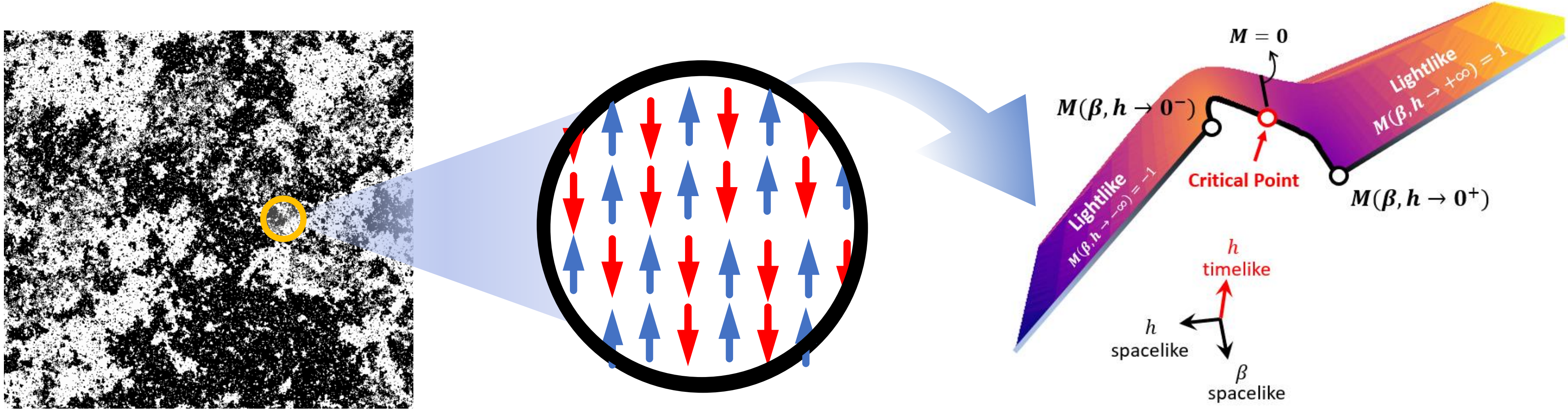
Visualizing statistical models in Minkowski space

isKL embedding

Han Kheng Teoh, Katherine N. Quinn, Jaron Kent-Dobias,
Colin B. Clement, Qingyang Xu and James P Sethna

