Exploring Lattice QCD Data with Symbolic Regression

Brannon Semp

Generalized Parton Distributions

- Describe the distribution of quarks and gluons in nucleons
- Their forward limits are PDFs
- Their moments are angular momentum and elastic form factors
- VERY difficult to determine appropriate functional form

$$\frac{1}{2P^+}\overline{u}(p',\Lambda')[\gamma^+ H^q(x,\xi,t) + \frac{i\sigma^{+i}\Delta_i}{2M}E^q(x,\xi,t)]u(p,\Lambda) = \frac{1}{2}\int\frac{\mathrm{d}z^-}{2\pi}\mathrm{e}^{ixP^+z^-} < p',\Lambda'|\overline{q}(\frac{-z}{2})\gamma^+\gamma_5 q(\frac{z}{2})|p,\Lambda > \Big|_{z^+=0,\mathbf{z}=0}$$

$$\frac{1}{2P^+}\overline{u}(p',\Lambda')[\gamma^+\gamma_5 \tilde{H}^q(x,\xi,t) + \frac{\Delta^+\gamma_5}{2M}\tilde{E}^q(x,\xi,t)]u(p,\Lambda) = \frac{1}{2}\int\frac{\mathrm{d}z^-}{2\pi}\mathrm{e}^{ixP^+z^-} < p',\Lambda'|\overline{q}(\frac{-z}{2})\gamma^+ q(\frac{z}{2})|p,\Lambda > \Big|_{z^+=0,\mathbf{z}=0}$$

Symbolic Regression

- Fits both a functional form and it's parameters
- Allows us to identify candidates for GPD functions using numerical LQCD data

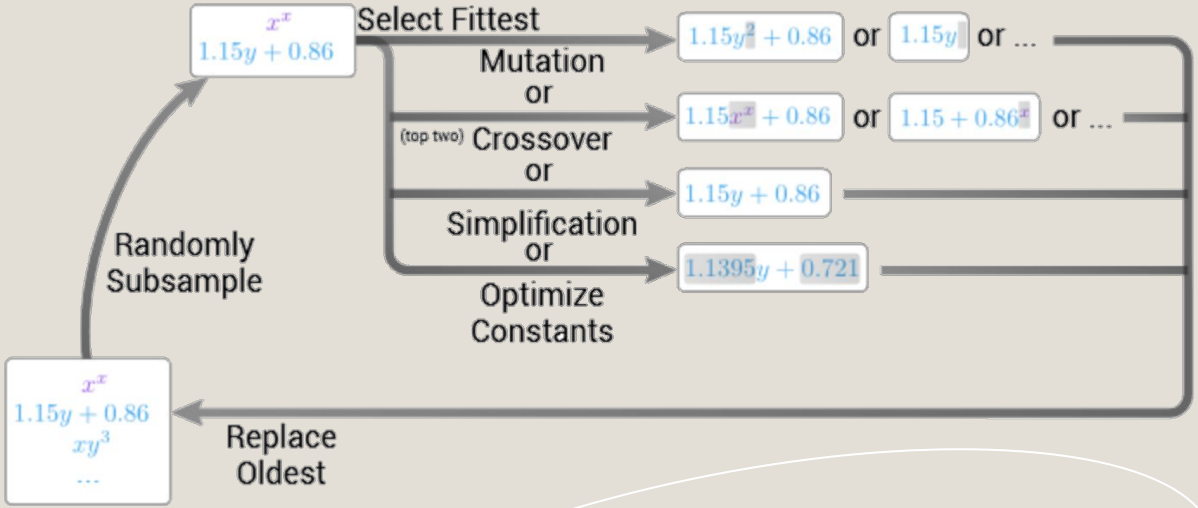


Figure from Cranmer, arXiv:2305.01582

Extraction

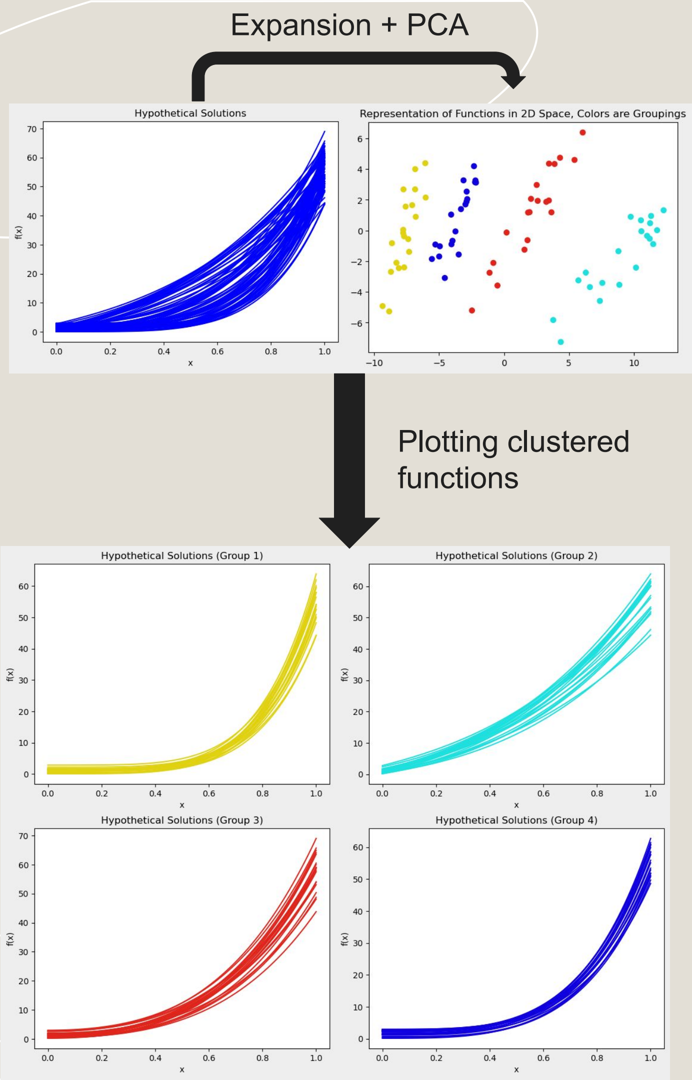
- 1000 fits of GPD H were performed on LQCD data as described in Dotson et. al. (arxiv:2504.13289) using PySR
- We analyze the fits as follows:
  - Expand each function to 20th order in a Legendre basis

