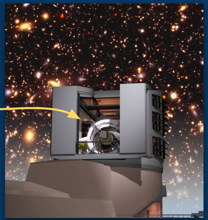
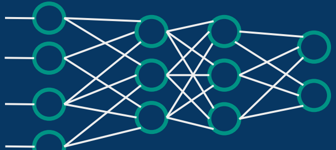
Machine Learning
for
Astrophysical Data Analysis

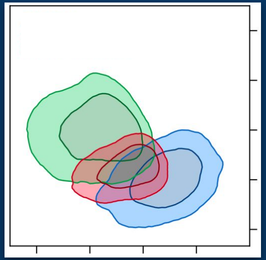
observables



data



machine learning



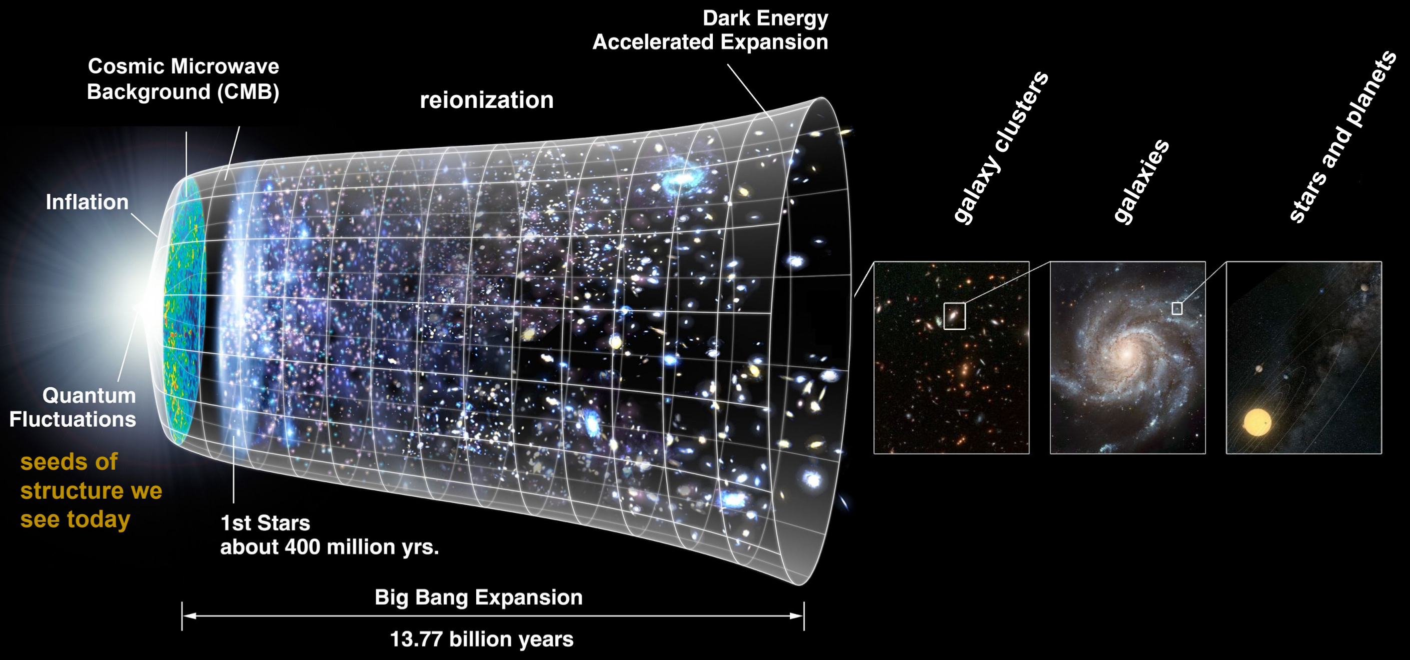
physics parameters ?

BERKELEY CENTER *for* COSMOLOGICAL PHYSICS

Vanessa Böhm

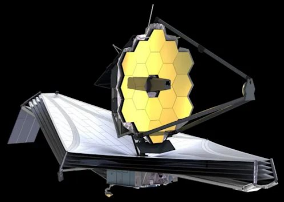
UC Berkeley

https://github.com/VMBoehm

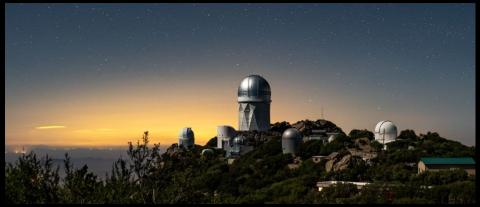

Dark Energy
Accelerated Expansion
Cosmic Microwave
Background (CMB)
reionization
galaxy clusters
galaxies
stars and planets
Inflation
Quantum
Fluctuations
seeds of
structure we
see today
1st Stars
about 400 million yrs.
Big Bang Expansion
13.77 billion years

Large-Scale Stage-4 Experiments

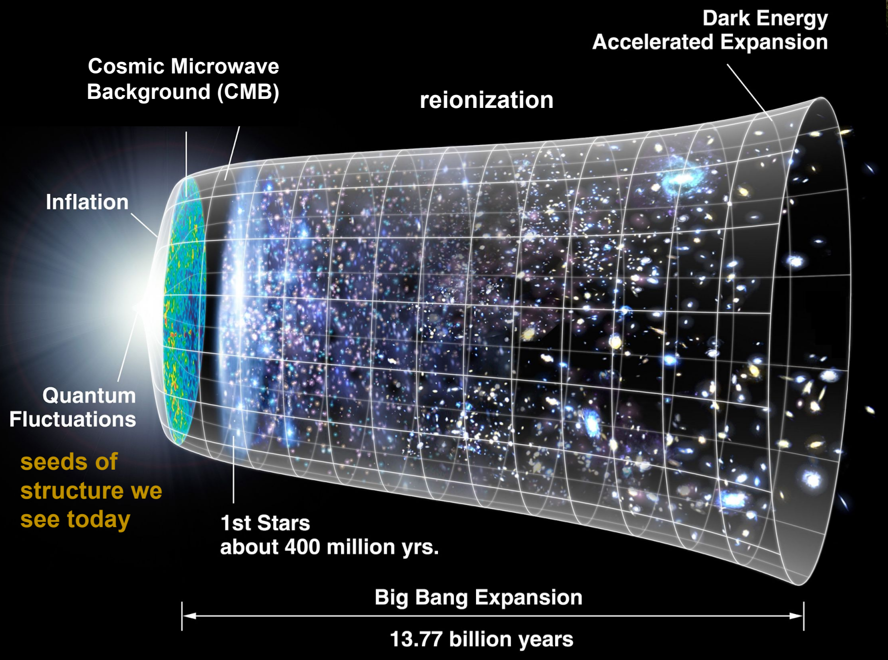
JWST



Dark Energy Spectroscopic Instrument (DESI)



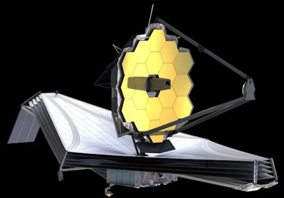
CMB-S4



Vera Rubin Observatory (LSST)

