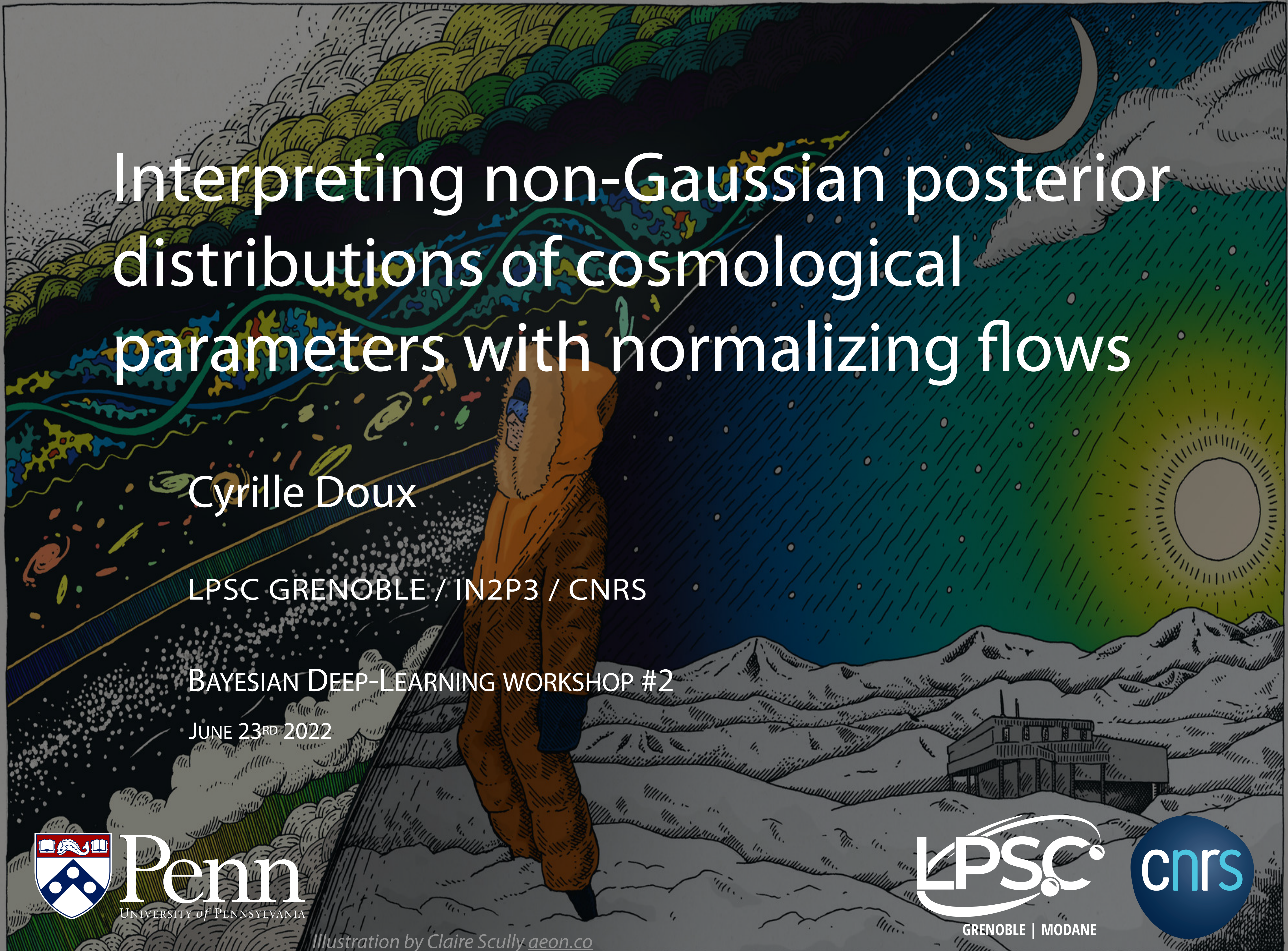


Interpreting non-Gaussian posterior distributions of cosmological parameters with normalizing flows

Cyrille Doux

LPSC GRENOBLE / IN2P3 / CNRS

BAYESIAN DEEP-LEARNING WORKSHOP #2

JUNE 23RD 2022

Penn
UNIVERSITY of PENNSYLVANIA

LPSC GRENOBLE | MODANE

cnrs

Illustration by Claire Scully aeon.co

Collaborators

Marco Raveri

Tara Dacuhna

Minsu Park

Bhuvnesh Jain

arXiv:2105.03324
10.1103/PhysRevD.104.043504
Non-Gaussian estimates of tensions in cosmological parameters
Authors: Marco Raveri, Cyrille Doux

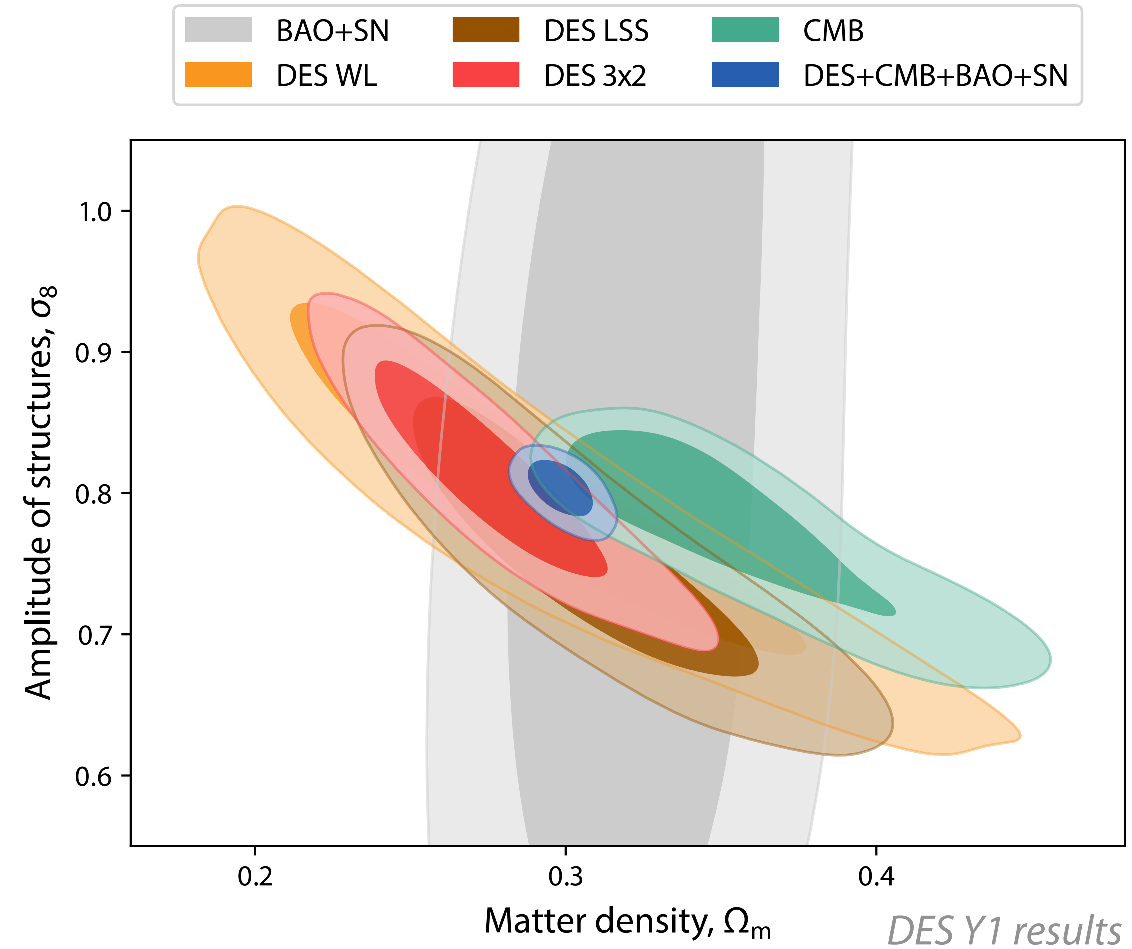
arXiv:2112.05737
10.1103/PhysRevD.105.063529
What does a cosmological experiment really measure? Covariant posterior decomposition with normalizing flows
Authors: Tara Dacunha, Marco Raveri, Minsu Park, Cyrille Doux, Bhuvnesh Jain

Penn
University of Pennsylvania

1. Non-Gaussian tension metrics for cosmology

ΛCDM and cosmic shear

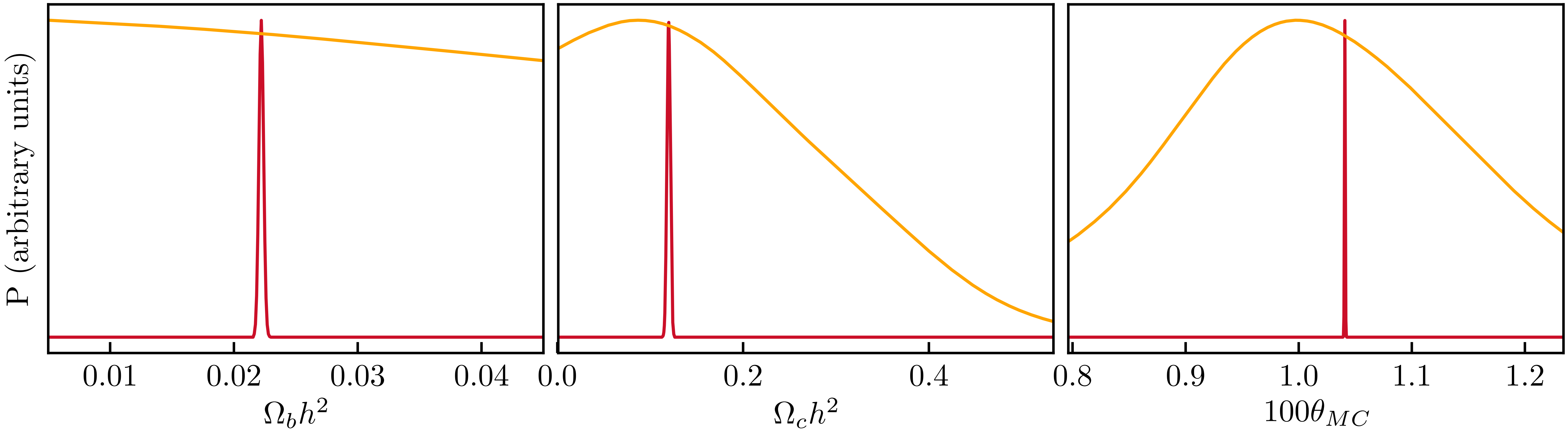
- **Tensions in current ΛCDM paradigm on H_0, σ_8**
  1. Early (CMB) *vs* late Universe (BAO, SNIa, LSS+WL)
  2. Geometry *vs* growth, aka background *vs* structure
- **Weak lensing by large-scale structure**
  - Ongoing precursor surveys : **Dark Energy Survey**, HSC, KiDS
  - Next-gen surveys : Rubin/LSST, Euclid, Roman



DES Y1 results

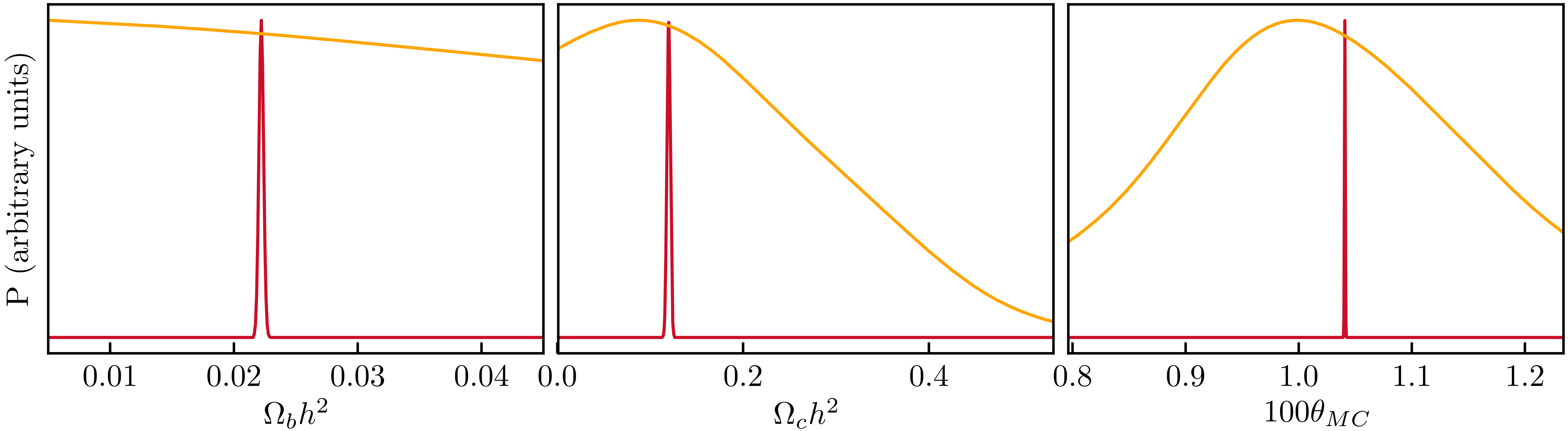
Agreement or disagreement?

Do these agree?

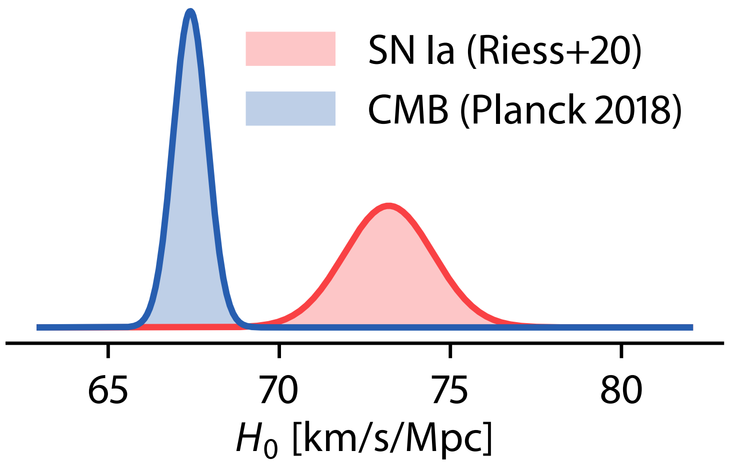


Agreement or disagreement?