$$f(x) = \sum_{n=0}^{\infty} c_n L_n(2x-1)$$

$$c_n = \frac{2n+1}{2} \int_0^1 f(x) L_n(2x-1)\ \mathrm{d}x$$

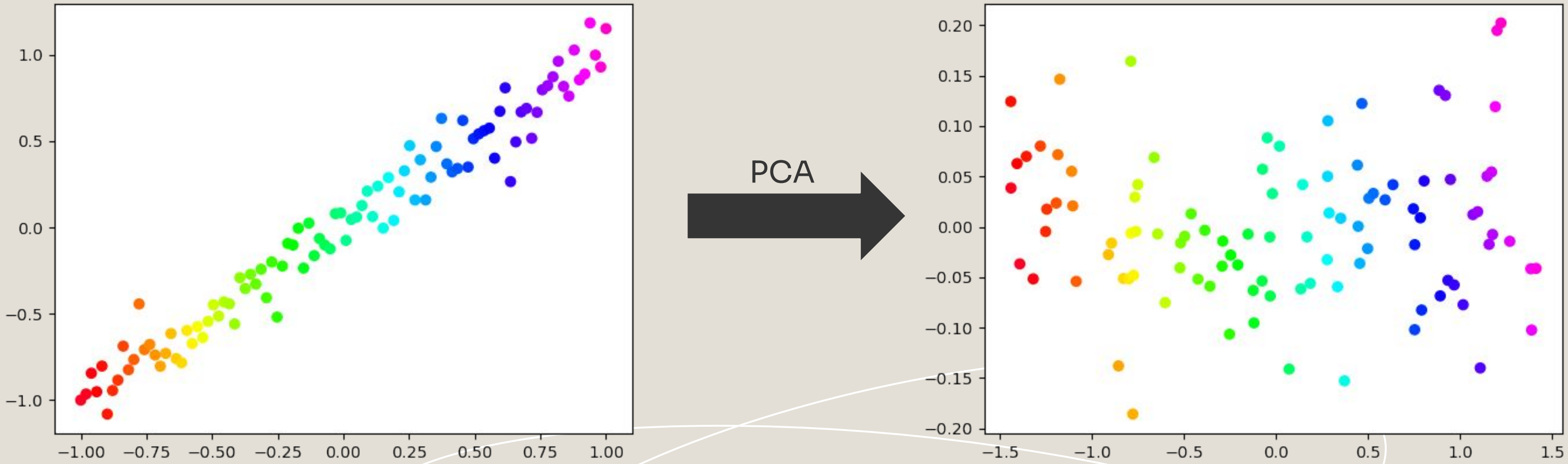
  - Principal component analysis is performed to further reduce dimensionality
  - Fits are clustered in PCA space using HDBSCAN algorithm