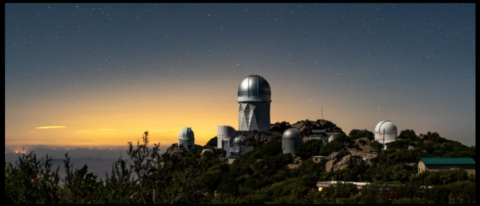
Large-Scale Stage-4 Experiments

JWST



35 Gigabits/day

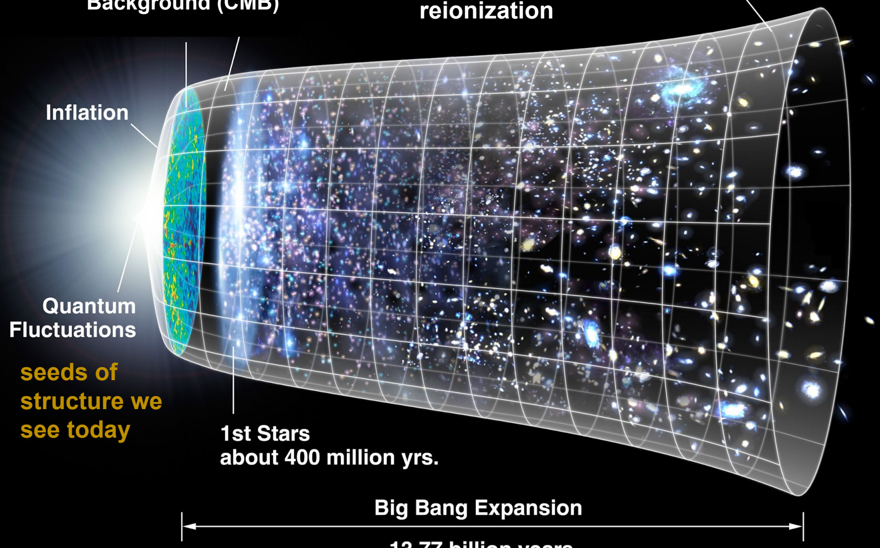
Dark Energy Spectroscopic Instrument (DESI)



35 million galaxies
10 million stars

CMB-S4

1 Gbs
50 PB total database



Vera Rubin Observatory (LSST)

20 ***billion*** galaxies
17 ***billion*** stars
20 ***terabyte*** data/day
15 ***petabyte*** total database

Machine Learning (ML) - the solution ?

An ML model is not told how to solve the problem at hand,

it *learns* how to solve it.

Machine Learning (ML) - the solution ?

An ML model is not told how to solve the problem at hand,

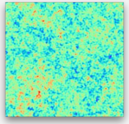
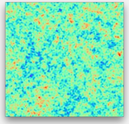
it *learns* how to solve it.

supervised learning

classification


spiral elliptical

regression

Ω=0.3 Ω=0.2

Machine Learning (ML) - the solution ?

An ML model is not told how to solve the problem at hand,

it *learns* how to solve it.

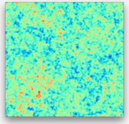
supervised learning

classification

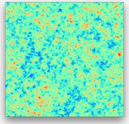
spiral

elliptical

regression



Ω=0.3



Ω=0.2

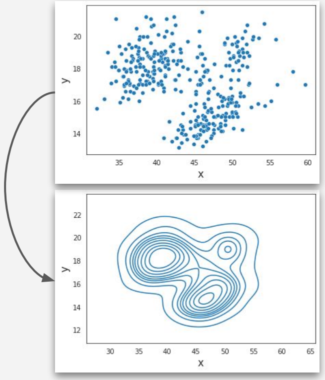
*un*supervised learning

compression

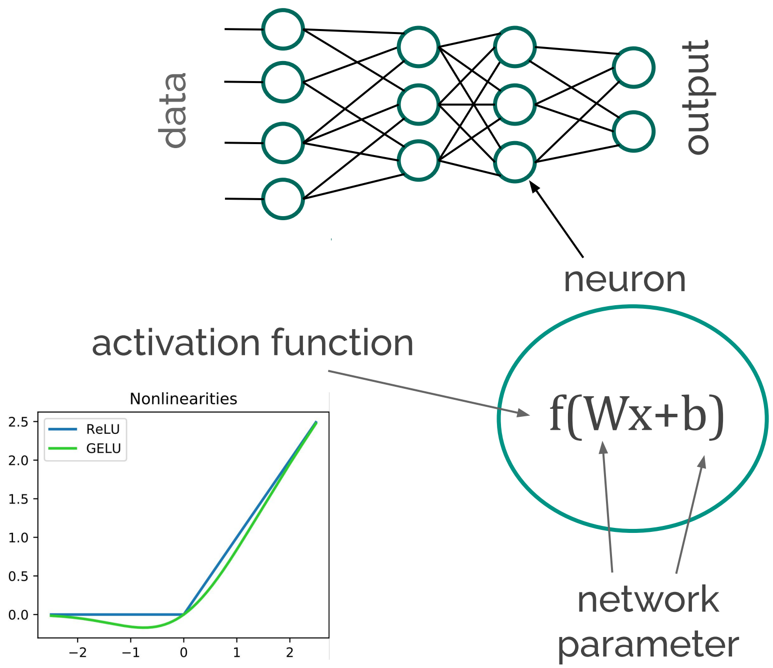
original

50%

density estimation



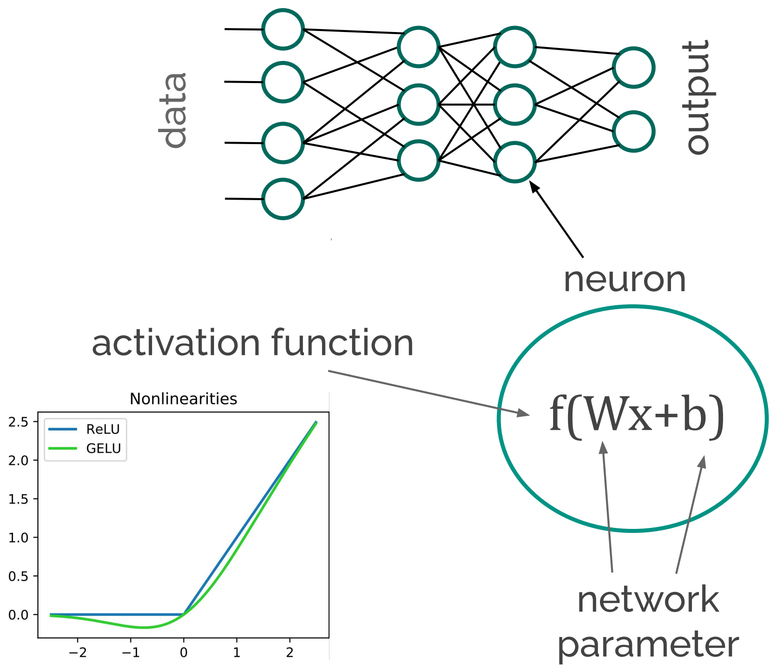
Deep Neural Networks & Related Innovations



“Universal function approximators”

Cybenco 1989, Leshno et al 1993

Deep Neural Networks & Related Innovations

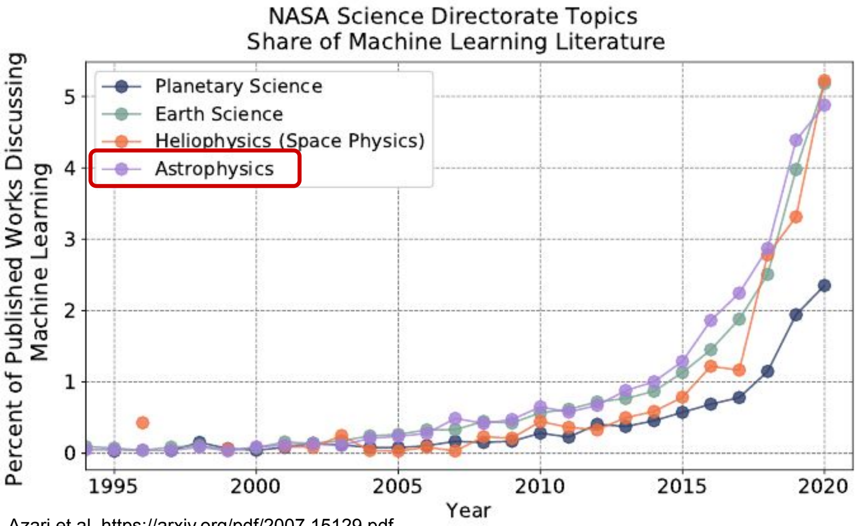


"Universal function approximators"

Cybenco 1989, Leshno et al 1993

- *Training:* Adapt **W,b** to minimize difference between current output and desired output
- *Gradients* through **backpropagation**
- *Computational feasibility* through **Graphical Processing Units (GPUs)**

Machine Learning for Astrophysical Data Analysis?



