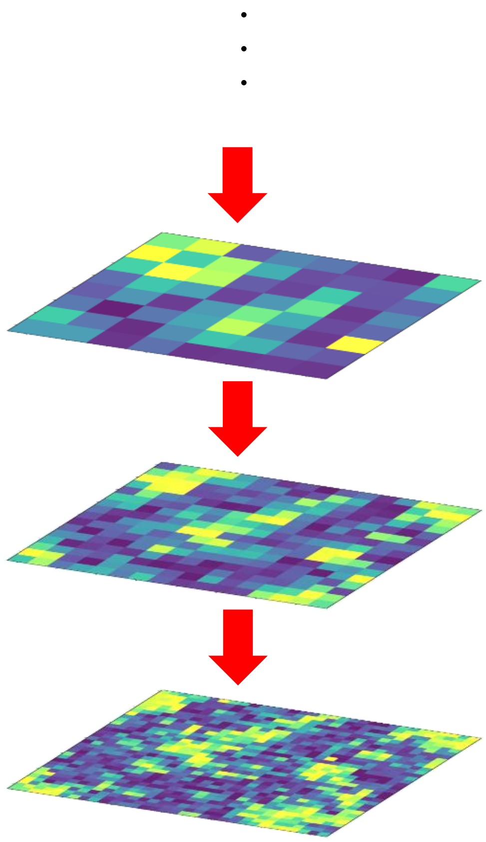
Renormalization Group Approach for Machine Learning Hamiltonian



Misaki Ozawa

Université Grenoble Alpes

Collaboration with

Tanguy Marchand, Giulio Biroli, and Stephane Mallat

Marchand, Ozawa, Biroli, and Mallat, Phy. Rev. X 2023

cnrs

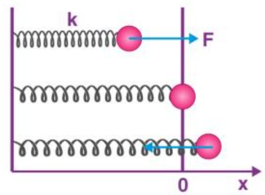
LIPhy

MIAI Grenoble Alpes
Multidisciplinary Institute In Artificial Intelligence

Machine Learning Hamiltonian

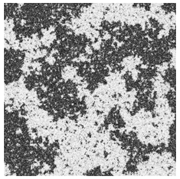
Hamonic oscillator

$$\mathcal{H} = \frac{1}{2m}p^2 + \frac{1}{2}kx^2$$



Ising model

$$\mathcal{H} = -J\sum_{ij} S_i S_j$$



φ^4 field model

$$\mathcal{H}(\varphi) = \int \mathrm{d}x \left(|\nabla\varphi|^2 + t\varphi^2 + \varphi^4\right)$$

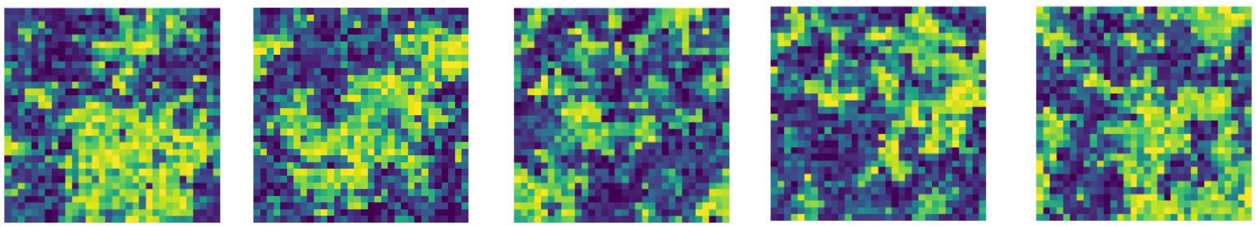
DATA

Machine learning

$\mathcal{H}$

Motivation

Field φ (Magnetization fields, dark matter distribution, etc.)



. . .

Training data

Standard physics approach

Inverse problem by machine learning

Cocco et al., Rep. Prog. Phys. 2018

$$p(\varphi) = \frac{e^{-\beta \mathcal{H}(\varphi)}}{Z}$$

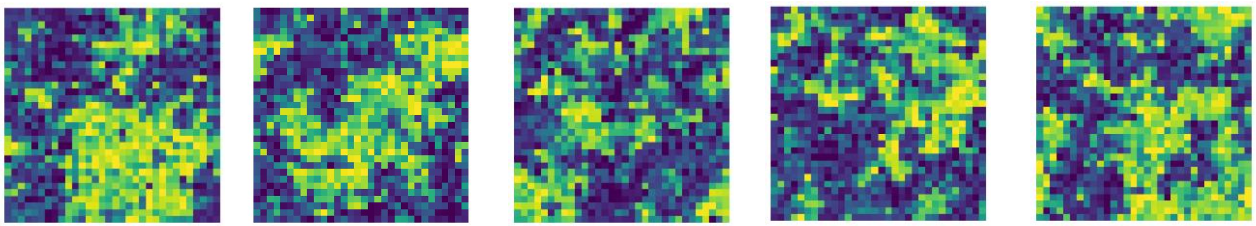
?

Lingxiao’s talk on Friday

Can we construct probability distribution or Hamiltonian from data?

Motivation

Field φ (Magnetization fields, dark matter distribution, etc.)



. . .

Training data

Generate new data

Gabrié, Rotskoff, Vanden-Eijnden, PNAS 2022

Inverse problem
by machine learning

Cocco et al., Rep. Prog. Phys. 2018

$$p(\varphi) = \frac{e^{-\beta \mathcal{H}(\varphi)}}{Z}$$

Once we learn $p(\varphi)$, we can use it as an efficient generative model

Decelle, Furtlehner, and Seoane, Adv Neural Inf Process Syst NeurIPS 2021

Main messages

- Learning microscopic Hamiltonian from data

(Generative model with high interpretability)

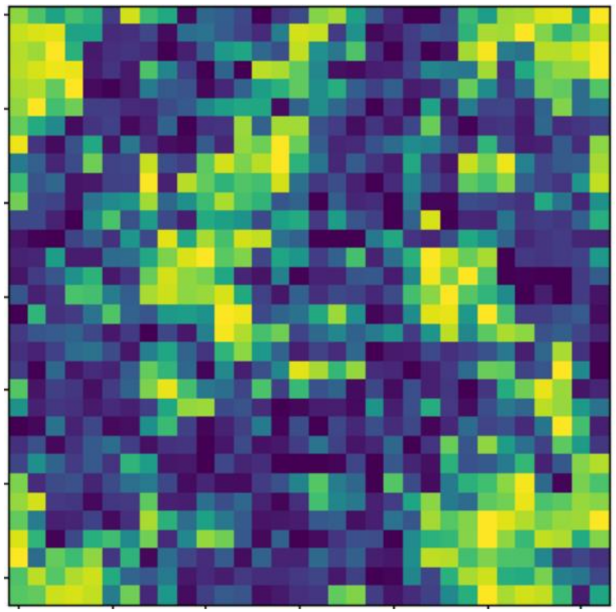
- Overcoming critical slowing down in training process

- Overcoming critical slowing down in generation process

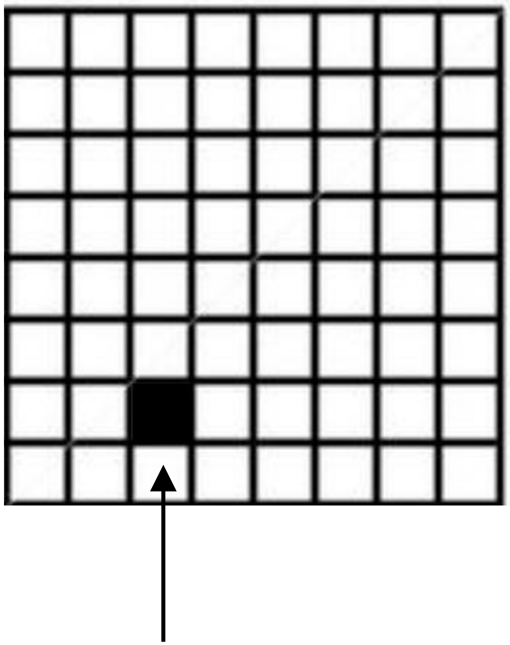
(RG is the key to achieve these goals)

Lattice field systems

Field φ



N sites



φ_i on site i

Naive approach

$$p_K(\varphi) = \frac{e^{-\mathcal{H}_K(\varphi)}}{Z_K}$$ is an estimate of true distribution, $$p(\varphi) = \frac{e^{-\mathcal{H}(\varphi)}}{Z}$$

Ansatz

Gaussian — Non-linear potential

$$\begin{aligned}\mathcal{H}_K(\varphi) &= \overbrace{\sum_{i,j} \varphi_i J_{ij} \varphi_j}^{\text{Gaussian}} + \overbrace{c_1 \sum_i \varphi_i + c_2 \sum_i \varphi_i^2 + c_3 \sum_i \varphi_i^3 + \cdots + c_m \sum_i \varphi_i^m}^{\text{Non-linear potential}} \\ &= K^\top O(\varphi)\end{aligned}$$

