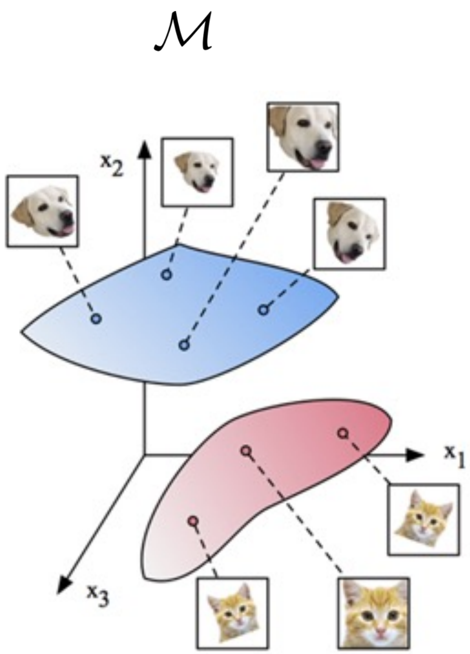
Estimating the intrinsic dimensionality using Normalizing Flows

NeurIPS 2022

Christian Horvat[b], Jean-Pascal Pfister[b]

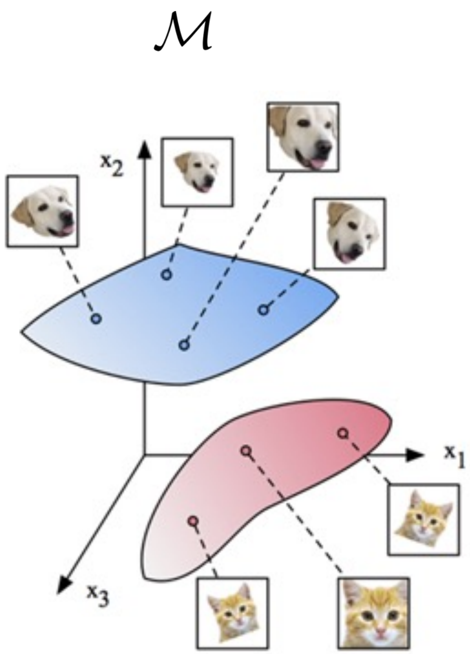
b UNIVERSITÄT BERN

Motivation



*Fefferman et. al (2016)

Motivation



*Fefferman et. al (2016)

Manifold hypothesis

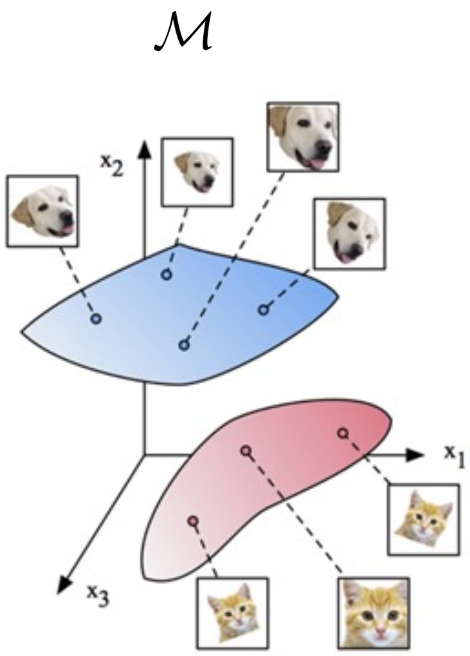
$\mathbb{R}^D$ $\quad D > d \quad$ $\mathbb{R}^d$

$g(x)$ →

$x \sim p(x)$ $\qquad$ $u \sim \mathcal{N}(u; 0, I)$

← $f(u)$

Motivation



*Fefferman et. al (2016)

Manifold hypothesis

$\mathbb{R}^D$ $\quad D > d \quad$ $\mathbb{R}^d$

$g(x)$

$x \sim p(x)$ $\quad\quad$ $u \sim \mathcal{N}(u; 0, I)$

$f(u)$

Latent variable models

- VAE:

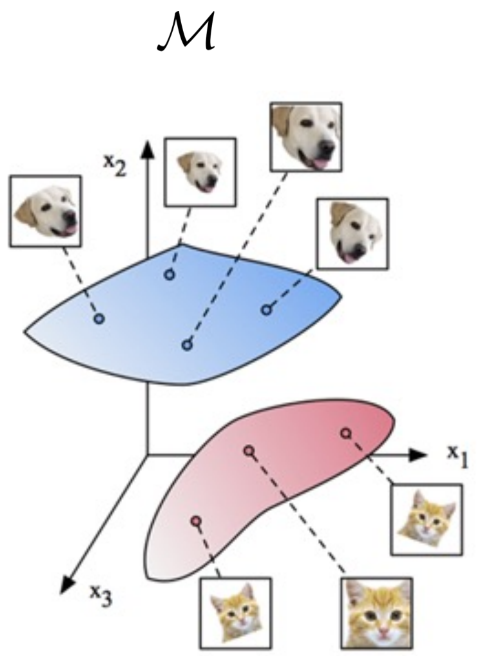
$$p(x|u) \sim \mathcal{N}(x; f(u), I)$$

$$p(u|x) \sim \mathcal{N}(u; g(x), I)$$

- Normalizing Flows (NF):

$$g = f^{-1}$$

Motivation



*Fefferman et. al (2016)