Azari et al. https://arxiv.org/pdf/2007.15129.pdf

Source: xkcd.com/

Inverse Problems

From observations to signals: probabilistic inversion with Bayes theorem

what we observe:

data

$$d = A(s) + n$$

forward model

signal

what we are interested in

likelihood prior

posterior

$$p(s|d) = \frac{p(d|s)p(s)}{p(d)}$$

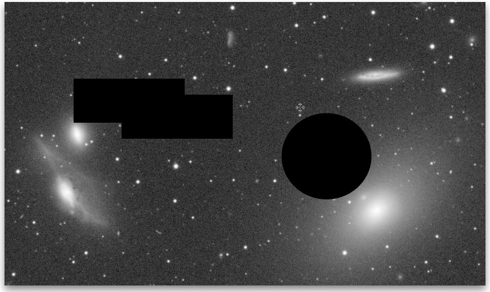
normalization

Inverse Problems in Astrophysics

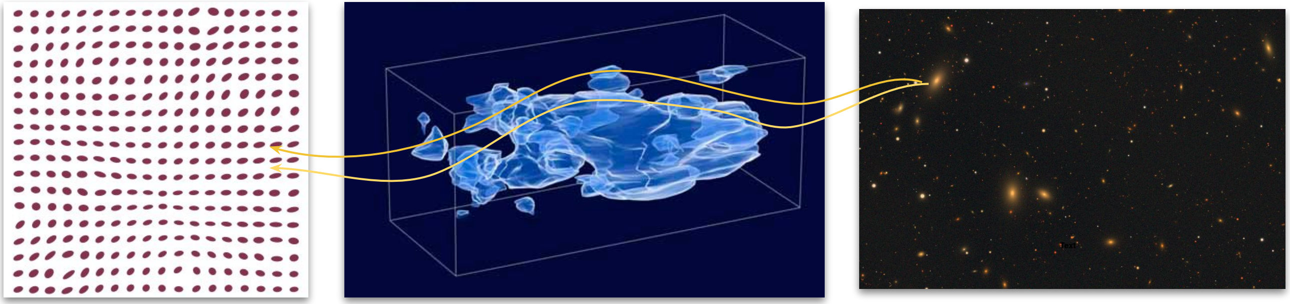
Credit: Event Horizon Telescope Coll.

Image Reconstruction from Interferometry

Lens Reconstruction
from strongly lensed Images



Data Denoising and
Inpainting



3D Dark Matter Tomography from Weak Lensing

Challenge #1: High Dimensionality

In posterior analysis we ask: what is the most likely signal?
How well can we reconstruct the signal?

- ❑ Find Maximum
- ❑ Sample from the Posterior

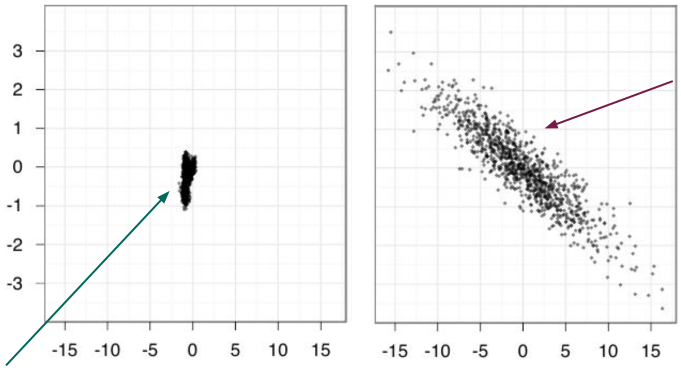
Challenge #1: High Dimensionality

In posterior analysis we ask: what is the most likely signal?
How well can we reconstruct the signal?

- ❏ Find Maximum
- ❏ Sample from the Posterior

Extremely difficult in high dimensions!

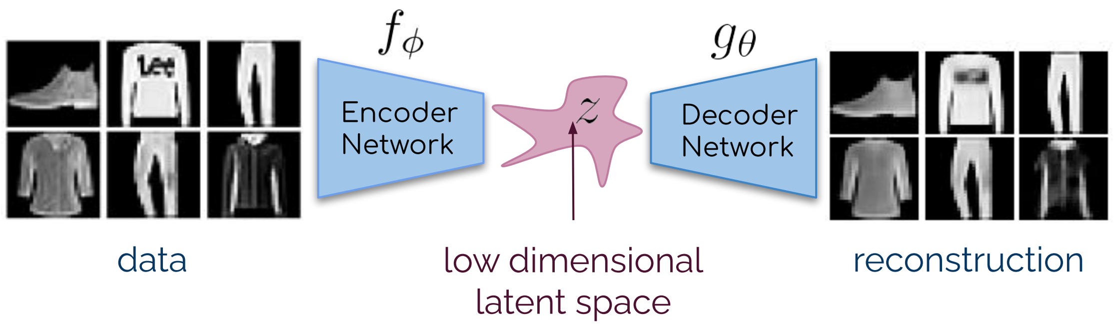
Sampling a highly correlated Gaussian in 250 dimensions



From Hoffmann & Gelman 2011
https://arxiv.org/abs/1111.4246

Challenge #1: High Dimensionality

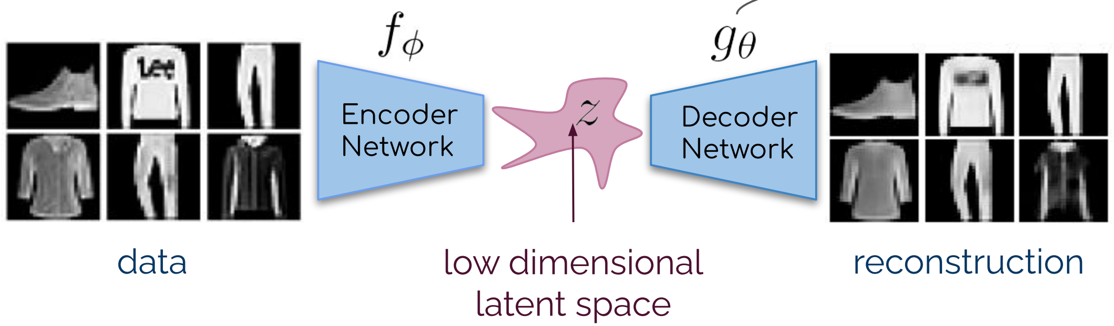
A deep Auto-Encoder finds a lower dimensional representation of the data