Coupling parameters: $K = (J,\ c_1,\ c_2,\ c_3, \cdots, c_m)$

Operators: $O(\varphi) = \left(\varphi\varphi^\top,\ \sum_i \varphi_i,\ \sum_i \varphi_i^2,\ \sum_i \varphi_i^3,\ \cdots, \sum_i \varphi_i^m\right)$

Task: Determination of K from the training dataset

Naive approach

Minimize Kullback-Leibler Distance between $p(\varphi)$ **and** $p_K(\varphi) = \frac{e^{-\mathcal{H}_K(\varphi)}}{Z_K}$

$$D(p||p_K) = \int \mathrm{d}\varphi \; p(\varphi) \ln \frac{p(\varphi)}{p_K(\varphi)}$$

with

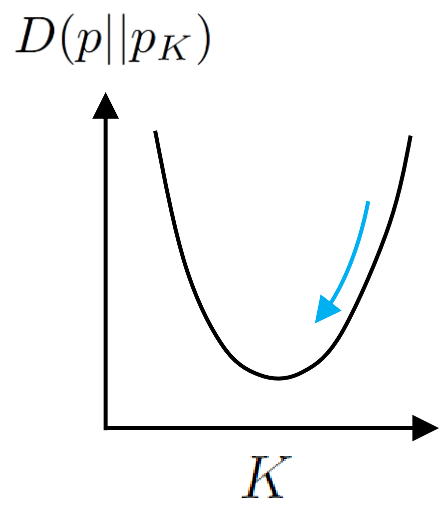
$$\mathcal{H}_K(\varphi) = K^\top O(\varphi)$$

Gradient Descent for Convex Optimization

$$\frac{\mathrm{d}K}{\mathrm{d}t} = -\nabla_K D(p||p_K) = \langle O \rangle_{p_K} - \langle O \rangle_p$$

Monte-Carlo by $p_K(\varphi)$ — Training data

$$\langle \cdots \rangle_{p_K} = \int \mathrm{d}\varphi \; p_K(\varphi)(\cdots)$$



Zhu, Wu, and Mumford, Neural Comput. 1997

Decelle, Furtlehner, and Seoane, Adv Neural Inf Process Syst NeurIPS 2021

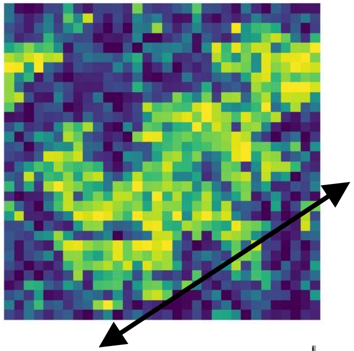
Naive approach

Problem

Convergences are very slow when long-range correlations appear

Timescale for Monte-Carlo (MC)

$$\tau_{\mathrm{MC}} \sim \max_k \mathcal{F}\left[\underbrace{\langle \varphi_i \varphi_j \rangle_p}_{\text{Long-range}} \right] \sim \xi^2$$



Correlation length ξ

Timescale for Gradient Descent (GD)

$$\tau_{\mathrm{GD}} \sim \left(\max_k \mathcal{F}\left[\underbrace{\langle \varphi_i \varphi_j \rangle_p}_{\text{Long-range}} \right] \right)^2 \sim \xi^4$$

(Results from Gaussian field theory)

Critical slowing down appears in Machine Learning!

Renormalization Group (RG)

One step RG

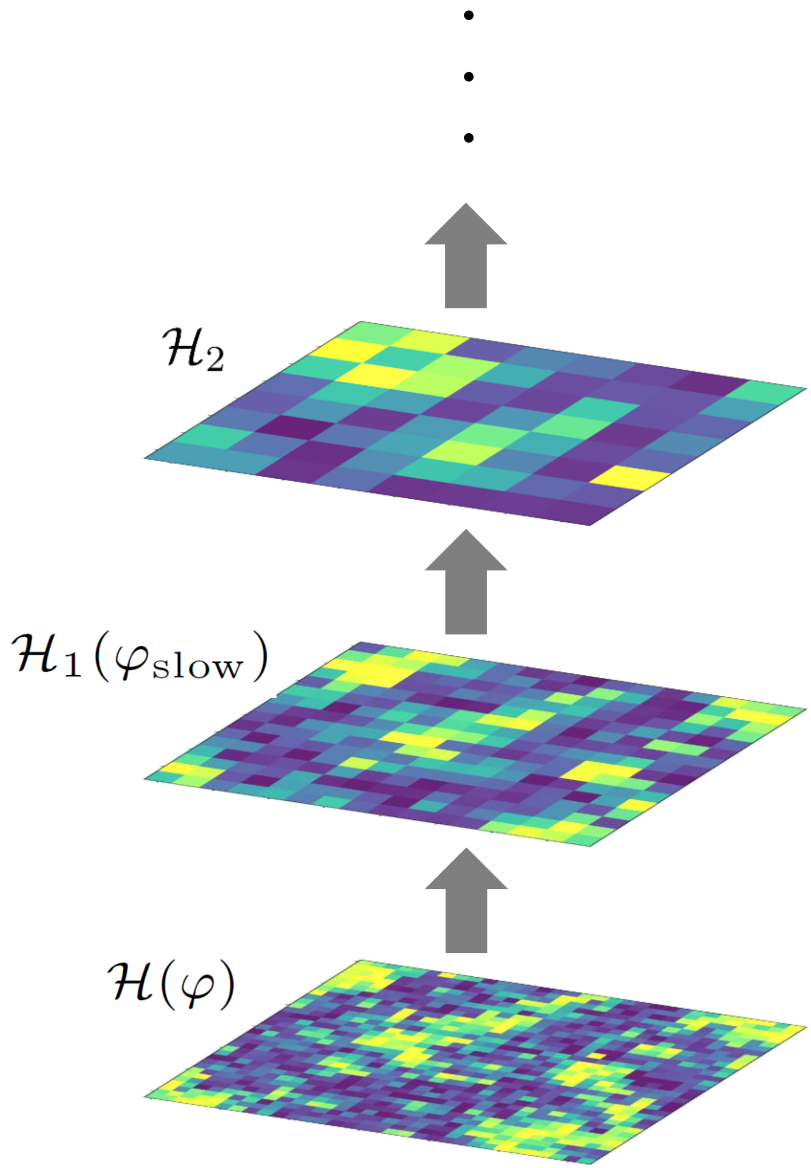
Wilson and Kogut, Phys. Rep. 1974

$$\varphi = \varphi_{\text{slow}} + \varphi_{\text{fast}}$$

$$e^{-\mathcal{H}_1(\varphi_{\text{slow}})} = \int \mathrm{d}\varphi_{\text{fast}}\; e^{-\mathcal{H}(\varphi)}$$