Manifold hypothesis

$\mathbb{R}^D$ $\quad D > d \quad$ $\mathbb{R}^d$

$g(x)$: $x \sim p(x) \rightarrow u \sim \mathcal{N}(u; 0, I)$

$f(u)$: $u \sim \mathcal{N}(u; 0, I) \rightarrow x \sim p(x)$

Latent variable models

- VAE:

$$p(x|u) \sim \mathcal{N}(x; f(u), I)$$

$$p(u|x) \sim \mathcal{N}(u; g(u), I)$$

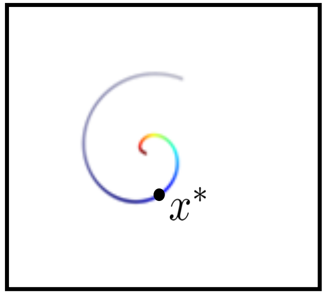
- Normalizing Flows (NF):

$$g = f^{-1}$$

Most latent variable models assume that data live on manifolds

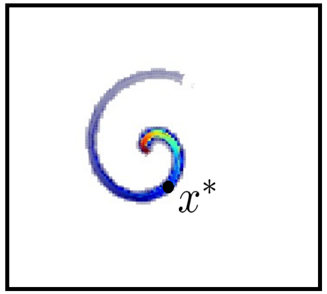
Method

$x \sim p(x)$

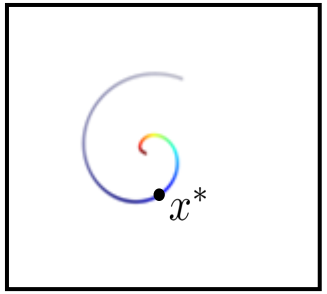


Method

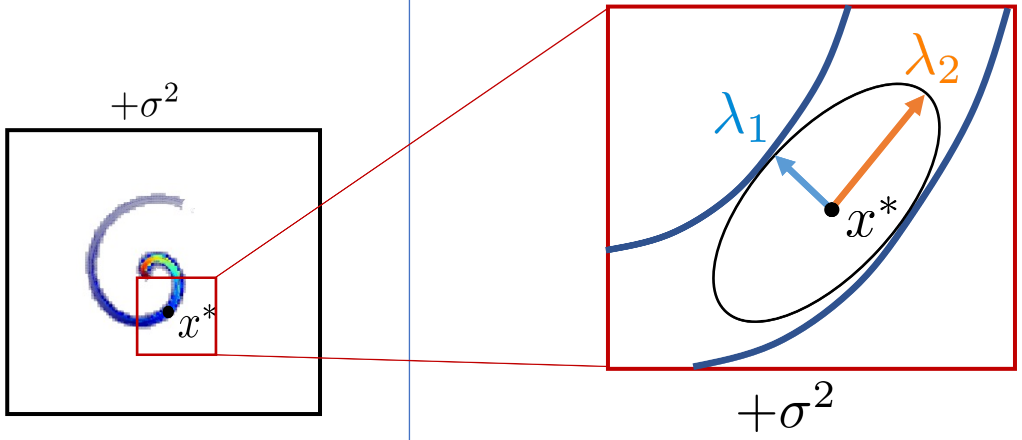
$+\sigma^2$



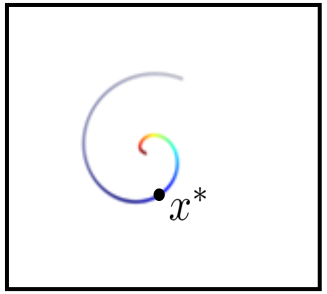
$x \sim p(x)$



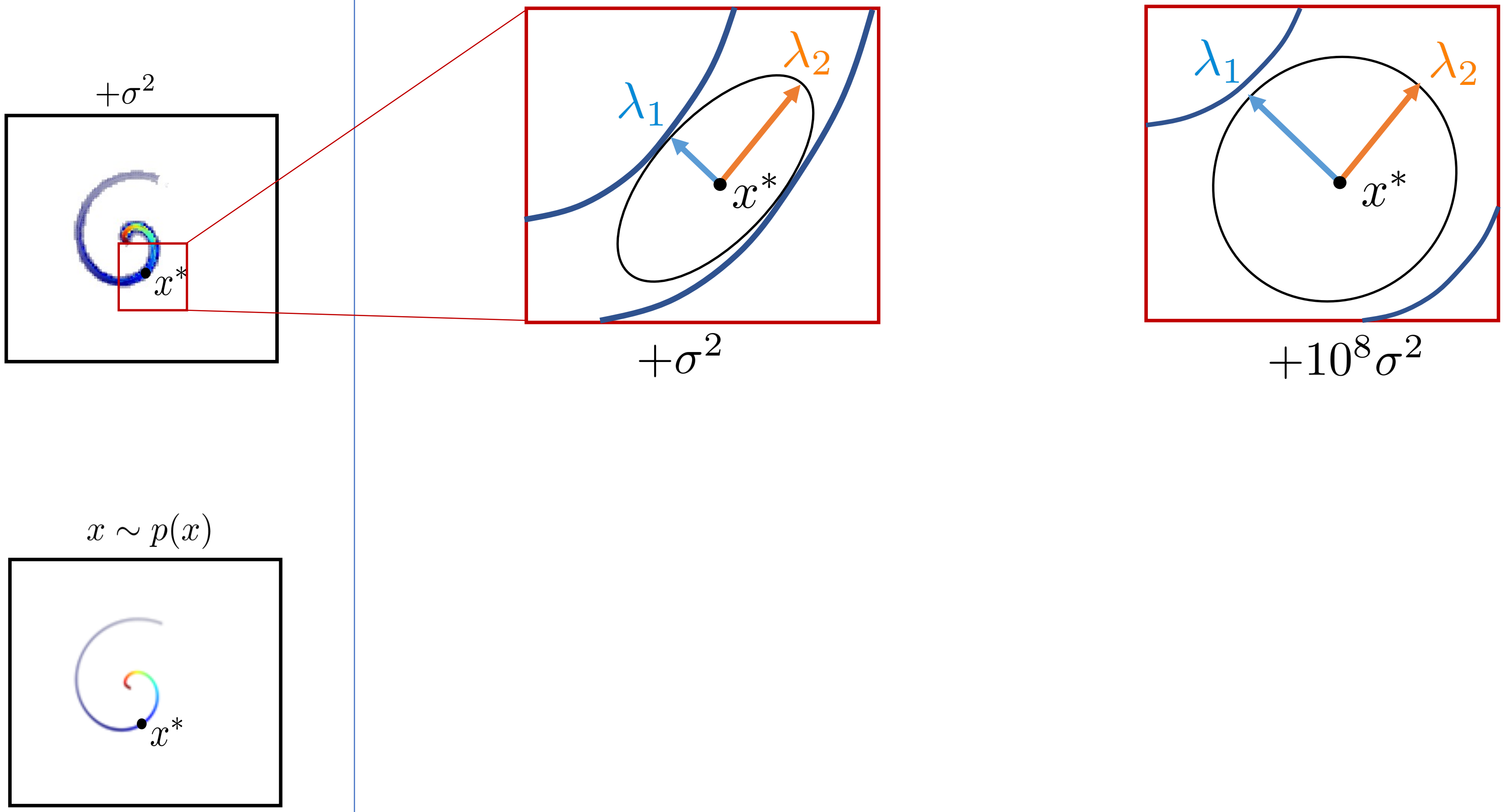
Method



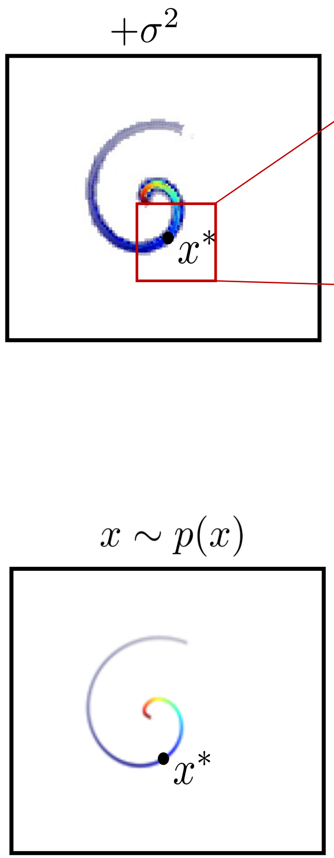
$x \sim p(x)$

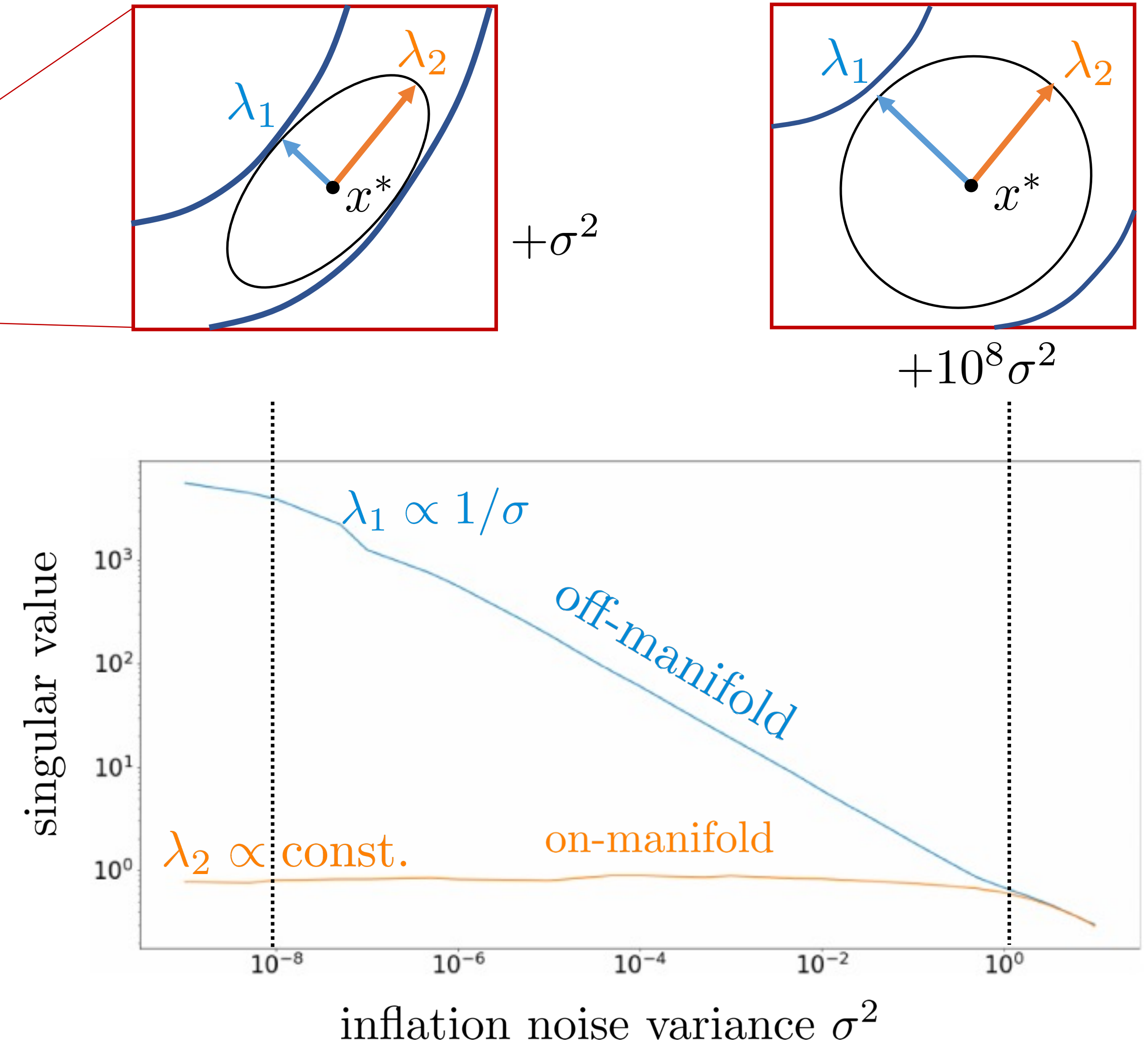


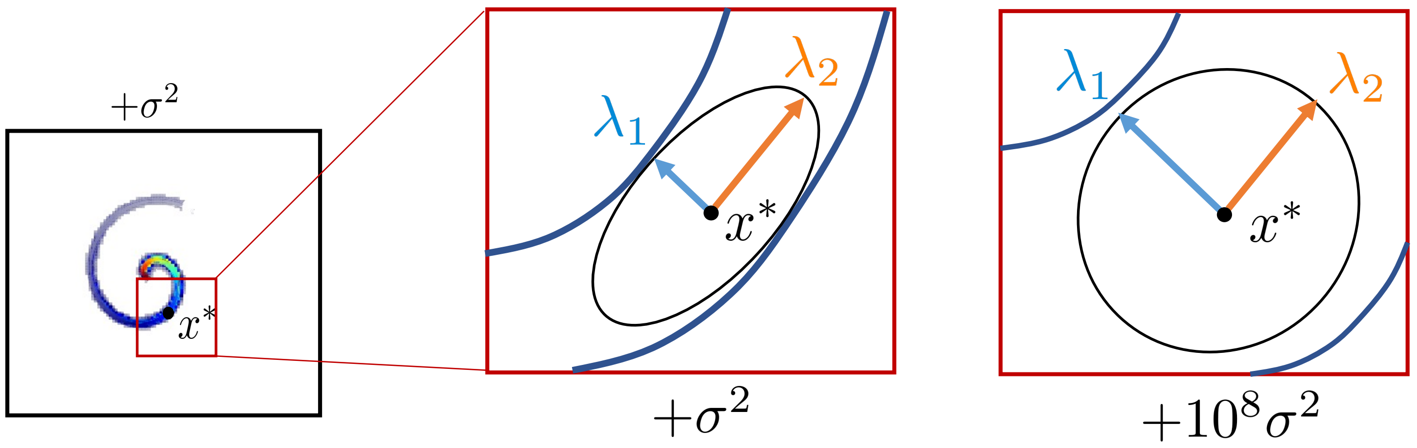