$$\cos(x) + i\sin(x) = \exp(ix)$$

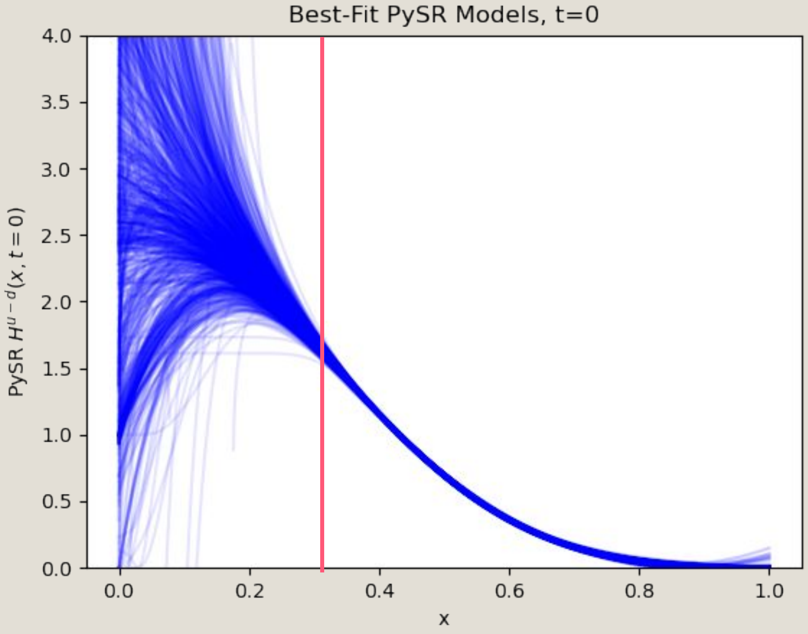


Principal Component Analysis

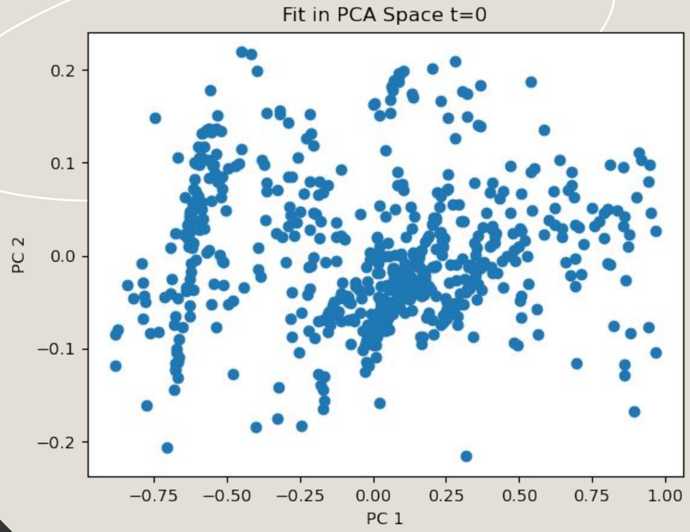
A statistical technique that reduces the dimensionality of a data set by rotating it such that most of the variance lies along the first axis (the principal component), second most variance along the second axis, and so on. Useful for identifying correlations between many different categories.



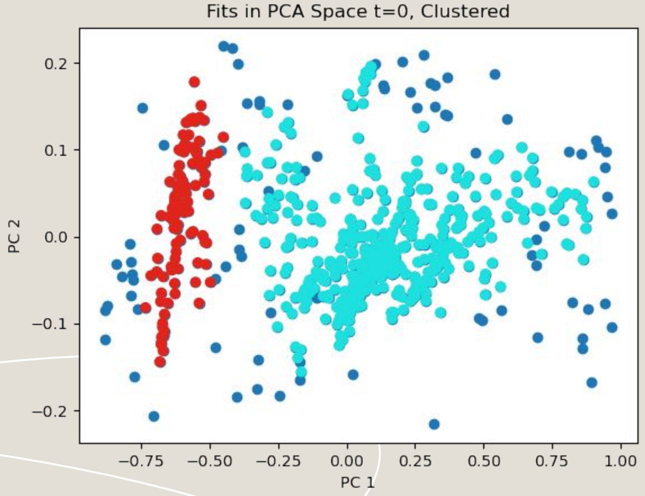
Analysis



Best-Fit PySR Models, t=0
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
PySR $H^{u-d}(x, t=0)$
0.0
0.2
0.4
0.6
0.8
1.0
x

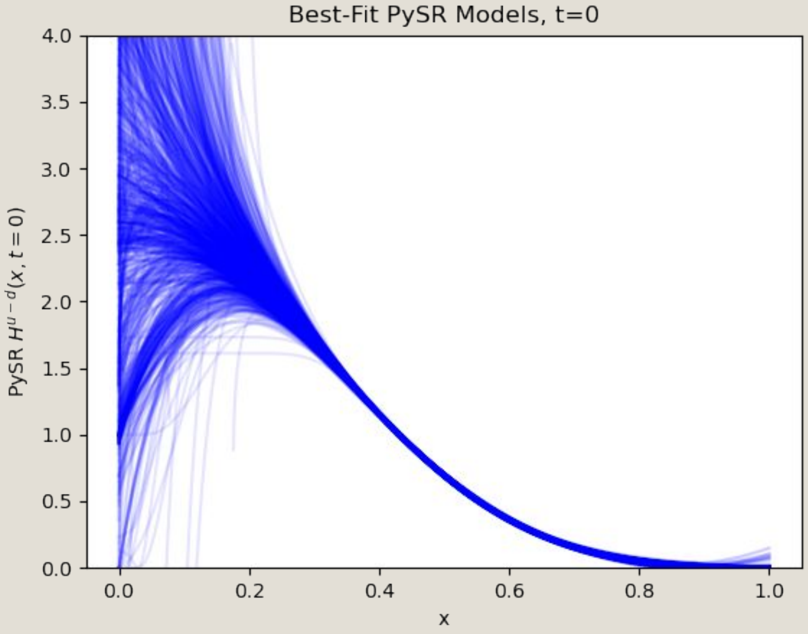


Fit in PCA Space t=0
0.2
0.1
0.0
−0.1
−0.2
PC 2
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
PC 1

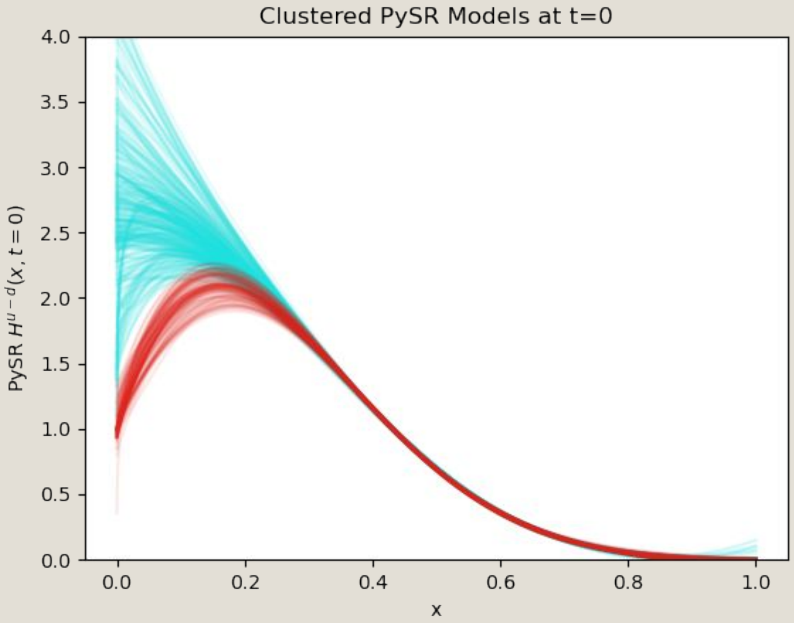


Fits in PCA Space t=0, Clustered
0.2
0.1
0.0
−0.1
−0.2
PC 2
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
PC 1