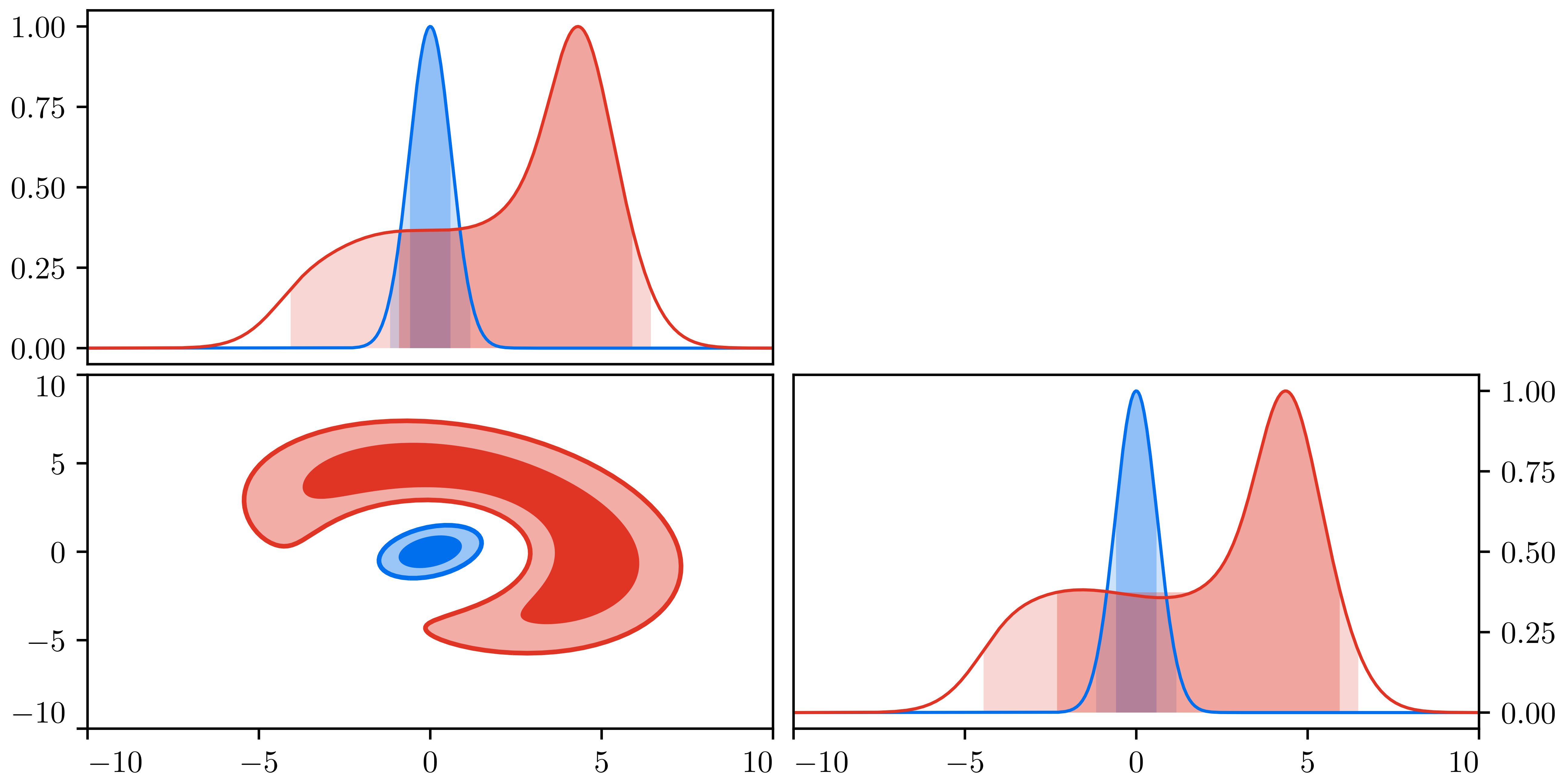
Do these agree?



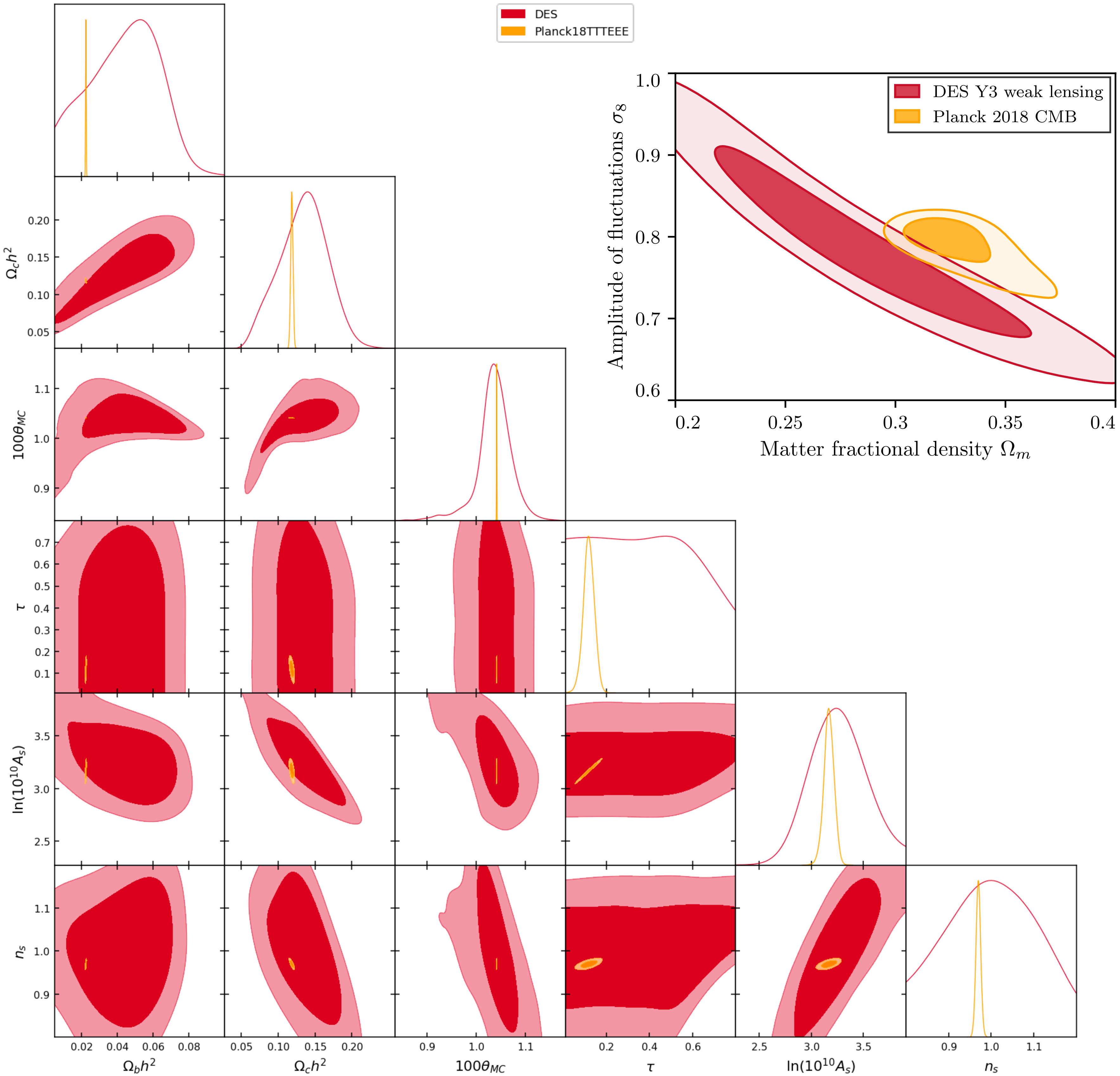
Nope, it's Planck and local Hubble constant measurements, which disagree at the 5σ level !



What's going on?



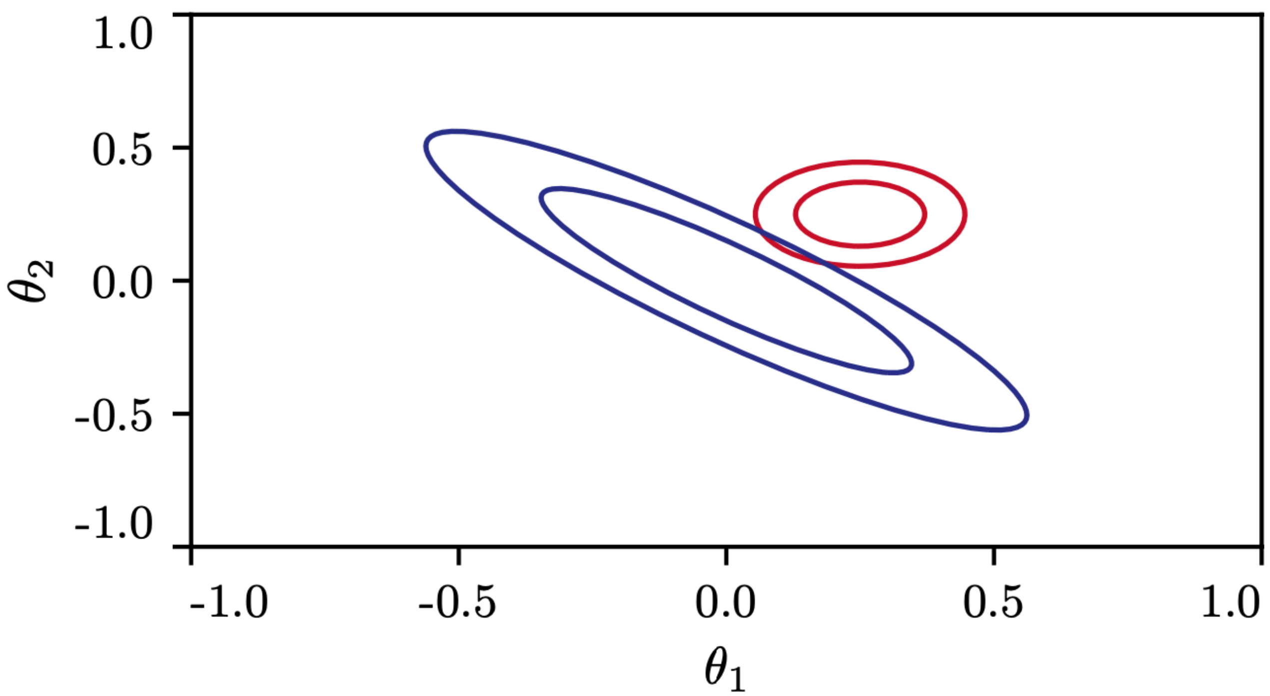
In practice : DES *vs* Planck

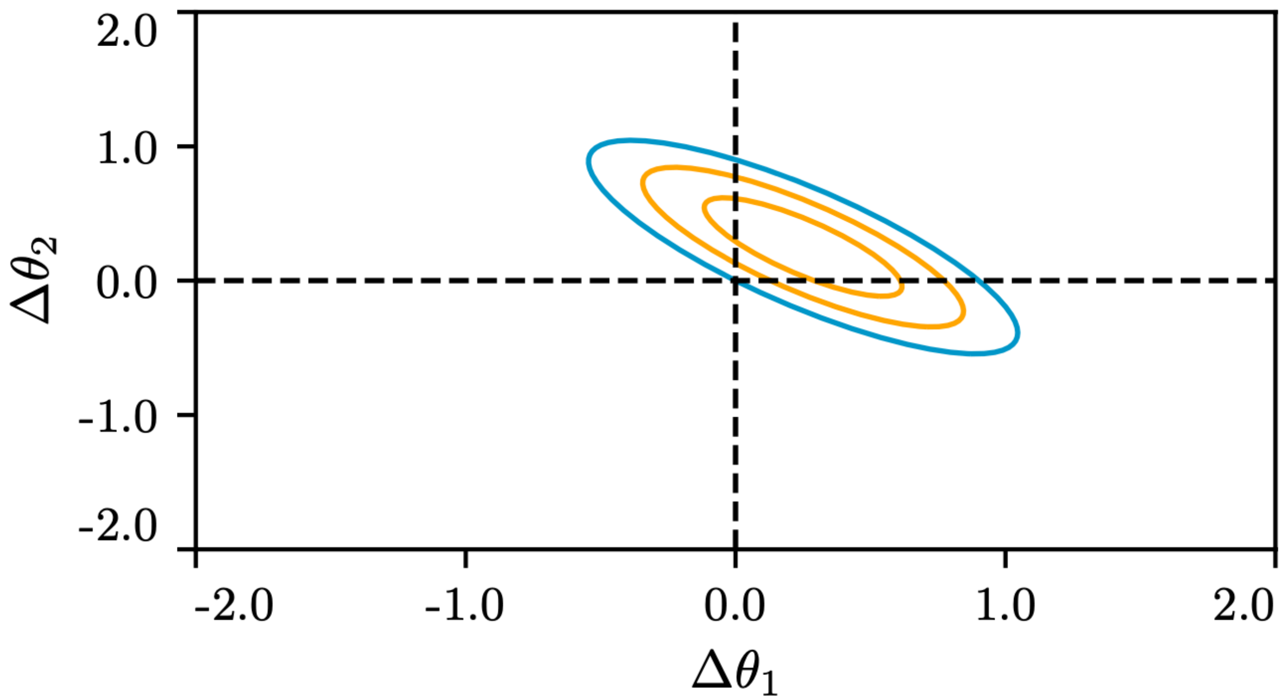


- **Why is it difficult?**
  - Data complexity: soon $O(10^4)$ data points
  - Model complexity: $O(50)$ parameters
  - Marginals *vs* full distribution
  - Projection effects (unconstrained params)
  - Gaussian approximation can be inadequate

Parameter differences

arXiv:2105.03324



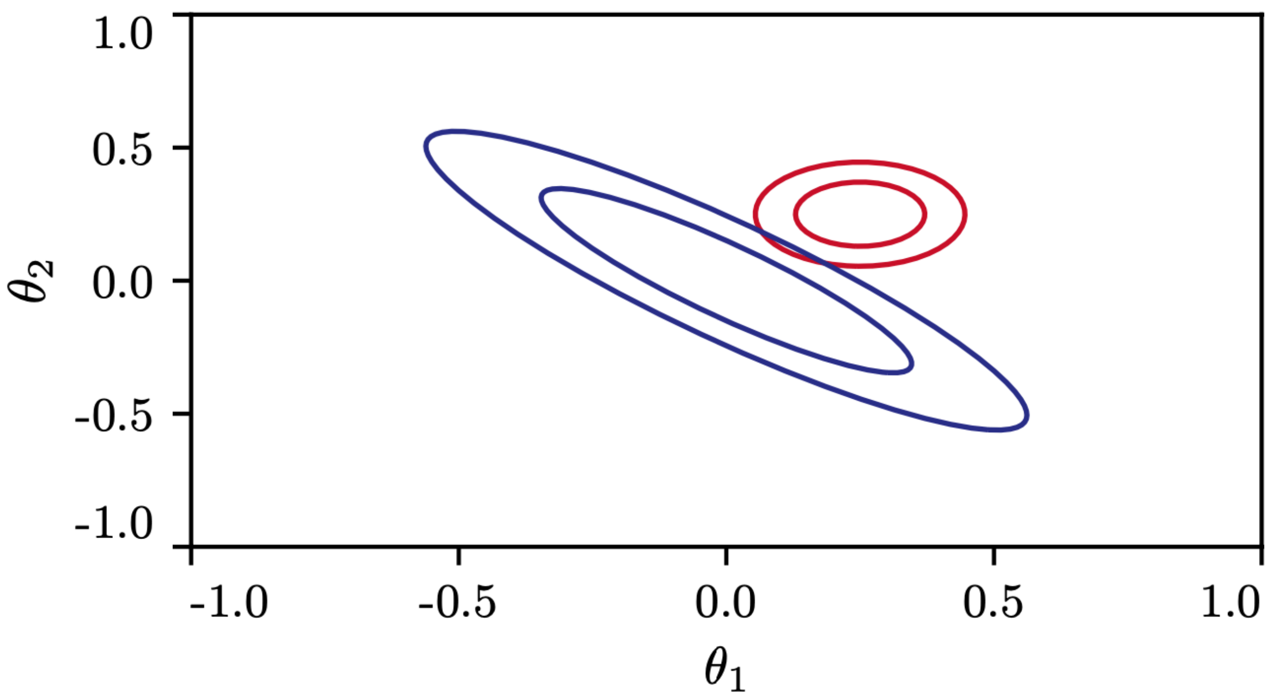


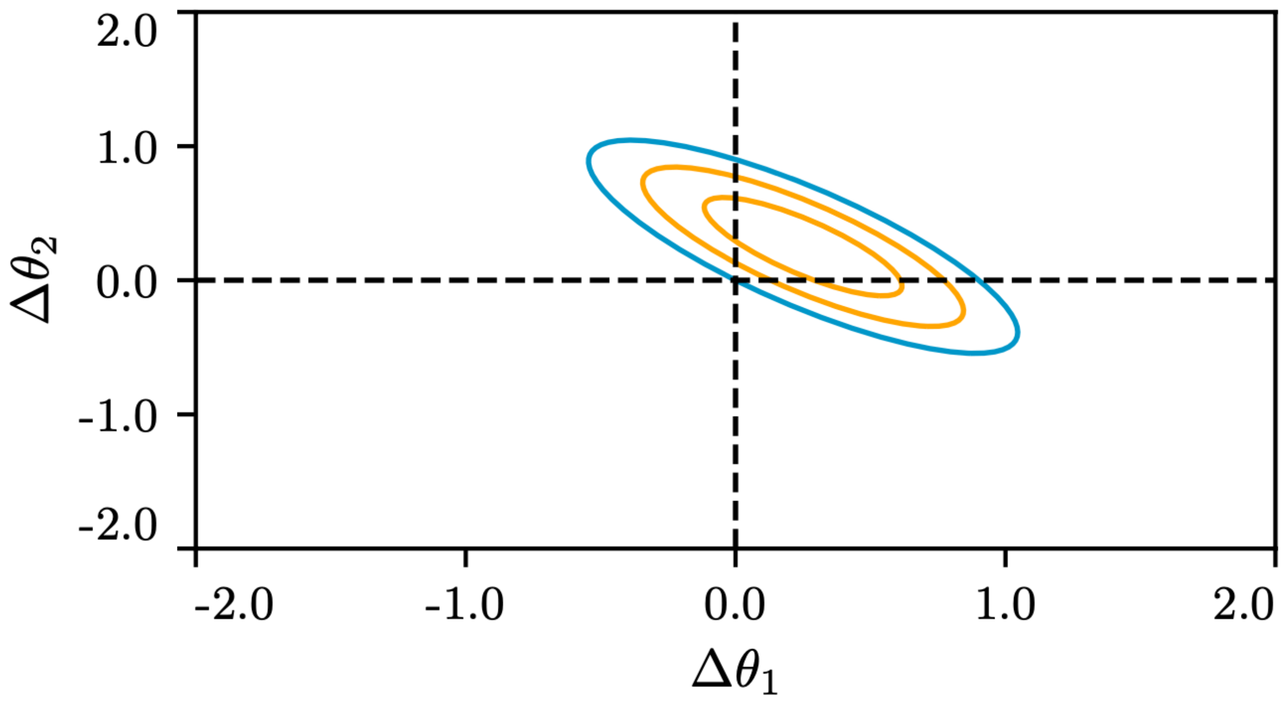
- **Distribution of parameter differences** for independent experiments 1 and 2

$$\mathcal{P}(\Delta\theta) = \int_{V_\pi} \mathcal{P}_1(\theta)\mathcal{P}_2(\theta - \Delta\theta)\,\mathrm{d}\theta$$