Method


$+\sigma^2$
x^*
$x \sim p(x)$
x^*
λ_1
λ_2
x^*
$+\sigma^2$
λ_1
λ_2
x^*
$+10^8\sigma^2$

Method

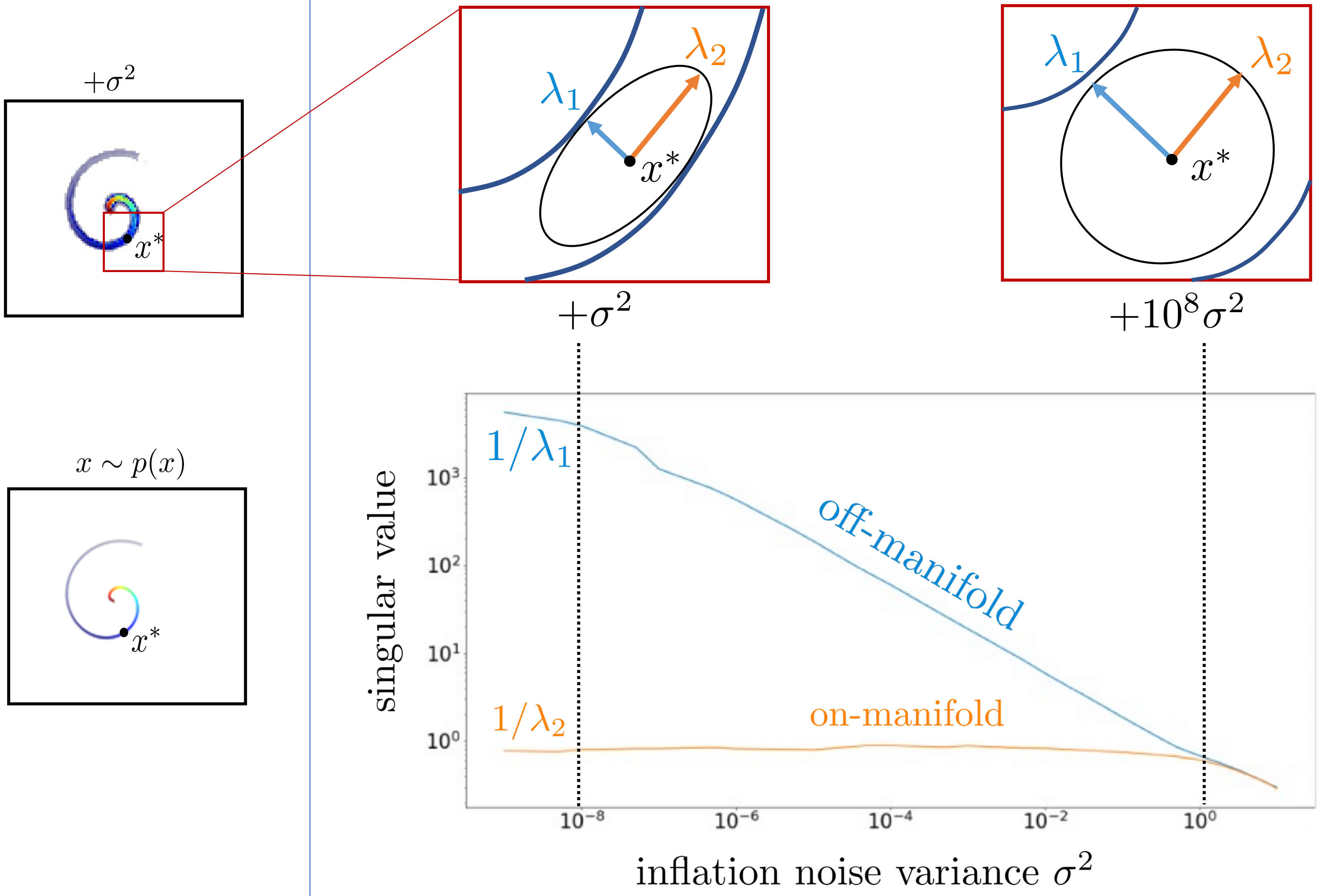

$+\sigma^2$
x^*
$x \sim p(x)$
x^*


λ_1
λ_2
x^*
$+\sigma^2$
λ_1
λ_2
x^*
$+10^8\sigma^2$
$\lambda_1 \propto 1/\sigma$
off-manifold
$\lambda_2 \propto$ const.
on-manifold
singular value
10^3
10^2
10^1
10^0
10^{-8}
10^{-6}
10^{-4}
10^{-2}
10^0
inflation noise variance σ^2