Analysis



Best-Fit PySR Models, t=0
PySR $H^{u-d}(x, t=0)$
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0
x



Clustered PySR Models at t=0
PySR $H^{u-d}(x, t=0)$
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0
x

Summary

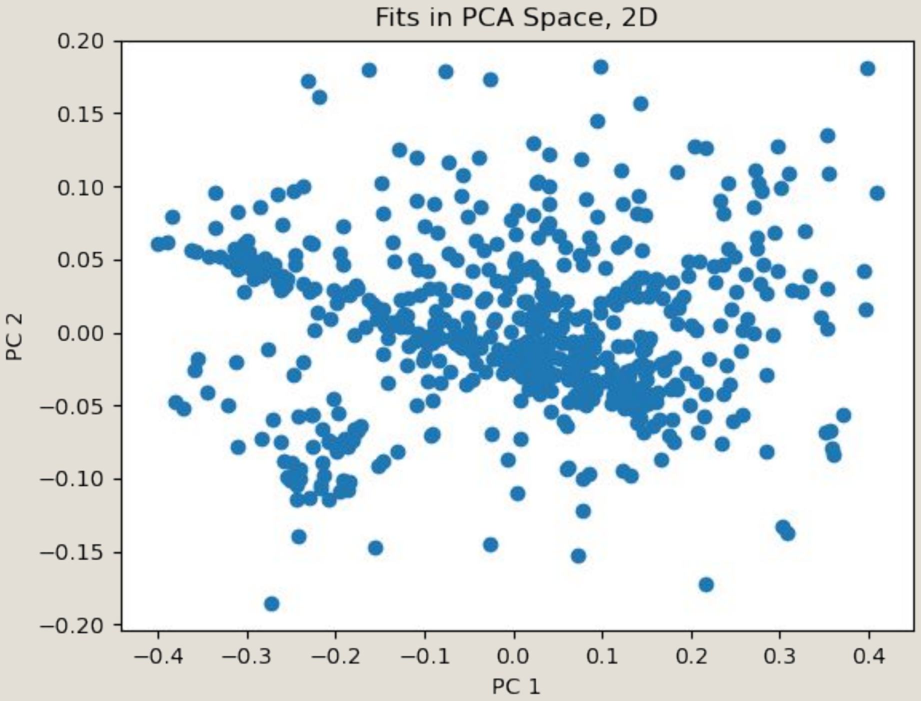
- PCA allows for easier visualization of clusters
- Fits appear to fall into one of two different clusters
- Physics constraints on GPDs may translate into simple cuts in PCA space
    - E.g. the relation between GPD H and proton radius

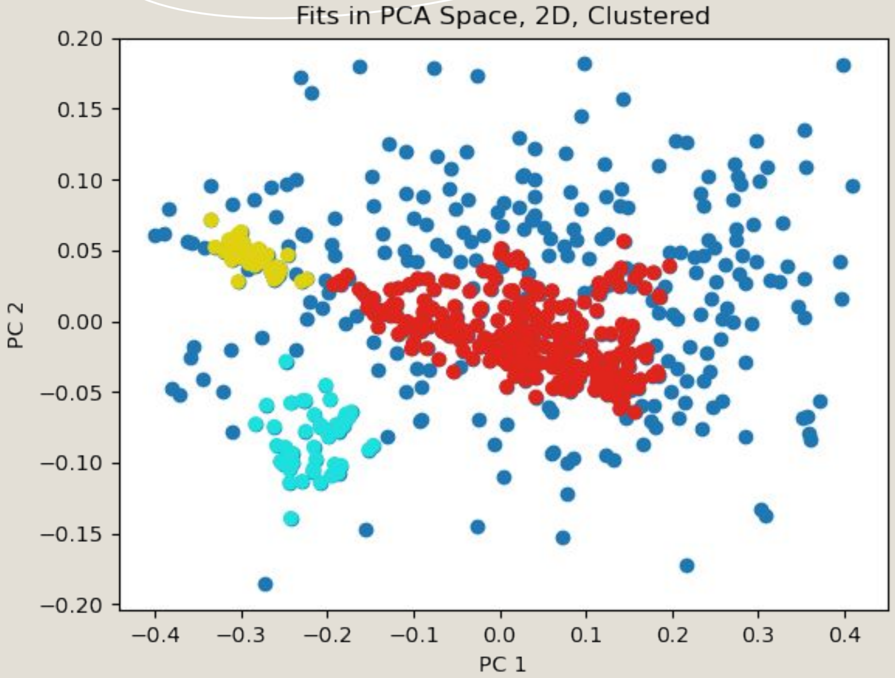
2D Analysis (Highly Preliminary)

