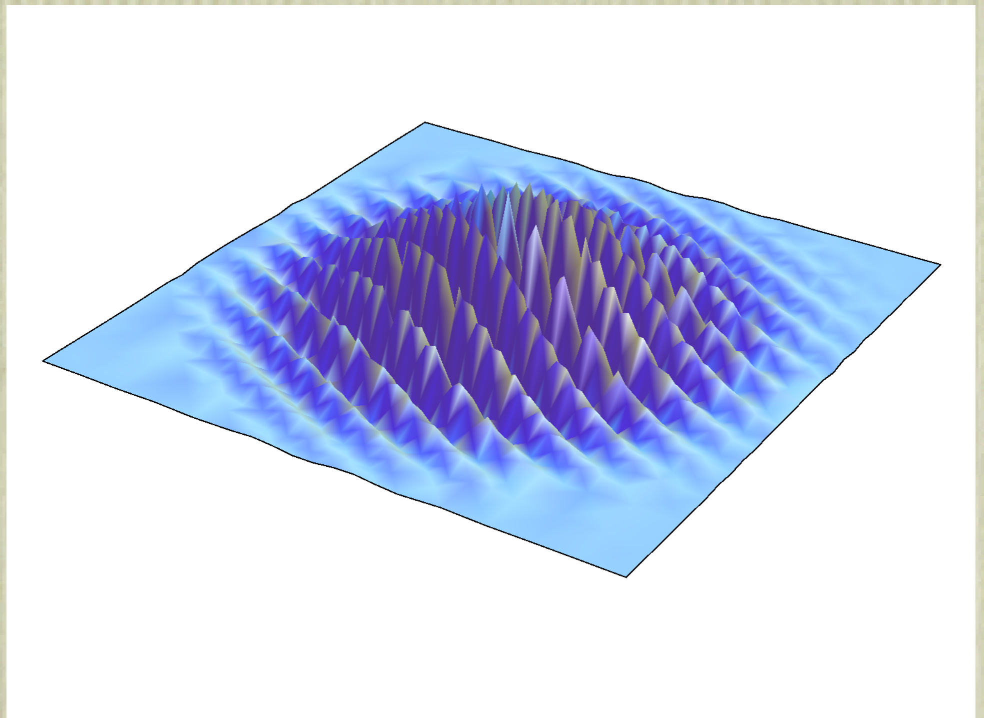

$$Z = \int_{-\infty}^{\infty} \mathrm{d}x \, e^{-S(x)}$$

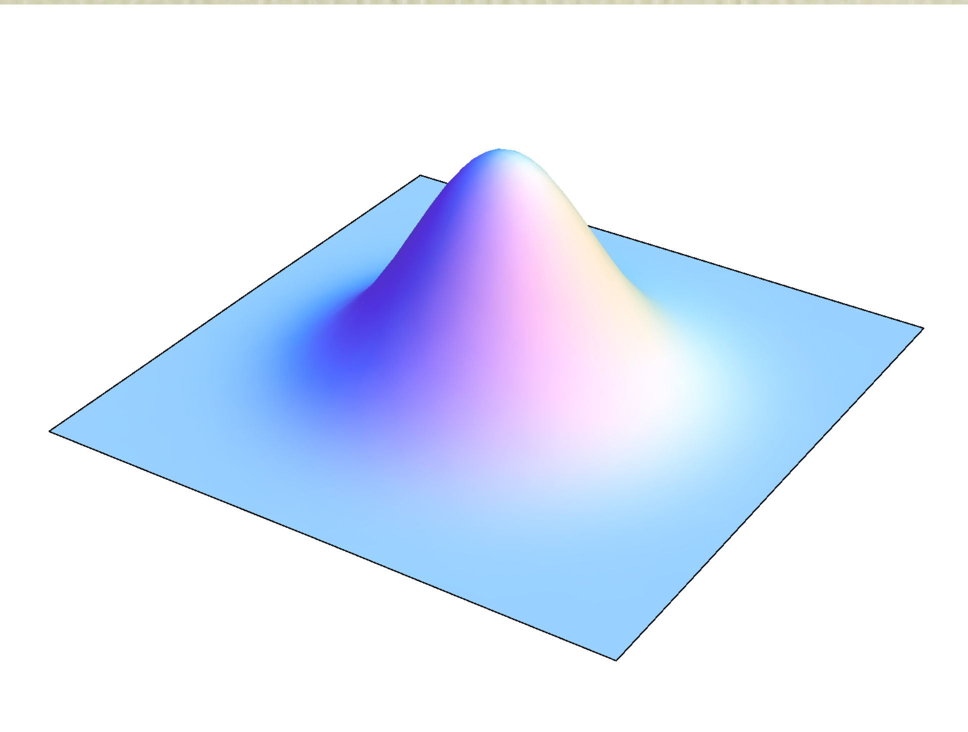
$$\boxed{\frac{dz}{d\tau} = \overline{\frac{\partial S}{\partial z}}}$$