networks are *trained to* minimize reconstruction error

Challenge #1: High Dimensionality

A deep Auto-Encoder finds a lower dimensional representation of the data



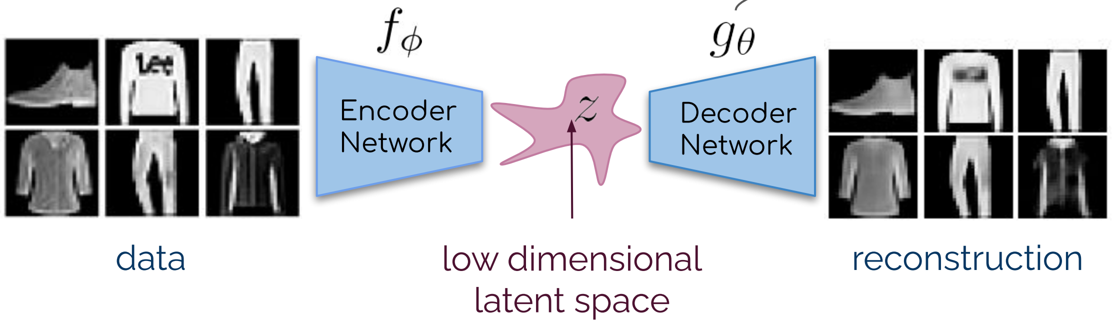
data model

$$d = A(s) + n$$

networks are *trained to* minimize reconstruction error

Challenge #1: High Dimensionality

A deep Auto-Encoder finds a lower dimensional representation of the data



data model

$$d = A(g_\theta(z)) + n$$

low dimensional posterior

$$p(z|d) = \frac{p(d|z)p(z)}{p(d)}$$

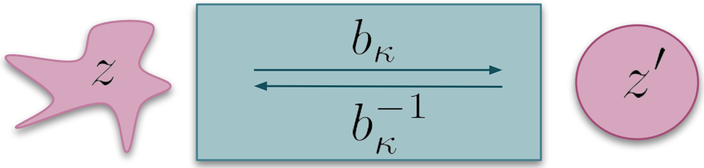
networks are *trained to* minimize reconstruction error

Challenge #2: Encoding prior knowledge

only in rare cases do we have analytic expressions for prior probabilities

Neural Density Estimation

Normalizing Flows: A bijective transformation Gaussizanizes the data



transformation is parameterized by neural network

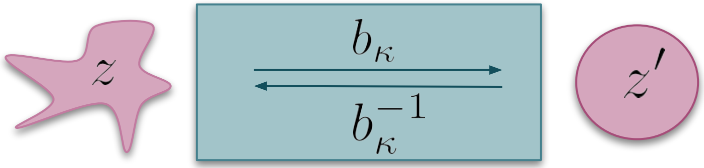
e.g. RealNVP (Dinh et al. 2019), Glow (Kingma et al 2018), NSF (Durcan 2019),SINF (Dai et al 2021)

Challenge #2: Encoding prior knowledge

only in rare cases do we have analytic expressions for prior probabilities

Neural Density Estimation

Normalizing Flows: A bijective transformation Gaussizanizes the data



transformation is parameterized by neural network

e.g. RealNVP (Dinh et al. 2019), Glow (Kingma et al 2018), NSF (Durcan 2019),SINF (Dai et al 2021)

data model

$$d = A(g_\theta(b_\kappa(z'))) + n$$

low dimensional posterior

Gaussian prior

$$p(z'|d) = \frac{p(d|z')p(z')}{p(d)}$$

An Illustrative Toy Example

dimensionality ~800 pixels

An Illustrative Toy Example

VB et al. 2019, BDL@NeurIPS

dimensionality ~800 pixels

most likely reconstruction

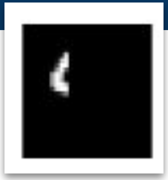
max p(z'|d)

underlying truth

An Illustrative Toy Example

VB et al. 2019, BDL@NeurIPS

dimensionality ~800 pixels



most likely reconstruction

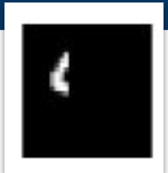
max p(z'|d)

underlying truth

An Illustrative Toy Example

VB et al. 2019, BDL@NeurIPS

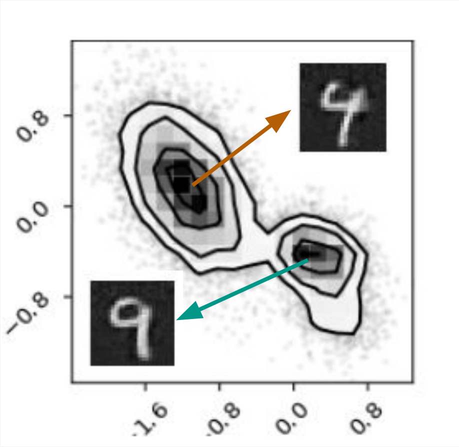
dimensionality ~800 pixels



most likely reconstruction
max p(z'|d)

underlying truth

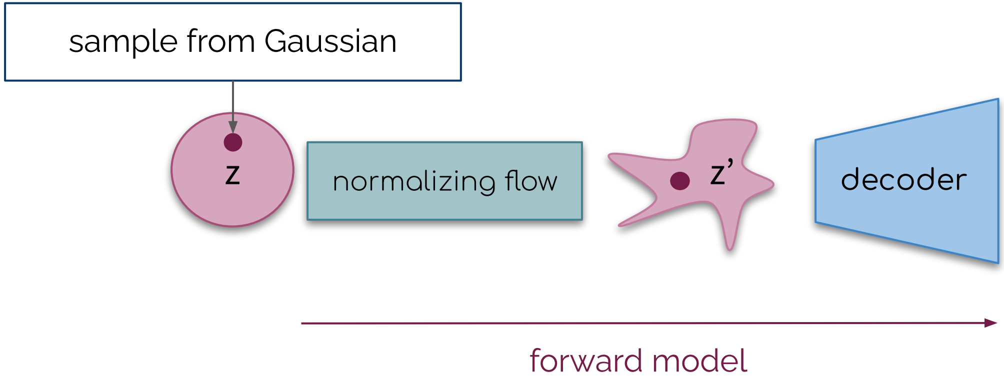
uncertainty quantification
sampling p(z'|d)



Artificial Data Generation

Probabilistic Autoencoder

VB & Seljak 2020 (arXiv:2006.05479)