- Expand as a function of both x and t

$$f(x,t) = \sum_{i=0}^{\infty} \sum_{n=0}^{i} c_{i,j} L_i(2x-1) L_j(2t-1)$$

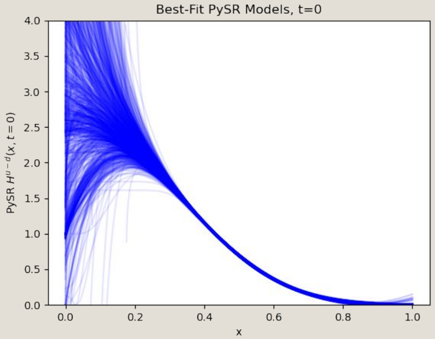
- Much higher complexity, could contribute to understanding of overall behavior of GPD
- Analysis is otherwise identical to the fixed t case

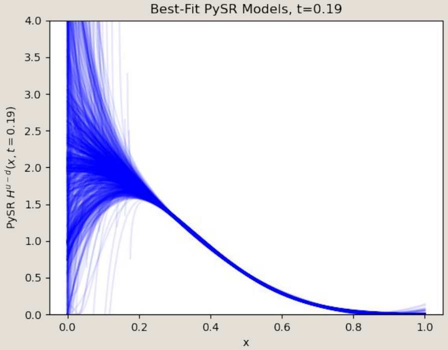

Fits in PCA Space, 2D
PC 2
PC 1
0.20
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
−0.20
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4

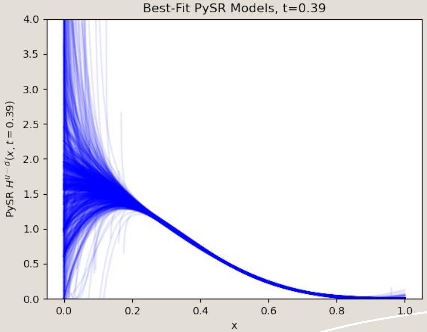

Fits in PCA Space, 2D, Clustered
PC 2
PC 1
0.20
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
−0.20
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4

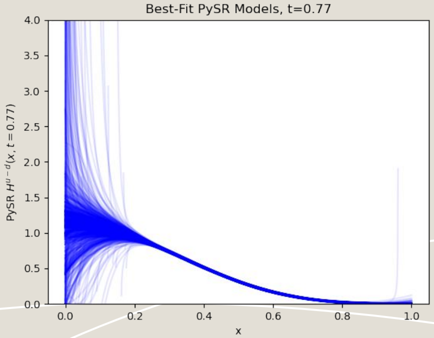
How does 2D compare to 1D?

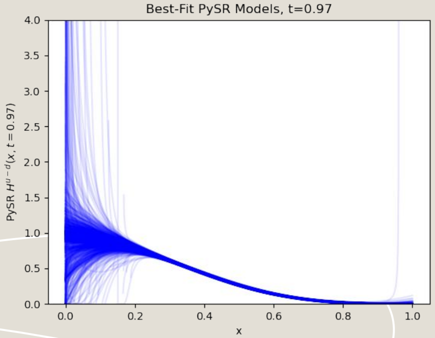
- Compare the 2D clustering to 1D clustering at different t