$$Z = \sum_{\sigma} n_\sigma \int_{J_\sigma} dz \, e^{-S(z)}$$

M. Cristoforetti, F. Di Renzo, and L. Scorzato, *High density QCD on a Lefschetz thimble*, Phys. Rev. D86 (2012) 074506

# Lefschetz thimble



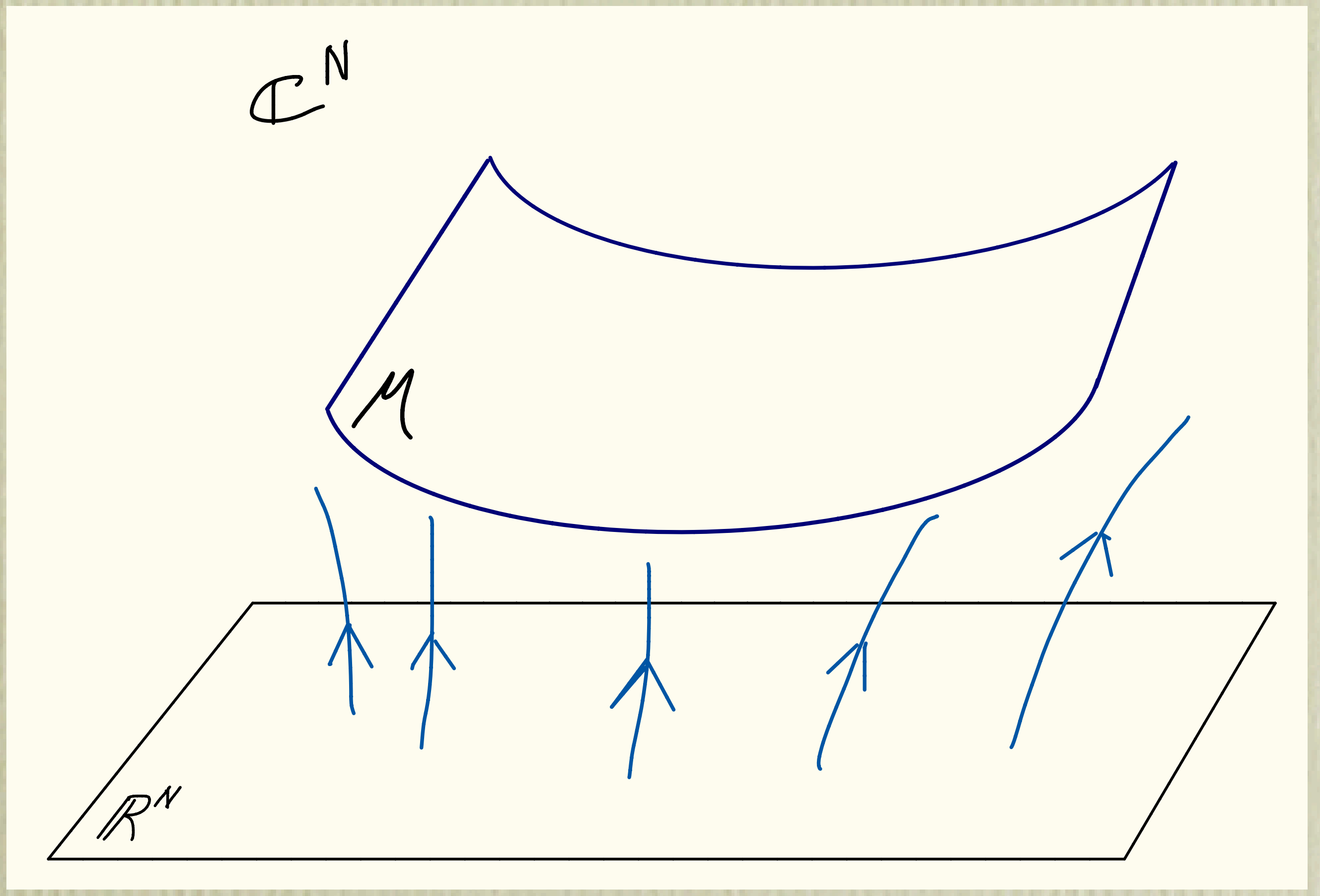


$e^{-S(x_1,x_2)}$ (real plane)

$e^{-S(z_1,z_2)}$ (gaussian thimble)

$$S(x_1, x_2) = x_1^2 + x_2^2 + 10ix_1 + 20ix_2 + ix_1x_2/3$$

# Generalized thimble method



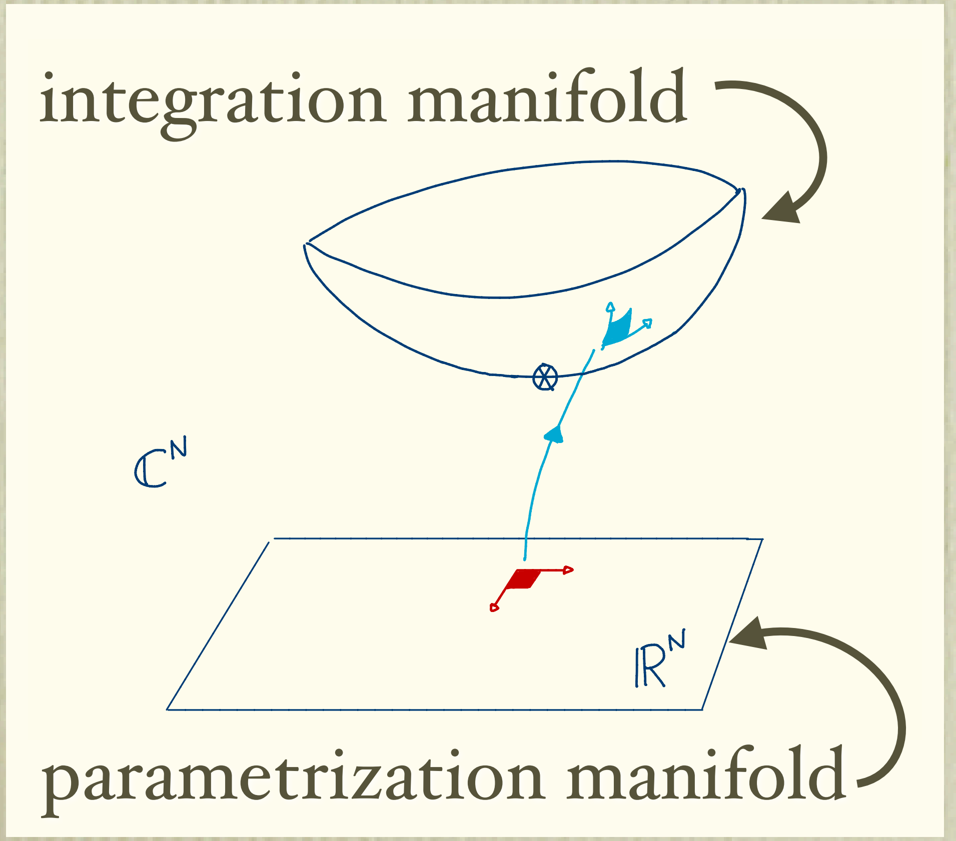

- Most systems require multiple thimble
- Thimble decomposition is hard
- Use the manifolds generated by the holomorphic flow

AA, G. Basar, P. F. Bedaque, G. W. Ridgway, and N. C. Warrington, *Sign problem and Monte Carlo calculations beyond Lefschetz thimble*s, JHEP 05 (2016) 053

# Basic idea

$$Z = \int_{\mathcal{M}} \mathrm{d}^N z \, e^{-S(z)} = \int_{\mathbb{R}^N} \mathrm{d}^N x \, \underbrace{\left\| \frac{\partial z_i}{\partial x_j} \right\|}_{\det J(x)} e^{-S(z(x))}$$

$$\det J(x) \longrightarrow \boxed{J_{ij} = \frac{\partial z_i}{\partial x_j}}$$




$$\boxed{S_{\text{eff}}(x) = S_R(z(x)) - \ln|\det J(x)|}$$

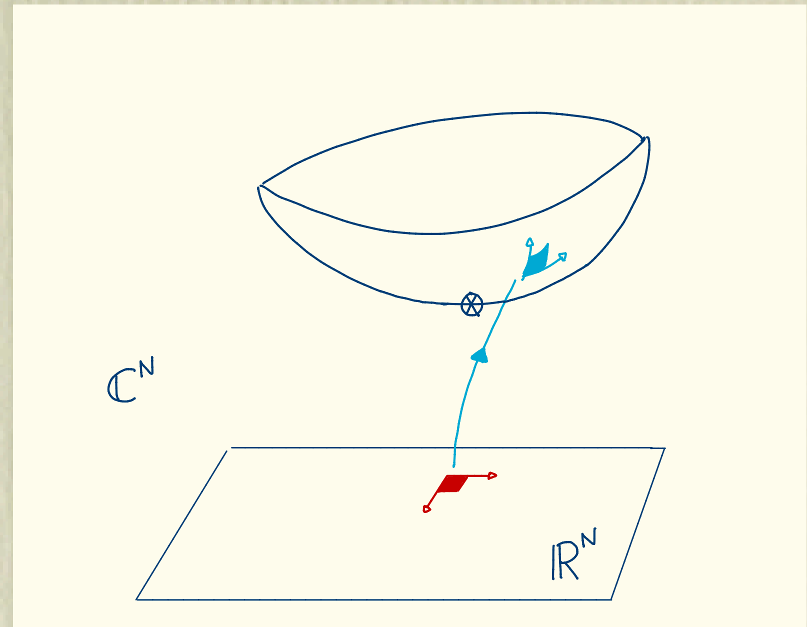
$$\Phi(x) = S_I(z(x)) - \mathrm{Im}\, \ln \det J(x)$$

$$\langle O \rangle = \frac{\left\langle O(z(x)) e^{-i\Phi(x)} \right\rangle_0}{\left\langle e^{-i\Phi(x)} \right\rangle_0}$$