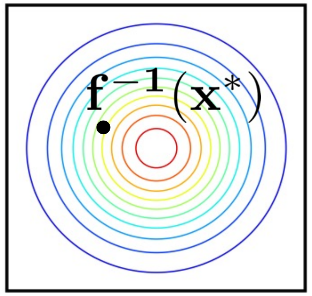
$+\sigma^2$
x^*
λ_1
λ_2
x^*
$+\sigma^2$
λ_1
λ_2
x^*
$+10^8\sigma^2$

Method



learn noisy $p(x)$ using

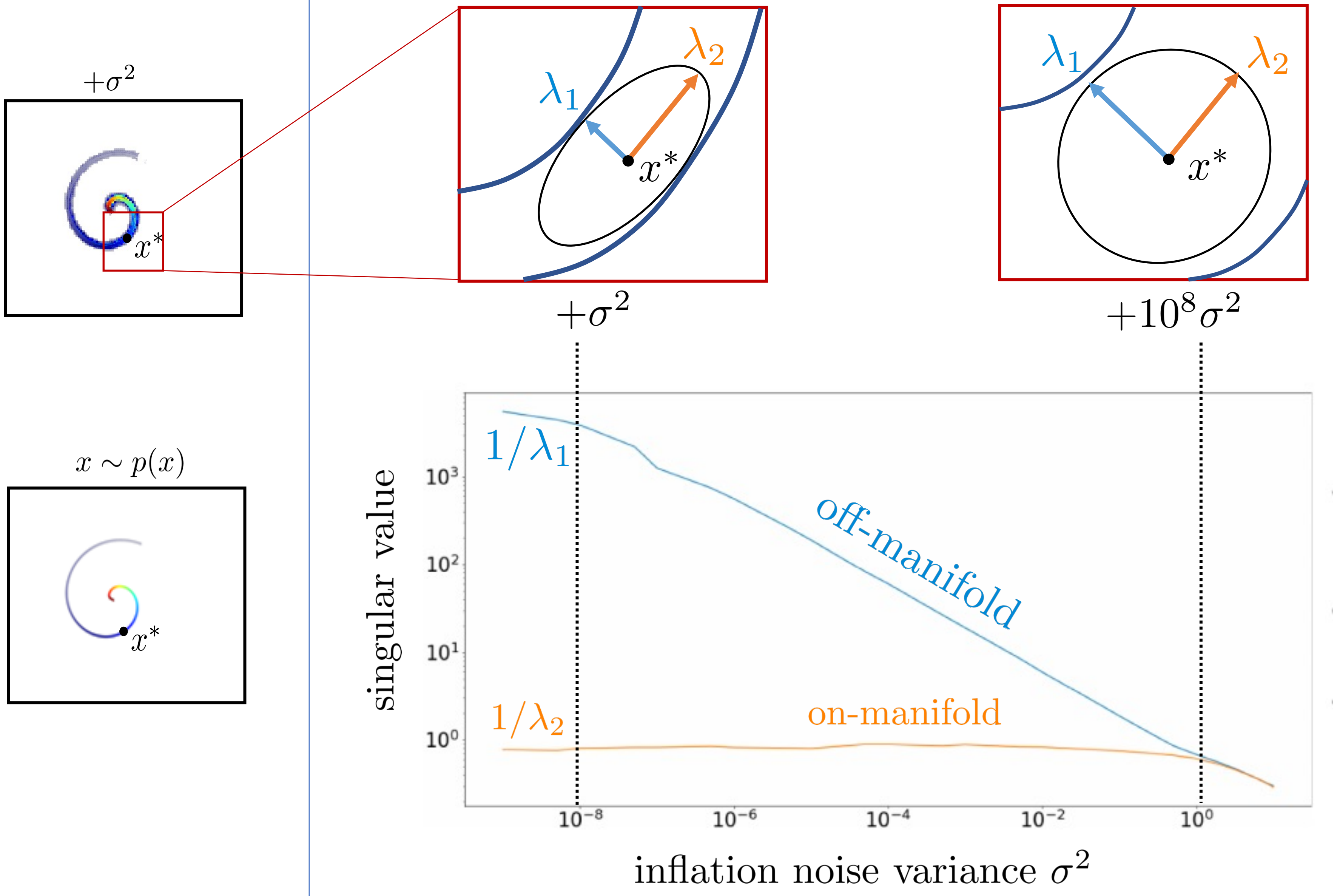
f^{-1}...Normalizing Flow



Singular values of $J_{f^{-1}}(x^*)$

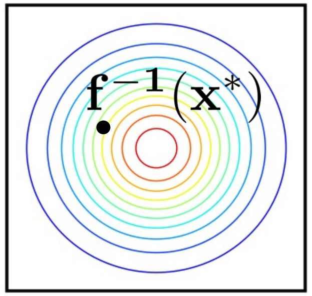
$\frac{1}{\lambda_1}$ and $\frac{1}{\lambda_2}$