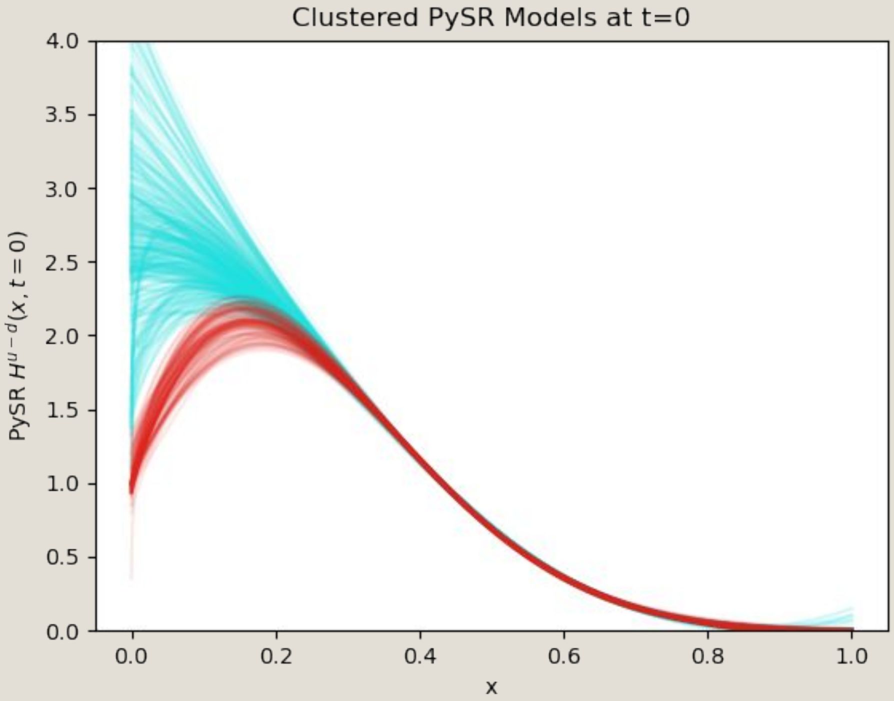


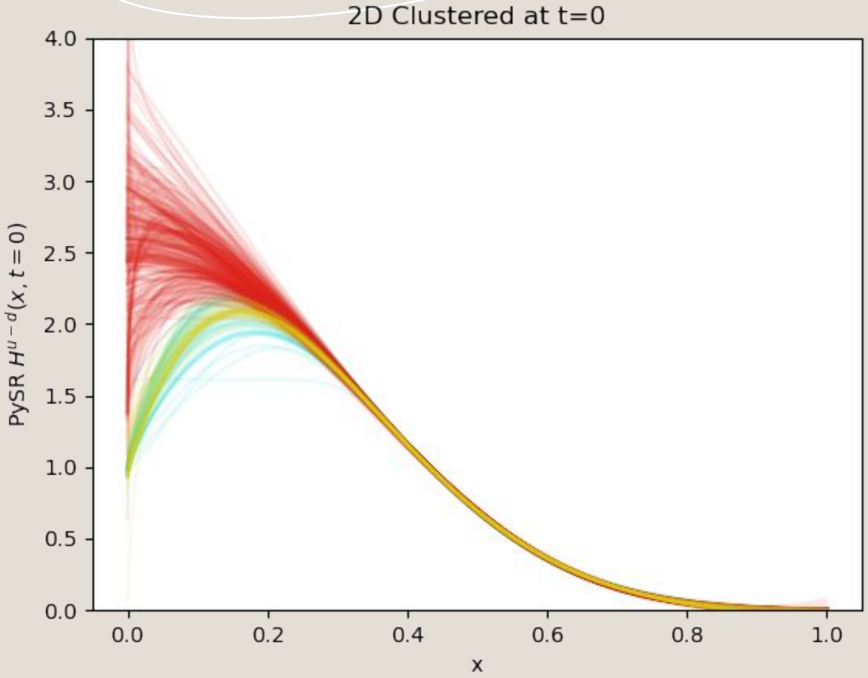


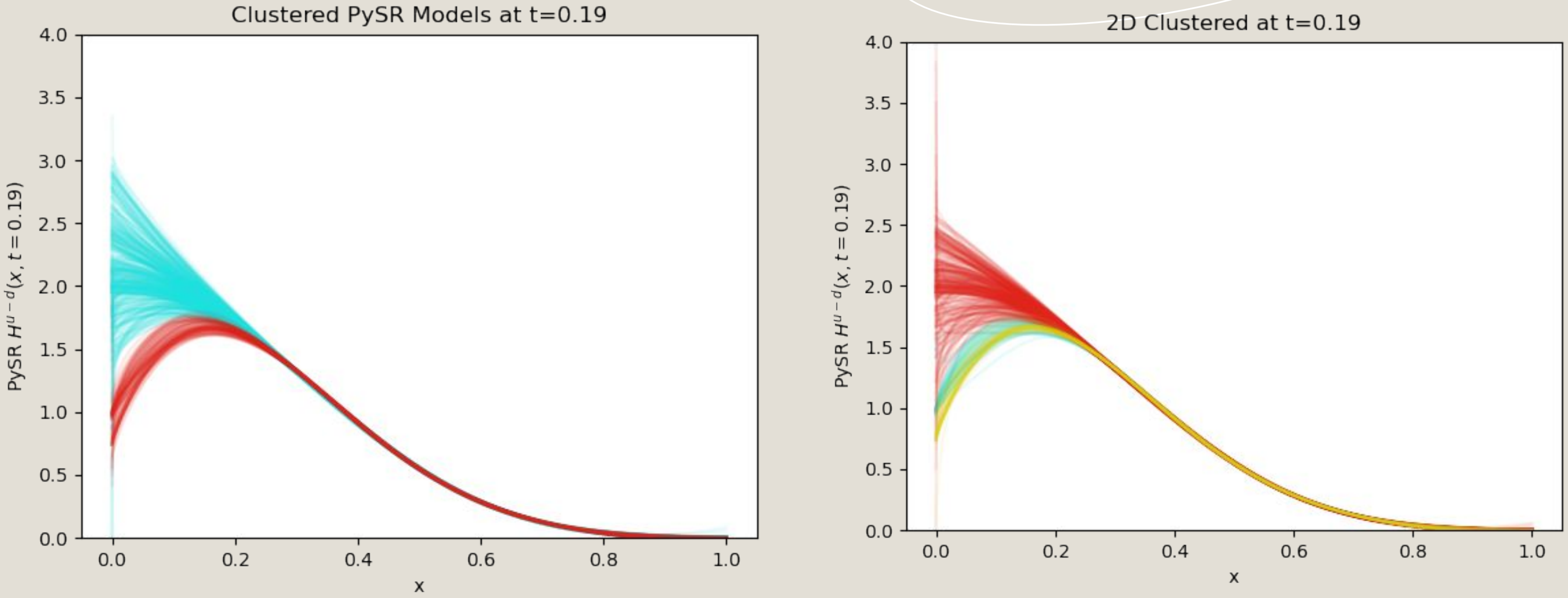


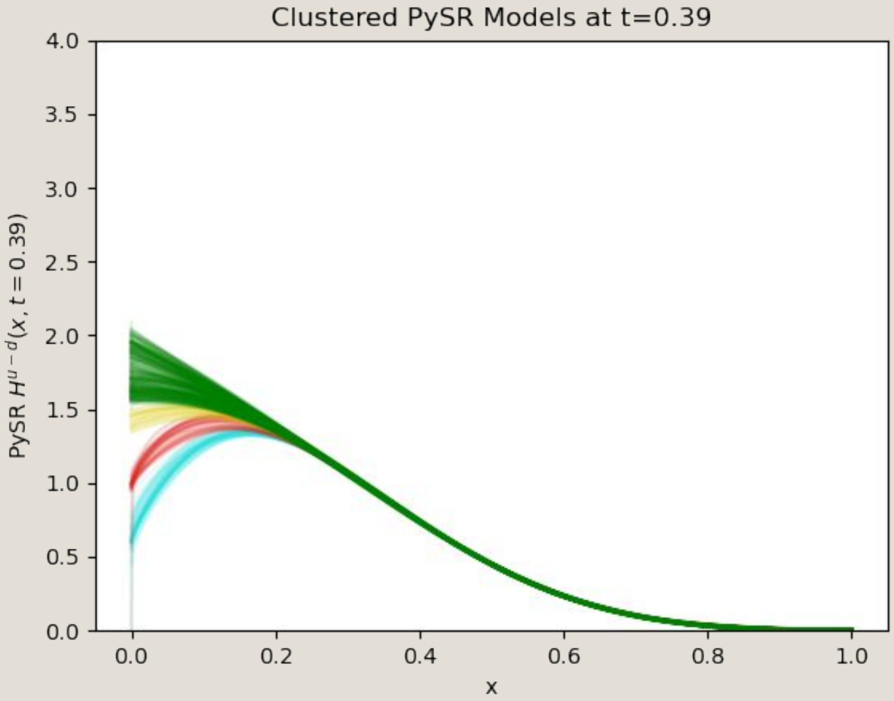




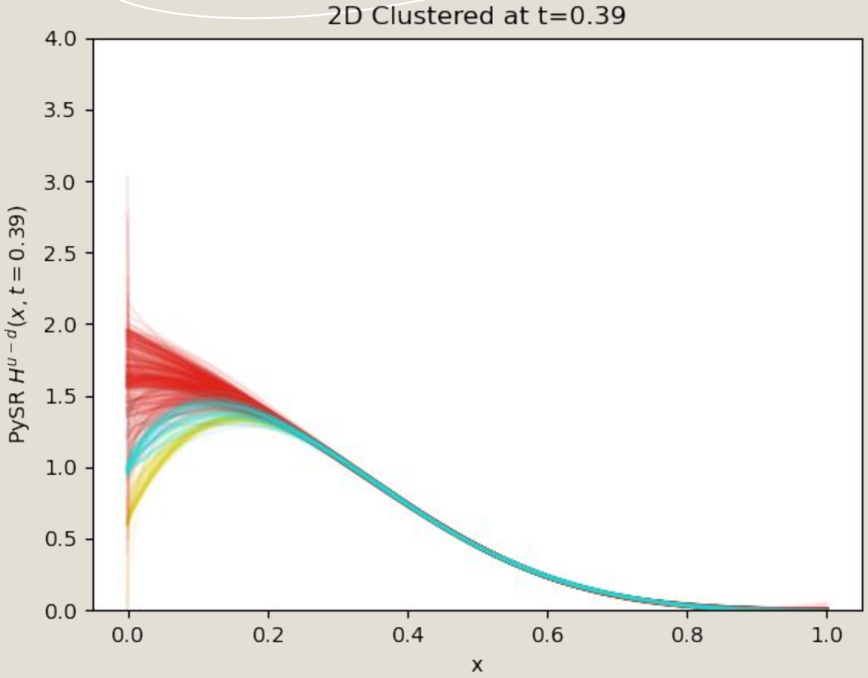

Clustered PySR Models at t=0
PySR $H^{u-d}(x, t=0)$
x
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
0.0
0.2
0.4
0.6
0.8
1.0


2D Clustered at t=0
PySR $H^{u-d}(x, t=0)$
x
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
0.0
0.2
0.4
0.6
0.8
1.0