# Basic idea

$$\frac{dz}{d\tau} = \overline{\frac{dS}{dz}}, \quad z(0) = x$$

$$\frac{dJ}{d\tau} = \overline{H(z)J}, \quad J(0) = I \quad H(z)_{ij} = \frac{\partial^2 S}{\partial z_i \partial z_j}$$




- The differential equations are integrated for a fixed amount of "time": $T_{\text{flow}}$
- This is expensive, especially the calculation of J
- Sampling is done based on the effective action and the phase is reweighted at the end

$$S_{\text{eff}}(x) = S_R(z(x)) - \ln|\det J(x)|$$

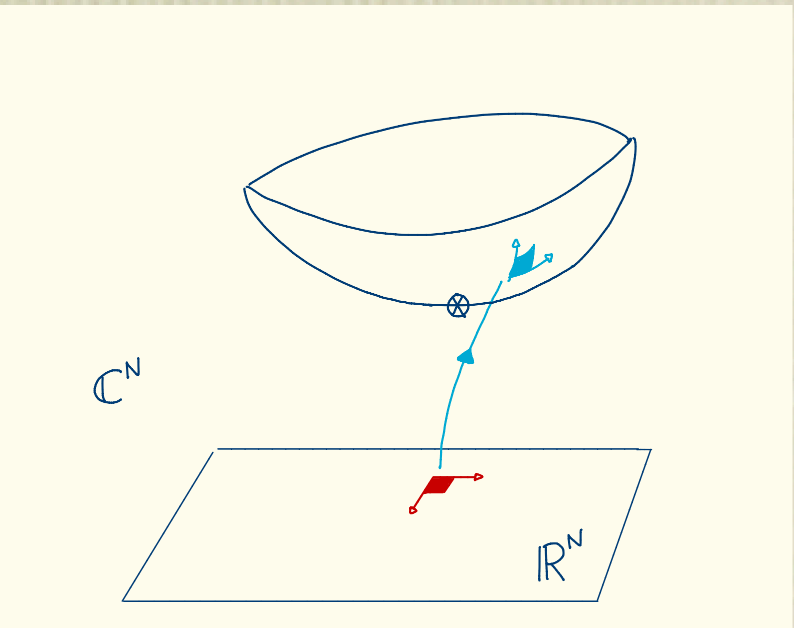
# Numerical challenge

- On the flow manifolds sampling is done based on the effective action

$$S_{\text{eff}}(x) = S_R(z(x)) - \ln|\det J(x)|$$

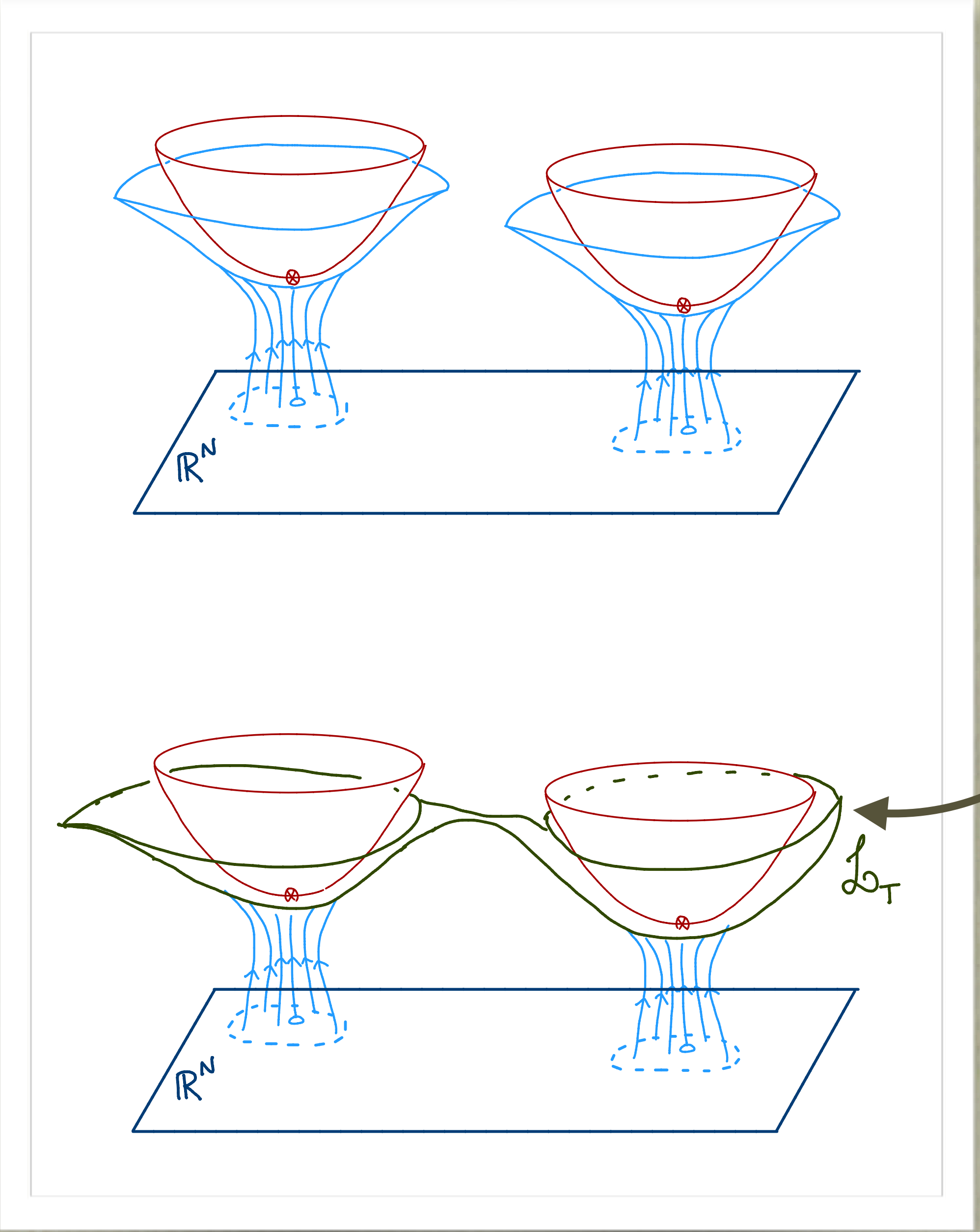
$$\frac{dz}{d\tau} = \overline{\frac{dS}{dz}}\,, \quad z(0) = x$$