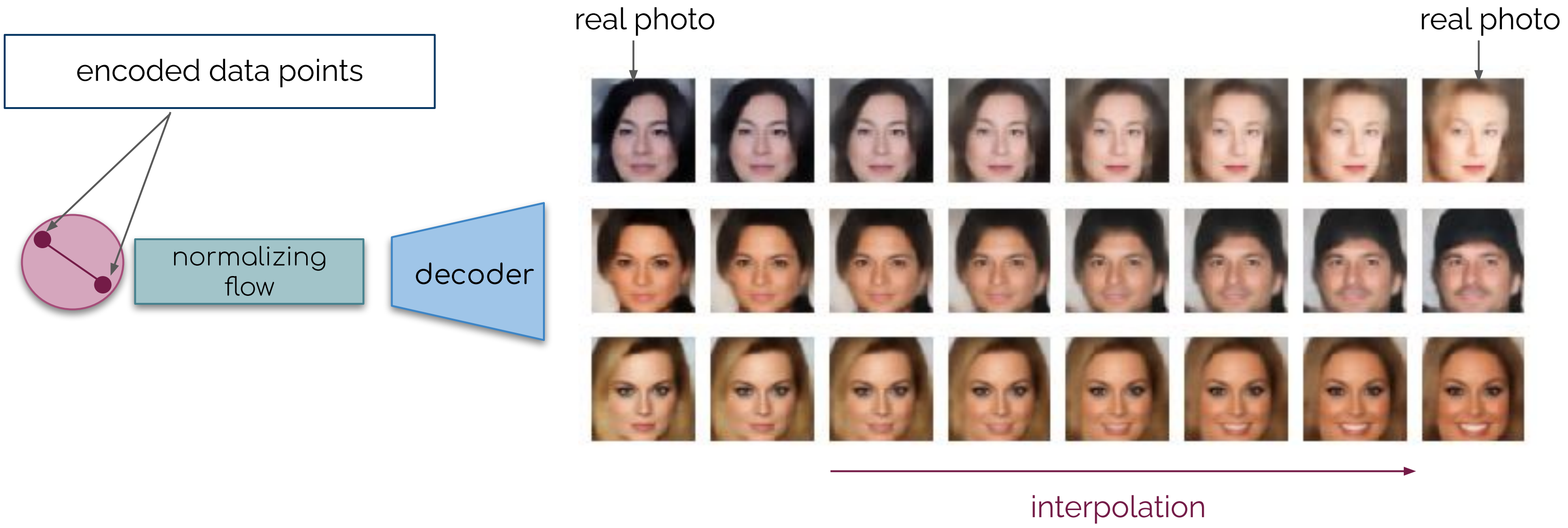
realistic artificial data

https://github.com/VMBoehm/PAE

Smooth Data Interpolation

Probabilistic Autoencoder

VB & Seljak 2020 (arXiv:2006.05479)

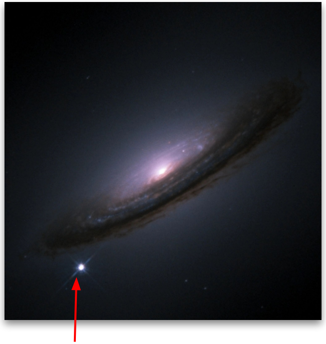


https://github.com/VMBoehm/PAE

Stein et al. (**VB**) 2022, submitted

Application in Astrophysics

Cosmology with Type 1a Supernovae

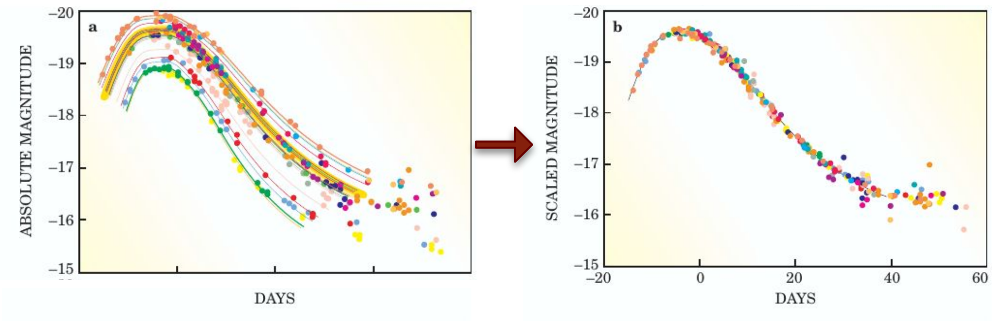


SN Type 1A

first significant detection of accelerated expansion

Perlmutter et al. 1998, Riess et al. 1998

Standardizable Candles



Source: S. Perlmutter, Physics today, 2003

limiting factors:

- sparse observations
- spectral diversity

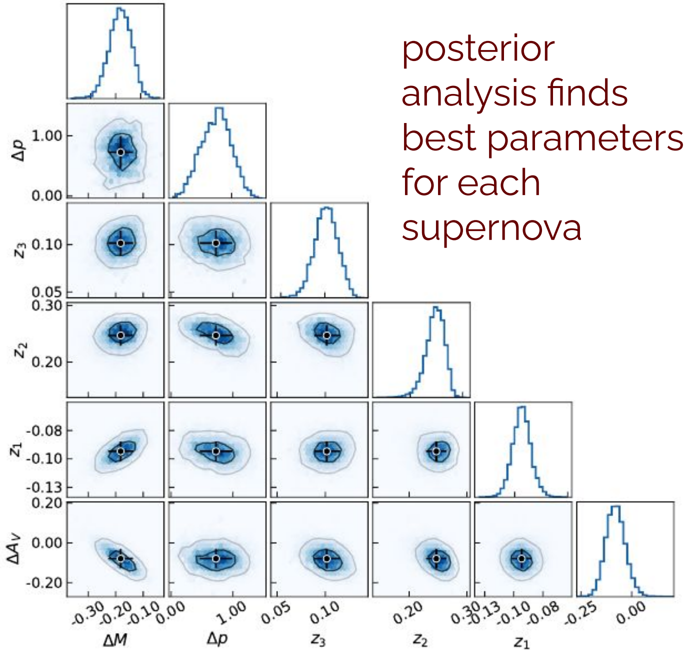
Application in Astrophysics

Probabilistic Autoencoder for Supernova Type 1a

- 3 latent parameters capture spectral diversity
- 3 latent parameters encode interpretable physics
  - Time after peak brightness
  - Magnitude
  - Extrinsic extinction

➔ Known physics is coded into the data model

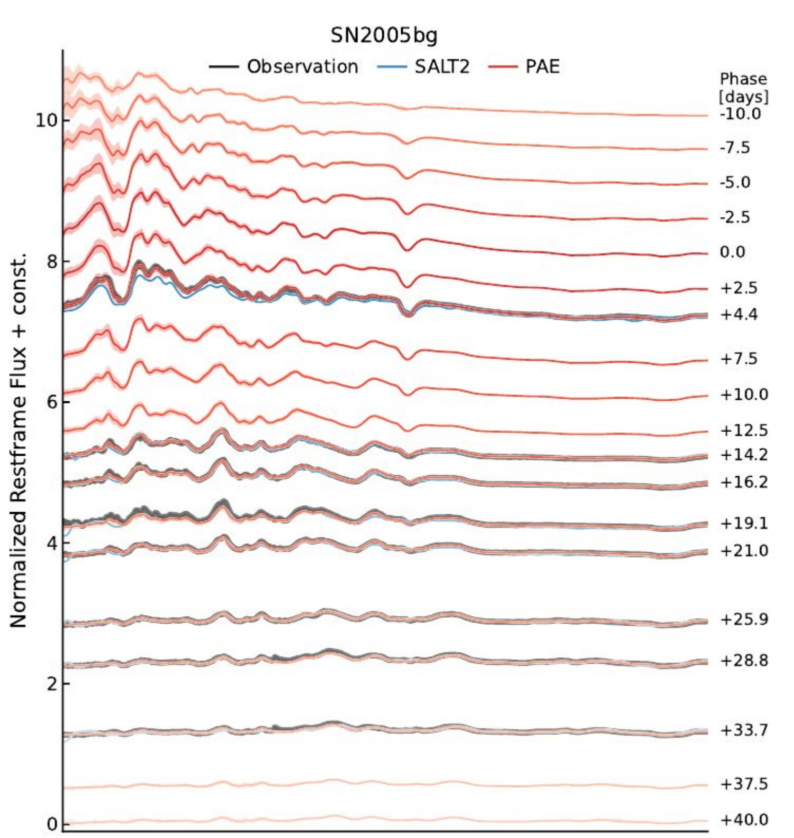
➔ Improvement over current models



posterior analysis finds best parameters for each supernova

Stein et al. (**VB**, The Nearby Supernova Factory) 2022, submitted

Filling in Missing Observations



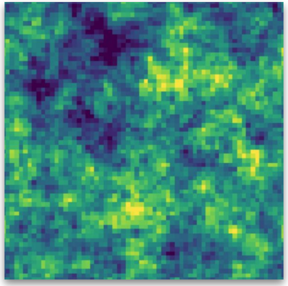
time

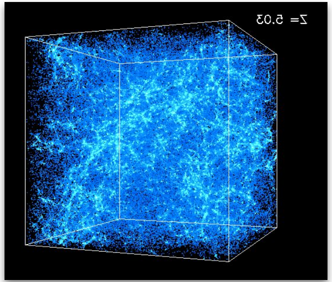
Stein et al. (**VB**, The Nearby Supernova Factory) 2022, submitted