RG steps scale by scale manner

$$\mathcal{H} \to \mathcal{H}_1 \to \mathcal{H}_2 \to \cdots$$



Inverse Renormalization Group

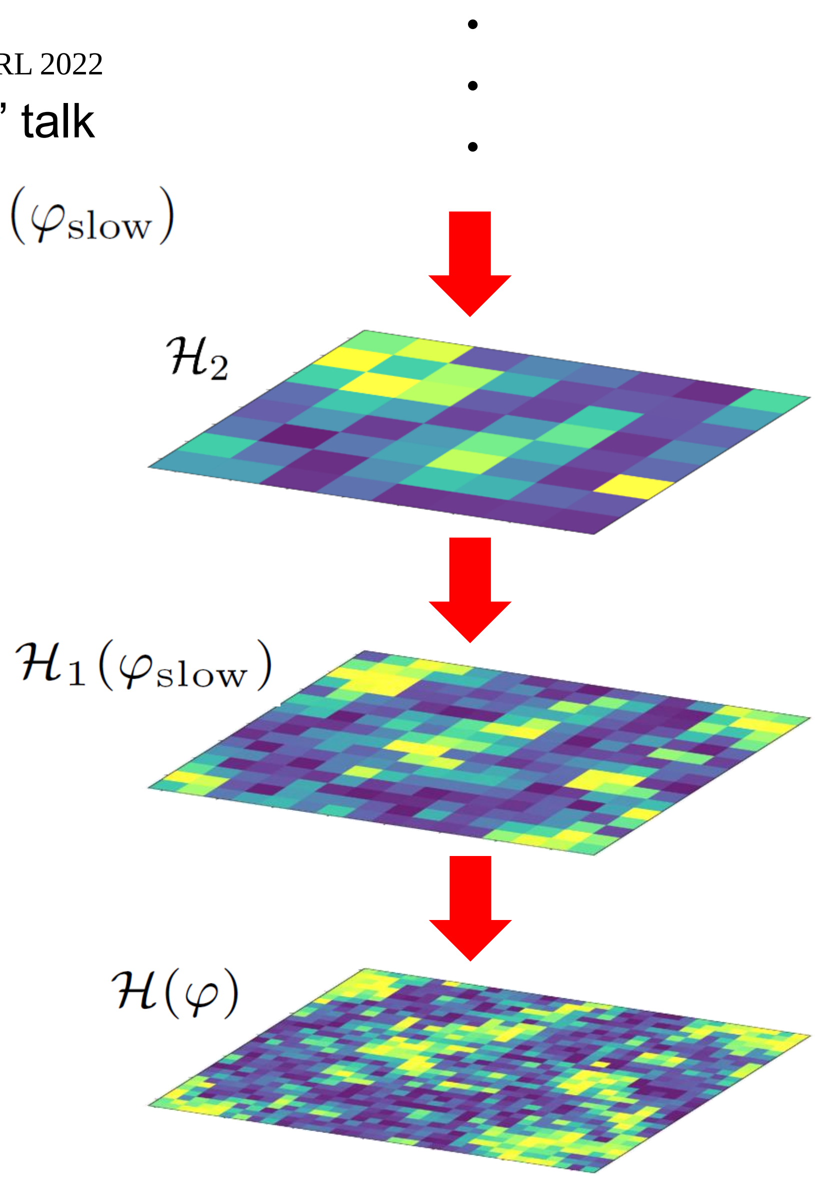
One step Inverse RG

Ron, Swendsen, and Brandt, PRL 2002

Bachtis, Aarts, Di Renzo, and Lucini, PRL 2022

Bachtis, arXiv 2024 Demitrios' talk

$$p(\varphi) = p(\varphi_{\text{slow}}, \varphi_{\text{fast}}) = p^{\text{c}}(\varphi_{\text{fast}}|\varphi_{\text{slow}})p_1(\varphi_{\text{slow}})$$



Ansatz in Conditional probability

$$p^{\text{c}}_K(\varphi_{\text{fast}}|\varphi_{\text{slow}}) = \frac{e^{-\mathcal{H}^{\text{c}}_K(\varphi_{\text{fast}},\varphi_{\text{slow}})}}{Z^{\text{c}}_K}$$

$$\mathcal{H}^{\text{c}}_K(\varphi_{\text{fast}}, \varphi_{\text{slow}}) = K^\top O(\varphi_{\text{fast}}, \varphi_{\text{slow}})$$

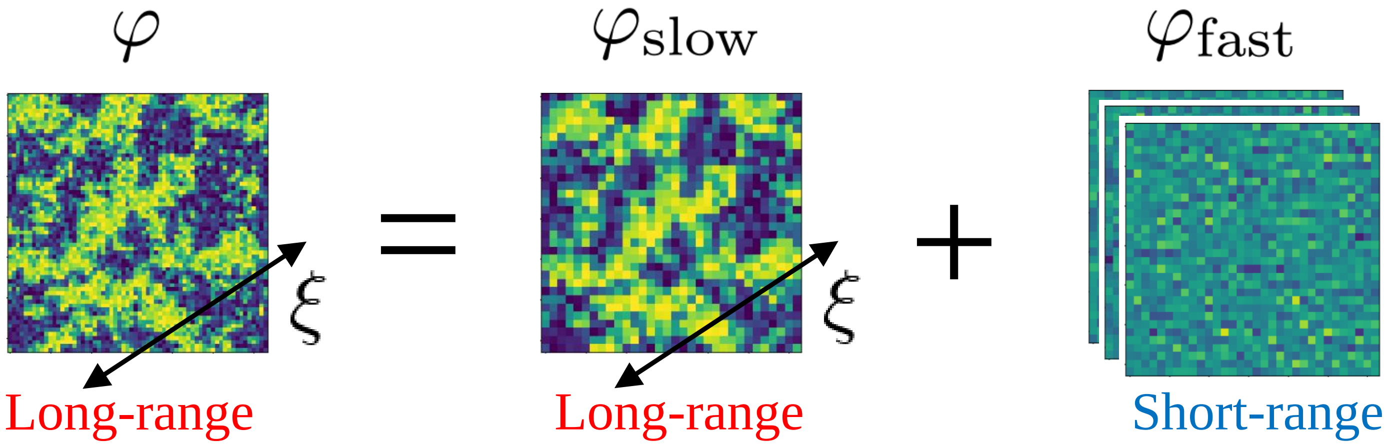
Gradient Descent minimizing KL distance

$$\frac{\mathrm{d}K}{\mathrm{d}t} = \langle O \rangle_{p^{\text{c}}_K} - \langle O \rangle_p$$

Monte-Carlo by $p^{\text{c}}_K(\varphi_{\text{fast}}|\varphi_{\text{slow}})$

Training data

Overcoming the critical slowing down



$$\tau_{\mathrm{MC}} \sim \max_k \mathcal{F}\left[\underbrace{\langle \varphi_{\mathrm{fast},i}\varphi_{\mathrm{fast},j} \rangle_p}_{\text{Short-range}} \right] \sim \mathrm{const.}$$

$$\tau_{\mathrm{GD}} \sim \left(\max_k \mathcal{F}\left[\underbrace{\langle \varphi_{\mathrm{fast},i}\varphi_{\mathrm{fast},j} \rangle_p}_{\text{Short-range}} \right] \right)^2 \sim \mathrm{const.}$$

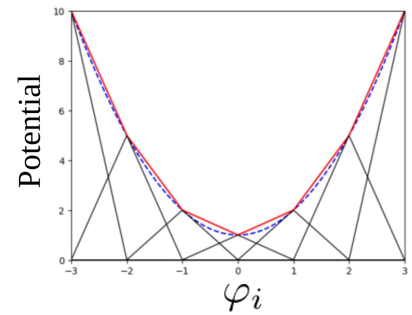
Timescales do not depend on ξ

Some technical issues

Non-linear potential

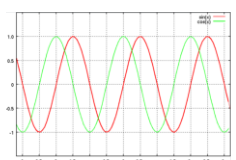
Polynomial functions → Linear combination of Hat functions

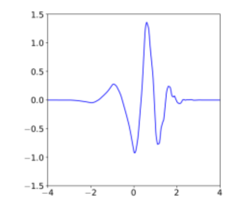
$$c_1\varphi_i + c_2\varphi_i^2 + c_3\varphi_i^3 + \cdots + c_m\varphi_i^m$$



Scale separation $\varphi = \varphi_{\text{slow}} + \varphi_{\text{fast}}$

Fourier transform → Wavelet transform





$$\varphi_{\text{slow}} = \int_{\text{Low }k} \mathrm{d}k\ e^{ikx}\hat{\varphi}$$

$$\varphi_{\text{fast}} = \int_{\text{High }k} \mathrm{d}k\ e^{ikx}\hat{\varphi}$$

$$\varphi_{\text{slow}} = \mathrm{L}^\top \mathrm{L}\varphi$$

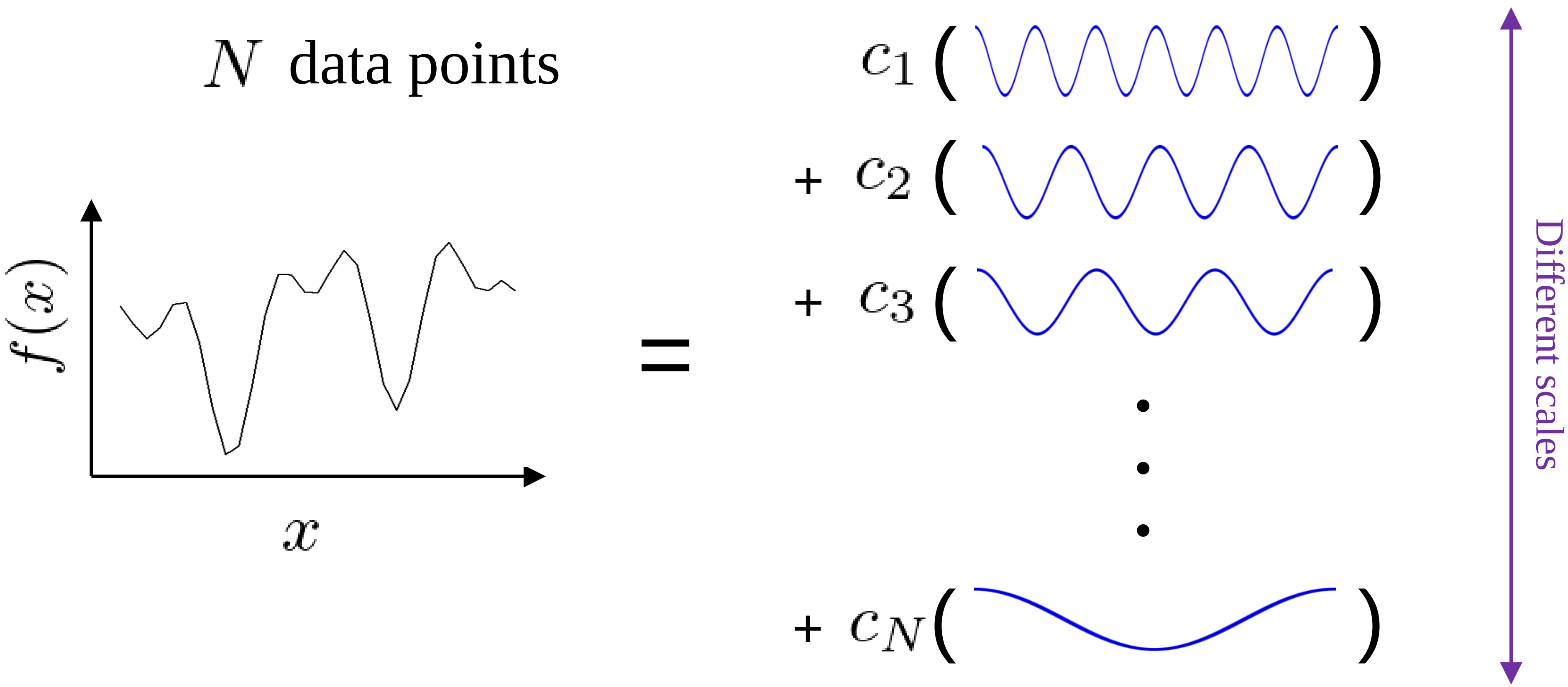
$$\varphi_{\text{fast}} = \mathrm{H}^\top \mathrm{H}\varphi$$