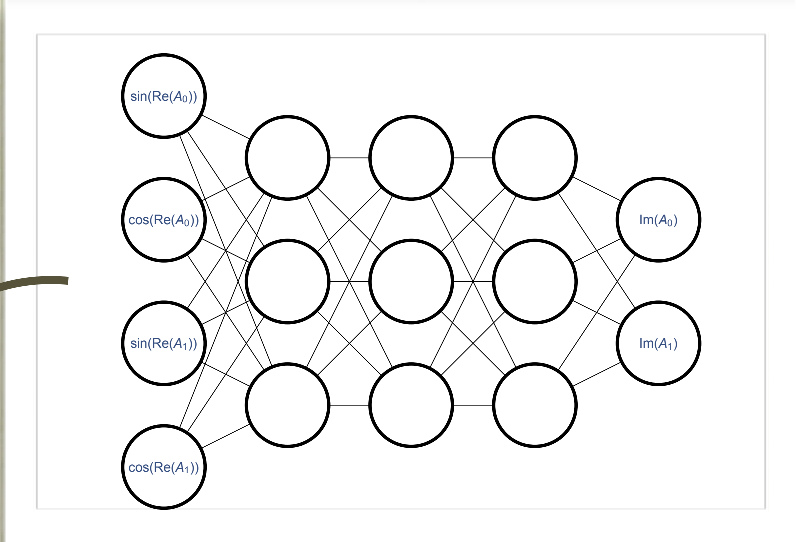
$$\frac{dJ}{d\tau} = \overline{H(z)J}\,, \quad J(0) = I \quad H(z)_{ij} = \frac{\partial^2 S}{\partial z_i \partial z_j}$$




- For each step integrate a set of differential equations to get z and J
- This is expensive, especially the calculation of J and det J
- To address this problem we used
  - improved sampling algorithms (avoid computing J or det J)
  - fast estimators for det J
  - **numerically cheaper integration manifolds**