Parameter differences

arXiv:2105.03324





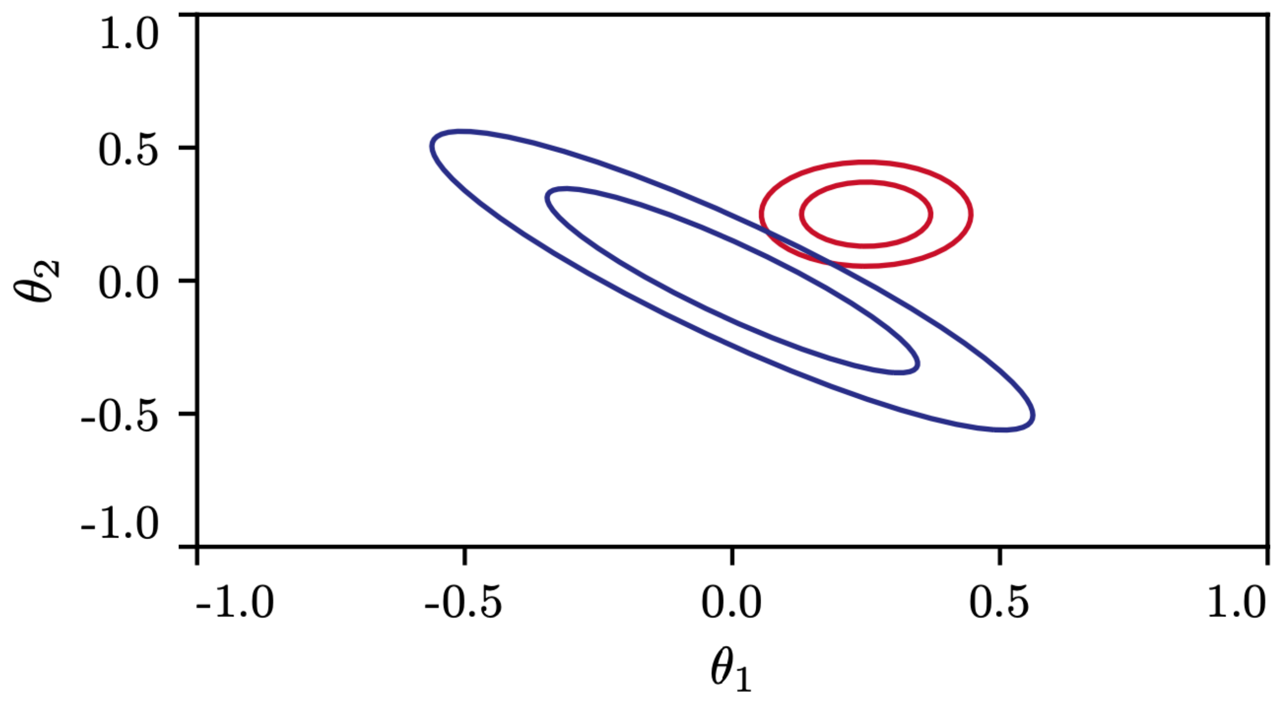
- **Distribution of parameter differences** for independent experiments 1 and 2

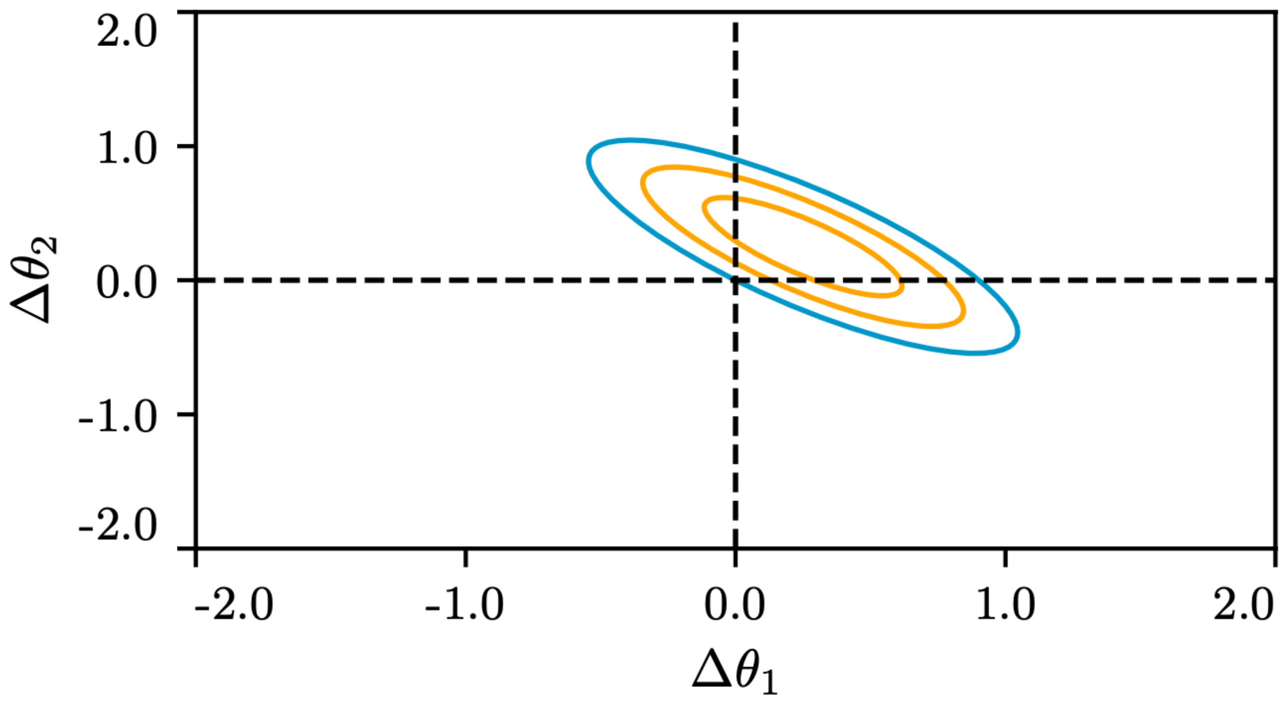
$$\mathcal{P}(\Delta\theta) = \int_{V_\pi} \mathcal{P}_1(\theta)\mathcal{P}_2(\theta - \Delta\theta)\, \mathrm{d}\theta$$

We have samples!

Parameter differences

arXiv:2105.03324





- **Distribution of parameter differences** for independent experiments 1 and 2

$$\mathcal{P}(\Delta\theta) = \int_{V_\pi} \mathcal{P}_1(\theta)\mathcal{P}_2(\theta - \Delta\theta)\, \mathrm{d}\theta$$

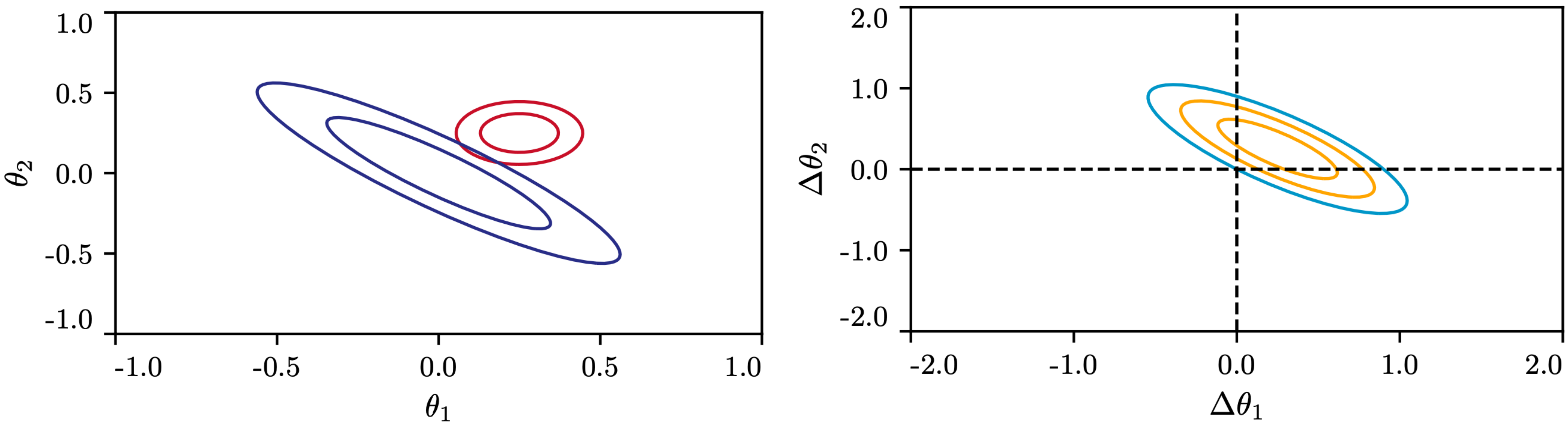
We have samples!

- Define a **tension metric**

$$\Delta = \int_{\mathcal{P}(\Delta\theta)>\mathcal{P}(\mathbf{0})} \mathcal{P}(\Delta\theta)\, \mathrm{d}\Delta\theta$$

Parameter differences

arXiv:2105.03324



- **Distribution of parameter differences** for independent experiments 1 and 2

$$\mathcal{P}(\Delta\theta) = \int_{V_\pi} \mathcal{P}_1(\theta)\mathcal{P}_2(\theta - \Delta\theta)\, \mathrm{d}\theta$$

We have samples!

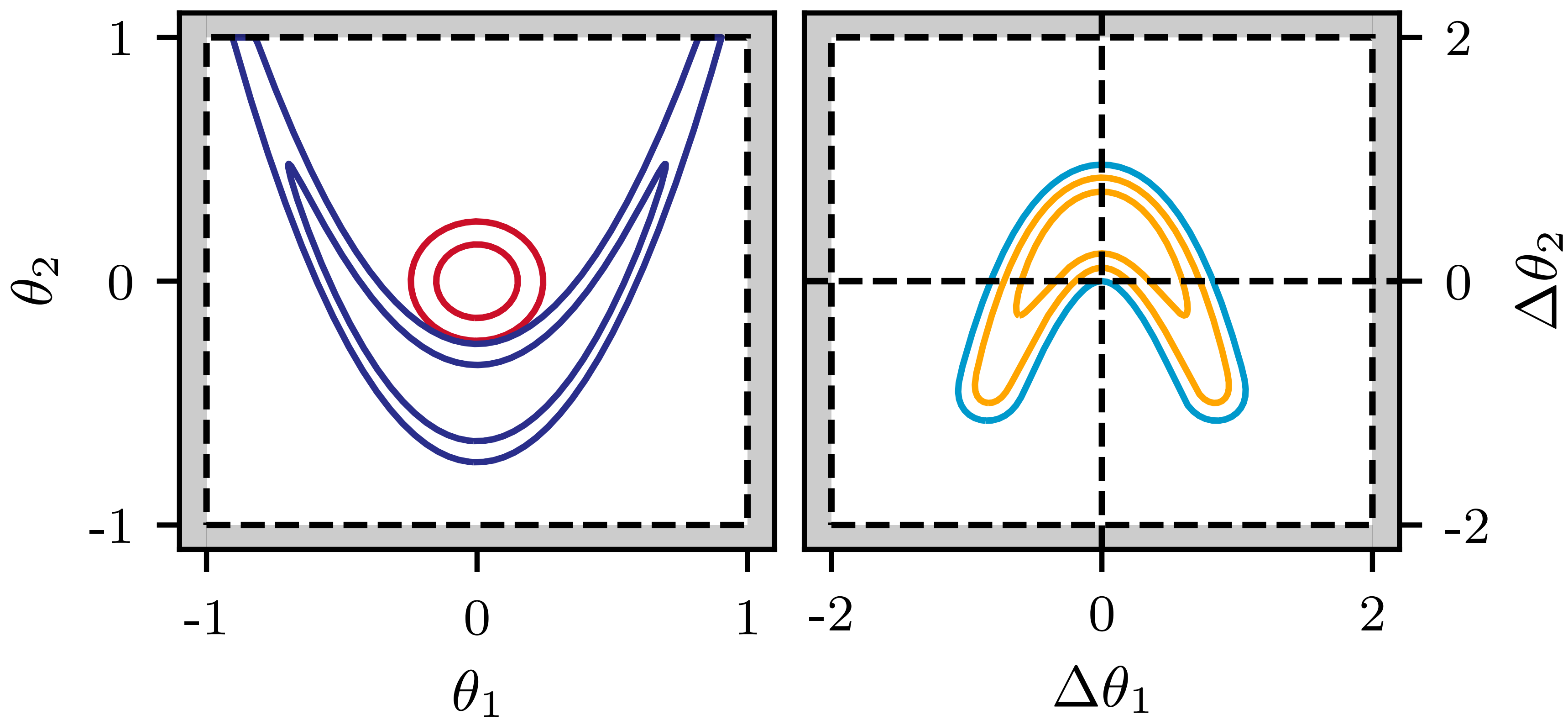
- Define a **tension metric**

$$\Delta = \int_{\mathcal{P}(\Delta\theta)>\mathcal{P}(\mathbf{0})} \mathcal{P}(\Delta\theta)\, \mathrm{d}\Delta\theta$$

Wait a minute! These are two high-dimensional integrals!

Parameter differences

arXiv:2105.03324



- **Distribution of parameter differences** for independent experiments 1 and 2

$$\mathcal{P}(\Delta\theta) = \int_{V_\pi} \mathcal{P}_1(\theta)\mathcal{P}_2(\theta - \Delta\theta)\,\mathrm{d}\theta$$

We have samples!