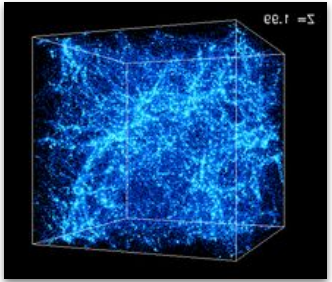
Computationally Complex Forward Models

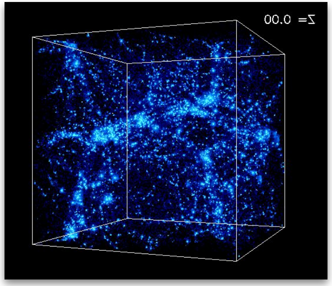
Forward models are often non-linear and computationally complex

Cosmic Structure Formation









source: http://cosmicweb.uchicago.edu/filaments.html

random initial fluctuations

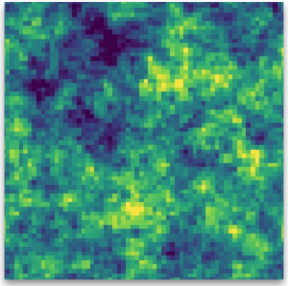
cosmic structure formation

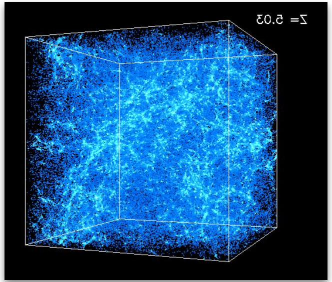
Data:
tracers of DM
e.g.
Galaxies & Lensing

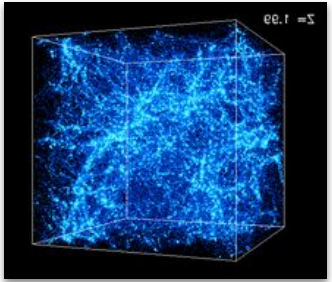
Computationally Complex Forward Models

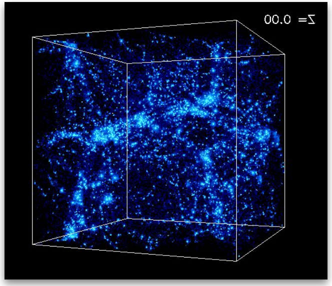
Forward models are often non-linear and computationally complex

Cosmic Structure Formation









source: http://cosmicweb.uchicago.edu/filaments.html

Signal
random initial fluctuations

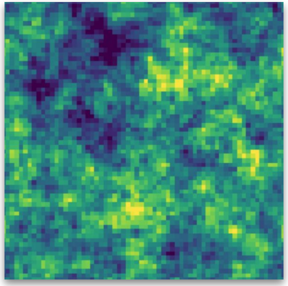
Forward Model:
cosmic structure formation

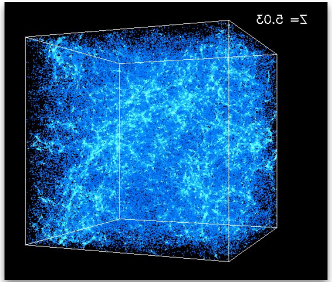
Data:
tracers of DM
e.g.
Galaxies & Lensing

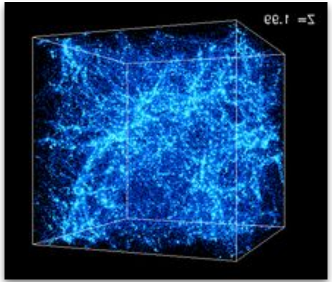
Computationally Complex Forward Models

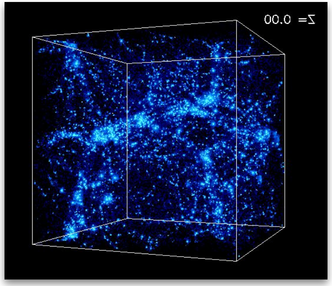
Forward models are often non-linear and computationally complex

Cosmic Structure Formation









source: http://cosmicweb.uchicago.edu/filaments.html

Signal
random initial fluctuations

Forward Model:
cosmic structure formation

Data:
tracers of DM
e.g.
Galaxies & Lensing

Cosmological Parameters:
amount of dark matter, dark energy, mass of neutrinos etc.

Computationally Complex Forward Models

a case that requires high-dimensional posterior analysis

extremely high-dimensional posterior

$$P(\eta, s|d) \sim P(d|\eta, s)P(\eta)P(s)$$

cosmological parameters

initial conditions

Gaussian prior

Computationally Complex Forward Models

a case that requires high-dimensional posterior analysis

extremely high-dimensional posterior

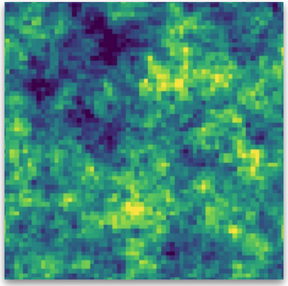
$$P(\eta, s|d) \sim P(d|\eta, s)P(\eta)P(s)$$

cosmological parameters

initial conditions

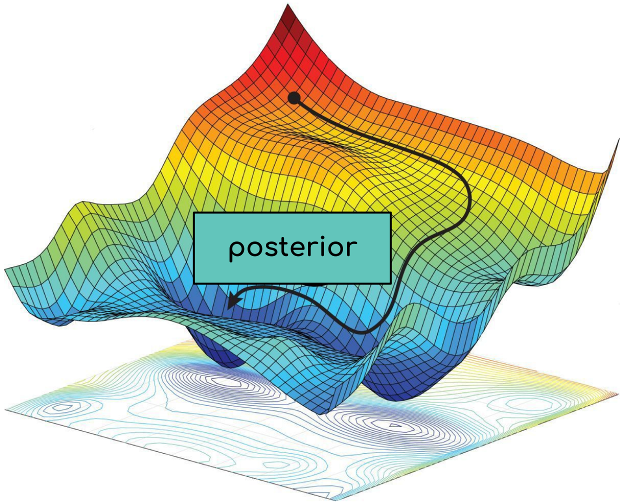
Gaussian prior

>10^6 additional parameters!