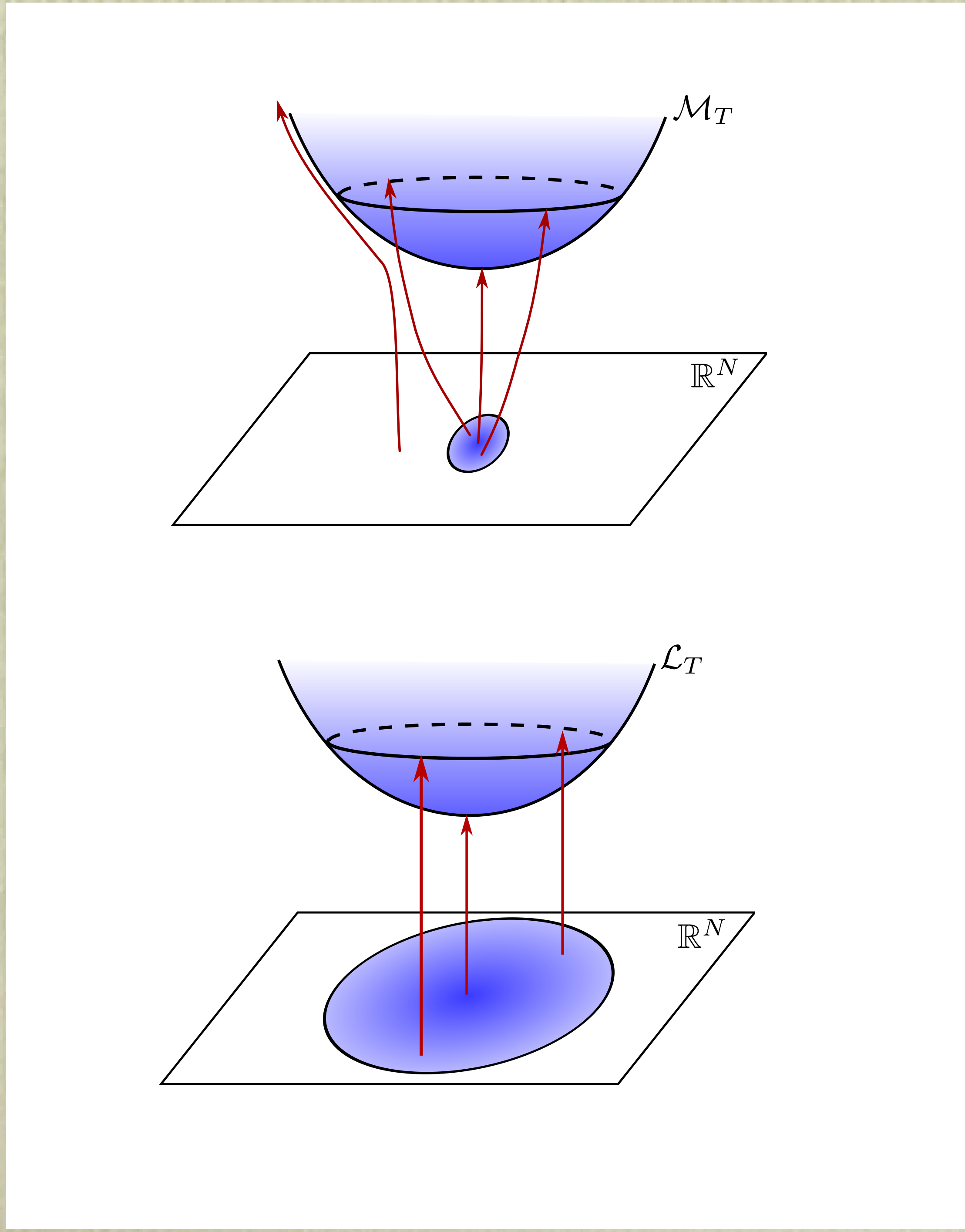
# Learnifold



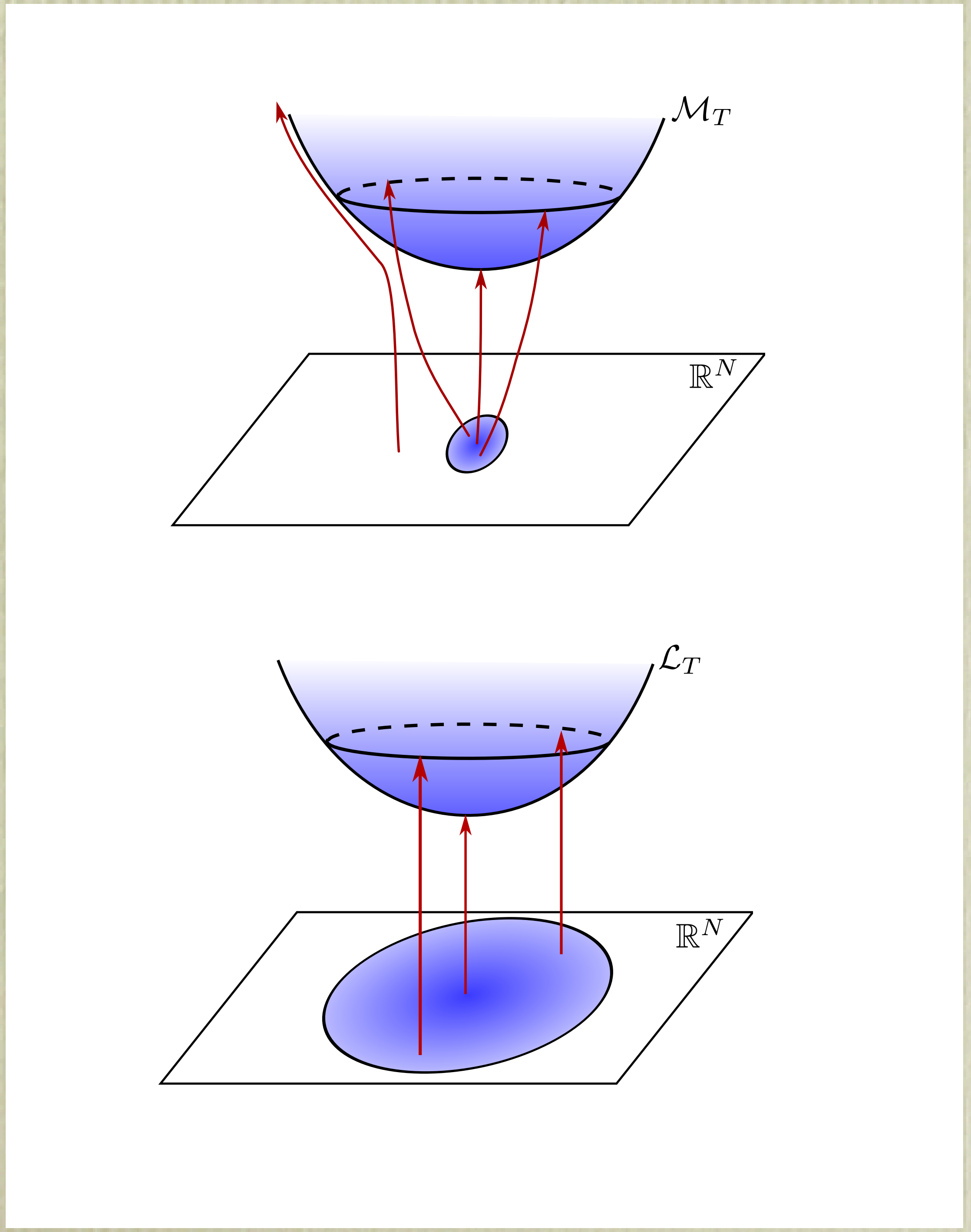

- Generate few configs on the generalized thimble manifold
- Use neural nets with appropriate symmetries to interpolate
- Integrate over the learnifold, the manifold defined by the trained neural net

# Learnifold



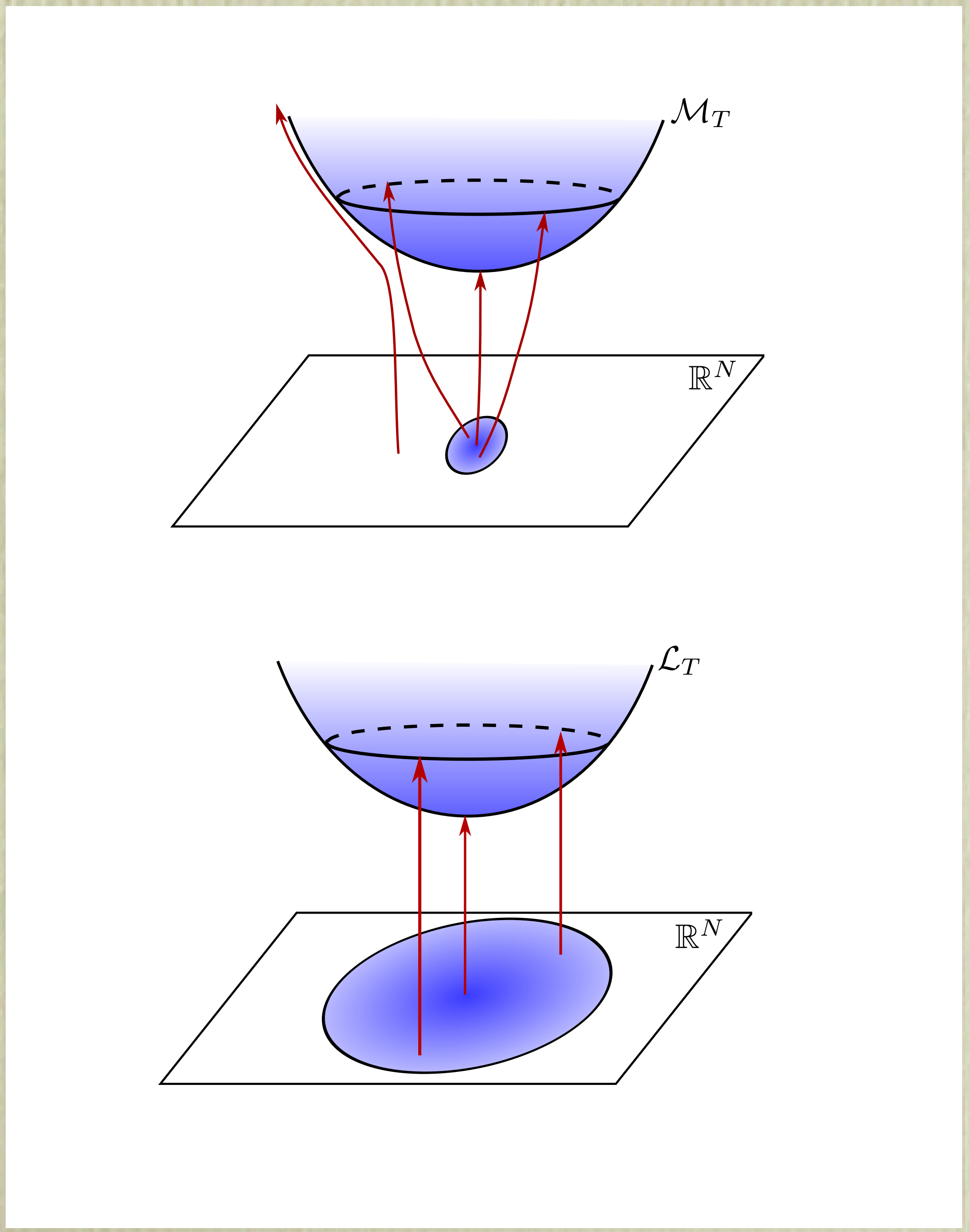

- We use a feed-forward network and train it using supervised learning
- The networks learns quickly about the constant shift, further improvements are slow
- Most of the cost is in generating the seed configurations with much less required for training
- Integrations over learnifold is fast

# Learnifold parametrization



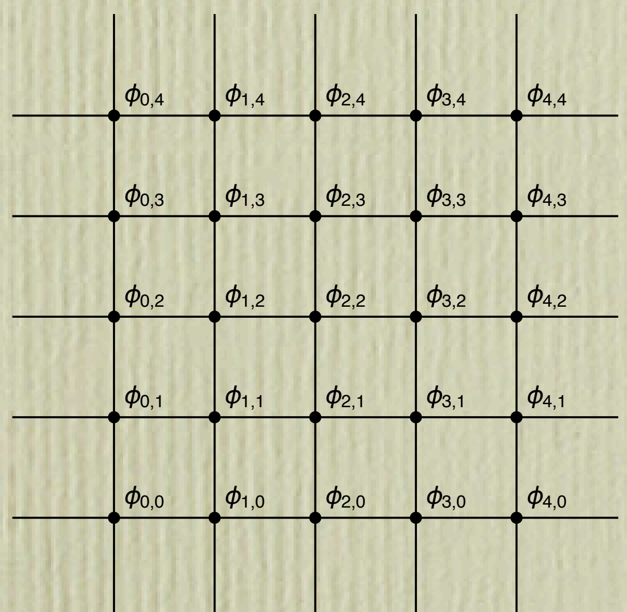

- The Learnifold map is
  $\varphi \to \tilde{\varphi} = \varphi + i\, f(\varphi)$
- For each real field a imaginary shift (elevation) is computed based on the current config
- The elevation function is determined by the neural network
- The elevation is position dependent
  $\tilde{\varphi}_i = \varphi_i + i\, f_i(\varphi)$