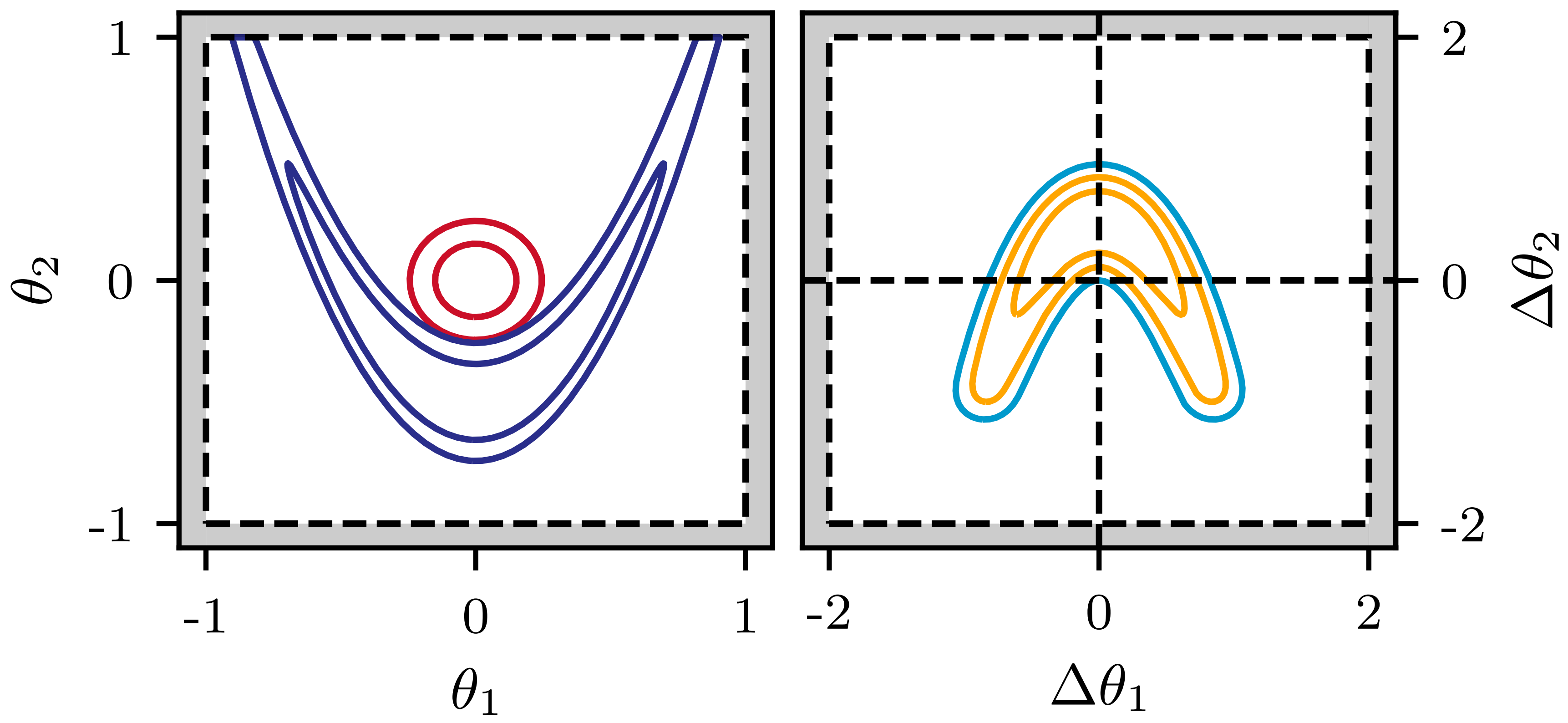
- Define a **tension metric**

$$\Delta = \int_{\mathcal{P}(\Delta\theta)>\mathcal{P}(\mathbf{0})} \mathcal{P}(\Delta\theta)\,\mathrm{d}\Delta\theta$$

Wait a minute! These are two high-dimensional integrals!

Parameter differences

arXiv:2105.03324



- **Distribution of parameter differences** for independent experiments 1 and 2

$$\mathcal{P}(\Delta\theta) = \int_{V_\pi} \mathcal{P}_1(\theta)\mathcal{P}_2(\theta - \Delta\theta)\,\mathrm{d}\theta$$

We have samples!

- Define a **tension metric**

$$\Delta = \int_{\mathcal{P}(\Delta\theta) > \mathcal{P}(\mathbf{0})} \mathcal{P}(\Delta\theta)\,\mathrm{d}\Delta\theta$$

- The solution? *Learn* the distribution, i.e. find a way to
  - 1) **sample** from and 2) **compute the density** of the distribution

MAF normalizing flows

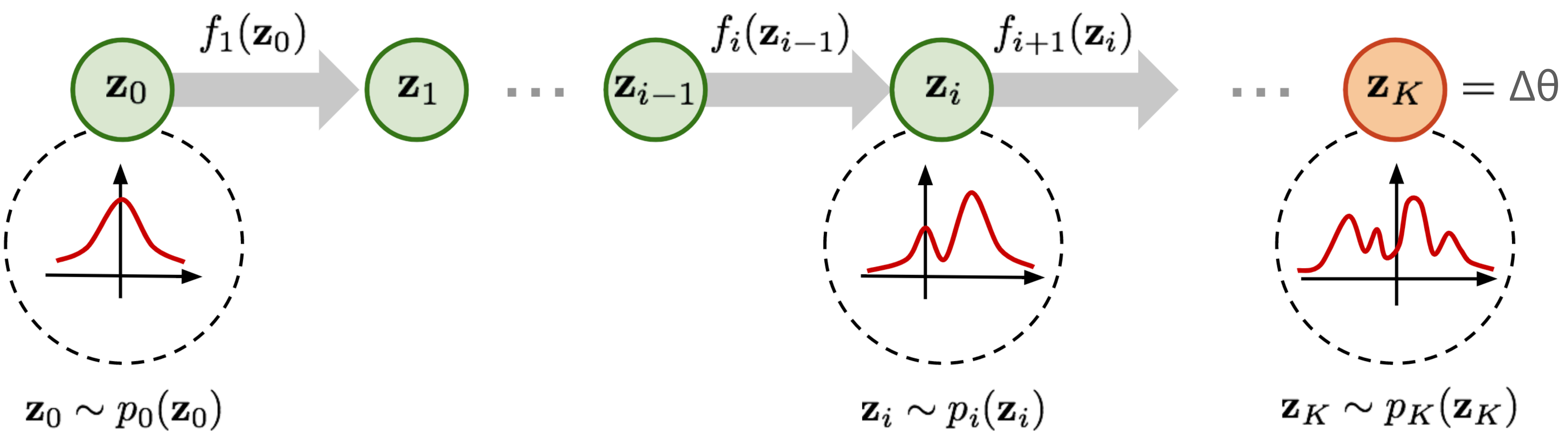
- **Normalizing flow**
  - An *invertible*, *differentiable* mapping $\Delta\theta \to z = T_\varphi(\Delta\theta)$ where we know the distribution of z

$$\mathcal{P}(\Delta\theta) \approx q(\Delta\theta) = \varphi_D(z)\left|\det\{\nabla_{\Delta\theta} T_\varphi(z)\}\right|$$

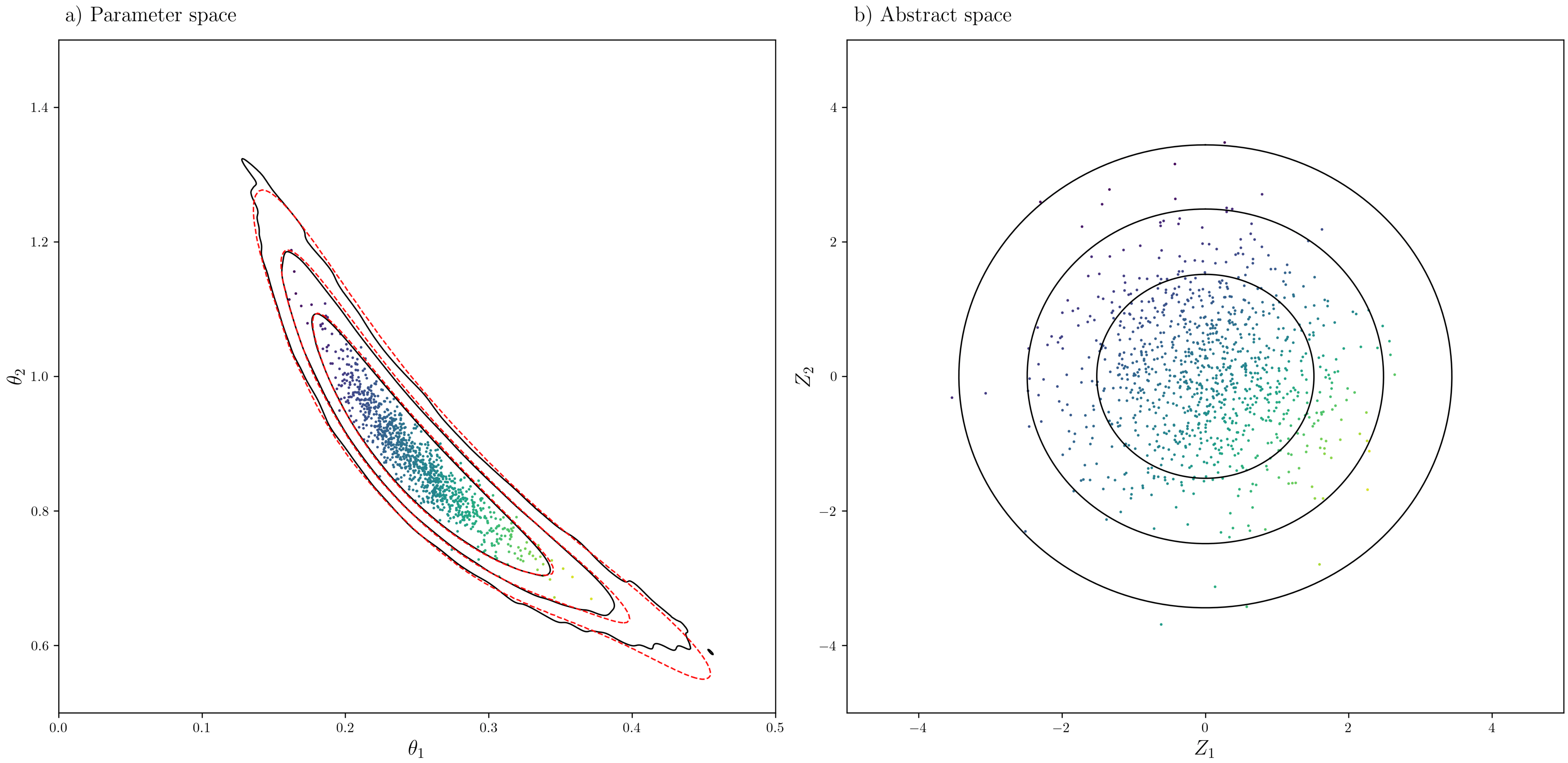
- **Masked Autoregressive Flows for density estimation**
  - Introduced by Papamakarios+17 [arXiv:1705.07057]
  - Autoregressive (*invertible*) transformations y ↦ z parametrised by *neural networks* μ and σ

$$y_1 = \mu_1 + \sigma_1 z_1$$
$$y_i = \mu(y_{1\ldots i-1}) + \sigma(y_{1\ldots i-1}) z_i$$

  - Stack a bunch of such transforms with random permutations of components



Normalizing flow training

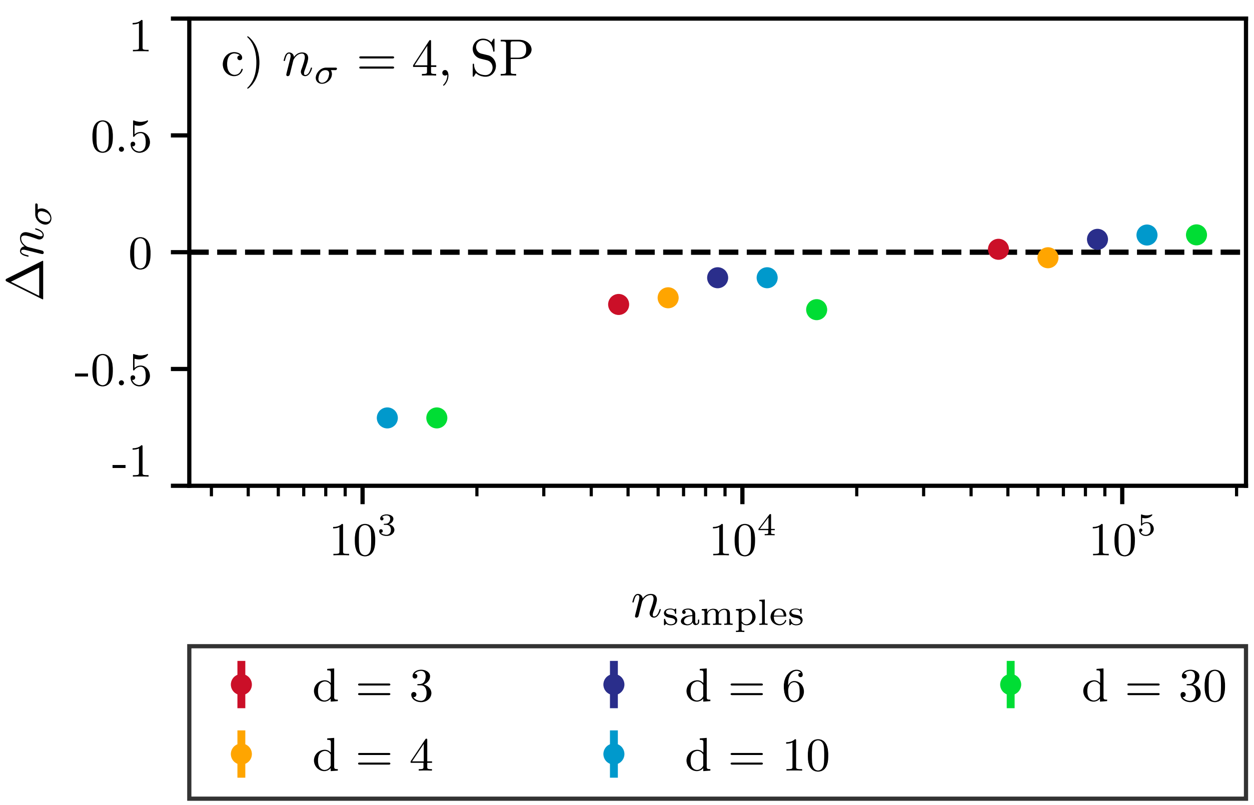


Benchmark tests

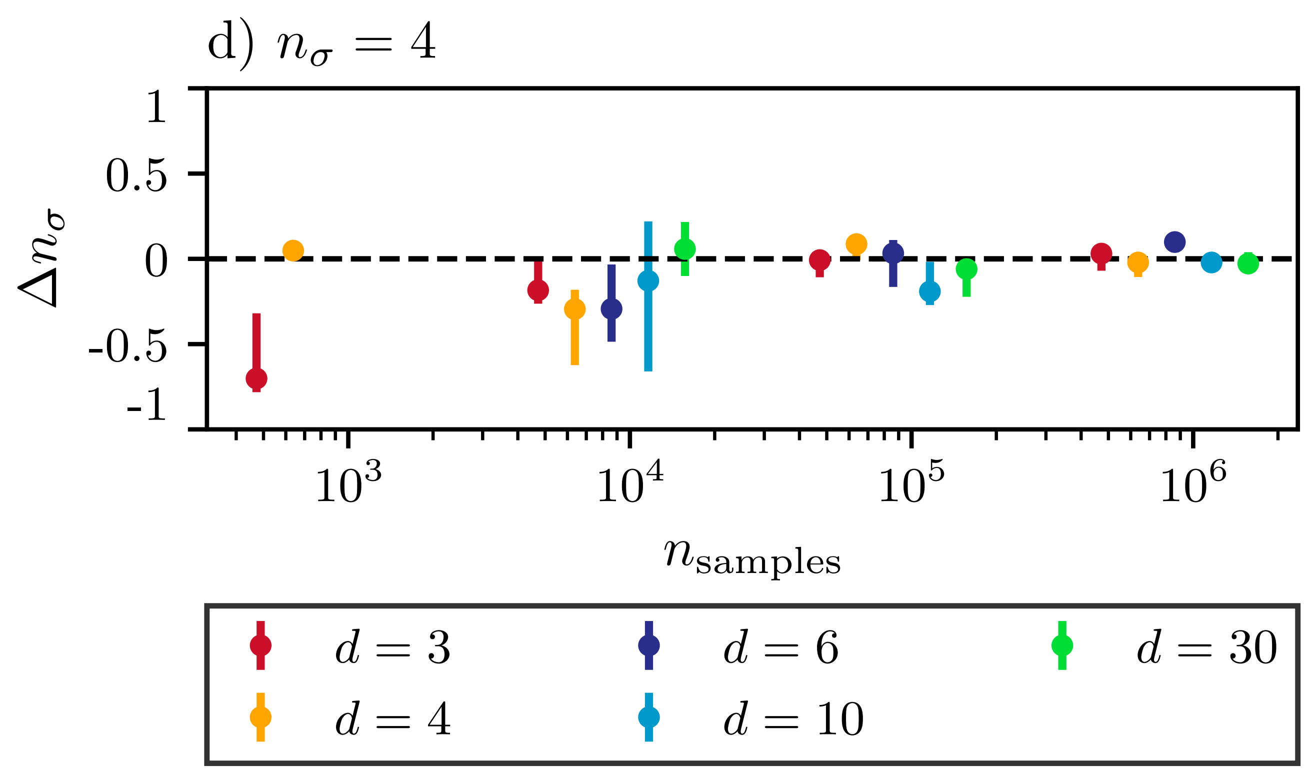
arXiv:2105.03324

- **We ran a whole bunch of tests with analytical distributions**
  - With complex non-Gaussian distributions in 2D where we can integrate numerically
  - With high-dimensional Gaussians where we have analytical results