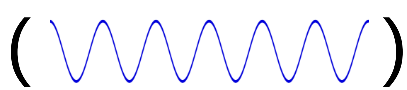
Mallat, IEEE Trans. Pattern Anal. Mach. Intell. 1989

Long-range interactions in the k-space

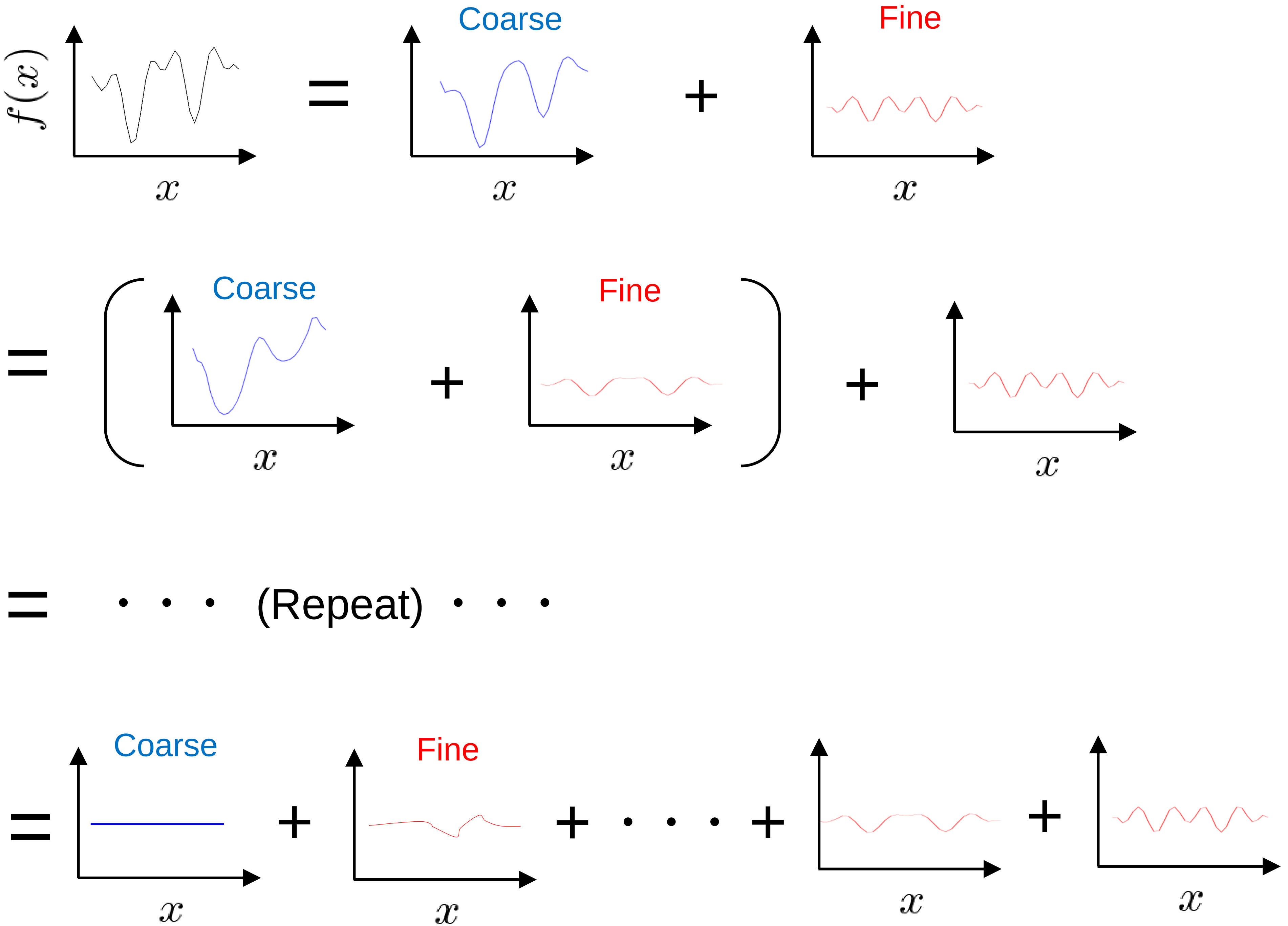
Localized in the real and k-spaces

Fourier transform



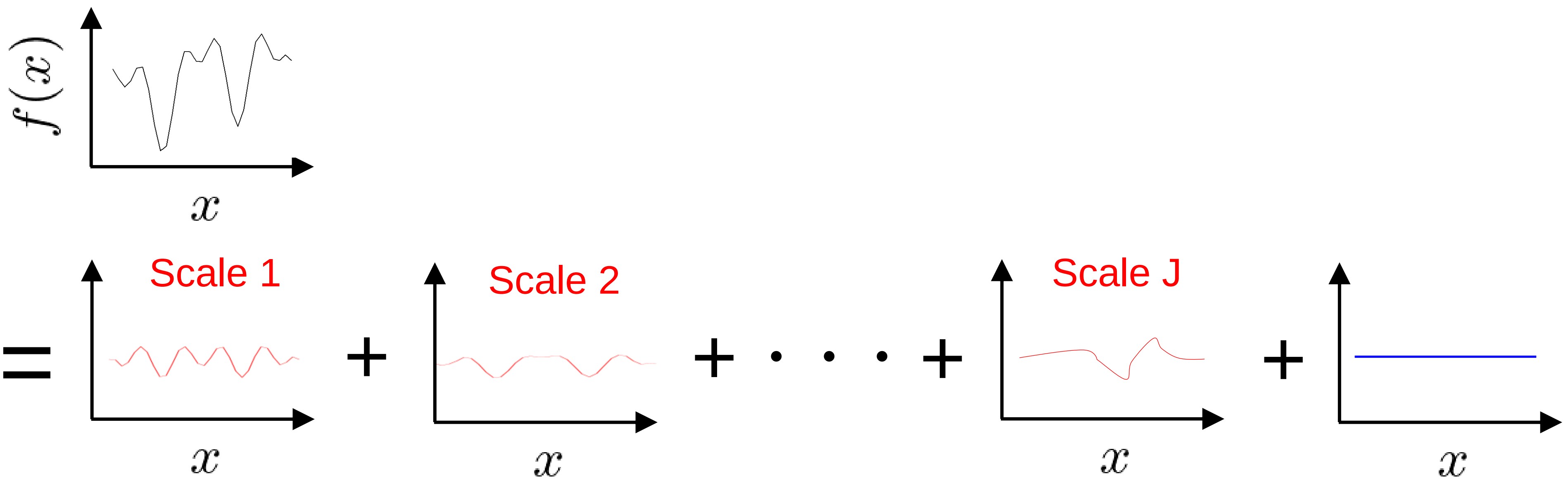
’s form an **orthonormal basis**

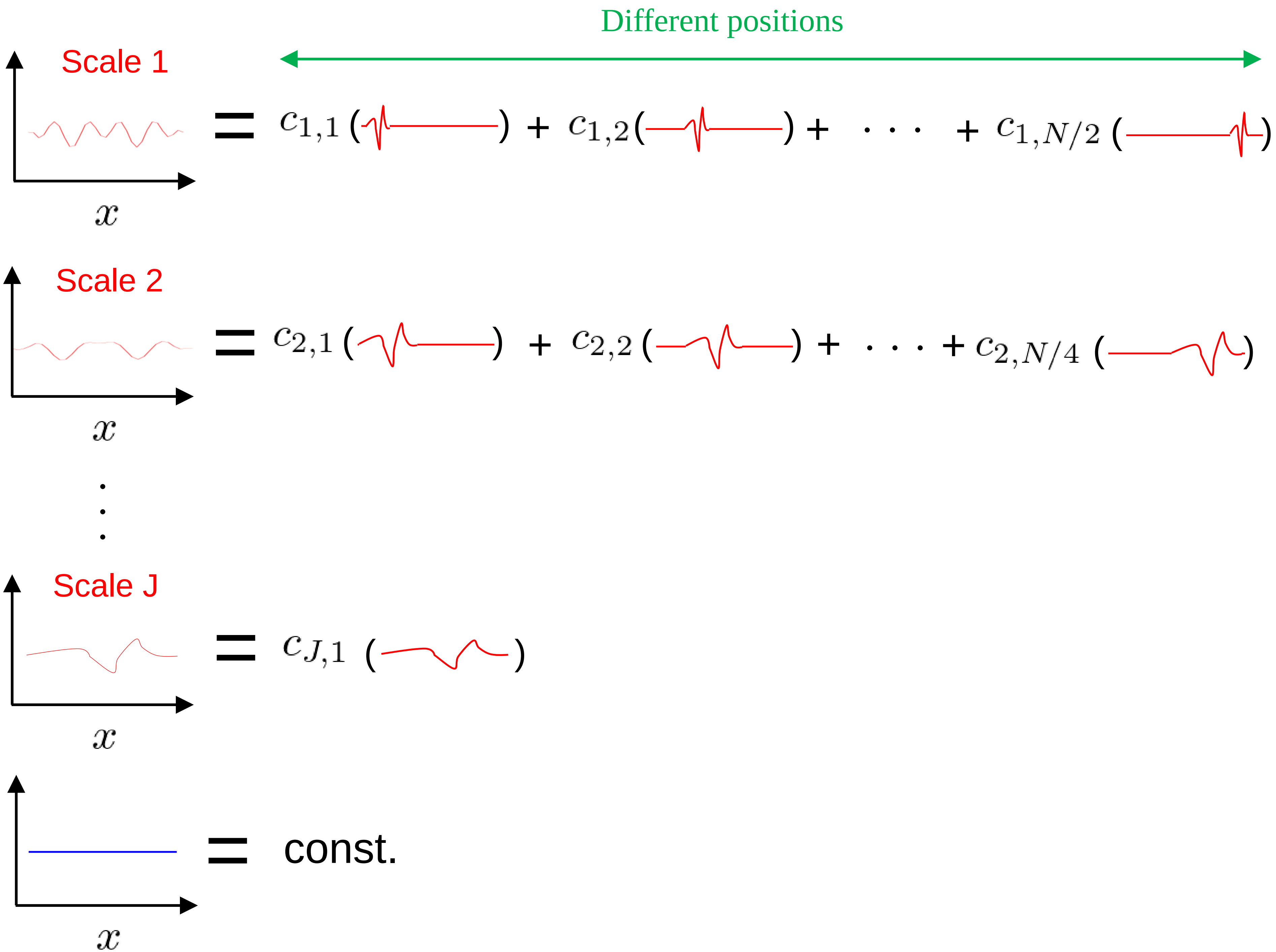
# of coefficients = N



$f(x)$
x
Coarse
Fine
(Repeat)

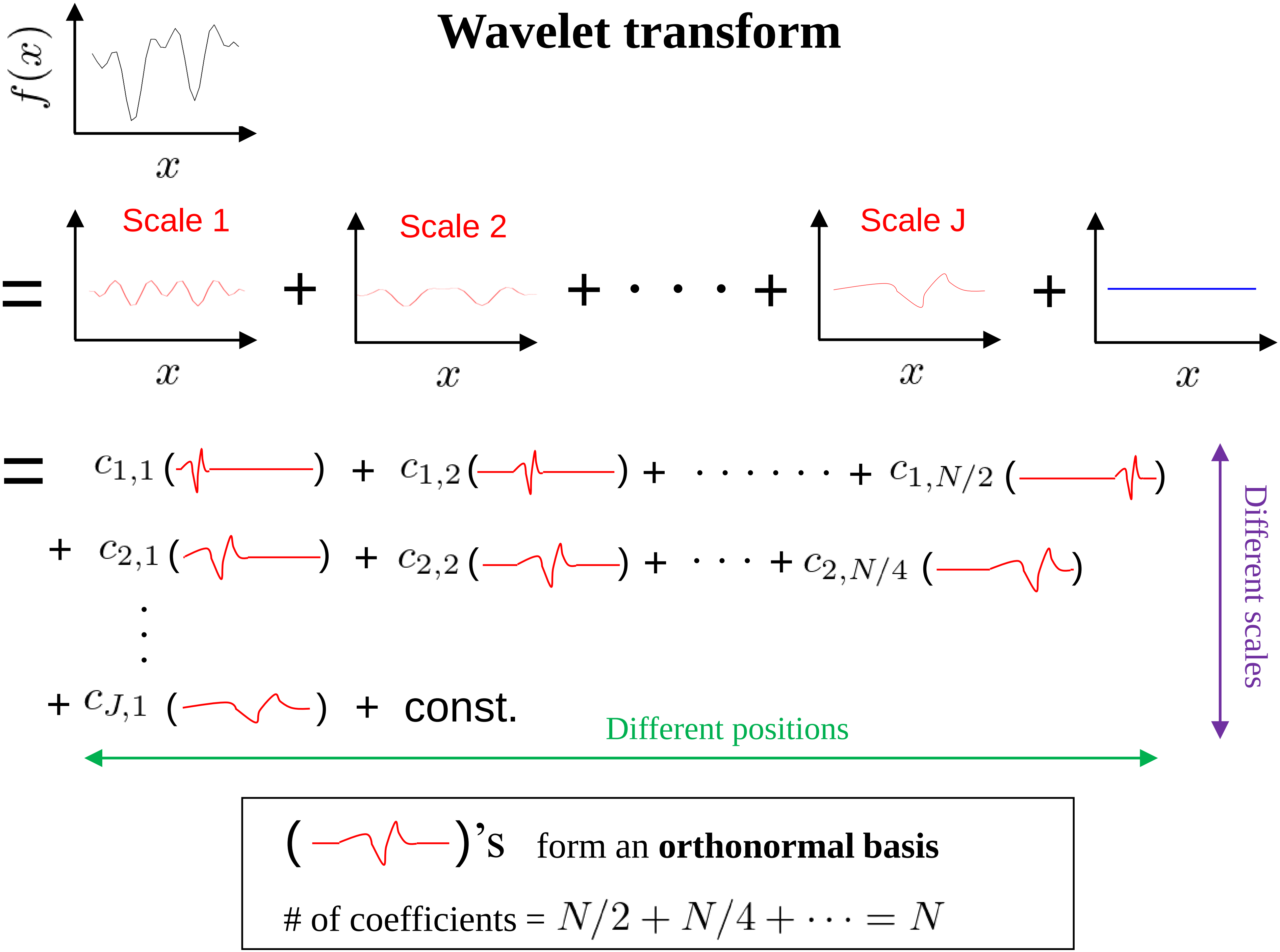
Wavelet transform



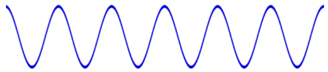


Different positions
Scale 1
$= c_{1,1}(\ ) + c_{1,2}(\ ) + \cdots + c_{1,N/2}(\ )$
x
Scale 2