Method



learn noisy $p(x)$ using

f^{-1}...Normalizing Flow

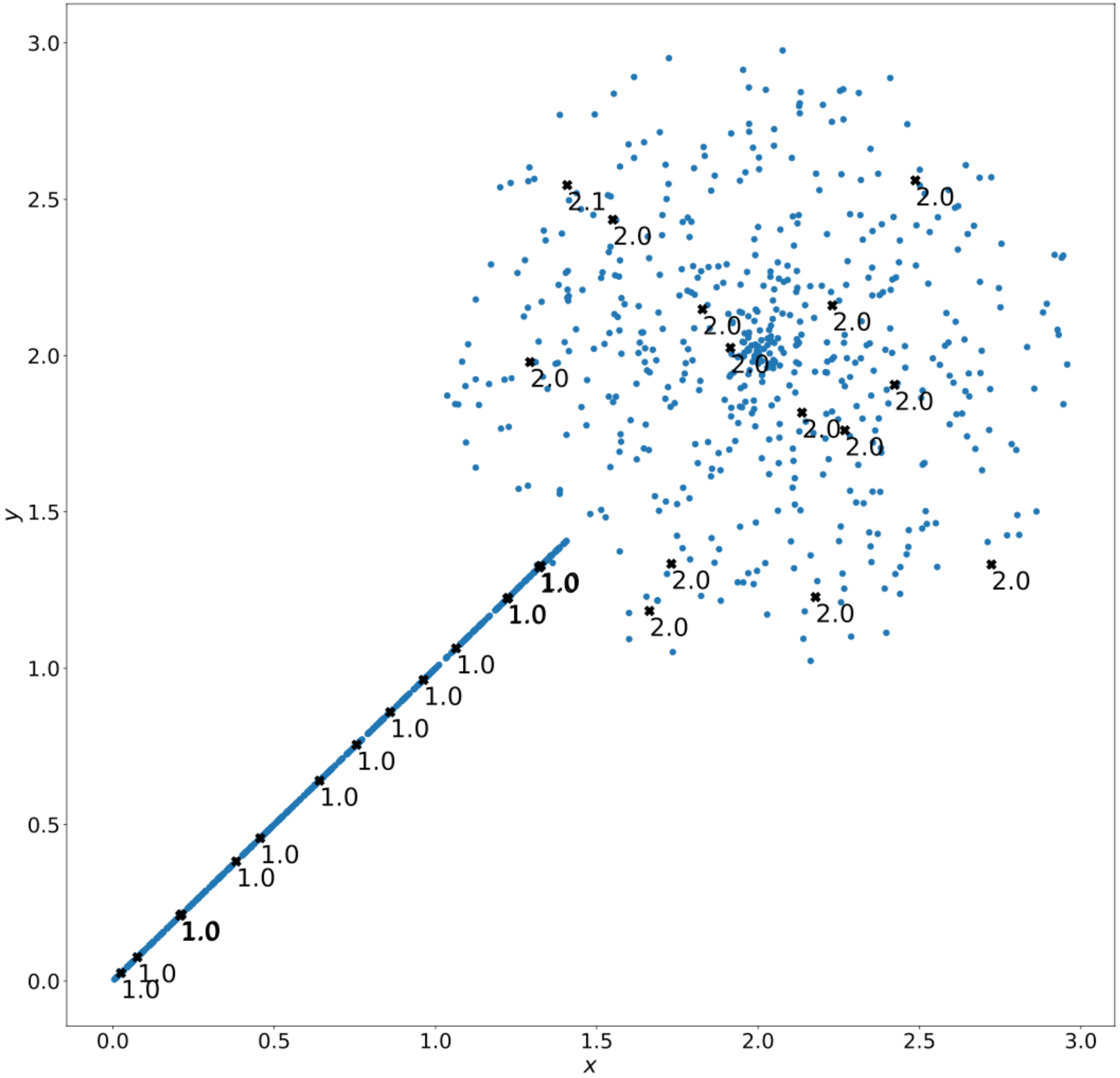


Singular values of $J_{f^{-1}}(x^*)$

$\frac{1}{\lambda_1}$ and $\frac{1}{\lambda_2}$

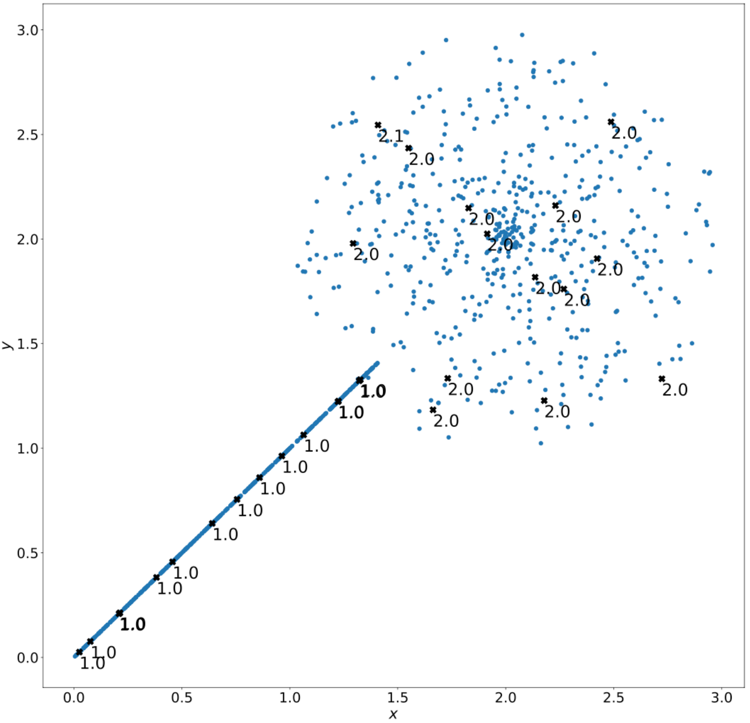
on-manifold
and
off-manifold
evolve differently
with σ^2