KDE



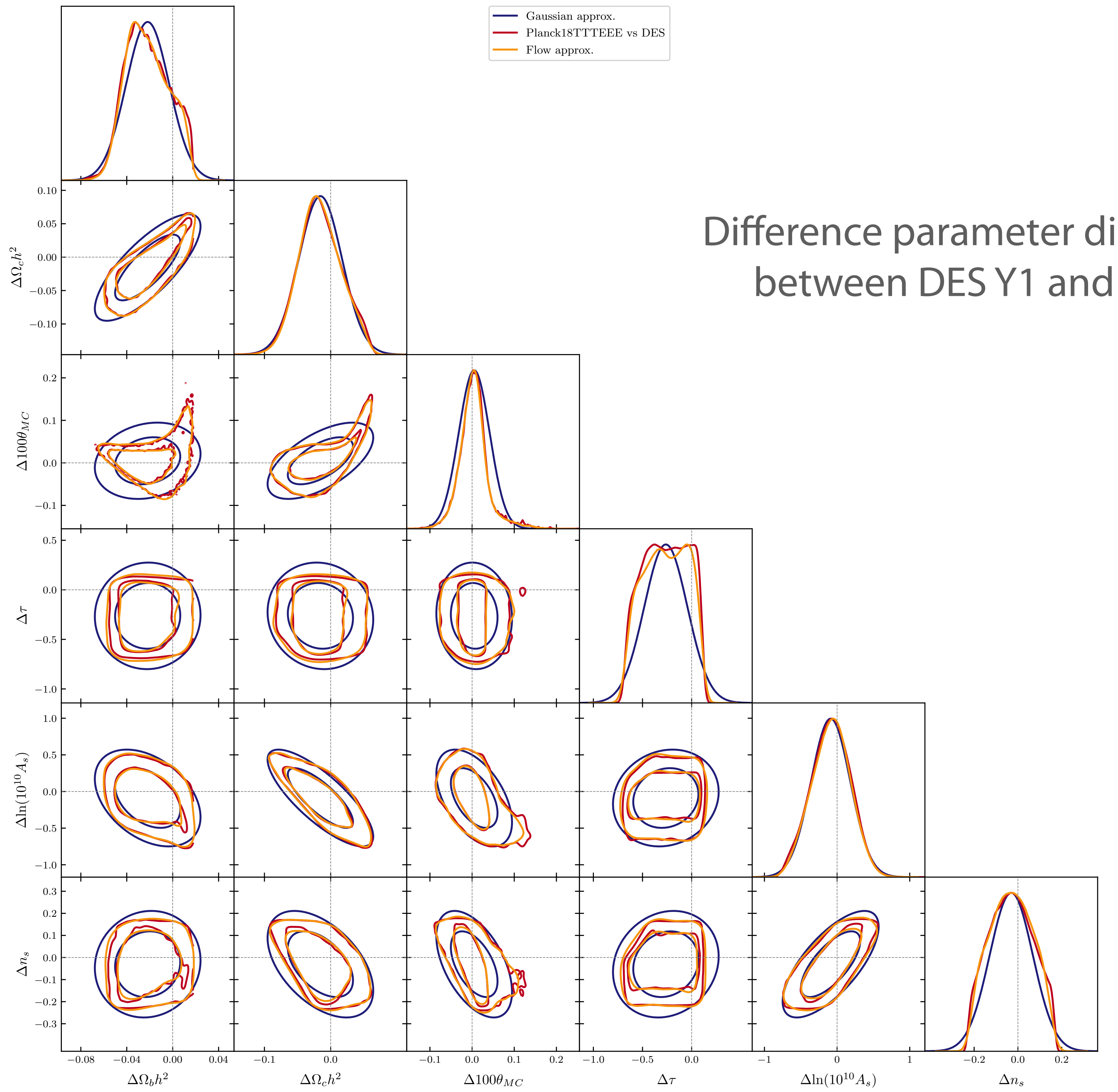
Flow



- Flows outperforms standard Monte Carlo estimations based on Kernel Density Estimates

Application to DES and Planck

arXiv:2105.03324

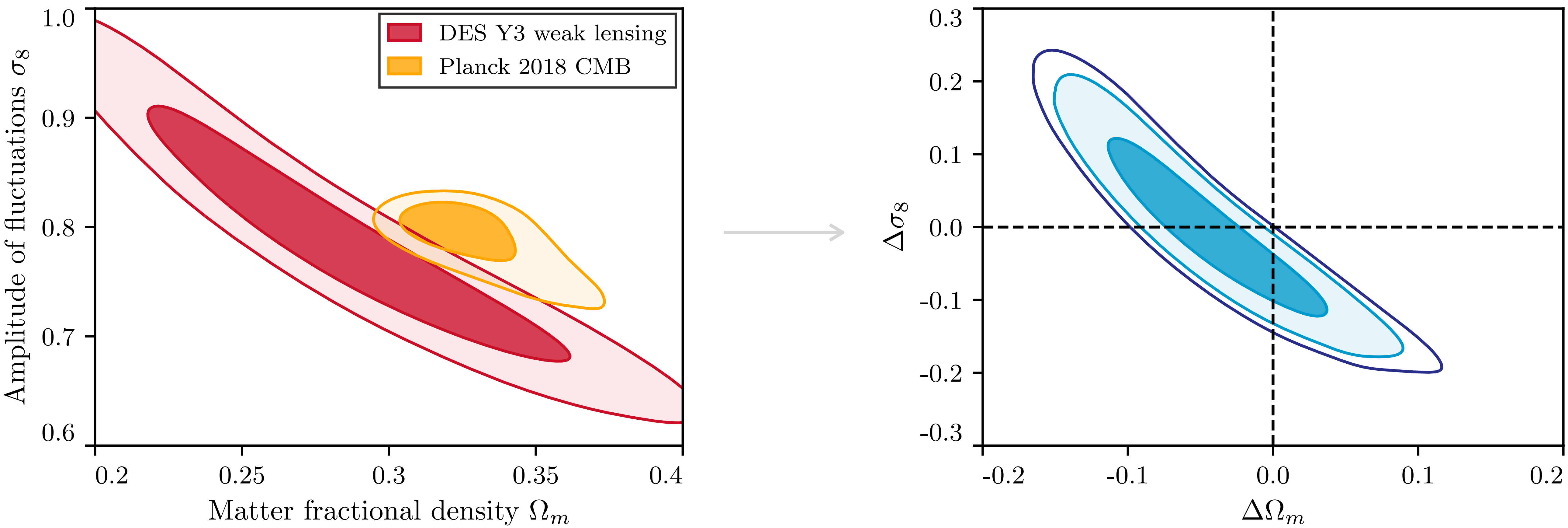


Difference parameter distribution between DES Y1 and Planck

Application to DES and Planck

arXiv:2105.03324

- **Application to DES Y1 vs Planck**
  - $n_\sigma = 3.0 \pm 0.1 \longrightarrow$ tension !



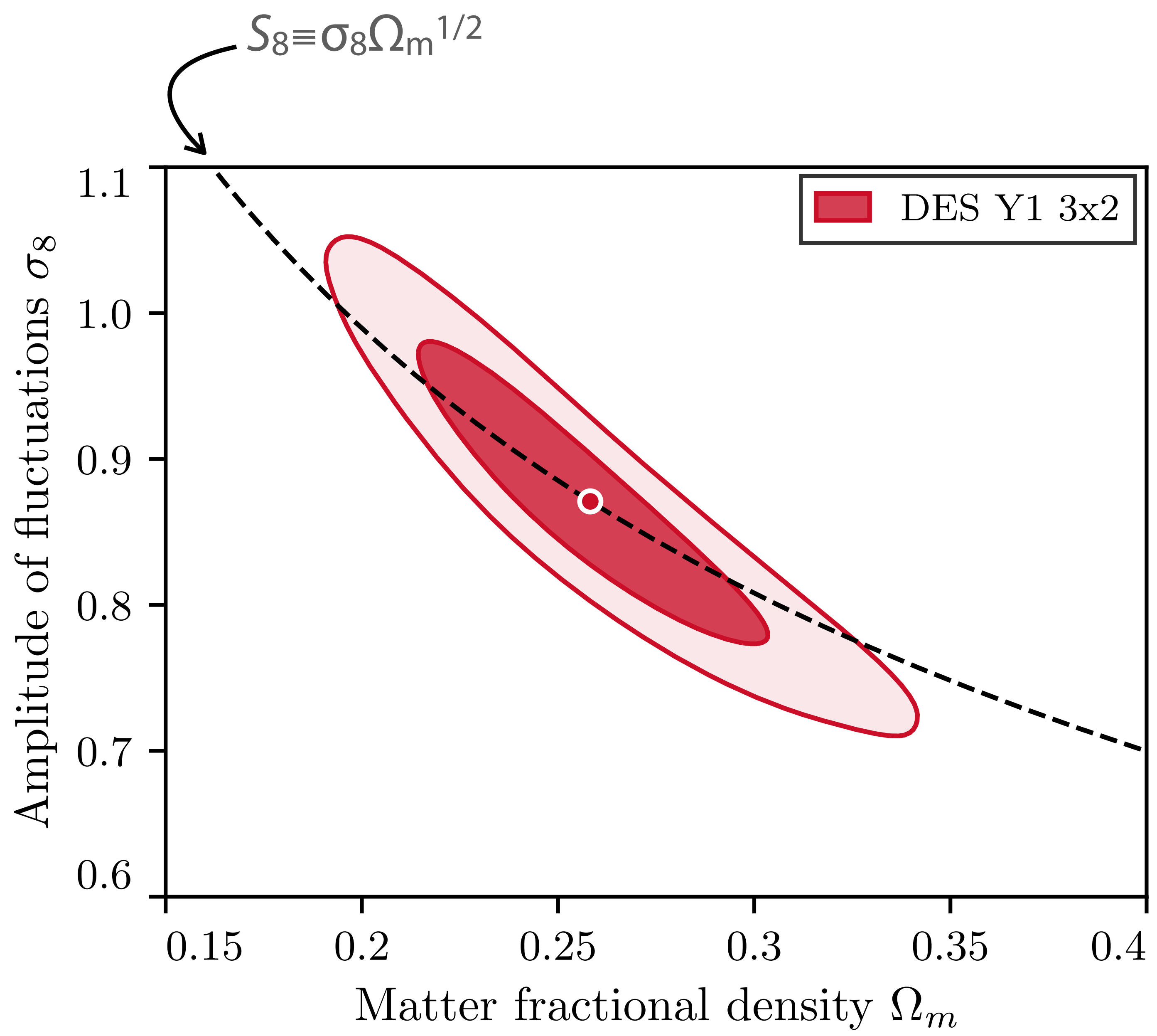
- **Application to DES Y1 lensing *vs* clustering**
  - $n_\sigma = 0.5 \pm 0.1 \longrightarrow$ internal agreement

2. What is an experiment really measuring?

What is an experiment measuring?

arXiv:2112.05737

- **Comparing results** from different experiments is tricky because they
    1. have different parameter *degeneracies*
    2. use different *parametrisation*, eg Ω_m *vs* $\omega_m \equiv \Omega_m h^2$
    3. use different *prior*(s)
- Can we automatically *learn* what experiments are measuring?



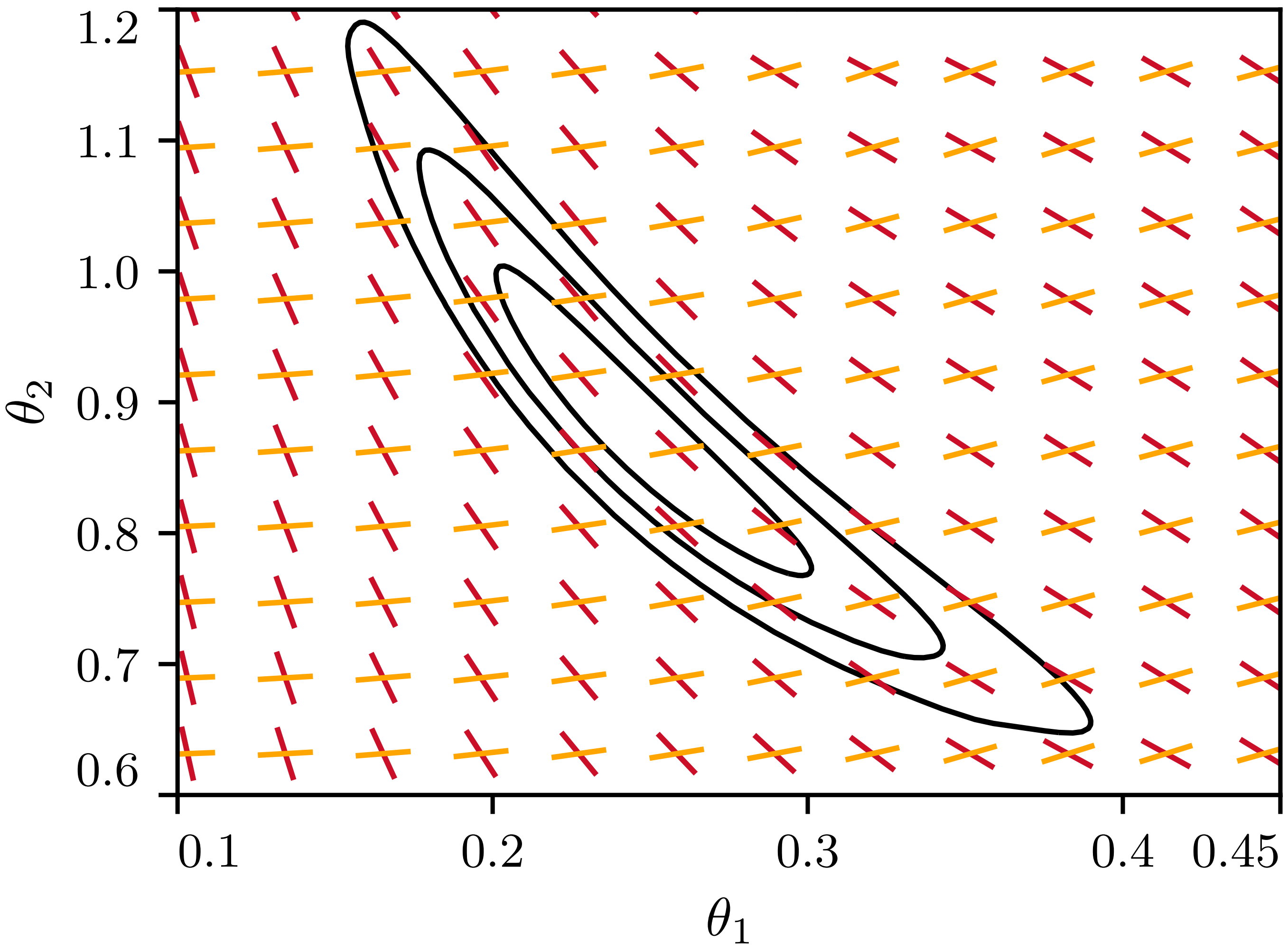
Information geometry with normalising flows

arXiv:2112.05737

- Local Fisher information matrix *

$$\mathcal{F}_{\mu\nu} = \left(\frac{\partial \phi^b}{\partial \theta^\nu}\right)^T \tilde{\mathcal{F}}_{ab} \frac{\partial \phi^a}{\partial \theta^\mu}$$

  - Defines a metric
  - NF+TFP to compute Jacobian



* (Fisher is *identity* everywhere in abstract space, ie both spaces are flat!)