$= c_{2,1}(\ ) + c_{2,2}(\ ) + \cdots + c_{2,N/4}(\ )$
x
Scale J
$= c_{J,1}(\ )$
x
= const.
x

Wavelet transform



Fourier transform

()'s form an **orthonormal basis**

Specified by **scales**

# of coefficients = N

Wavelet transform

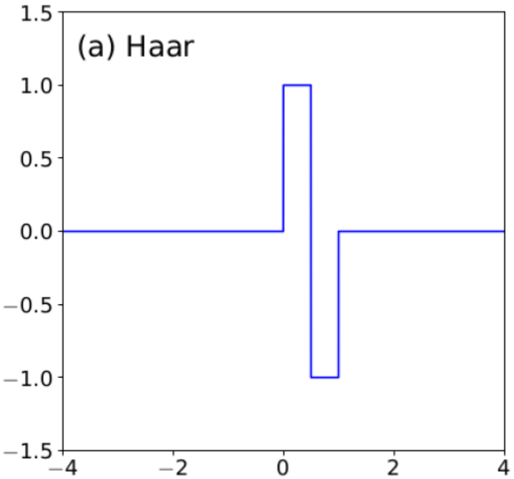
()'s form an **orthonormal basis**

Specified by **scales** and **positions**

# of coefficients = $N/2 + N/4 + \cdots = N$

Sparsity: Most of coefficients are nearly zero (e.g., JPEG 2000)

Haar wavelet (1909)

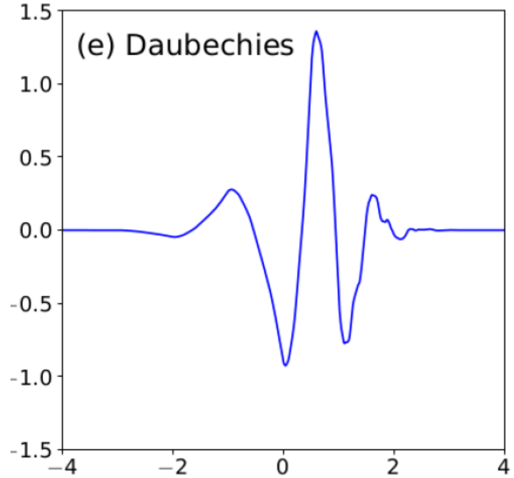


Kadanoff's block averaging

Real space (**localized**)

k-space (**delocalized**)

Modern wavelets (1980's-)

