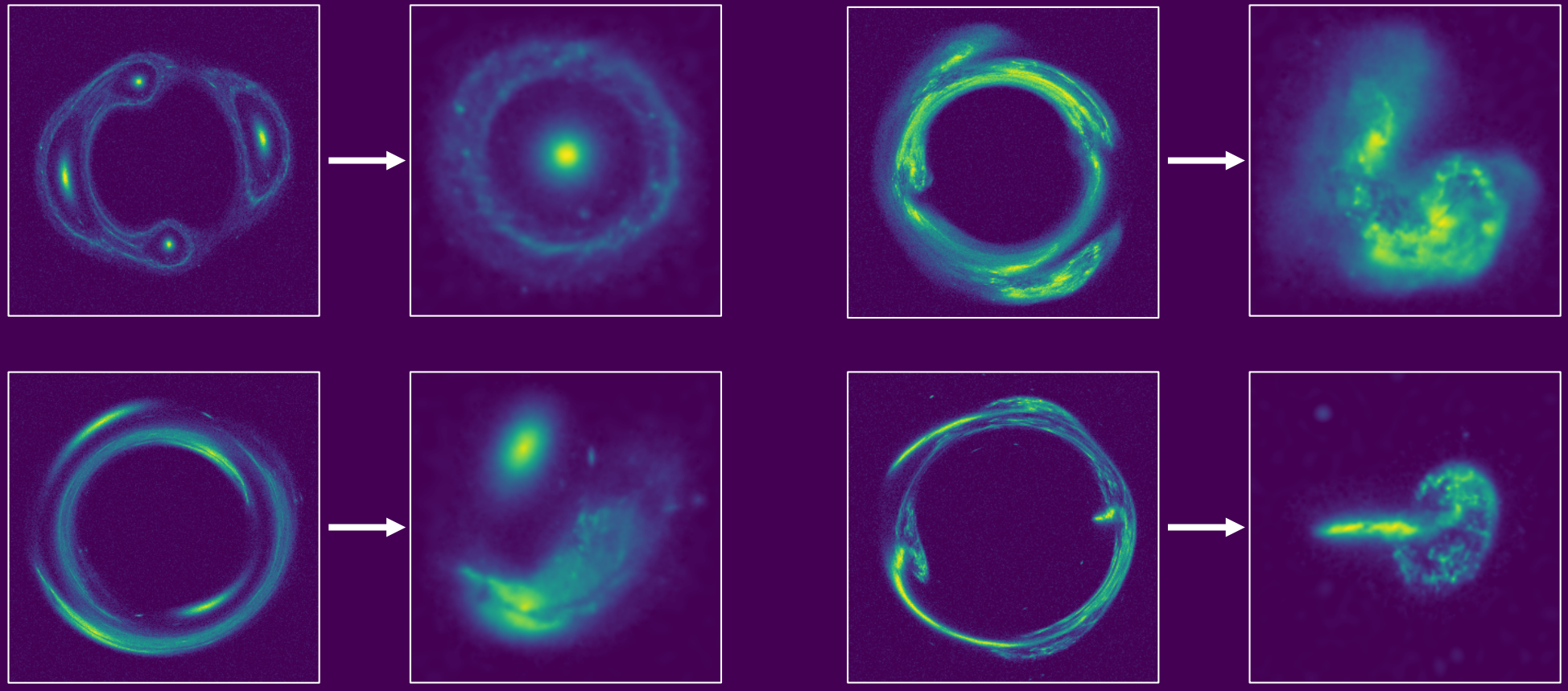
Strong-lensing source reconstruction with variationally optimised Gaussian processes

Konstantin Karchev[1,2], Adam Coogan[1], Christoph Weniger[1]