* (Technically, we're forgetting about the data dependence here, but it matches in the Gaussian case, and it's way more practical)

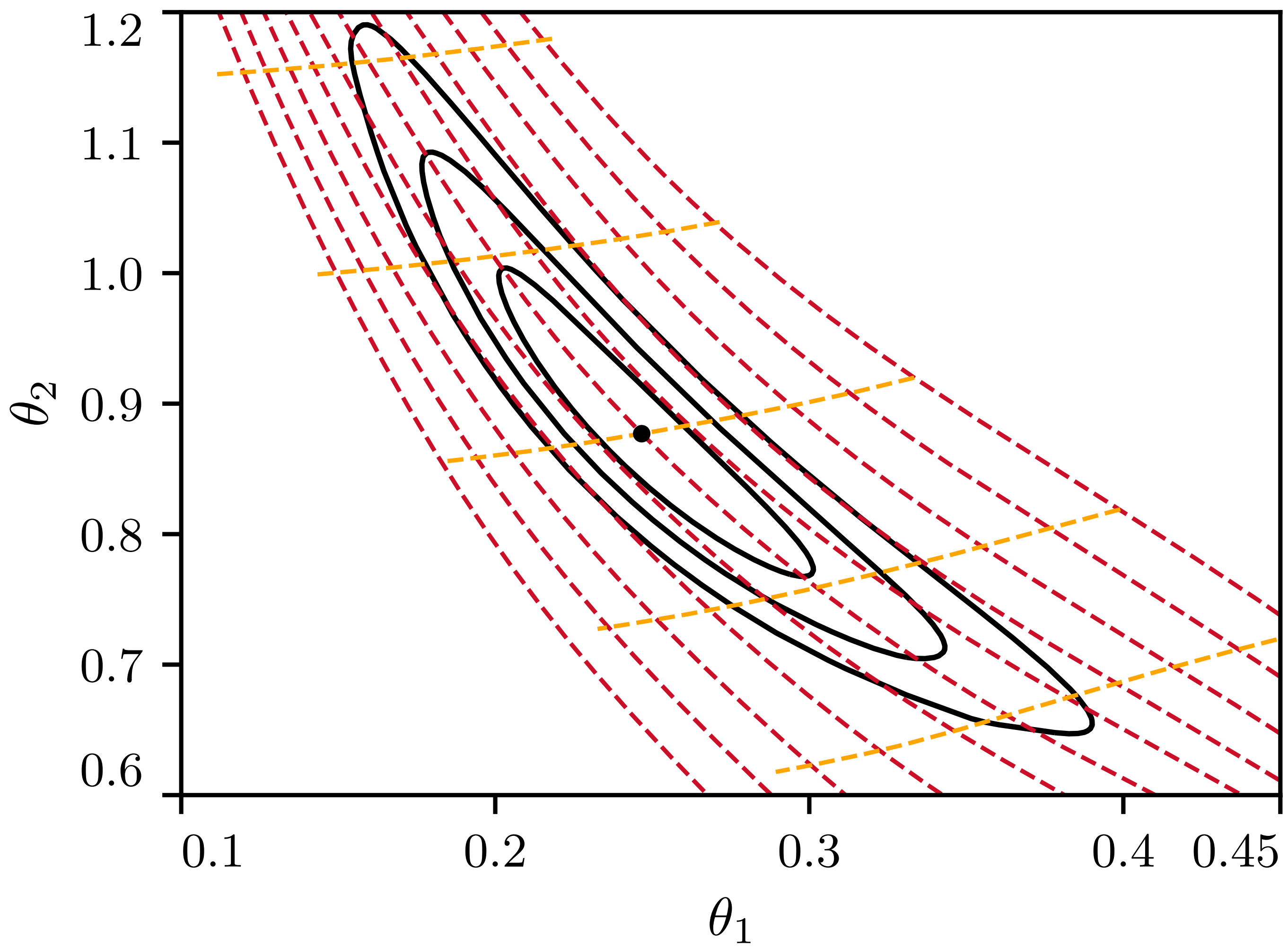
Information geometry with normalising flows

arXiv:2112.05737

- Local Fisher information matrix *

$$\mathcal{F}_{\mu\nu} = \left(\frac{\partial \phi^b}{\partial \theta^\nu}\right)^T \tilde{\mathcal{F}}_{ab} \frac{\partial \phi^a}{\partial \theta^\mu}$$

  - Defines a metric
  - NF+TFP to compute Jacobian
- **Constrained directions** in parameter space given by
  - PCA ?



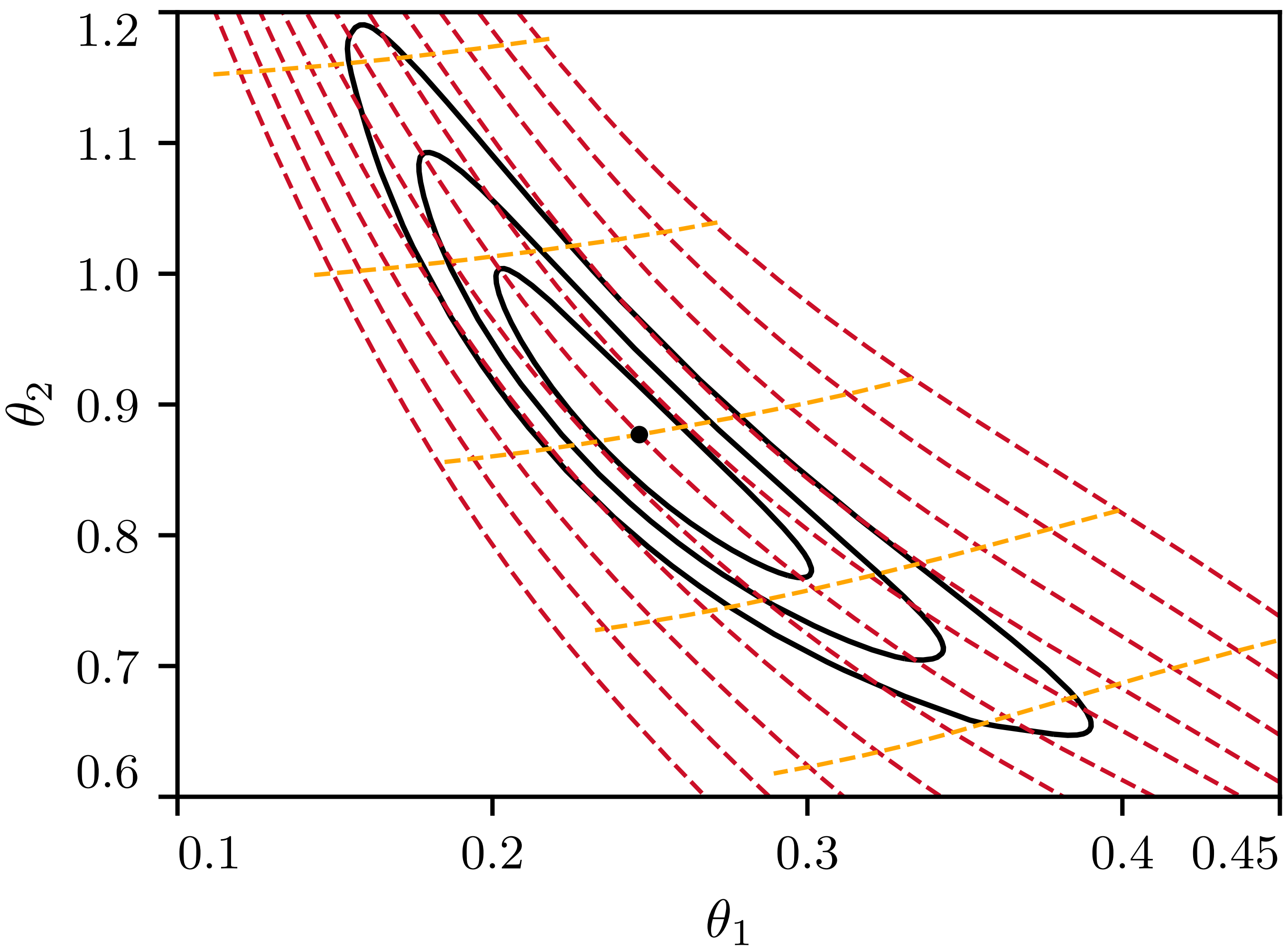
Information geometry with normalising flows

arXiv:2112.05737

- Local Fisher information matrix *

$$\mathcal{F}_{\mu\nu} = \left(\frac{\partial \phi^b}{\partial \theta^\nu}\right)^T \tilde{\mathcal{F}}_{ab} \frac{\partial \phi^a}{\partial \theta^\mu}$$

  - Defines a metric
  - NF+TFP to compute Jacobian
- **Constrained directions** in parameter space given by
  - PCA ? No, depends on scale



Information geometry with normalising flows

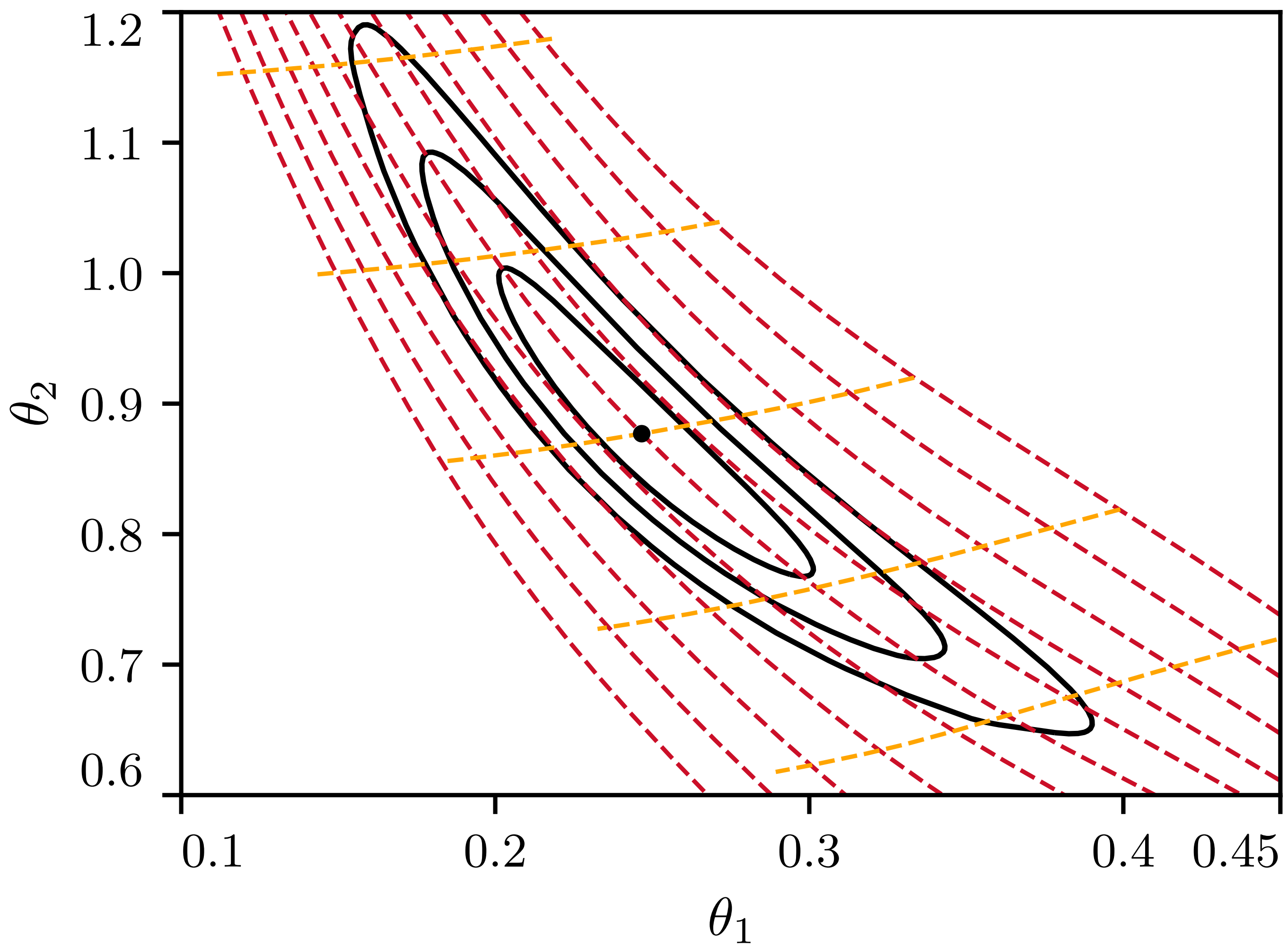
arXiv:2112.05737

- Local Fisher information matrix *

$$\mathcal{F}_{\mu\nu} = \left(\frac{\partial \phi^b}{\partial \theta^\nu}\right)^T \tilde{\mathcal{F}}_{ab} \frac{\partial \phi^a}{\partial \theta^\mu}$$

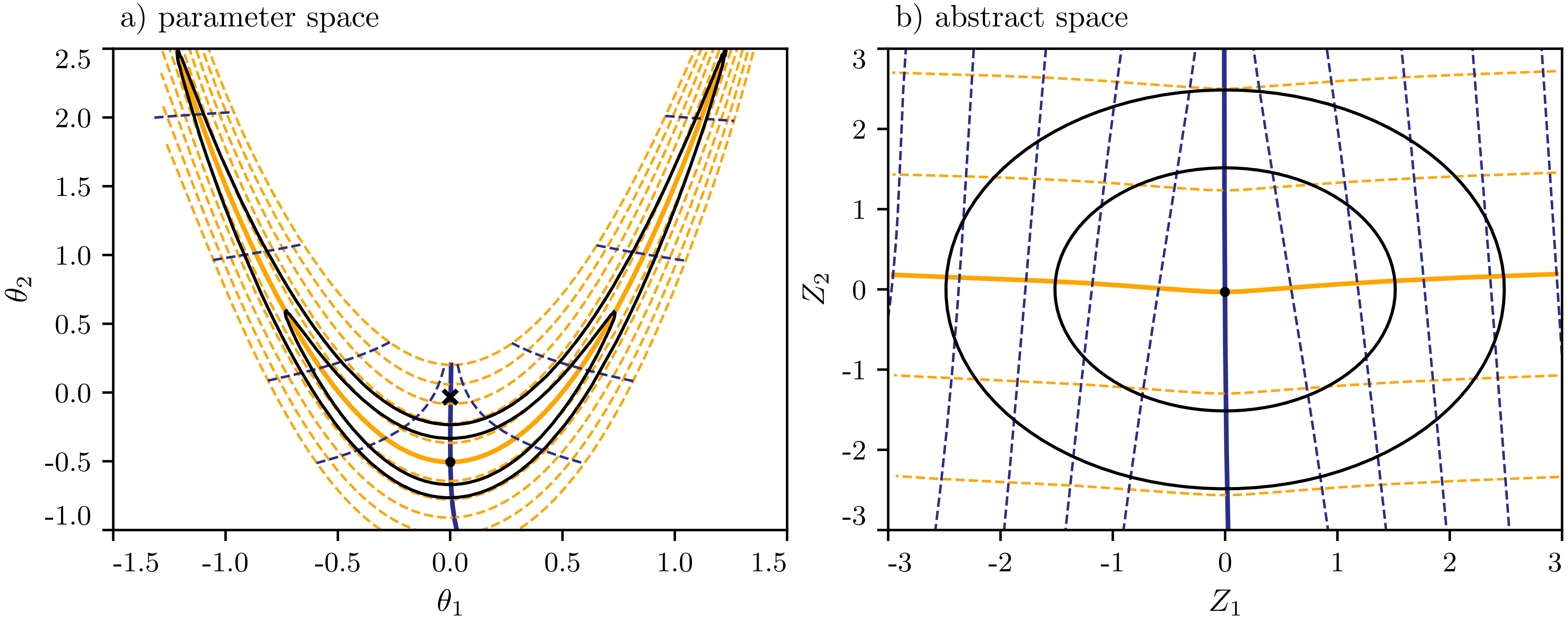
  - Defines a metric
  - NF+TFP to compute Jacobian
- **Constrained directions** in parameter space given by
  - PCA ? No, depends on scale
  - CPCA: *joint diagonalisation* of prior and posterior Fisher metrics