Results - toysets



Results - toysets

our method

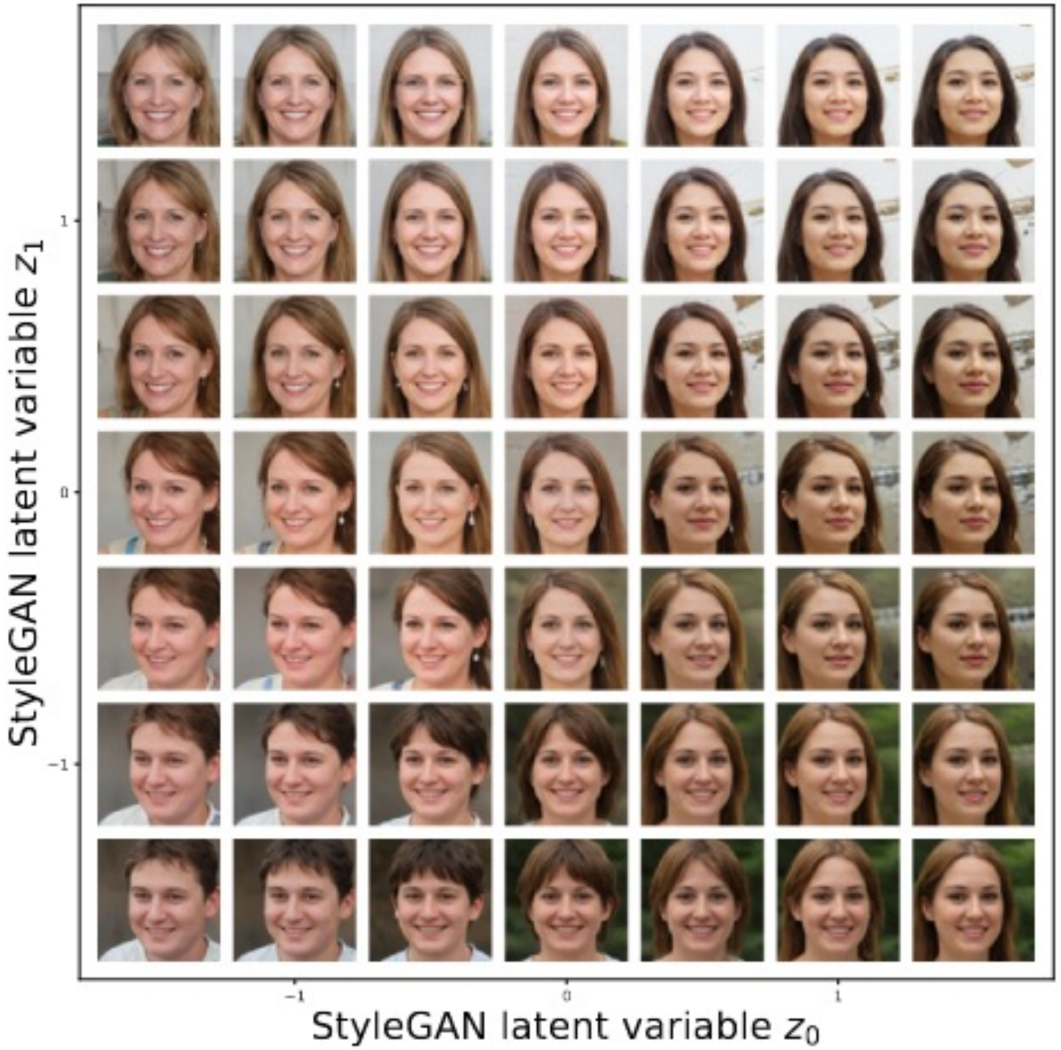


Distribution	D	ID	ID-NF	LIDL	twoNN
mixture on sphere	3	2	2.01	1.91±0.06	1.98
correlated on sphere	3	2	2.04	1.66±0.07	1.99
mixture on torus	3	2	2.02	2.05±0.04	1.97
correlated on torus	3	2	2.02	2.07±0.05	2.02
correlated on hyperboloid	3	2	2.02	2.01±0.07	1.99
unimodal on hyperboloid	3	2	2.01	1.92±0.1	1.96
exponential on thin spiral	2	1	1	1.08±0.06	1
mixture on swiss roll	3	2	2.02	2.26±0.03	1.98
correlated on swiss roll	3	2	2.02	2.48±0.03	1.94
mixture on stiefel	4	1	1.07	1.19±0.01	0.99