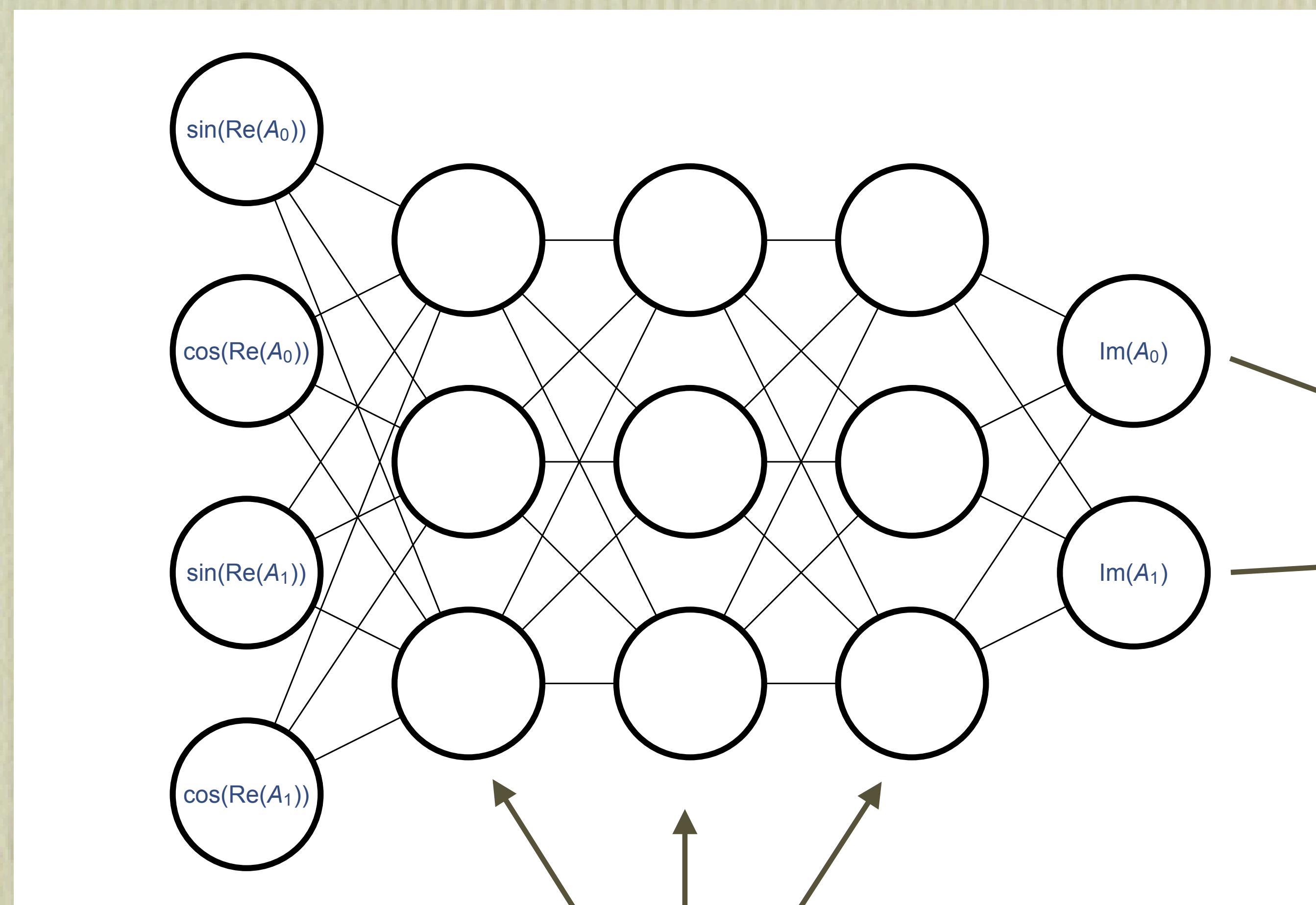
# Learnifold parametrization

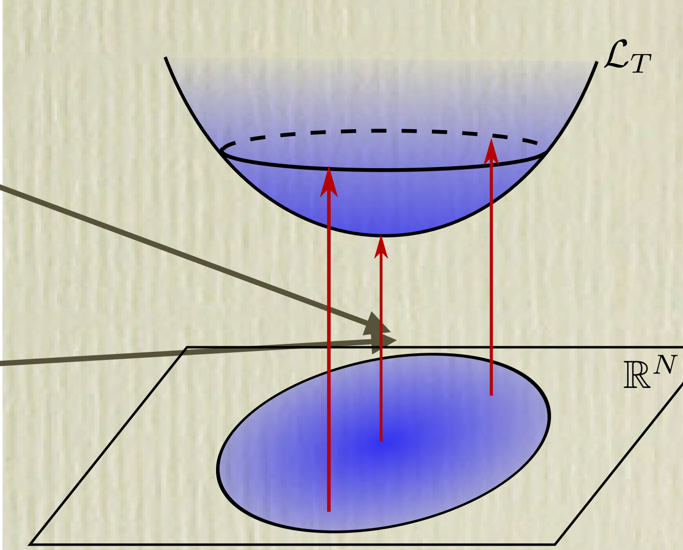



- The parametrization does not allow for "folding" manifolds, but this is not too restrictive
- The Jacobian for this map is better behaved than the generalized thimble manifold — we can disregard it during sampling and "reweight" its contribution
- The map induced by the flow creates "pockets" when multiple thimbles contribute and the distribution is usually multimodal
- The Learnifold map is smoother and avoids trapping

# Feed forward network




input
real field values

hidden layers

output
imaginary part for $A_\mu(0)$

# Feed forward network

- The feed forward network is defined by its topology and parameters: on each link we have a weight $w^{(k)}_{i,j}$, connecting node i on layer k with node j on layer k+1
- At each node we have a bias $b^{(k)}_i$
- The result is computed iteratively, starting from the input layer towards the output

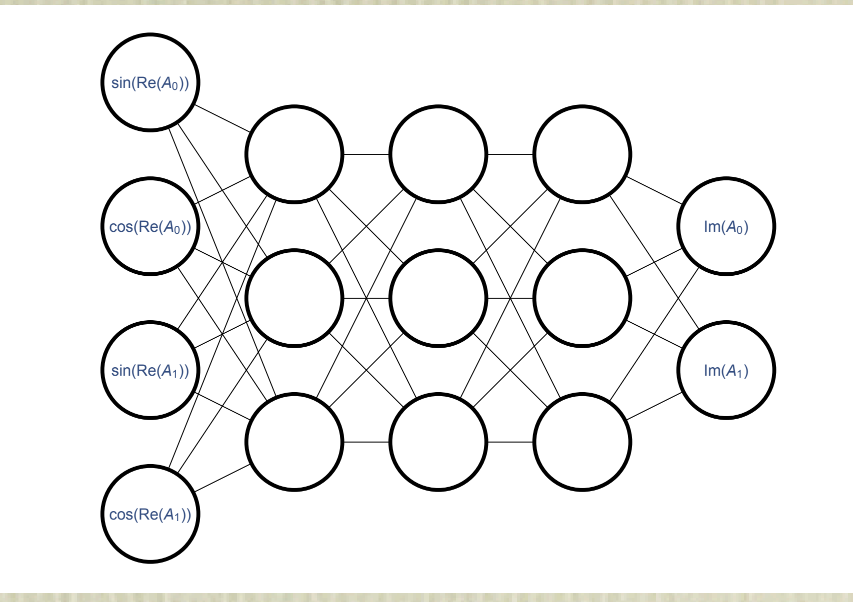
$$v^{(k+1)}_i = \sigma(b^{k+1}_i + \sum_j w^{(k)}_{j,i} v^{(k)}_j)$$