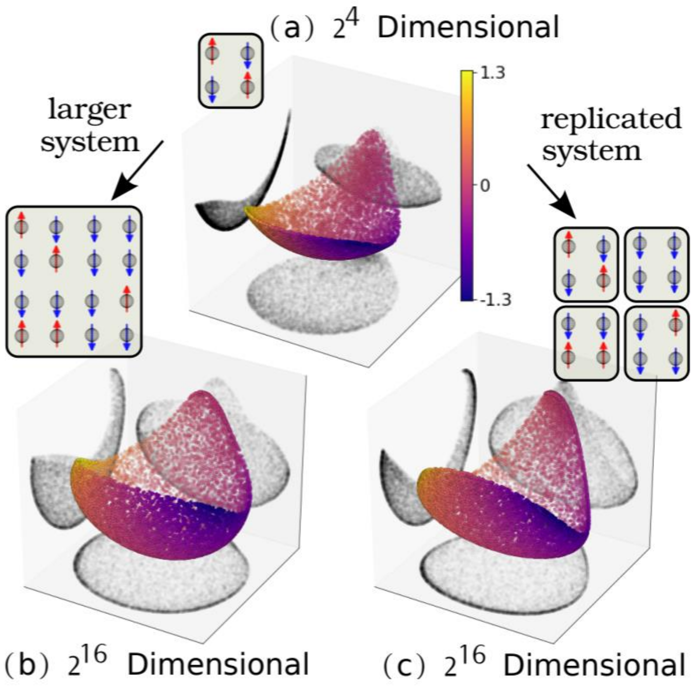


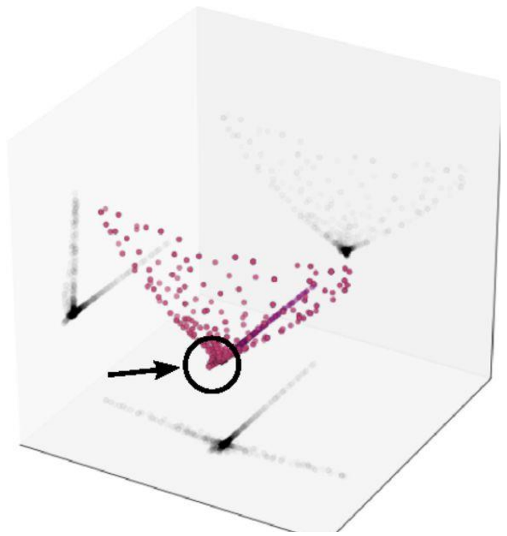
Basic Training 2024

Recap : Visualizing Ising model with Hellinger distance

Hellinger distance

$$D_{Hel}^2(P,Q) = \sum_i \left(\sqrt{P_i} - \sqrt{Q_i}\right)^2$$



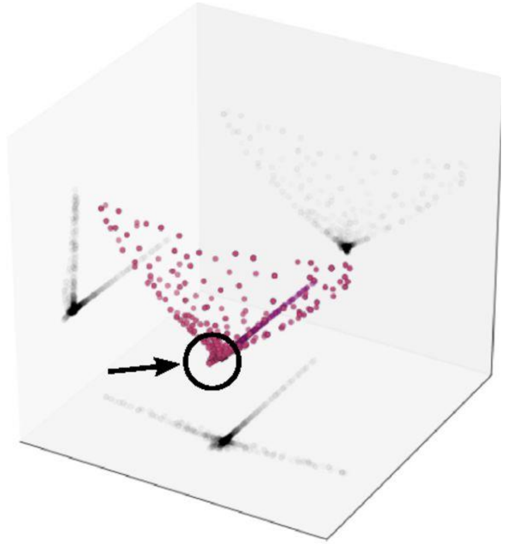


2^{128} Dimensional

Recap : Visualizing Ising model with inPCA

Hellinger distance

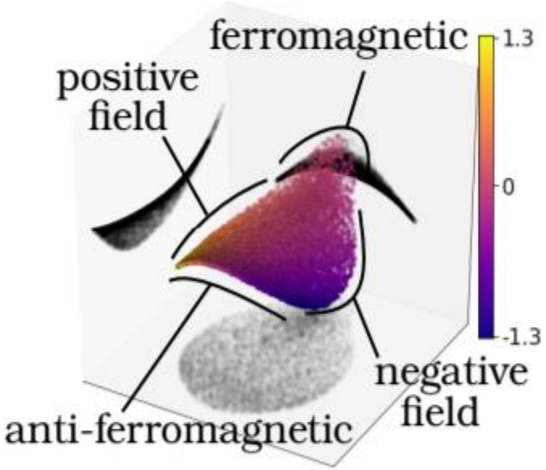
$$D^2_{Hel}(P,Q) = \sum_i \left(\sqrt{P_i} - \sqrt{Q_i}\right)^2$$