Local and global intrinsinc dimensionality estimator

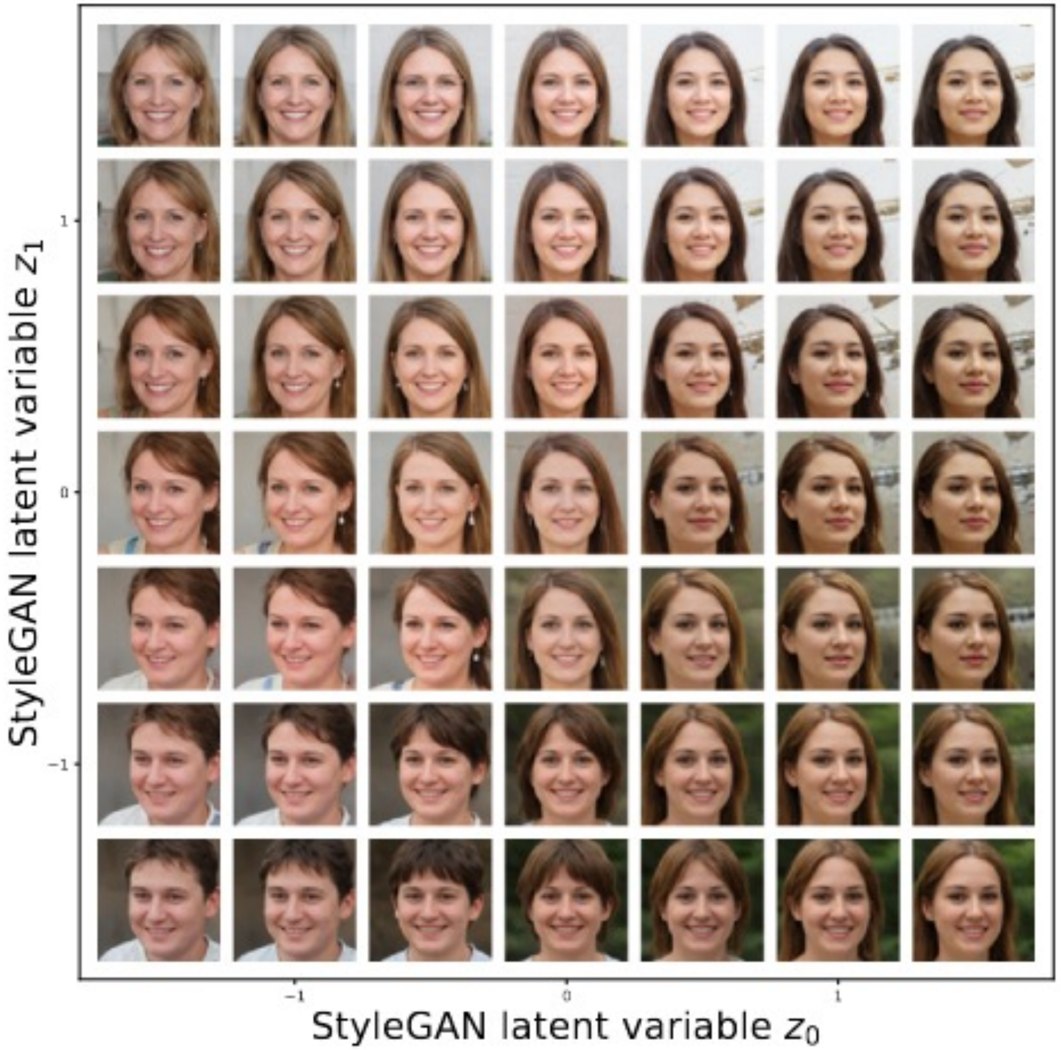
Results - images

$$x \in [0,1]^D, \quad D = 3 \cdot 255^2 = 12288$$



Results - images

$$x \in [0,1]^D, \quad D = 3 \cdot 255^2 = 12288$$



Datasets	ID
StyleGan2d	4.06 ± 1.75
StyleGan64d	62.24 ± 18.64

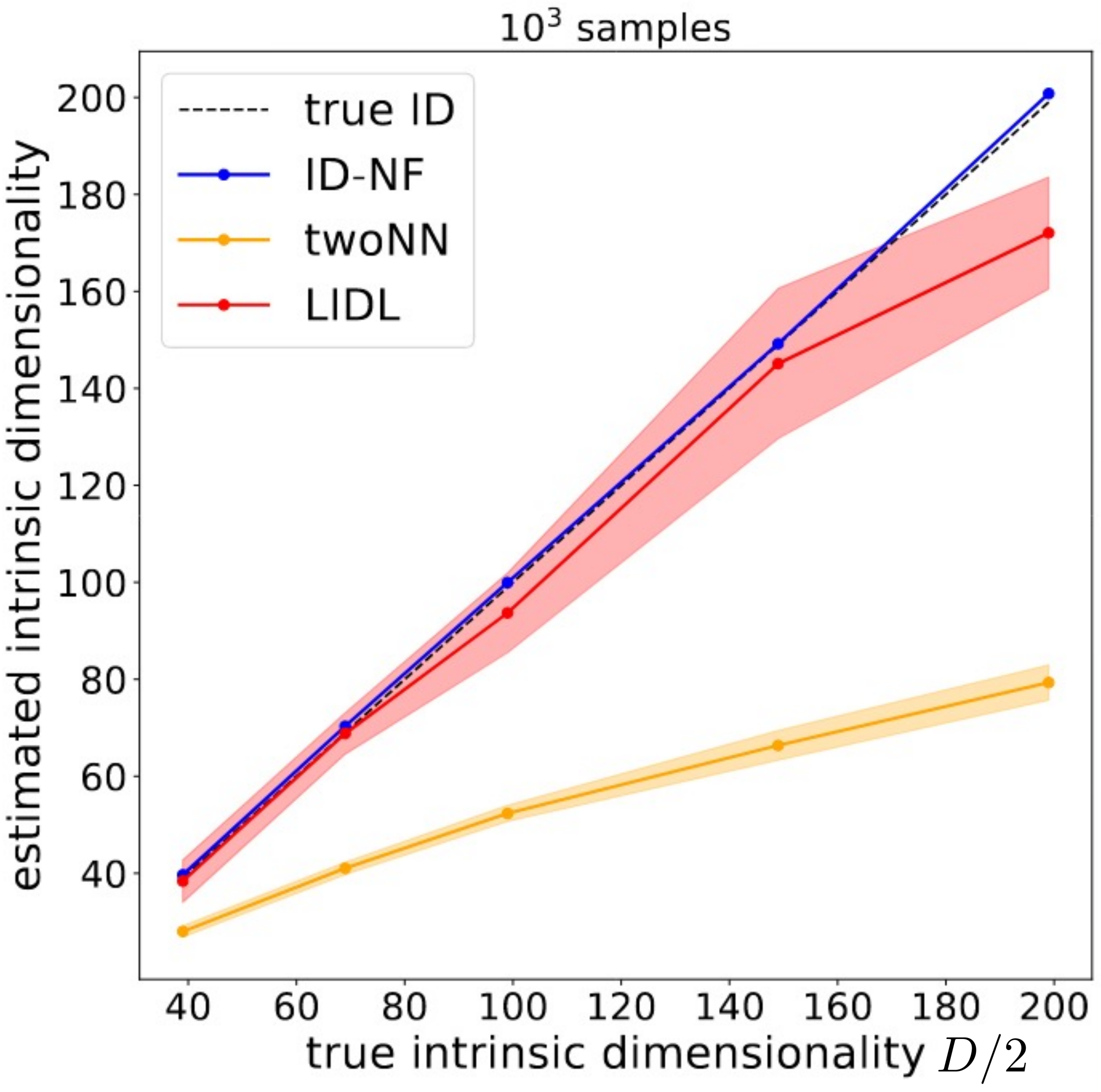
First good estimator for images!

Take home messages

- We developed a novel algorithm to estimate the intrinsic dimensionality based on Normalizing Flows
- on-manifold singular values of the flow's Jacobian evolve differently than the off-manifold singular values as a function of the inflation noise
- We found that our method correctly estimates the intrinsic dimensionality of all the tested toy models (locally and globally)
- It is the first good dimensionality estimator for images

Results - toysets

Sphere $S(D/2)$ embedded in $\mathbb{R}^D$



Why should you read the paper

- Generelizations:
  - Singular values are directions of large and small variability (Lemma 1)
  - Algorithm how to estimate d
  - method for unbounded data (such as images)
- Improving denoising normalizing flow (DNF) → same generative performance with significantly fewer latent variables
- ID for OOD samples is significantly higher
- More experiments, Discussions, limitations, open questions