$$\mathcal{F}^{p}_{\mu\nu} u^\nu = \alpha \mathcal{F}^{\Pi}_{\mu\nu} u^\nu$$

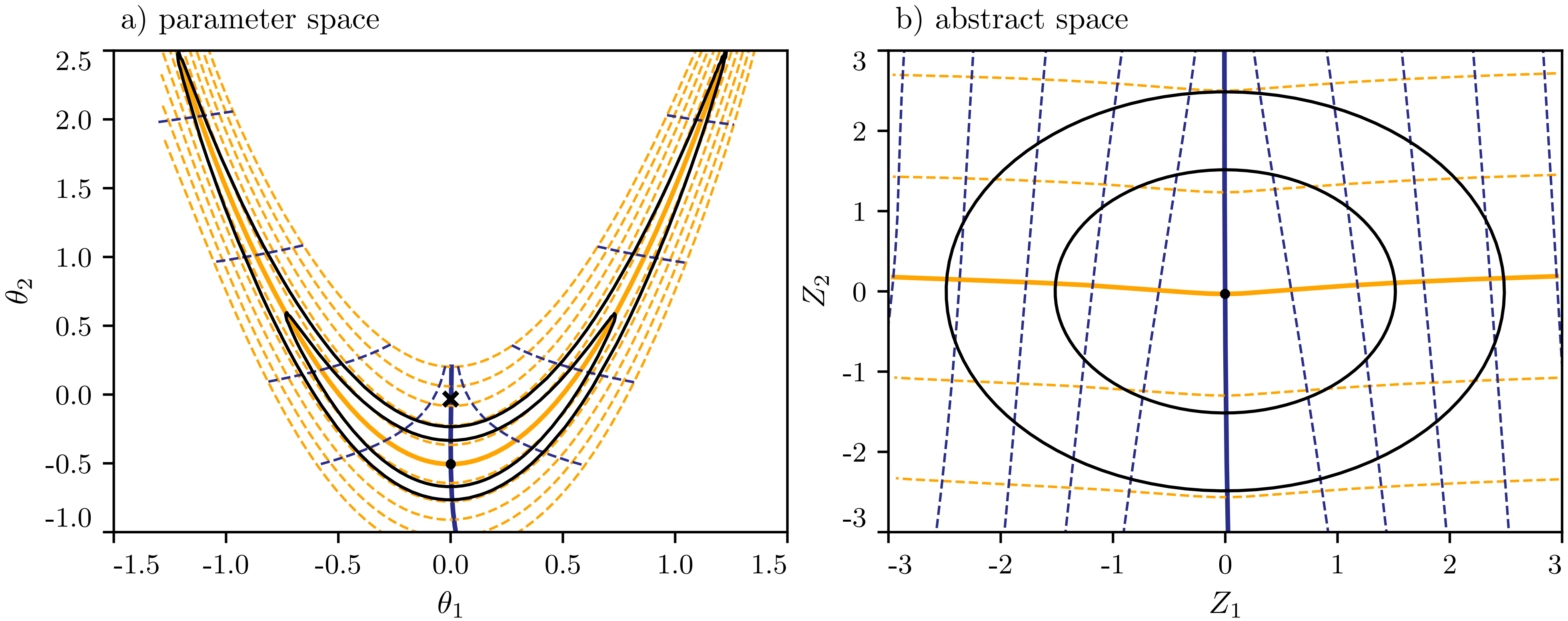


Toy model 1

arXiv:2112.05737



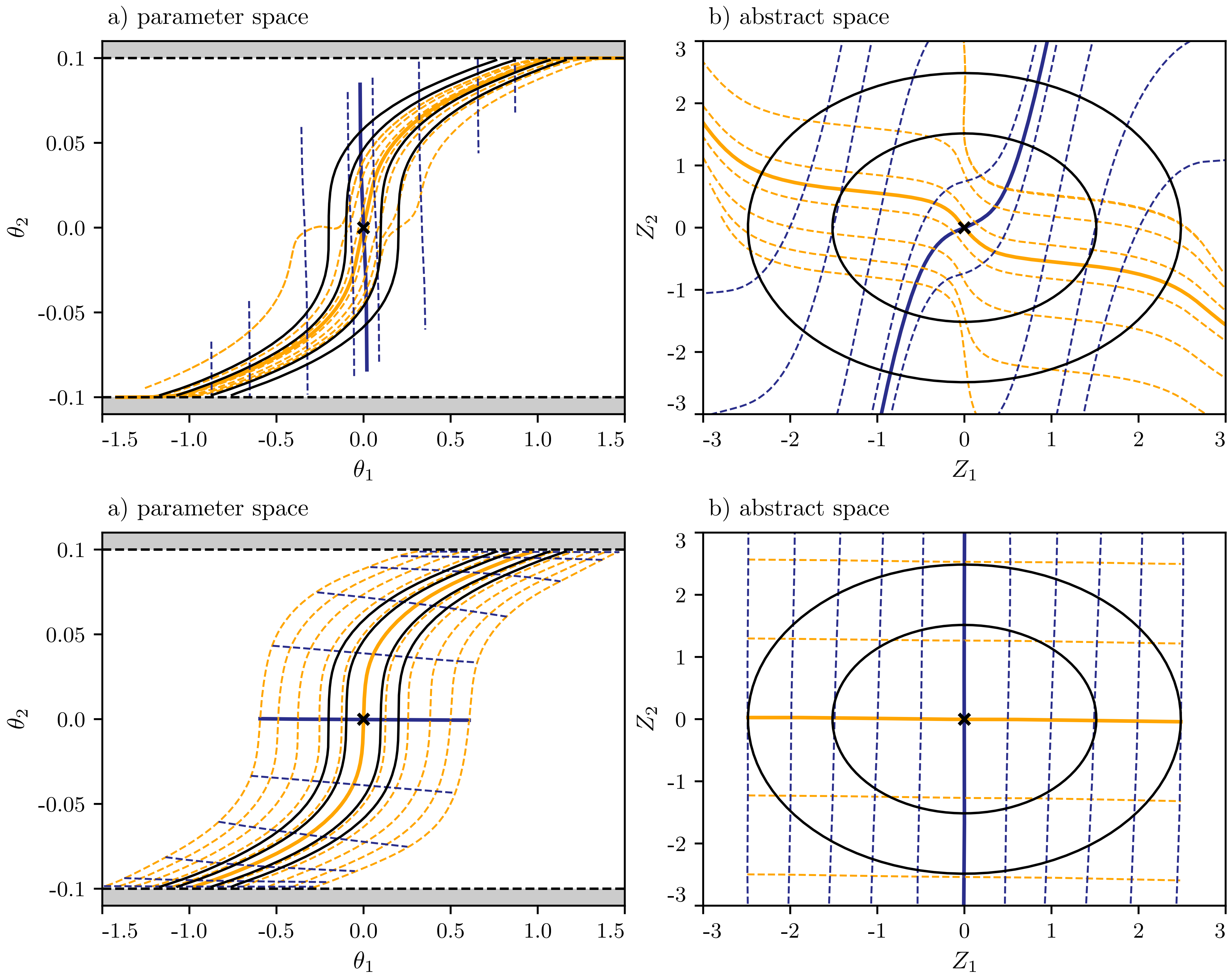
PCA



CPCA

Toy model 2

arXiv:2112.05737



PCA

CPCA

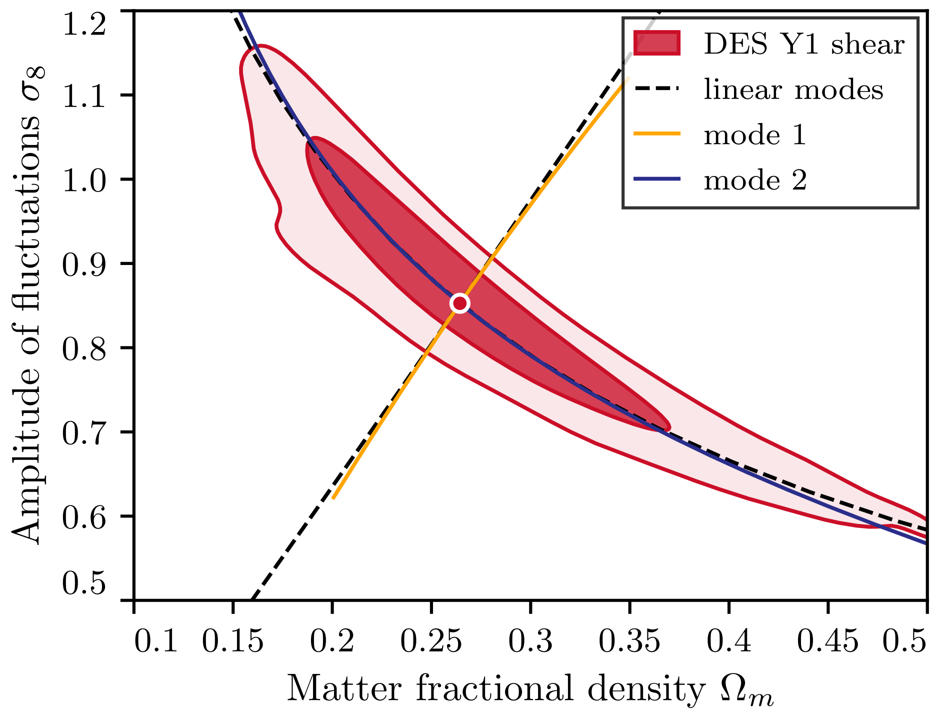
Applications

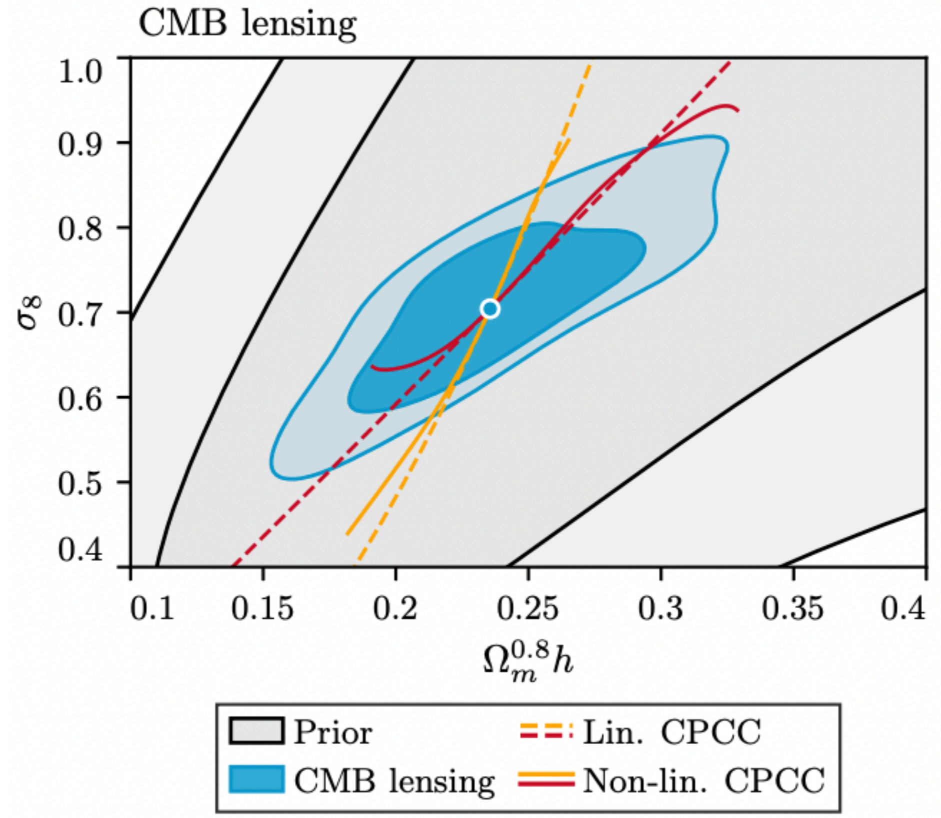
- CPCA recovers power law degeneracies automatically!

- Hubble from LSS alone ?
  1. DES clustering + lensing constrains $\sigma_8 \, \Omega_m^{+0.9}$
  2. CMB lensing constrains $\sigma_8 \, \Omega_m^{-0.8} H_0^{-1}$
  3. Add RSD from BOSS (to get σ_8), we get $H_0 = 73.8 \pm 7.5$ km/s/Mpc

arXiv:2112.05737

DES lensing



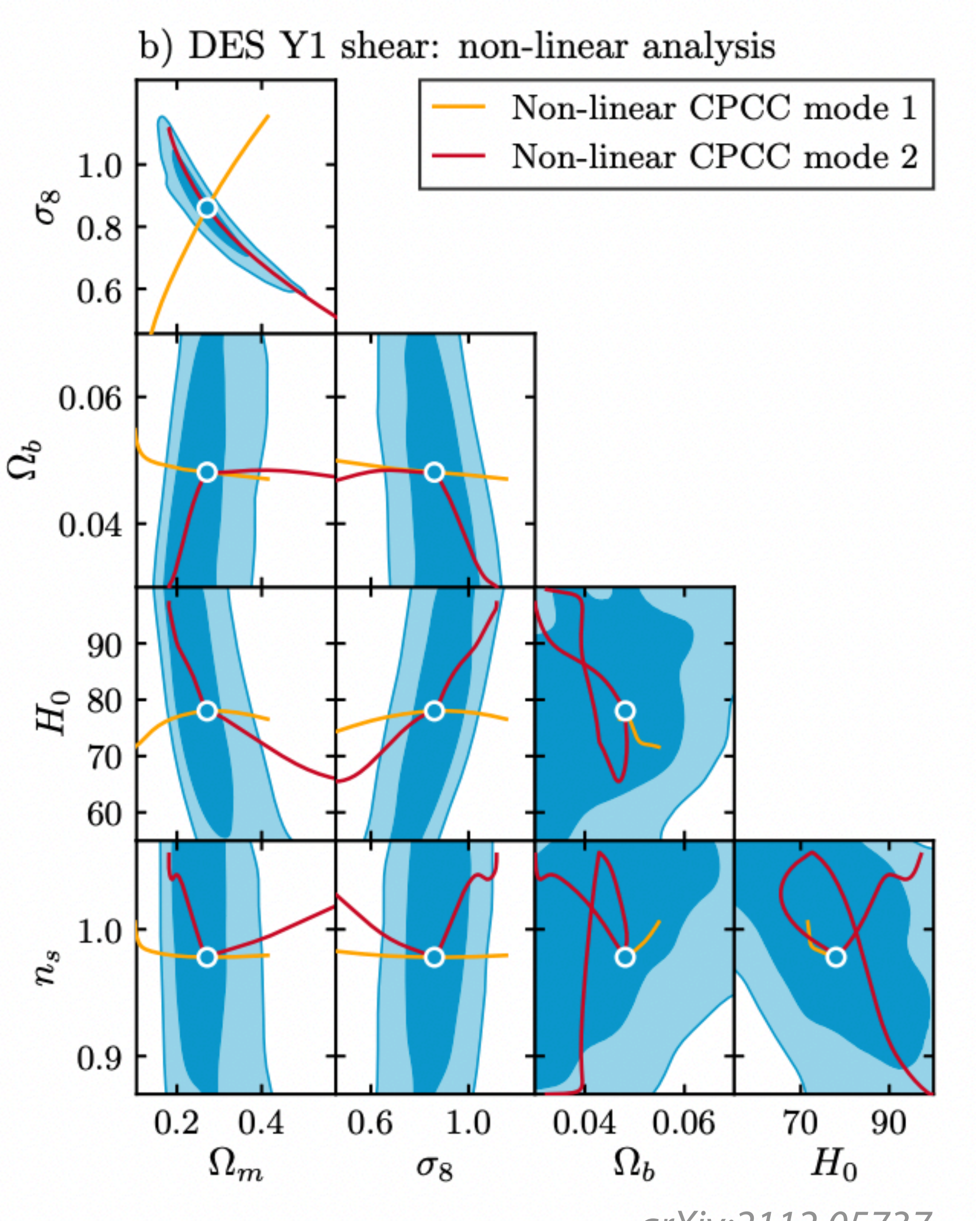


Take-away messages

- Normalizing flows can be used to tackle otherwise difficult statistical inference questions (arXiv:2105.03324, arXiv:2112.05737)
  - Tensions between experiments
  - Opens up information geometry questions