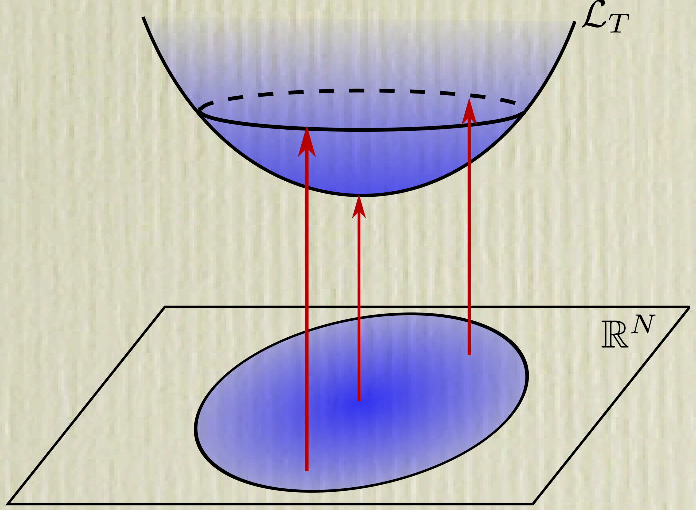
where $\sigma(x) = \log(1 + e^x)$ is a smooth step function

- The parameters are tuned using steepest descent to minimize the cost function

$$C(w, b) = \sum_i |\tilde{\varphi}(\varphi) - \hat{\varphi}_i|$$

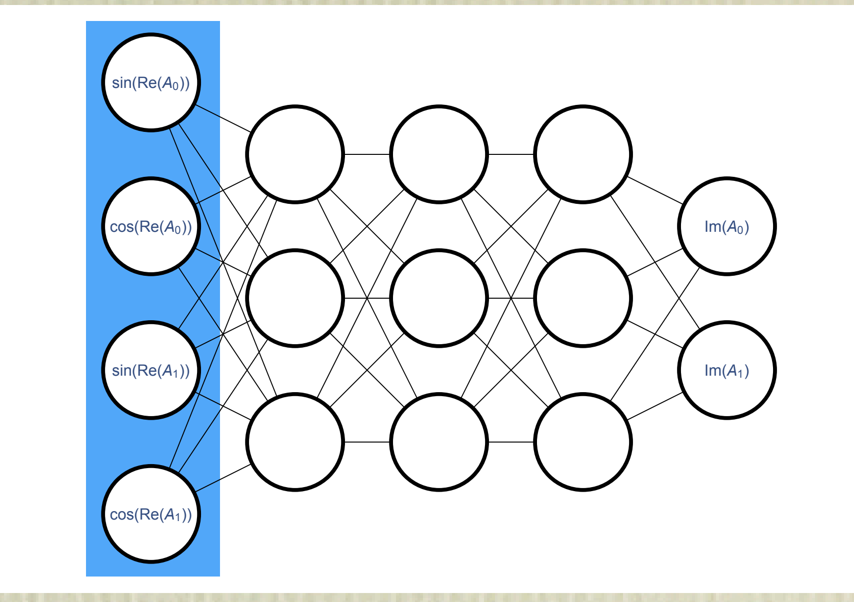
with $\hat{\varphi}_i$ the configurations in the training set.

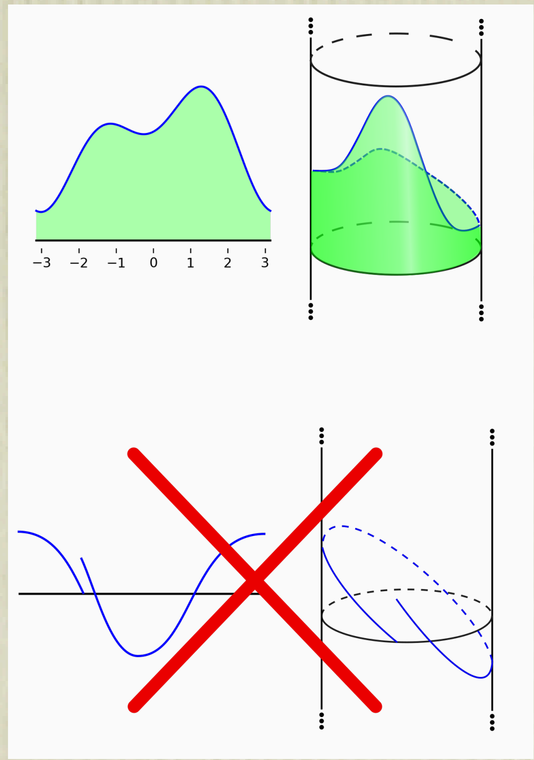


# Learnifold symmetries

- Two important considerations for the design of the learnifold:
  - respect symmetries
  - insure correctness
- For symmetries we force the network to respect translation symmetry using a single scalar function f(φ), that computes the elevation for the field located at position o on the lattice. The elevation for position x is then computed using f(Tφ), where T is the translation that brings x to o.
- To insure that the manifold is equivalent to the original manifold the asymptotic/periodicity conditions need to be satisfied. We did this by using as input periodic functions of the field.



# Case study: massive Thirring model in 1+1D

$$\mathcal{L} = \bar{\psi}^a \left(\gamma^\mu \partial_\mu + m + \mu\gamma^0\right) \psi^a + \frac{g^2}{2N_f}(\bar{\psi}^a \gamma^\mu \psi^a)(\bar{\psi}^b \gamma_\mu \psi^b)$$