Replica trick →

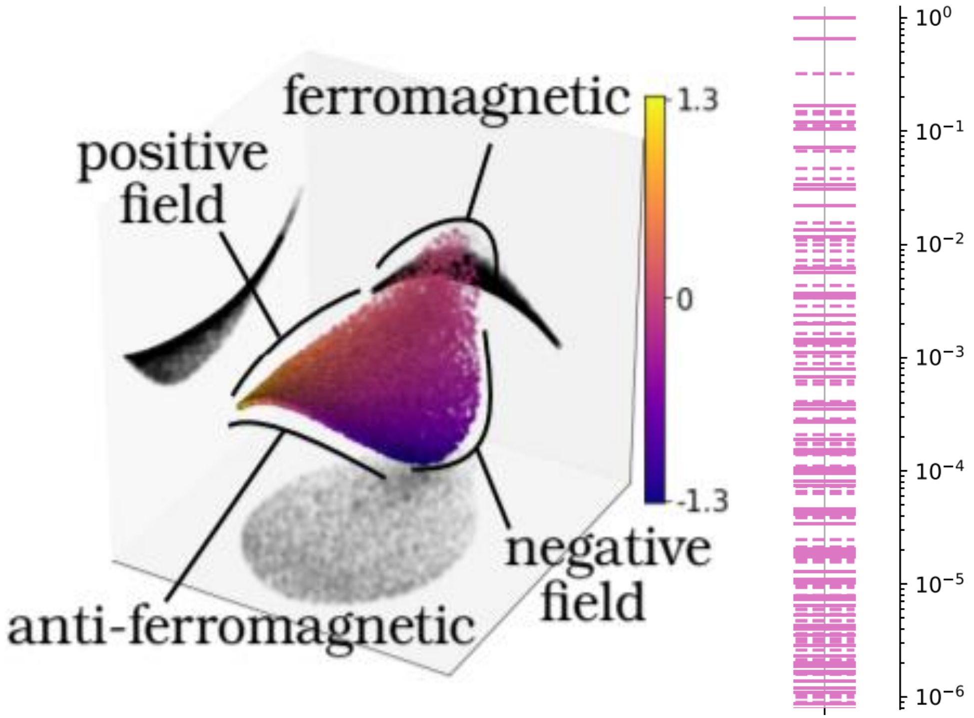
Bhattacharyya distance

$$D^2_{Bhat}(P,Q) = \sum_i \log \sqrt{P_i} \cdot \sqrt{Q_i}$$



InPCA of Ising model

Manifold widths



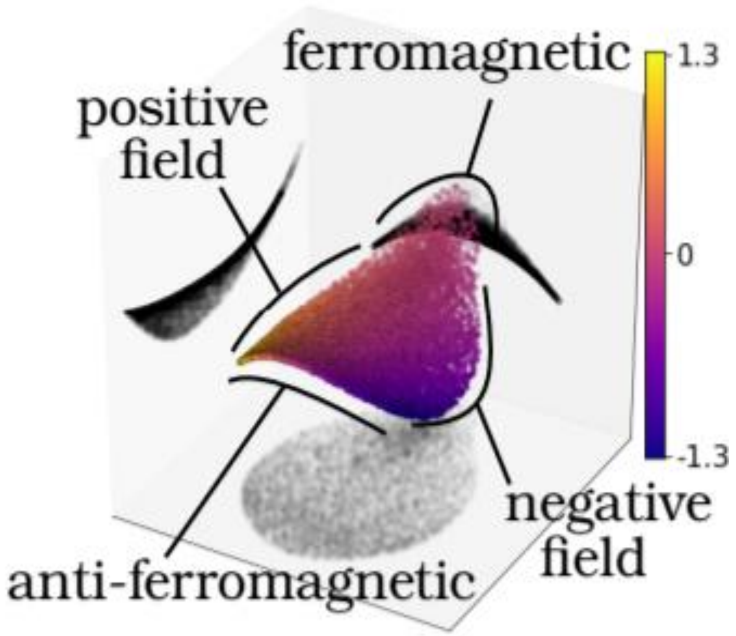
Large embedding dimensions

Meaning of each axis of projection?

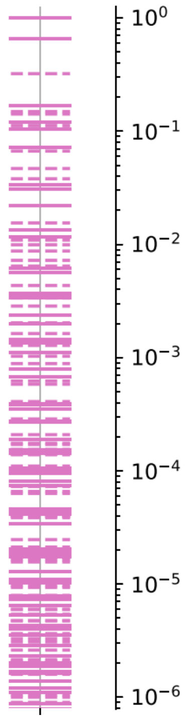
Computationally expensive for large system

Critical point has a singularity in the scalar curvature. Is there a cusp?

InPCA of Ising model



Manifold widths



Large embedding dimensions

Meaning of each axis of projection?

Computationally expensive for large system

Critical point has a singularity in the scalar curvature. Is there a cusp?

Can we do better?

Alternative distance measure : symmetrized KL divergence

Relative entropy (KL divergence)

Given distribution P and Q

$$\mathrm{KL}(P|Q) = \sum_i P_i \log P_i - P_i \log Q_i$$

Locally, corresponds to Fisher Information Metric

$$\mathrm{KL}(P(\boldsymbol{\theta})|P(\boldsymbol{\theta} + \delta\boldsymbol{\theta})) \approx \mathrm{FIM}$$

Symmetrize it

$$D_{SKL}^2(P, Q) = \sum_i (P_i - Q_i) \log\left(\frac{P_i}{Q_i}\right)$$

**One can check that symmetrized Kl divergence is intensive as well

Back to Ising 2D model

$$P(\boldsymbol{s}|\beta, h) = \frac{\exp\left(\beta \sum_{\langle i,j \rangle} s_i s_j + h \sum_i s_i\right)}{Z(\beta, h)}$$

Assignment 4.2

$$D_{SKL}^2\left(P_{\boldsymbol{\theta_1}}, P_{\boldsymbol{\theta_2}}\right) = \sum_s \left(P_{\boldsymbol{\theta_1}}(\boldsymbol{s}) - P_{\boldsymbol{\theta_2}}(\boldsymbol{s})\right) \log \frac{P_{\boldsymbol{\theta_1}}(\boldsymbol{s})}{P_{\boldsymbol{\theta_2}}(\boldsymbol{s})}$$