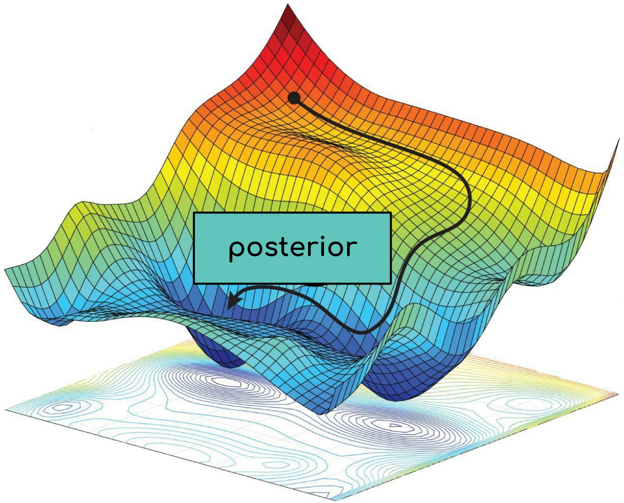
Derivative-aided Optimization and Sampling

Gradient Descent



Derivative-aided Optimization and Sampling

Gradient Descent

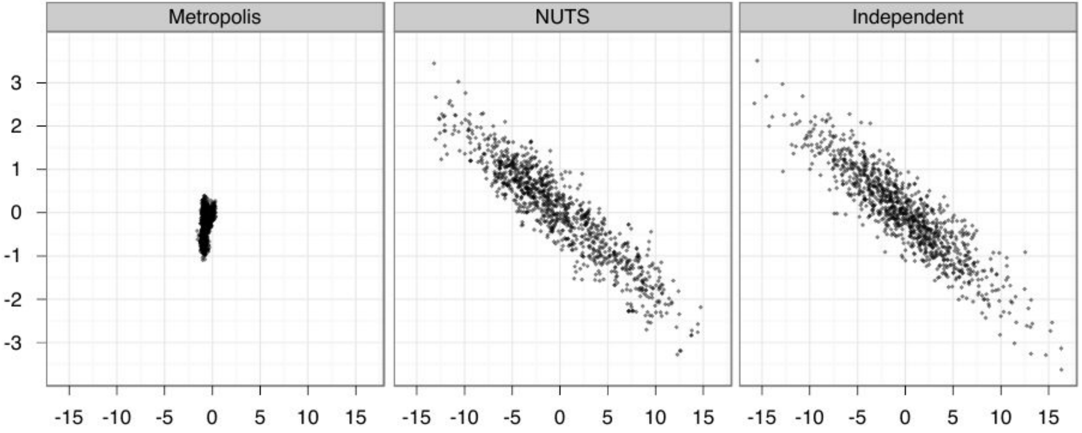


Derivative-aided Sampling

common sampler

derivative-aided

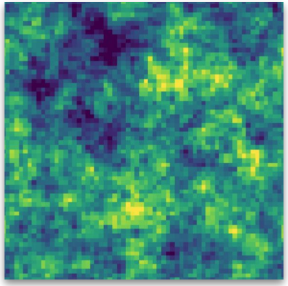
drawn from original

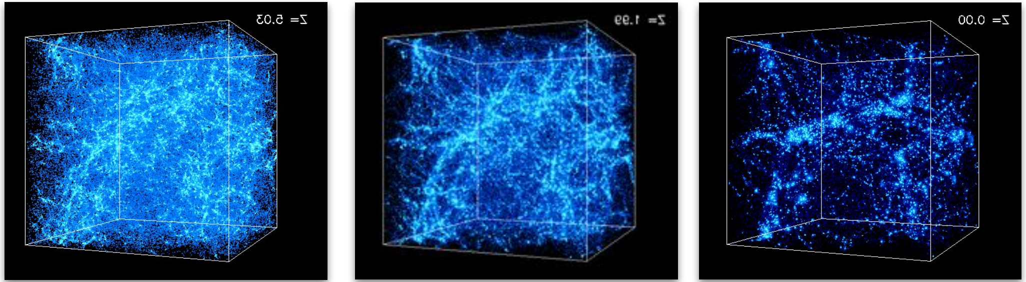


From Hoffmann & Gelman 2011
https://arxiv.org/abs/1111.4246

Computationally Complex Forward Models

Can this be made differentiable?

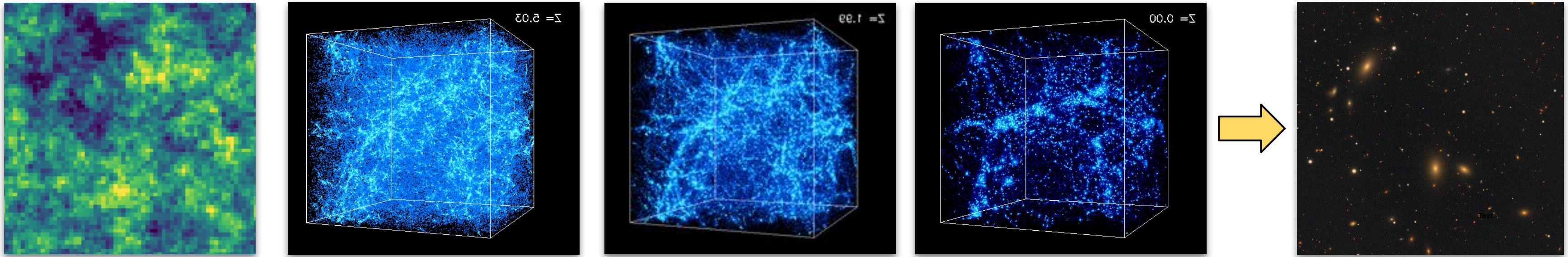




source: http://cosmicweb.uchicago.edu/filaments.html

Computationally Complex Forward Models

Can this be made differentiable?



source: http://cosmicweb.uchicago.edu/filaments.html

Yes, with backpropagation!

Backpropagation

Memory efficient derivatives of scalar functions with *reverse mode differentiation*

$$y = F(x)$$

scalar
(posterior probability)

high-dimensional vector
(initial conditions)

$$F = A \circ B \circ C$$

consecutive operations
in forward model

Forward-mode differentiation:

huge matrix

$$\frac{\partial F}{\partial x}\Big|_{x_0} = \frac{\partial A}{\partial a}\Big|_{a_0}\left(\frac{\partial B}{\partial b}\Big|_{b_0}\frac{\partial C}{\partial x}\Big|_{x_0}\right) = J_A \cdot (J_B \cdot J_C)$$

Backpropagation

Memory efficient derivatives of scalar functions with *reverse mode differentiation*

$$y = F(x)$$

scalar
(posterior probability)

high-dimensional vector
(initial conditions)

$$F = A \circ B \circ C$$

consecutive operations
in forward model

Reverse-mode differentiation:

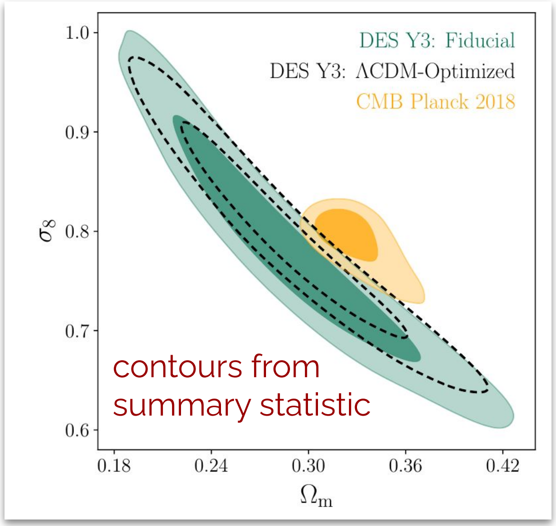
vector!

$$\frac{\partial F}{\partial x}\Big|_{x_0} = \left(\frac{\partial A}{\partial a}\Big|_{a_0} \frac{\partial B}{\partial b}\Big|_{b_0} \right) \frac{\partial C}{\partial x}\Big|_{x_0} = (J_A \cdot J_B) \cdot J_C$$

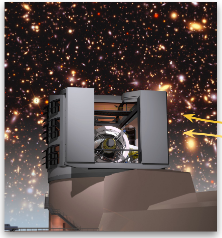
requires saving results of forward pass on a tape