Auxiliary field A's

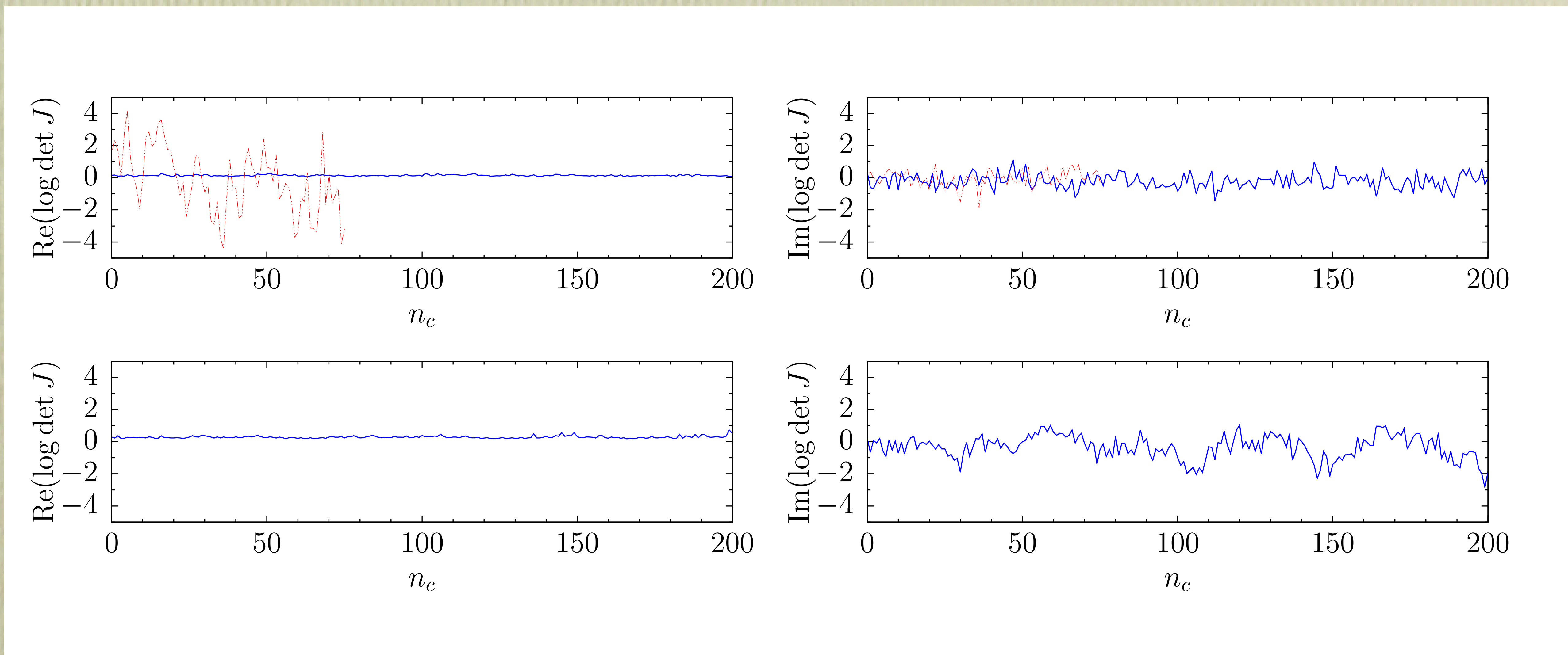
$$S = \int d^2x \, \left[\frac{N_F}{2g^2} A_\mu A_\mu + \bar{\psi}^\alpha (\not\partial + \mu\gamma_0 + i\not A + m)\psi^\alpha\right]$$

Discretization (compact A's)

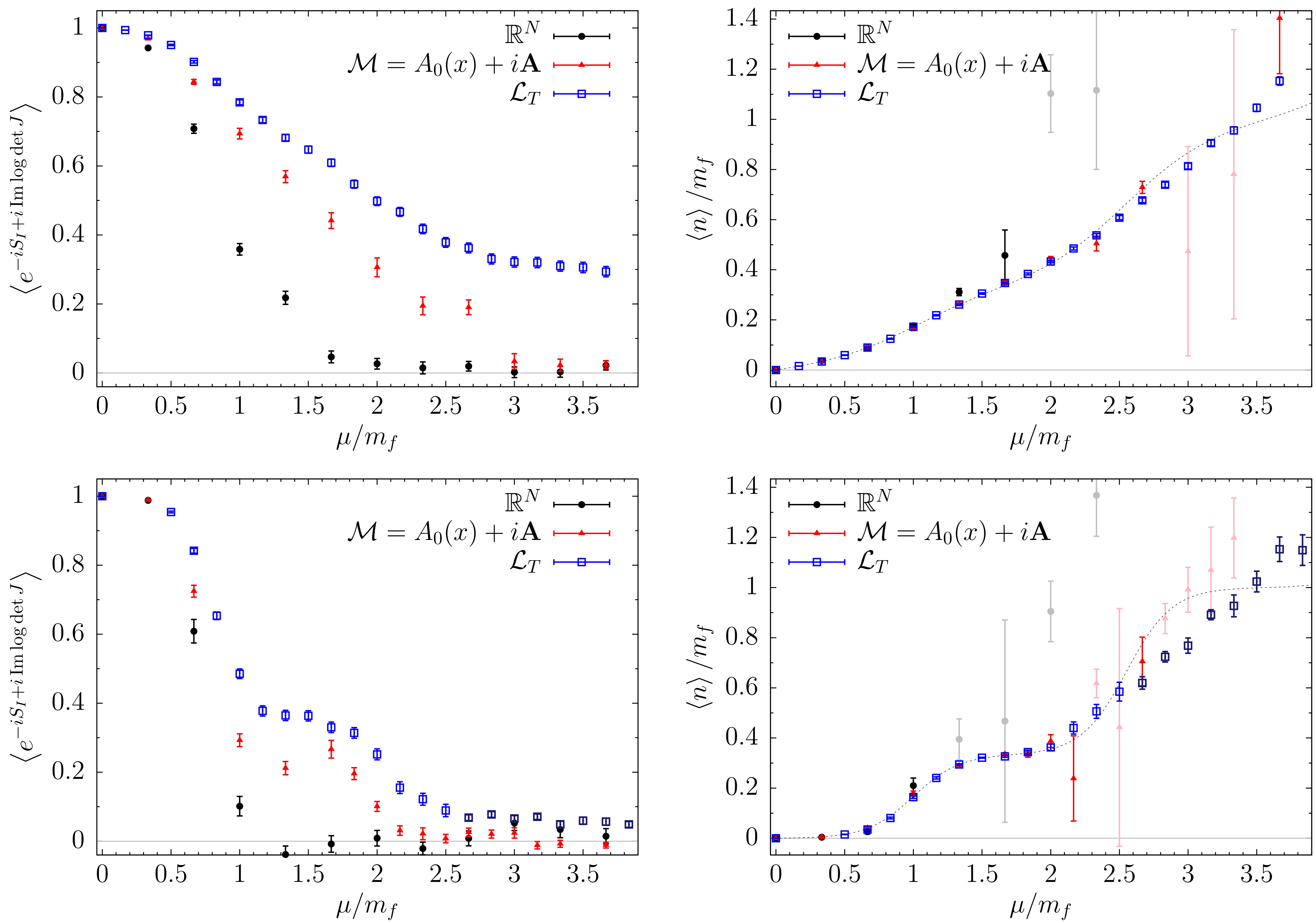
complex for μ ≠ 0

$$S = N_F \left(\frac{1}{g^2}\sum_{x,\nu}(1 - \cos A_\nu(x)) - \gamma \log \det D(A)\right)$$

# Jacobian fluctuations



# Results



$\mathbb{R}^N$
$\mathcal{M} = A_0(x) + i\mathbf{A}$
$\mathcal{L}_T$
$\left\langle e^{-iS_I + i\,\mathrm{Im}\log\det J} \right\rangle$
$\langle n \rangle / m_f$
$\mu/m_f$

# Conclusions

- Thimbles and holomorphic flow manifolds are only one option for complex deformations. There is a large degree of freedom in choosing complex deformations to address the numerical challenges specific to the system of interest (with new challenges and opportunities).

- Machine learning can be used to define such a manifold and match it to manifolds with reduced sign fluctuations. The resulting manifolds are correct by design and the success can be measured by the size of the sign fluctuations, which is easy to test aposteriori.

- For the test case we showed that the sign problem is alleviated allowing us to investigate new parameter space. The physical results confirm that we correctly sample the configurations space. Sampling is fast, the Jacobian can be reweighed, and the cost is dominated by generating training set.