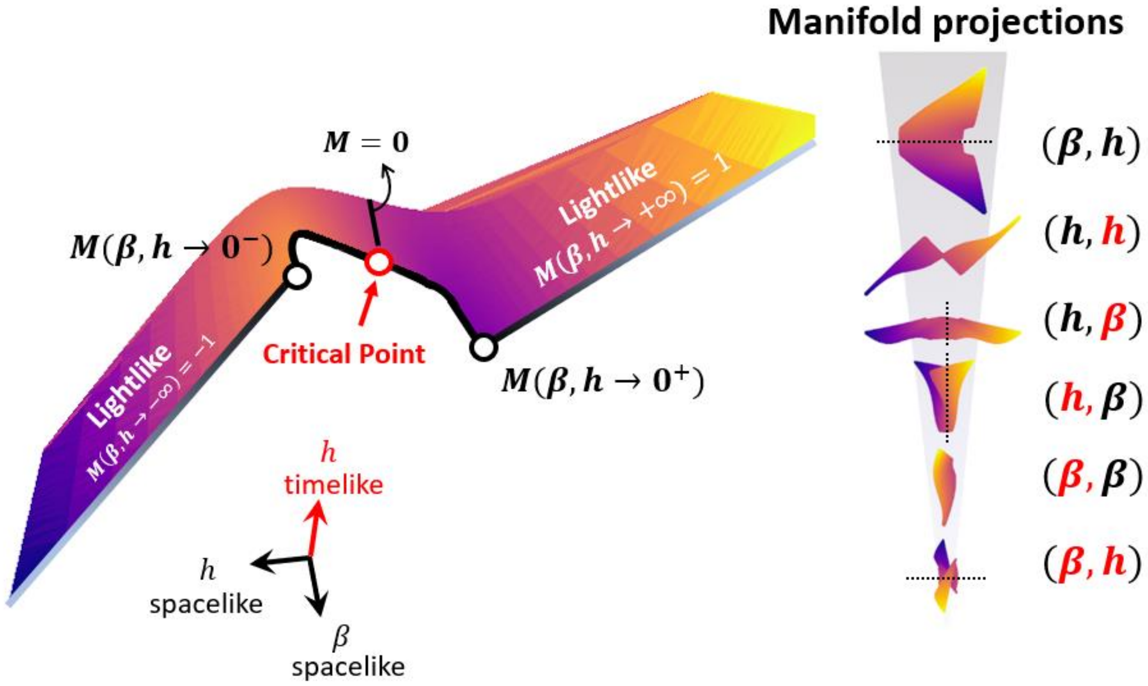
$$= -(\beta_1 - \beta_2)(e_1 - e_2) + (h_1 - h_2)(m_1 - m_2)$$

$$\vdots$$

$$= \left(S_{\beta_2} - S_{\beta_1}\right)^2 - \left(T_{\beta_2} - T_{\beta_1}\right)^2 + \left(S_{h_2} - S_{h_1}\right)^2 - \left(T_{h_2} - T_{h_1}\right)^2$$

One can show that $S_{(\cdot)}$ and $T_{(\cdot)}$ are the axis of projections in Multidimensional Scaling (MDS)

Ising 2D model



2+2 dimension

Spacelike

$$S_\beta(\boldsymbol{\theta}) = \frac{1}{2}\left(\lambda_\beta\, \beta + \frac{1}{\lambda_\beta}\left(\frac{Mh}{\beta} - E\right)\right)$$

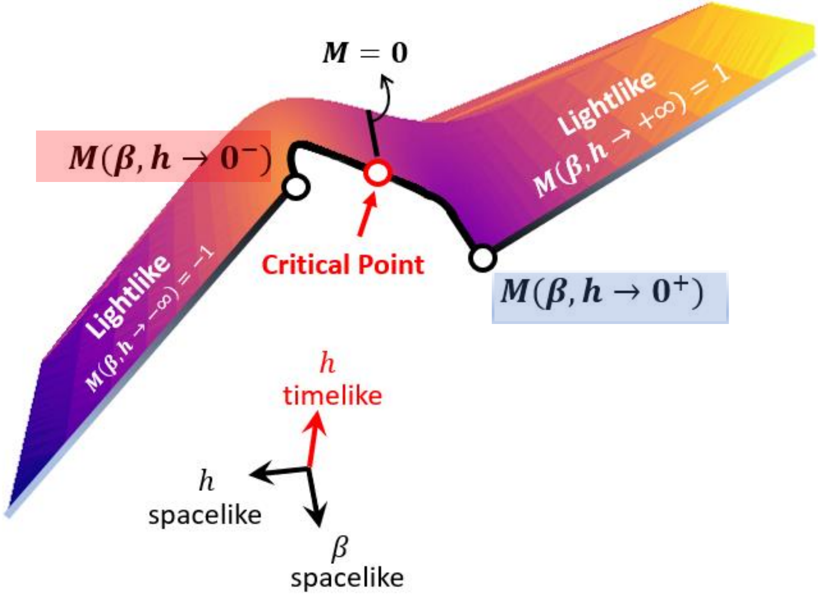
$$S_h(\boldsymbol{\theta}) = \frac{1}{2}\left(\lambda_h\, h + \frac{1}{\lambda_h} M\right)$$