- It's all in tensiometer.readthedocs.io

- More to come !
  - Symbolic regression of local/non-linear CPCA



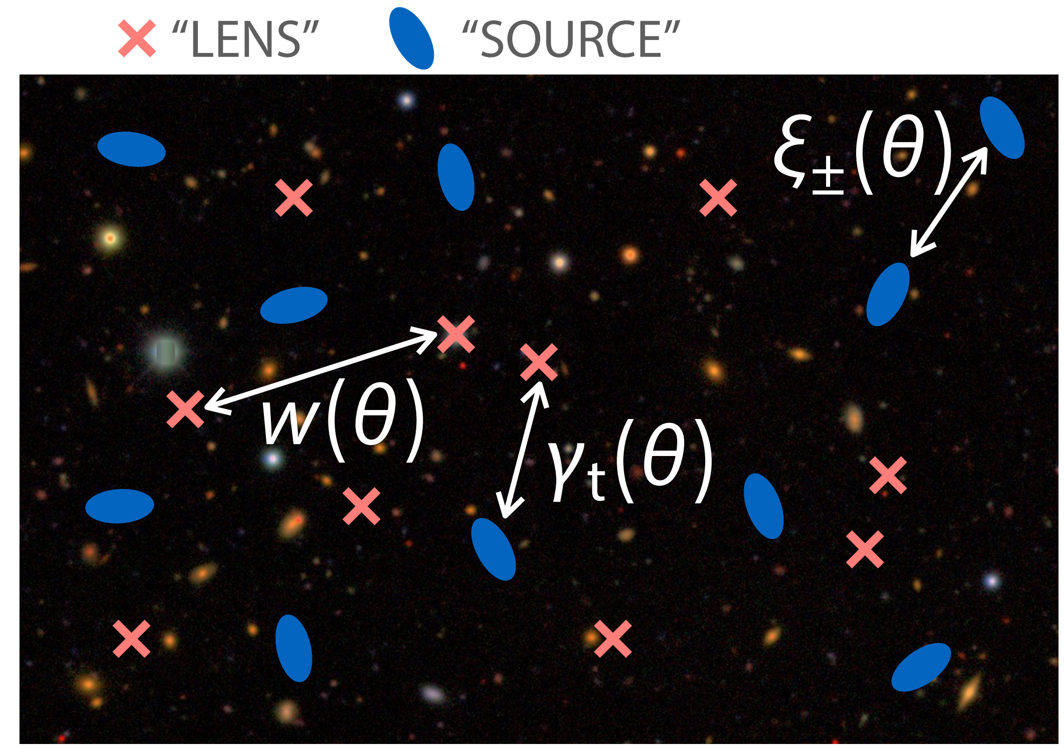
arXiv:2112.05737

THANKS!

Thanks!

Dark Energy Survey

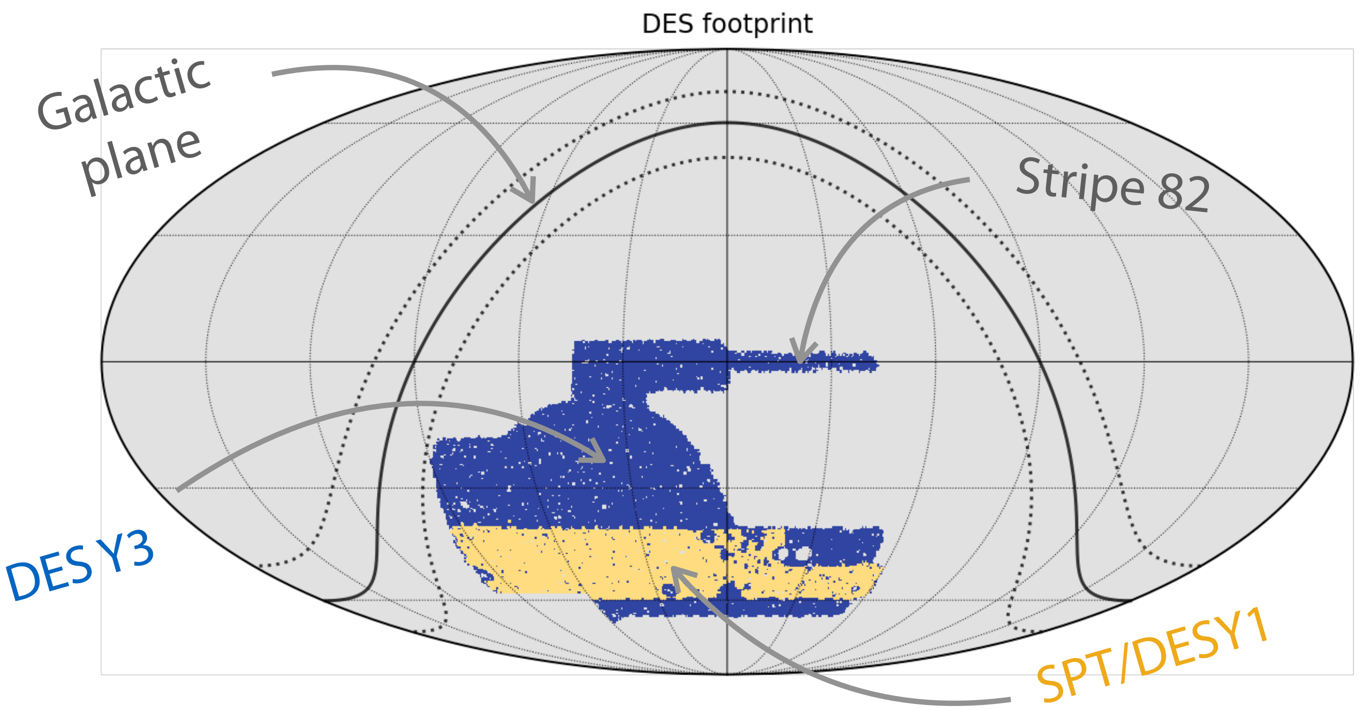
- **Blanco 4-meter telescope** at Cerro Tololo (CTIO) in Chile
- **Dark Energy Camera** (DECam)
  - 3.0 deg² field-of-view, 70 CCD chips, 570 Mpix, *griz(Y)* filters
  - Seeing ~0.9′ in *r*-band, magnitude i_{AB}<23.0, *r*<23.5
- **Survey(s)**
  - 5000 deg² footprint + deep fields, observed 2013-2019
  - Overlaps with SPT, BOSS and COSMOS
  - DR2 (6 years) of 543M galaxies + 145M stars to *i*~23.8



$\xi_\pm(\theta)$ = COSMIC SHEAR

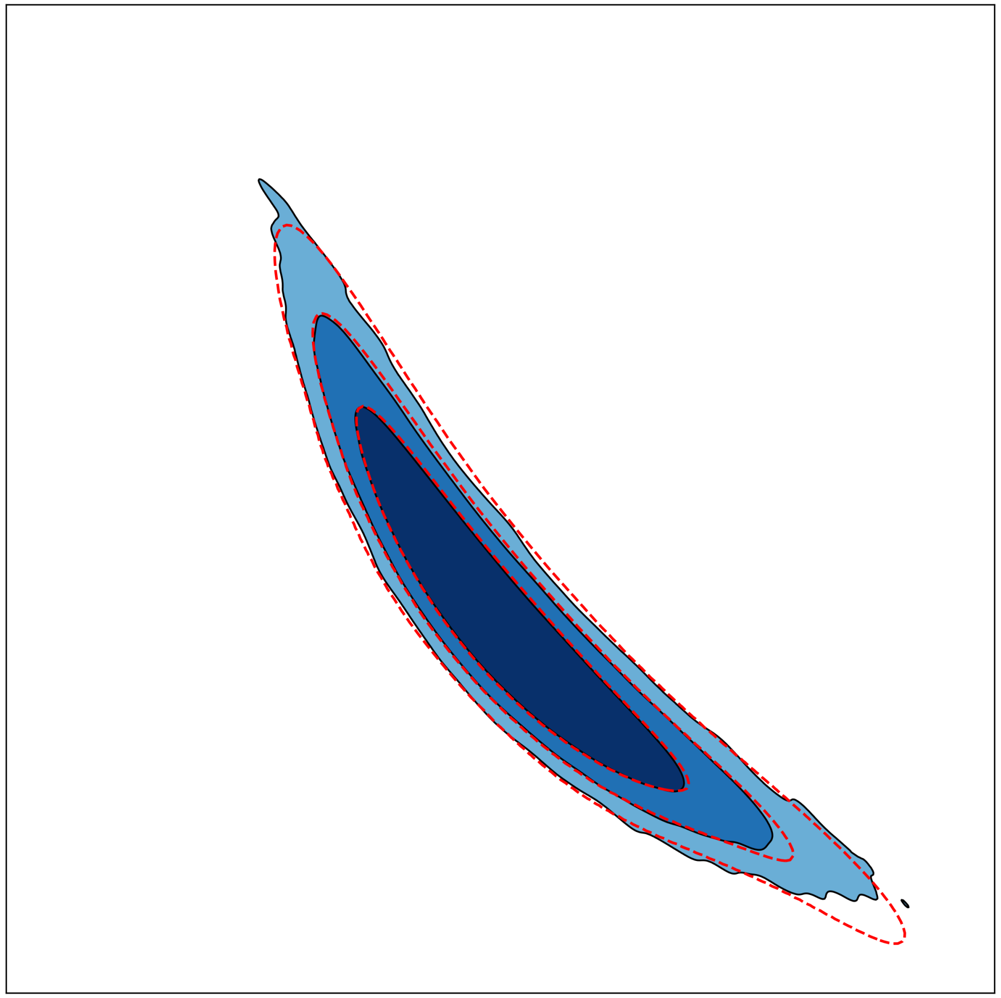
$\gamma_t(\theta)$ = GALAXY-GALAXY LENSING

$w(\theta)$ = CLUSTERING



The solution? *Learn* that distribution!

$$\Delta = \int_{\mathcal{P}(\Delta\theta)>\mathcal{P}(\mathbf{0})} \mathcal{P}(\Delta\theta)\,\mathrm{d}\Delta\theta$$



- **What do we need to compute this?**
  - Being able to *sample* from the distribution
  - Compute its *density* for every sample (quickly if possible)
- **Sampling the distribution**

A. Uncorrelated experiments

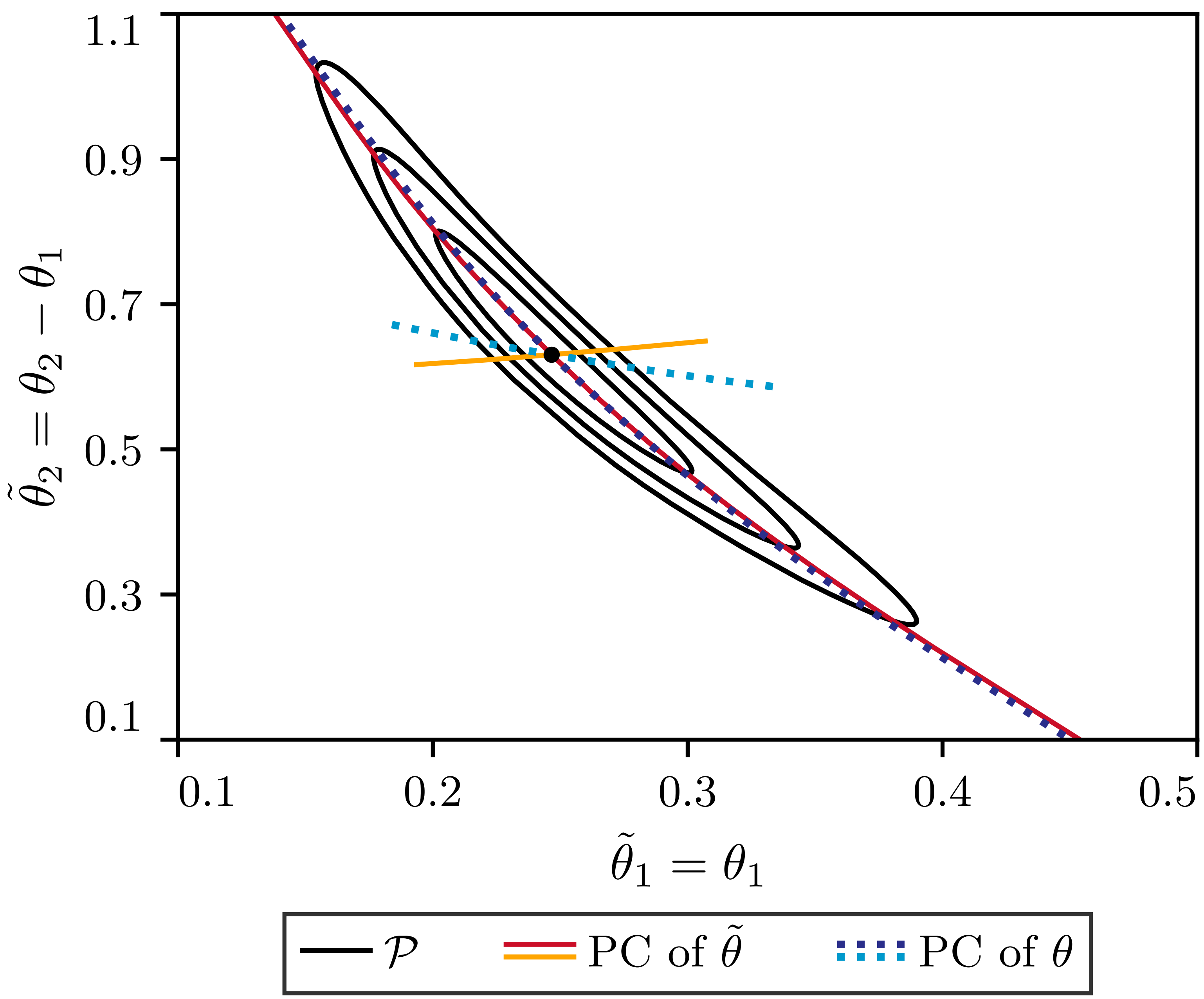
$$\mathcal{P}(\Delta\theta) = \int_{V_\pi} \mathcal{P}_1(\theta)\mathcal{P}_2(\theta - \Delta\theta)\,\mathrm{d}\theta$$

1. Run two chains
2. Take differences

B. Correlated case (*eg* DES/LSST lensing and clustering)

1. Duplicate all common parameters
2. Compute d_1 with θ_1, d_2 with $\theta 2$
3. Run one bigger chain for $\mathcal{L}(\theta_1, \theta_2)$

Local principal components



- **Principal component analysis**
  - Eigenvectors of the covariance, or its inverse, ie the Fisher matrix
  - Locally, this reads

$$\mathcal{F}_{\mu\nu} u^\nu = \alpha \eta_{\mu\nu} u^\nu$$

with

$$u^\mu \eta_{\mu\nu} u^\nu = 1$$

  - Connect the dot with… parallel transport ! (note they're not geodesics!)
  - …but $\eta_{\mu\nu}$ is not a tensor…
  - PCA depends on parametrisation and units!