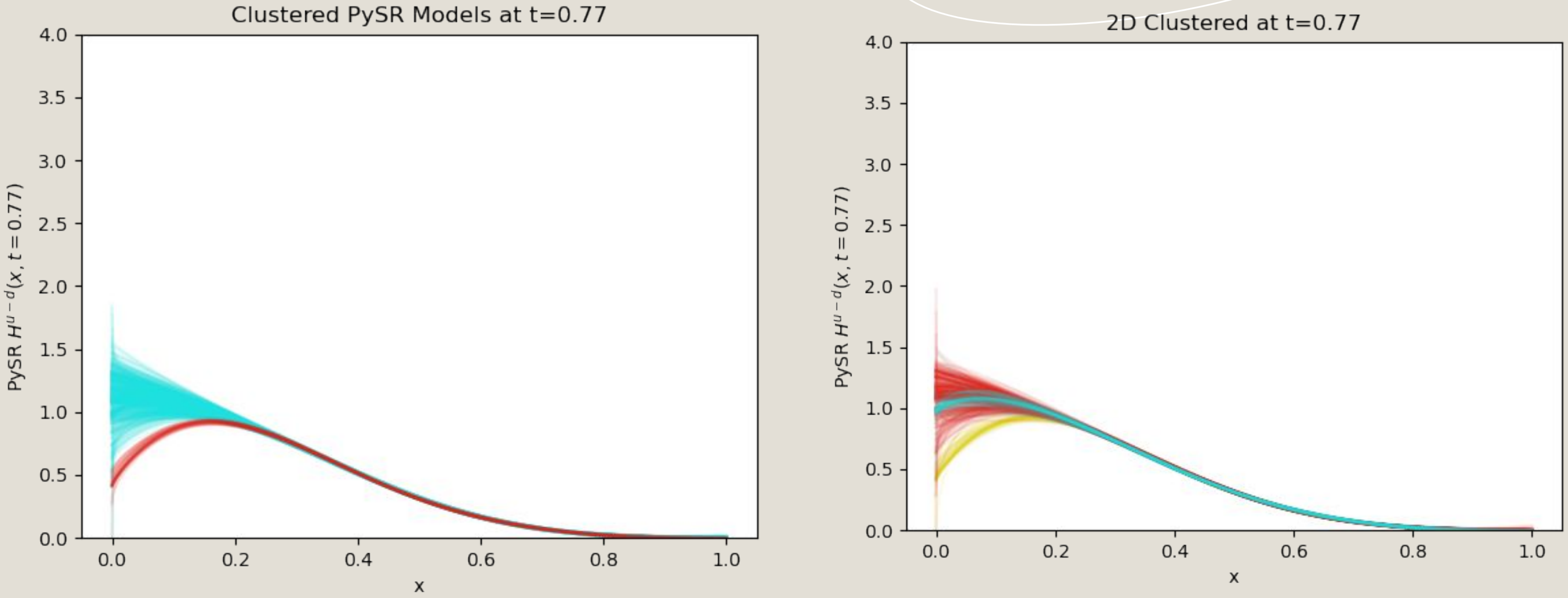

Clustered PySR Models at t=0.19
PySR $H^{u-d}(x, t=0.19)$
x
2D Clustered at t=0.19
PySR $H^{u-d}(x, t=0.19)$
x

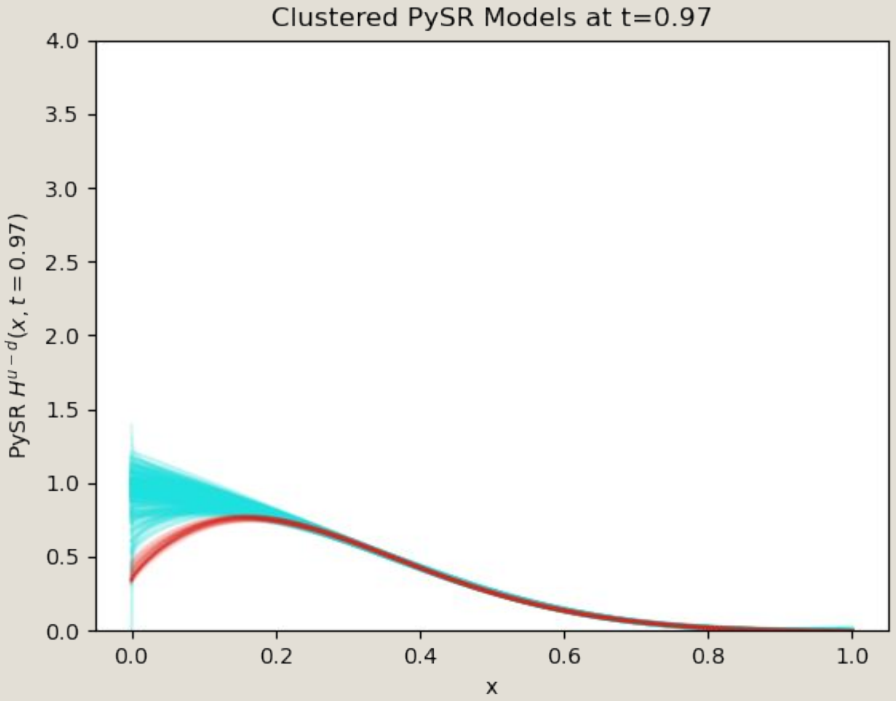


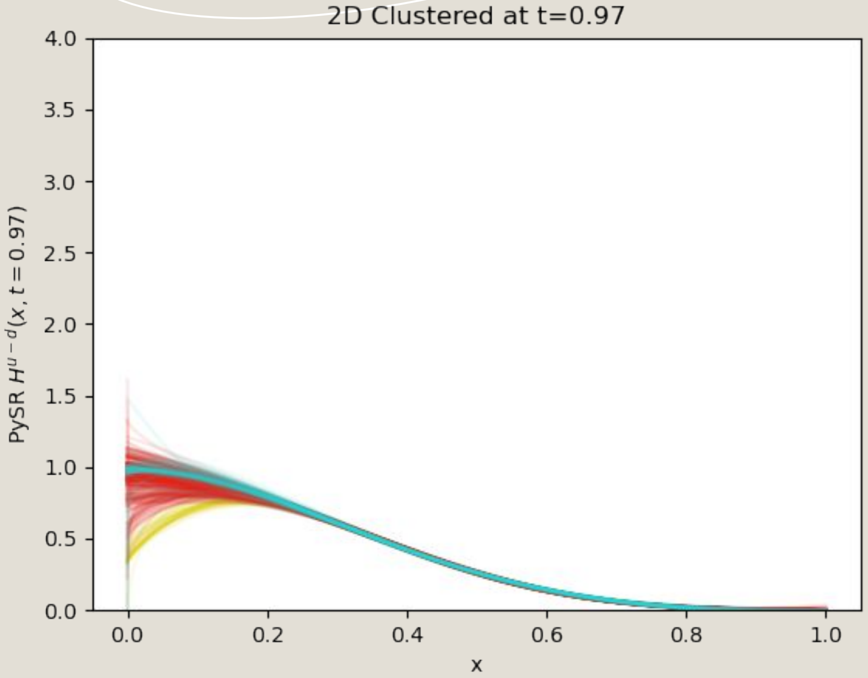
Clustered PySR Models at t=0.39
PySR $H^{u-d}(x, t=0.39)$
x
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0



2D Clustered at t=0.39
PySR $H^{u-d}(x, t=0.39)$
x
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0


Clustered PySR Models at t=0.77
PySR $H^{u-d}(x, t=0.77)$
x
2D Clustered at t=0.77
PySR $H^{u-d}(x, t=0.77)$
x


Clustered PySR Models at t=0.97
PySR $H^{u-d}(x, t = 0.97)$
x
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0


2D Clustered at t=0.97
PySR $H^{u-d}(x, t = 0.97)$
x
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0