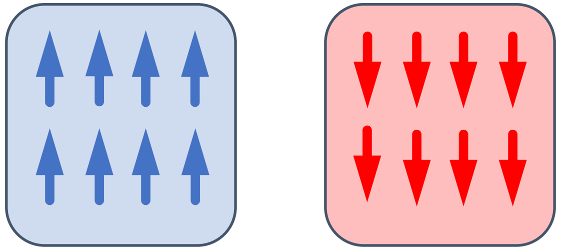
Timelike

$$T_\beta(\boldsymbol{\theta}) = \frac{1}{2}\left(\lambda_\beta\, \beta - \frac{1}{\lambda_\beta}\left(\frac{Mh}{\beta} - E\right)\right)$$

$$T_h(\boldsymbol{\theta}) = \frac{1}{2}\left(\lambda_h\, h - \frac{1}{\lambda_h} M\right)$$

Ising 2D model

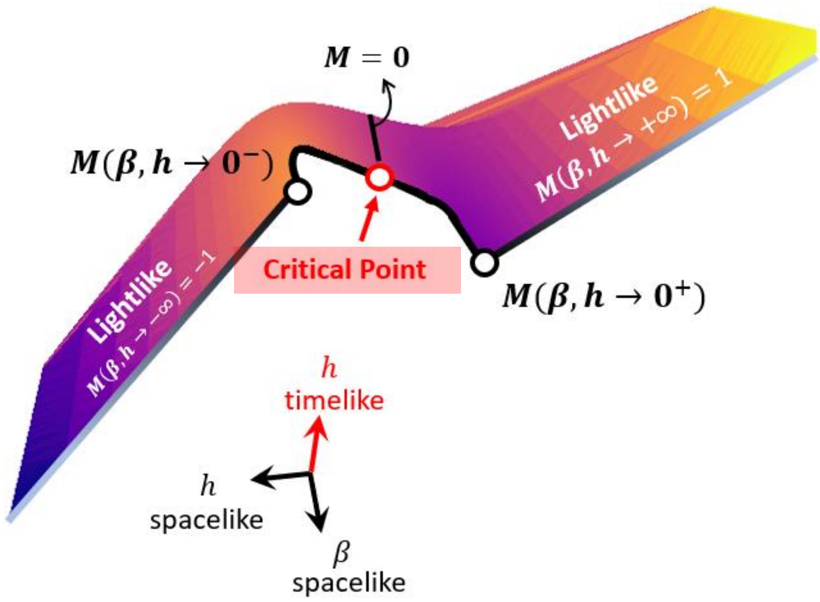




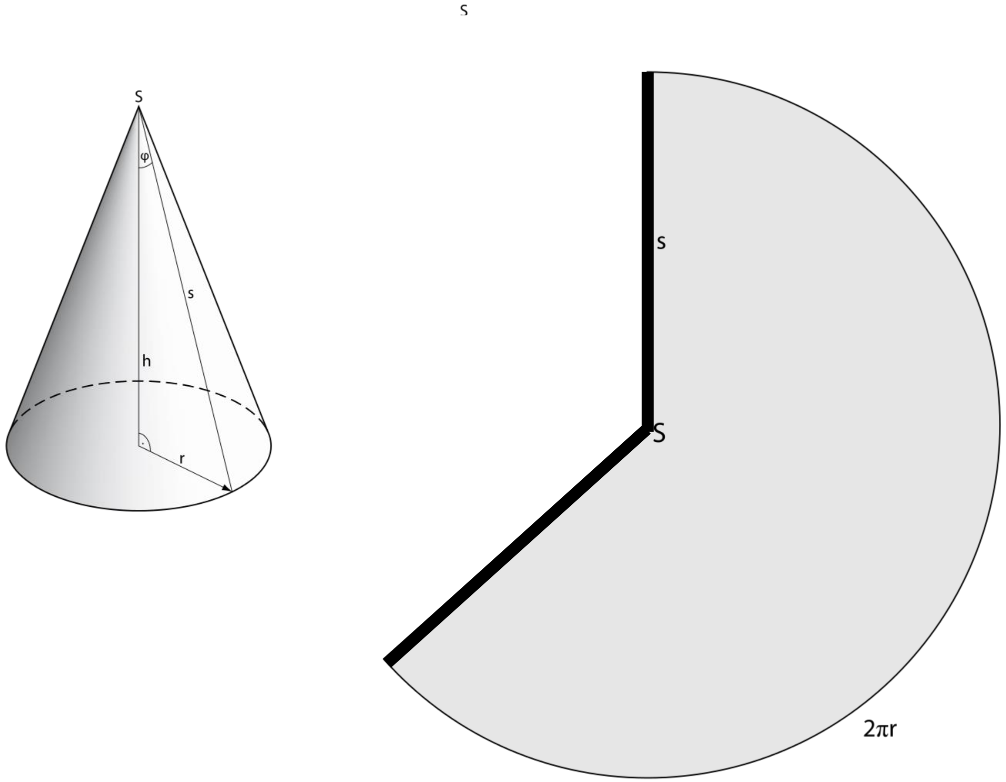
Fisher distance = 0

Lightcone – separate physically distinct systems that are distributionally similar