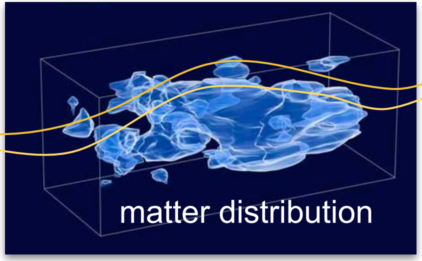
Back to Astrophysics

gravitational lensing is a direct probe of (dark) matter



DES 3-year results, Amon et al. 2021



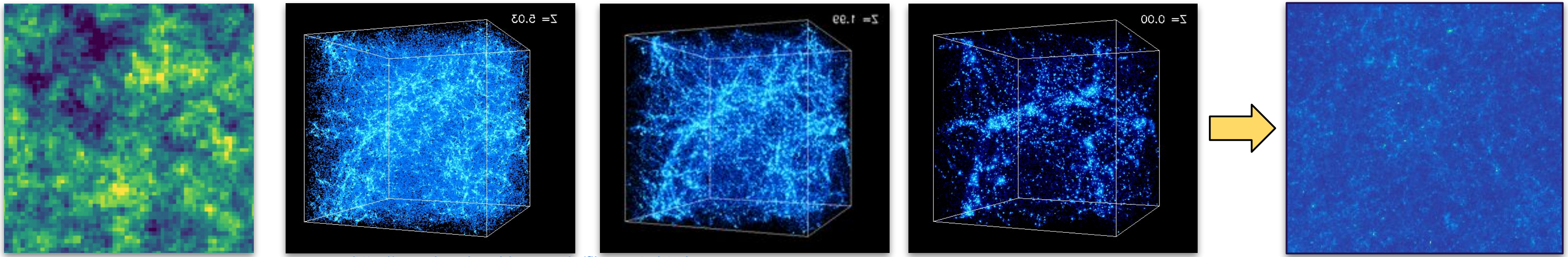


- Sensitivity to a variety of cosmological parameters
- Tightest constraint on combined mass of neutrinos to date
- Emerging tension with high-redshift probe?
- Full-field analysis could significantly tighten constraints!

The *MADLens* Code

VB, Y. Feng, M. Lee, B. Dai 2021

A fully differentiable lensing code



source: http://cosmicweb.uchicago.edu/filaments.html

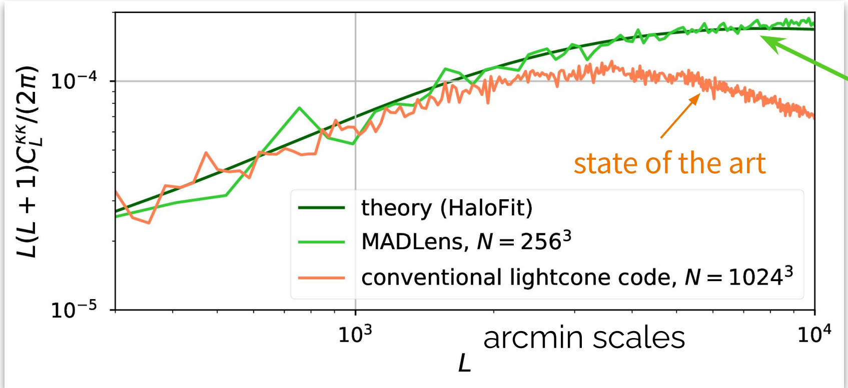
https://github.com/VMBoehm/MADLens

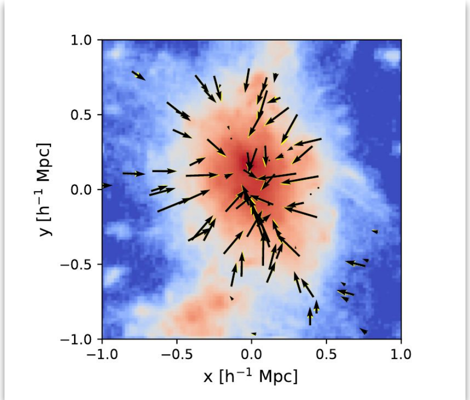
The *MADLens* Code

VB, Y. Feng, M. Lee, B. Dai 2021

A fully differentiable lensing code

- Runs an entire N-body simulation in tens of seconds
- Reaches extremely high-accuracy, even on small scales





Particle Gradient Descent[2] (PGD) displaces particles to recover positions of high-resolution simulation.

https://github.com/VMBoehm/MADLens

[2]Dai et al. 2016 (arxiv:1603.00476)

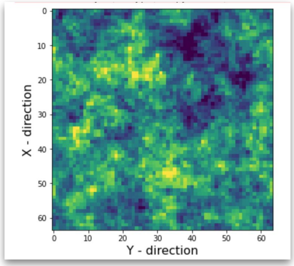
Lensing Reconstruction with MADLens

Lee et al. (**VB**) in prep.

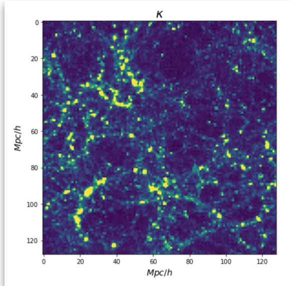
truth

reconstruction

initial fluctuations



data



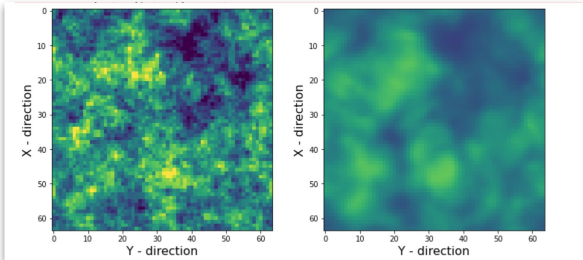
VB et al. 2017 (arxiv:1701.01886)

Lensing Reconstruction with MADLens

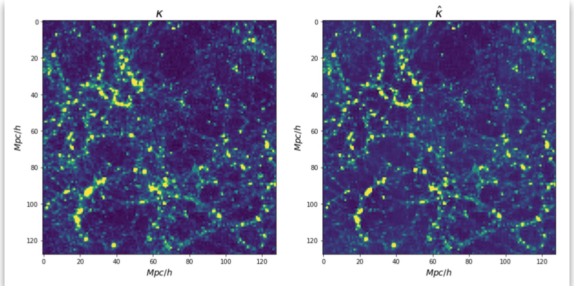
Lee et al. (**VB**) in prep.
"*MADMuse*"

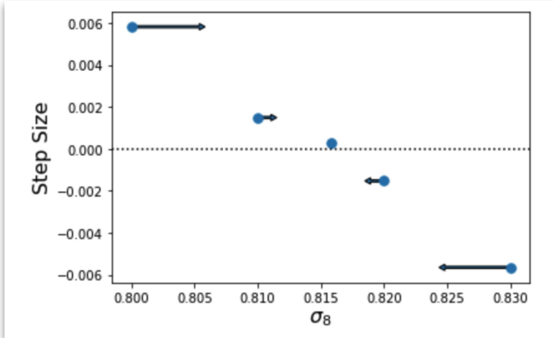
truth reconstruction

initial fluctuations



data





"Differentiable Universe Initiative"

VB et al. 2017 (arxiv:1701.01886)

Millea & Seljak 2021 (arxiv:2112.09354)

Conclusions

Things I have talked about:

- Machine Learning is more than just classification.
- It can be part of a sound statistical data analysis.
- Differentiability is great and we should make all our codes differentiable!

Things I didn't have time to talk about:

- Using Machine Learning for Anomaly Detection and Discovery