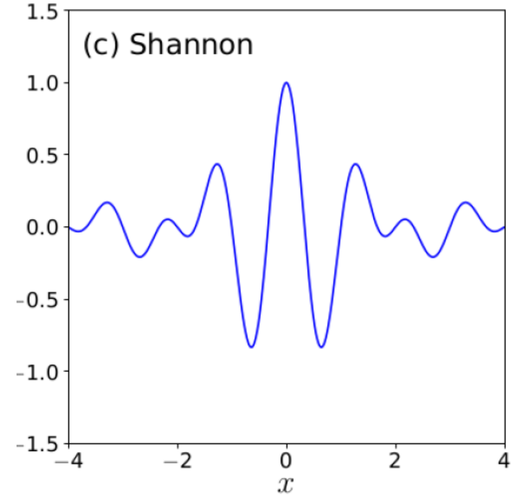


Many different wavelets were developped

Real space (**localized**)

k-space (**localized**)

Shannon wavelet (1940's)



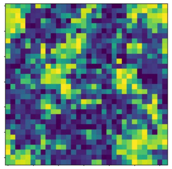
Wilson, PRB 1971

Real space (**delocalized**)

k-space (**localized**)

In practice, we can use *PyWavelets* (Software)

Test case study



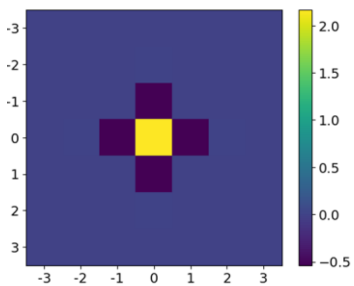
Lattice φ^4 field model

Milchev, Heermann, and Binder, J. Stat. Phys. 1986

$$\mathcal{H}(\varphi) = \int \mathrm{d}x \left(|\nabla \varphi|^2 + t\varphi^2 + \varphi^4 \right)$$

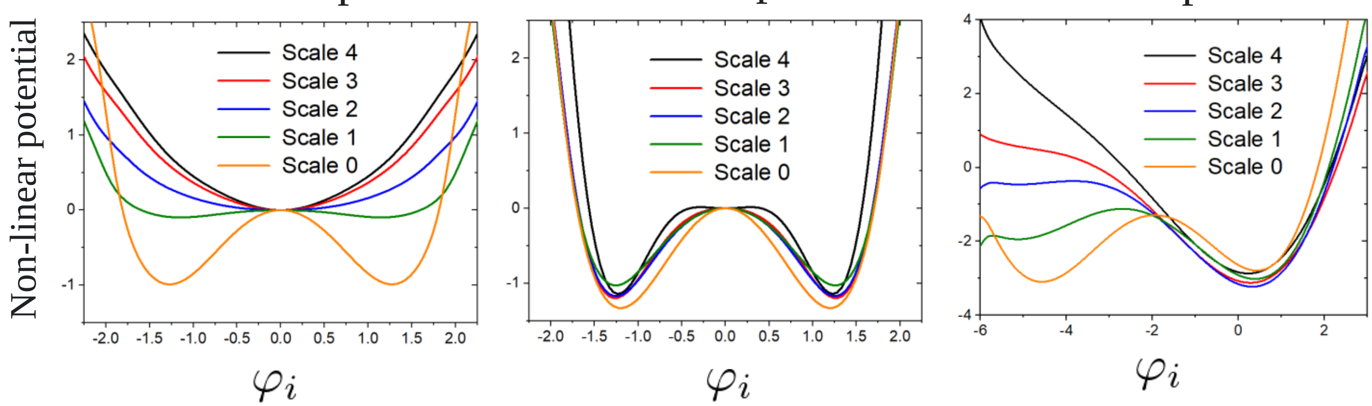
Reconstruction of Hamiltonian

Gaussian term



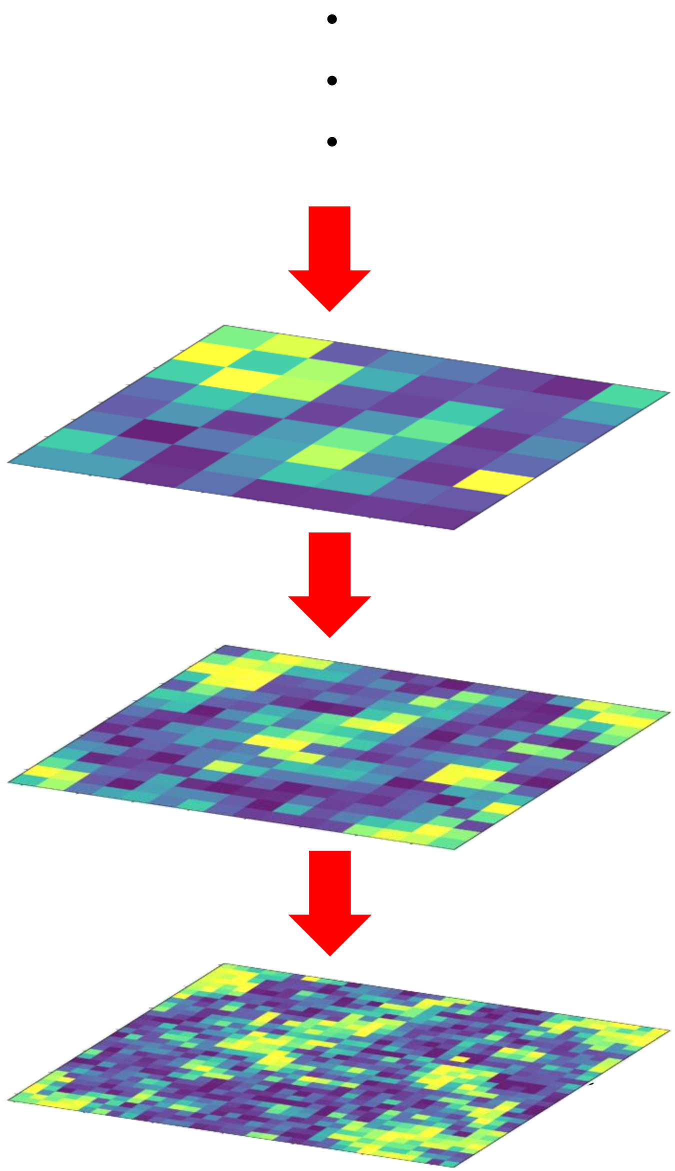
Non-linear potentials

Disordered phase | Critical point | Ordered phase

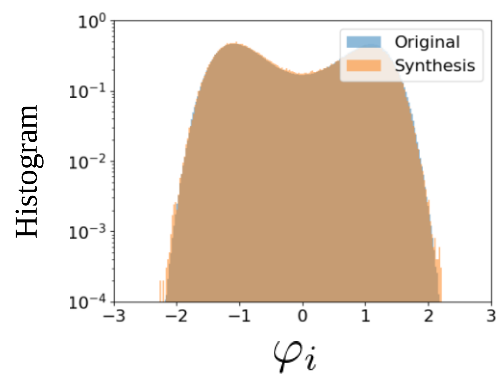


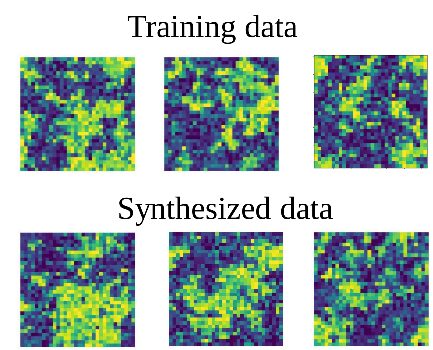
Self-similarity of the potential appears near the critical point

New fast Monte-Carlo sampling algorithm



Once we learn $p_K^{c}(\varphi_{\text{fast}}|\varphi_{\text{slow}})$ across scales,
we use them for a fast MC sampling method





$$\tau_{\text{MC}} \sim \xi^{z}$$

Standard algorithms: $z \simeq 2$

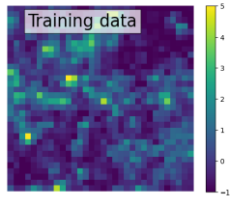
Cluster algorithms: $z \simeq 0.2 - 0.3$

Swendsen and Wang, PRL 1987

Wolff, Phys. Lett. B 1989

Our algorithm: $z = 0$

Application for Astrophysics



Weak Gravitational Lensing map

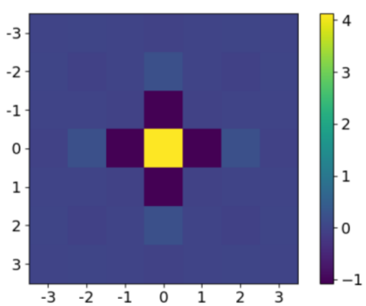
Manuel Zorrilla Matilla, Himan, Hsu, Gupta, and Petri, PRD 2016

Inherently non-equilibrium field where we do not know $p(\varphi)$ and $\mathcal{H}(\varphi)$ *a priori*