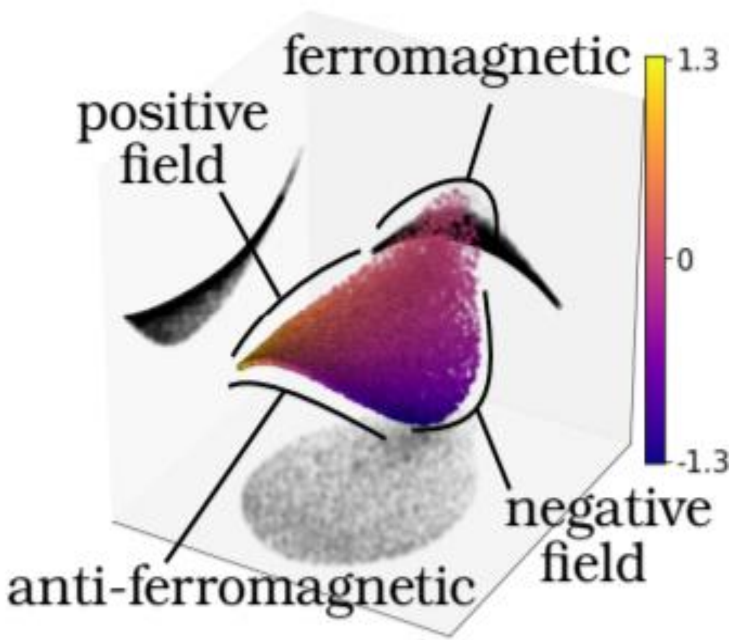
Ising 2D model



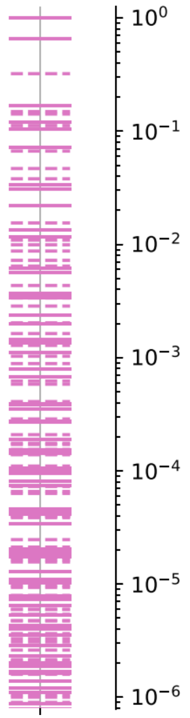
Curvature at critical point diverges , $R \sim -\frac{1}{|t|^2 \log(t^2)}$



Issues raised previously – Fixed!



Manifold widths



Large embedding dimensions

Meaning of each axis of projection?

Computationally expensive for large system

Critical point has a singularity in the scalar curvature. Is there a cusp?

Exponential Family

$$p(x|\theta) = h(x)\exp(\eta(\theta)\Phi(x) - A(\theta))$$

Base Measure

Natural Parameter

Sufficient Statistics

Log partition

Coin toss

$$P(x|p) = p^x(1-p)^{1-x}$$

$$h(x) = 1$$

$$\eta(p) = \log\frac{p}{1-p}$$

$$\Phi(x) = x$$

$$A(p) = -\ln(1-p)$$

Ideal Gas

$$p(\mathbb{P}, \mathbb{Q}, V | P, \beta) = Z^{-1}(P, \beta)\exp\left(-\beta\mathbb{P}^2/2m - \beta PV\right)$$

$$h(x) = 1$$

$$\boldsymbol{\eta}(\boldsymbol{\theta}) = (-\beta, -\beta P)$$

$$\boldsymbol{\Phi}(\boldsymbol{x}) = \left(\frac{\mathbb{P}^2}{2m}, V\right)$$

$$A(\boldsymbol{\theta}) = -\ln\big(Z(P, \beta)\big)$$

Nonlinear least square

$$P(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_i \frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left(-\frac{(f_i(\boldsymbol{\theta})-x_i)^2}{2\sigma_i^2}\right)$$

$$h(x) = -\sum_i x_i^2/\sigma_i^2$$

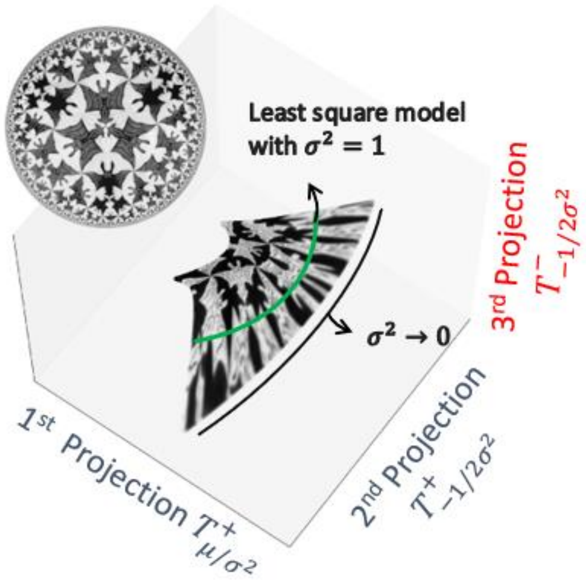
$$\eta_i(\boldsymbol{\theta}) = f_i(\boldsymbol{\theta})/\sigma_i$$

$$\Phi(x_i) = x_i/\sigma_i$$

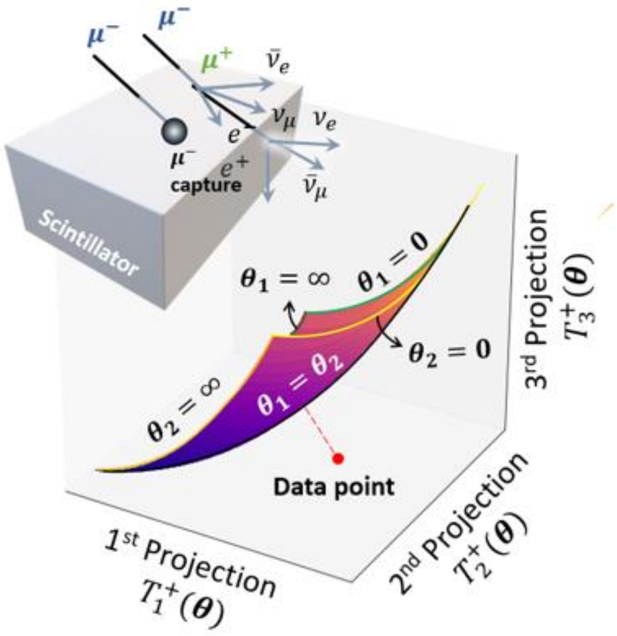
$$A(\boldsymbol{\theta}) = \sum_i f_i^2(\boldsymbol{\theta})/2\sigma_i^2 + \ln(2\pi\sigma_i^2)/2$$

Exponential Family Examples

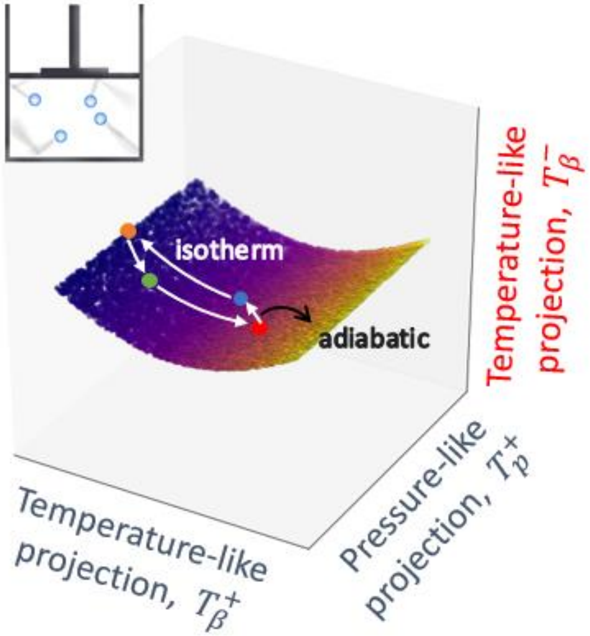
Gaussian Fits



Muon lifetime



Ideal Gas



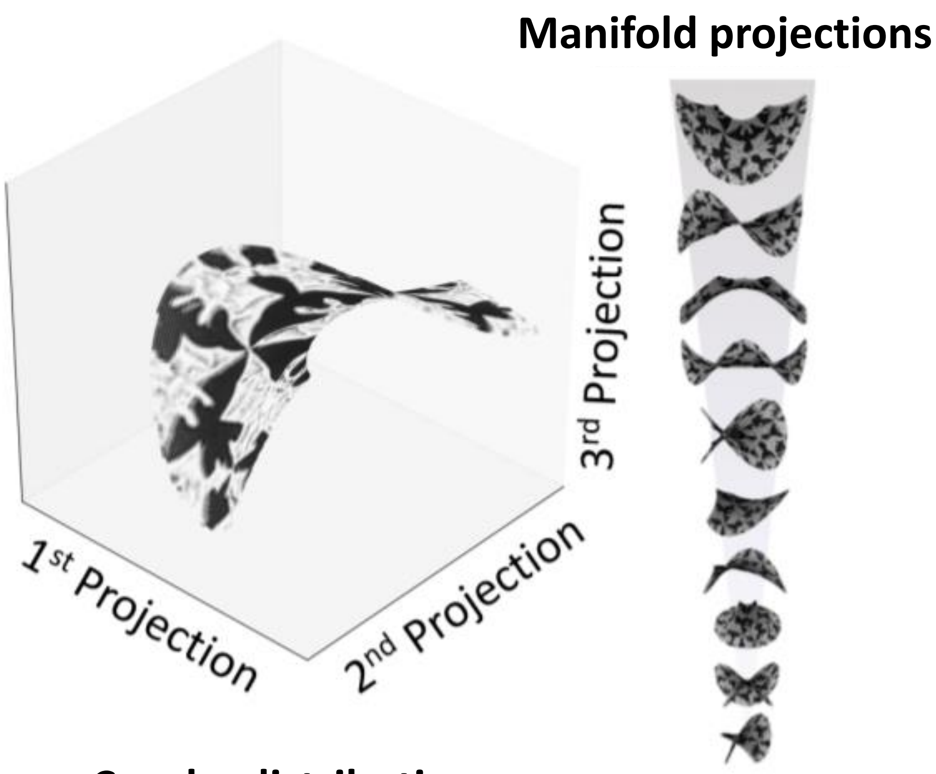
For any n parameter statistical model that fits into the exponential family