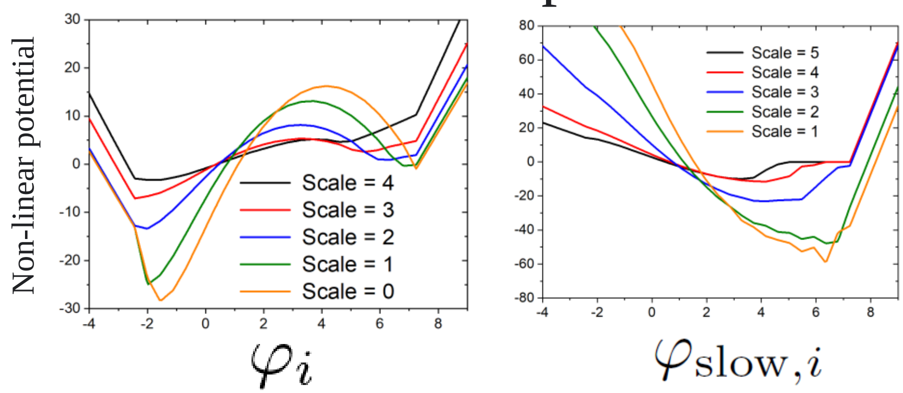
$\mathcal{H}(\varphi)$ is a sort of effective Hamiltonian to represent $p(\varphi)$ compactly

Construction of Hamiltonian

Gaussian term



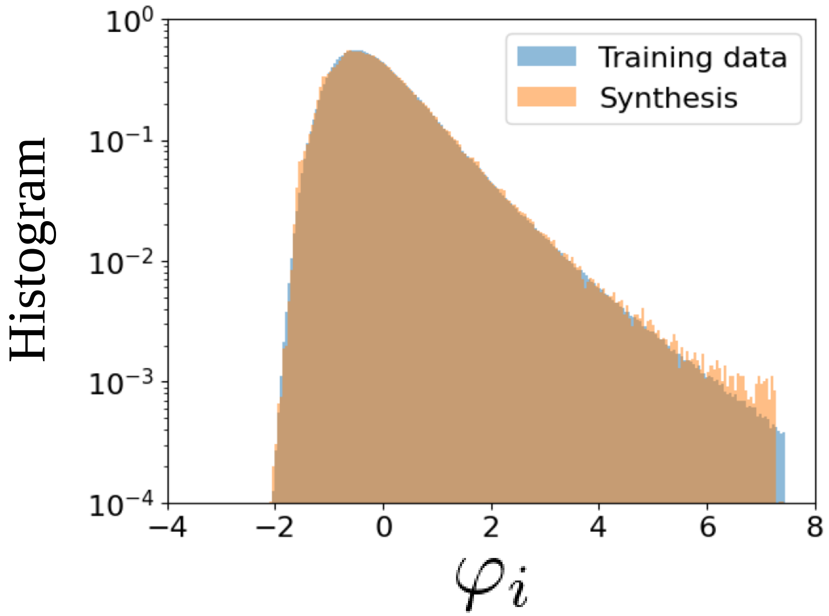
Non-linear potentials

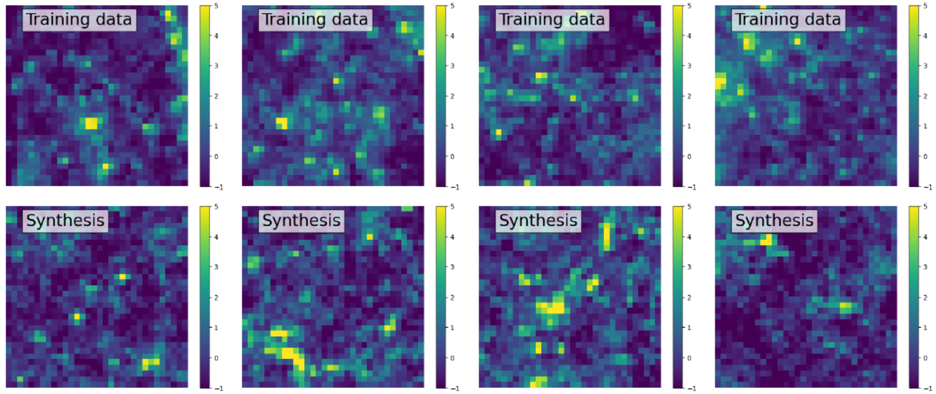


The emergence of non-linear potentials for φ_{slow} → Long-range interactions

Application for Astrophysics

Fast MC sampling from coarse to fine





Conclusions

- Learning microscopic Hamiltonian from data

(Generative model with high interpretability)

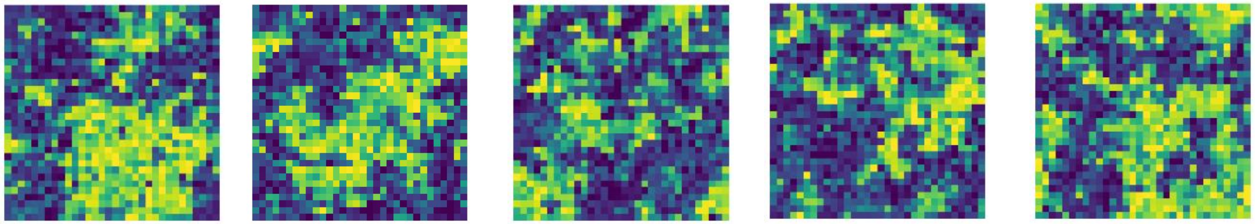
- Overcoming critical slowing down in training process

- Overcoming critical slowing down in generation process

(RG is the key to achieve these goals)

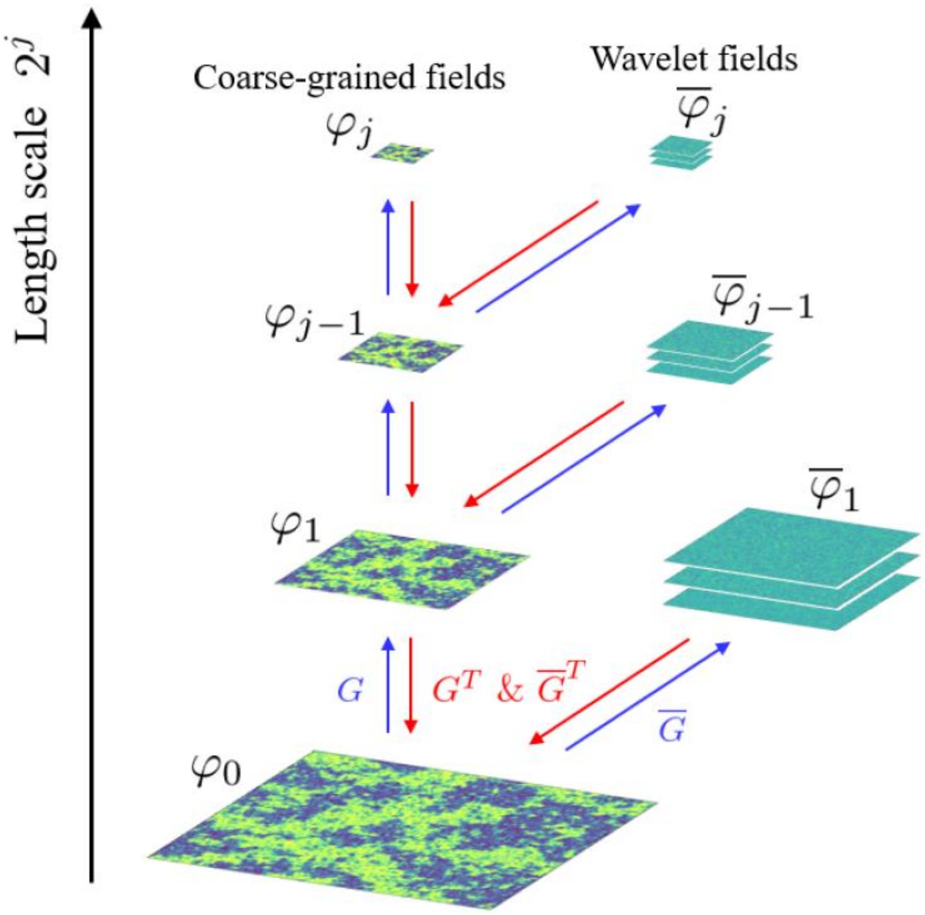
1) Prepare training data set

Field φ_0



. . .