$$p(x|\theta) = h(x)\exp(\eta(\theta)\Phi(x) - \log(A(\theta)))$$

isKL embedding gives $(n + n)$ embedding dimension

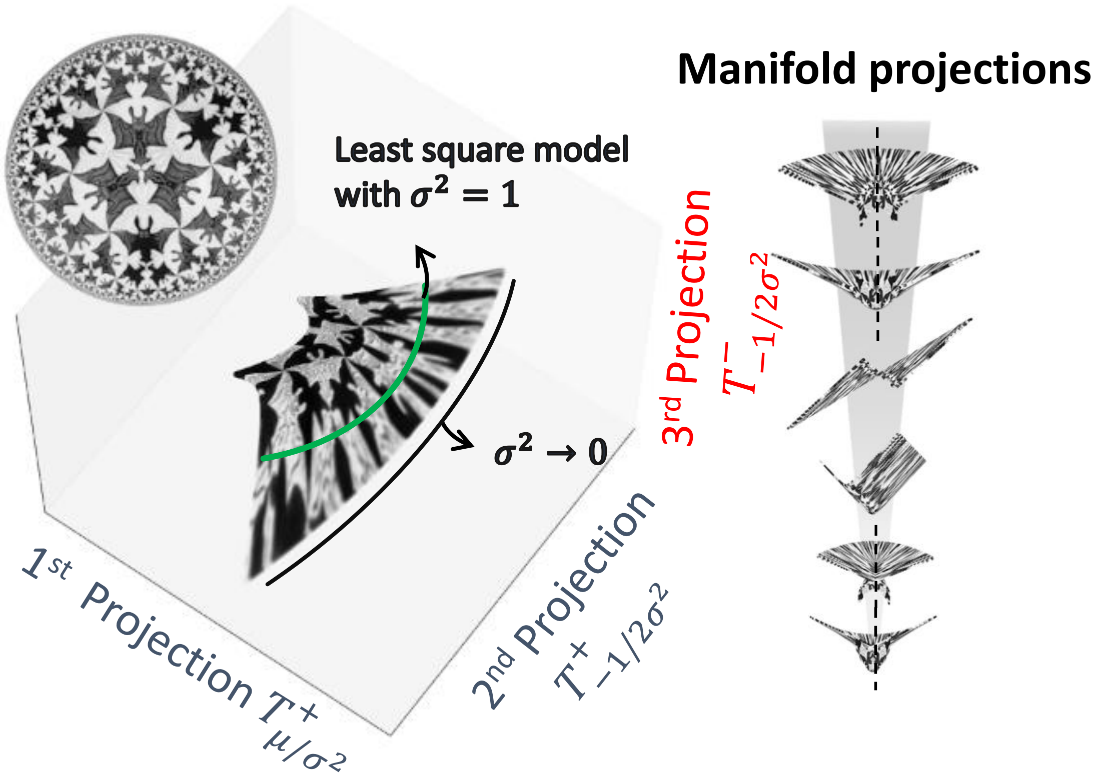
$$S_i(\boldsymbol{\theta}) = \frac{1}{2}\left(\lambda\,\eta_i(\boldsymbol{\theta}) + \frac{1}{\lambda}\langle\Phi_\mathrm{i}(\boldsymbol{x})\rangle_{\boldsymbol{\theta}}\right) \qquad T_i(\boldsymbol{\theta}) = \frac{1}{2}\left(\lambda\,\eta_i(\boldsymbol{\theta}) - \frac{1}{\lambda}\langle\Phi_\mathrm{i}(\boldsymbol{x})\rangle_{\boldsymbol{\theta}}\right)$$

Non exponential family : Cauchy distribution



Cauchy distribution

∞ dimensional embedding



Gaussian distribution

(2+2) embedding