2) Decompose fields



Orthogonal wavelet transform

Coarse-grained field

$$\varphi_j = \gamma_j^{-1} \, G \, \varphi_{j-1}$$

Wavelet field

$$\overline{\varphi}_j = \gamma_j^{-1} \, \overline{G} \, \varphi_{j-1}$$

Reconstruction

$$\varphi_{j-1} = \gamma_j G^T \varphi_j + \gamma_j \overline{G}^T \overline{\varphi}_j$$

3) Conditional probabilities

Chain rule

$$p_{j-1}(\varphi_{j-1}) = \overline{p}_j(\overline{\varphi}_j|\varphi_j)p_j(\varphi_j)$$

Applying the chain rule many times

$$\begin{aligned} p_0(\varphi_0) &= \overline{p}_1(\overline{\varphi}_1|\varphi_1)\,\overline{p}_2(\overline{\varphi}_2|\varphi_2)\,\overline{p}_3(\overline{\varphi}_3|\varphi_3)\cdots\overline{p}_J(\overline{\varphi}_J|\varphi_J)\,p_J(\varphi_J) \\ &= \left[\Pi_{j=1}^{J}\overline{p}_j(\overline{\varphi}_j|\varphi_j)\right]p_J(\varphi_J) \end{aligned}$$

Tasks: Estimation of $p_J(\varphi_J)$ and $\overline{p}_j(\overline{\varphi}_j|\varphi_j)$

4) Estimations

Estimation of $p_J(\varphi_J)$:

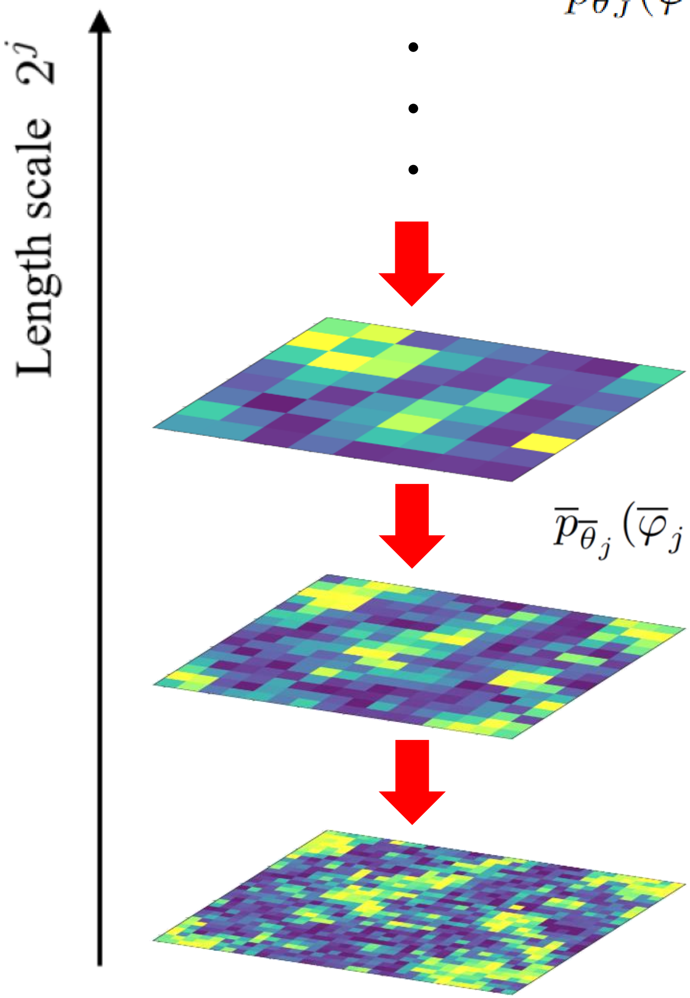
$$p_{\theta_J}(\varphi_J) = \frac{1}{Z_J} e^{-\theta_J^T U_J(\varphi_J)} \quad \text{by} \quad \min_{\theta_J} D_{\mathrm{KL}}(p_J || p_{\theta_J})$$

Estimation of $\overline{p}_j(\overline{\varphi}_j|\varphi_j)$:

$$\overline{p}_{\overline{\theta}_j}(\overline{\varphi}_j|\varphi_j) = \overline{Z}_j^{-1} \exp\left[-\overline{\theta}_j^T \overline{U}_j(\varphi_{j-1}) + \overline{F}_j(\varphi_j)\right] \quad \text{by} \quad \min_{\overline{\theta}_j} D_{\mathrm{KL}}(p_{j-1} || \overline{p}_{\overline{\theta}_j} p_j)$$

5) Sampling from coarse to fine

$$p_{\theta_J}(\varphi_J) = \frac{1}{Z_J} e^{-\theta_J^T U_J(\varphi_J)}$$



$$\overline{p}_{\overline{\theta}_j}(\overline{\varphi}_j|\varphi_j) = \overline{Z}_j^{-1} \exp\left[-\overline{\theta}_j^T \overline{U}_j(\varphi_{j-1}) + \overline{F}_j(\varphi_j)\right]$$

6) Estimation of the microscopic Hamiltonian

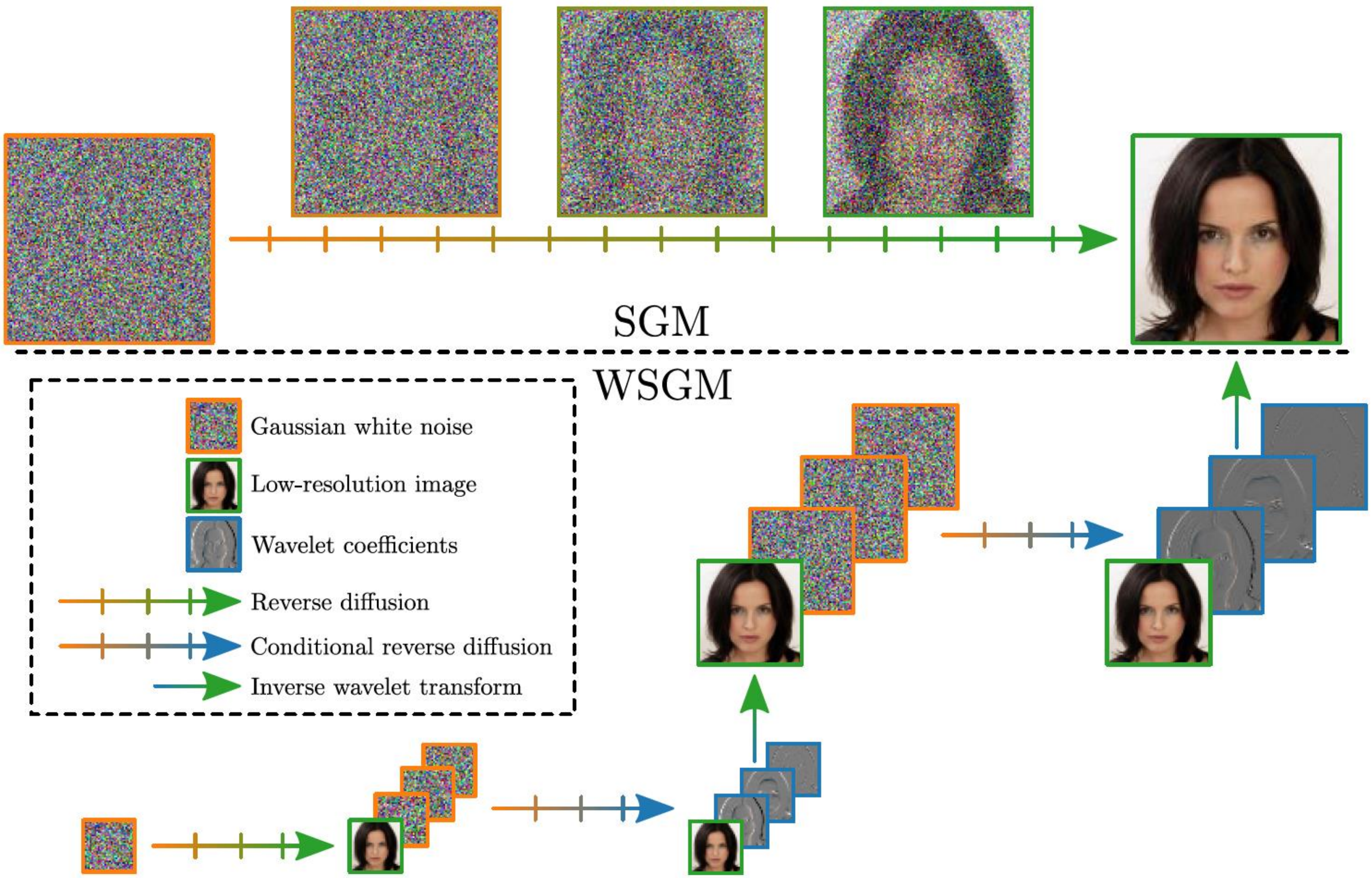
Estimation of $\overline{F}_j(\varphi_j)$:

$$\overline{F}_j(\varphi_j) \approx \widetilde{\theta}_j^T \widetilde{U}_j(\varphi_j) \quad \text{by} \quad \min_{\widetilde{\theta}_j} \left\langle \left(\overline{F}_j - \widetilde{\theta}_j^T \widetilde{U}_j \right)^2 \right\rangle_{p_j}$$

(Linear regression)

Estimation of $p_0(\varphi_0) = \frac{1}{Z_0} e^{-\mathcal{H}_0(\varphi_0)}$

$$\mathcal{H}_0 \approx \theta_J^T U_J(\varphi_J) + \sum_{j=1}^{J} \left(\overline{\theta}_j^T \overline{U}_j(\varphi_{j-1}) - \widetilde{\theta}_j^T \widetilde{U}_j(\varphi_j) \right) + c_0$$



Guth, Coste, De Bortoli, and Mallat, NeurIPS 2022