NAM 2021

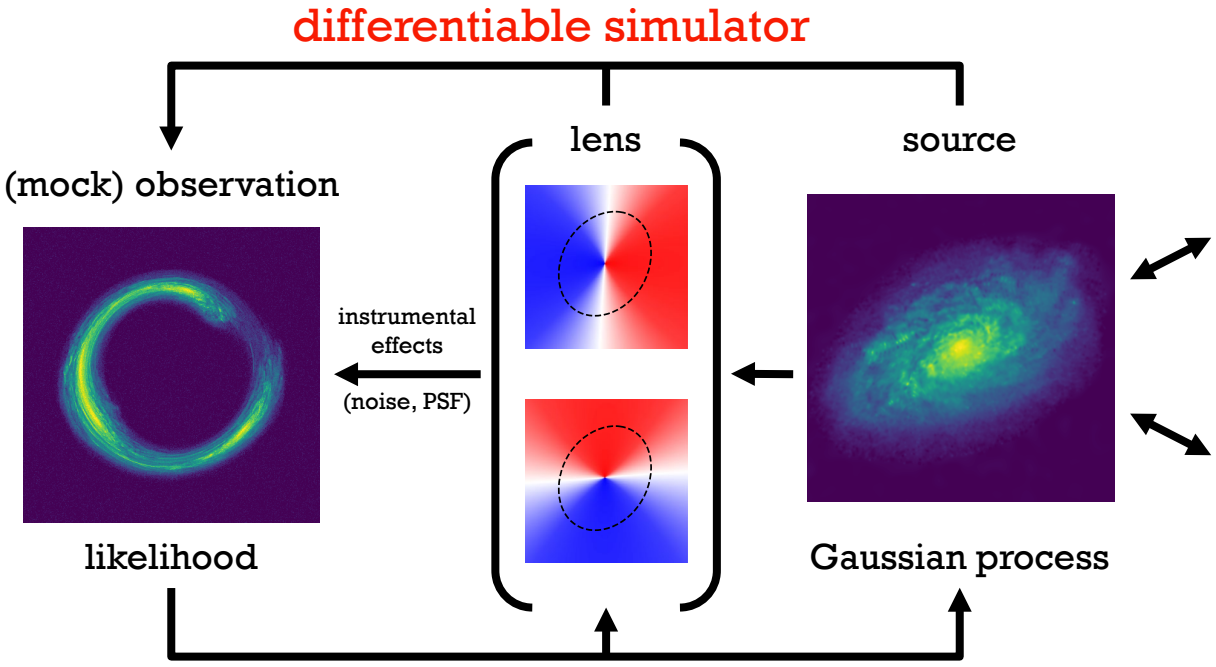


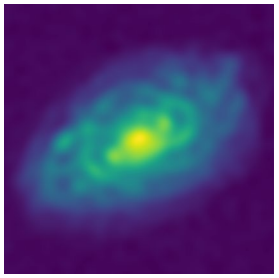
1 GRAPPA
GRavitation AstroParticle Physics Amsterdam

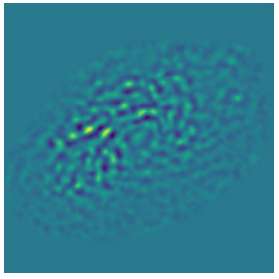
preprint: arXiv:2105.09465
previous: arXiv:2010.07032

Towards substructure inference:
- Adam Coogan's poster
- Noemi Anau Montel's talk

2 SISSA







On **large scales** the source is modelled on a grid of "inducing points". This first stage is rapid: fitting 1000 parameters takes about 10 minutes.

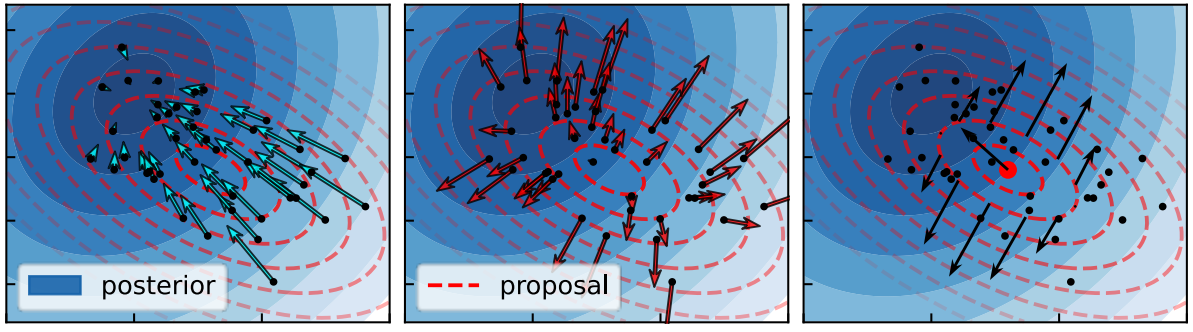
Small details are reconstructed by inducing points attached to the image pixels (which thus adapt to the magnification). Several layers containing $\mathcal{O}(10^5)$ parameters in total require a few hours to converge.

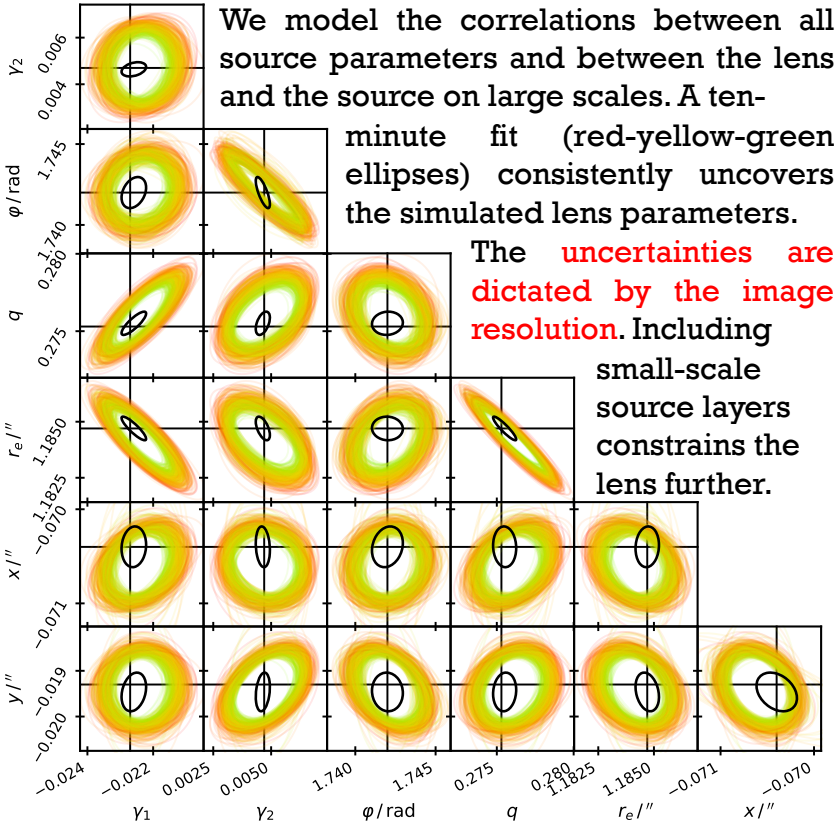
Variational Inference

In stochastic variational inference (SVI) a parametrised proposal distribution $q_\phi(\theta)$ is fit to the multi-dimensional posterior $p(\theta|x)$ of a large number of parameters θ, given the observation x. The target of optimisation via gradient descent is the **evidence lower bound (ELBO)**, which balances a good fit to the data (left panel below) and the appropriate uncertainty (middle panel). It is also useful in optimising hyperparameters via a relation to the Bayesian evidence.

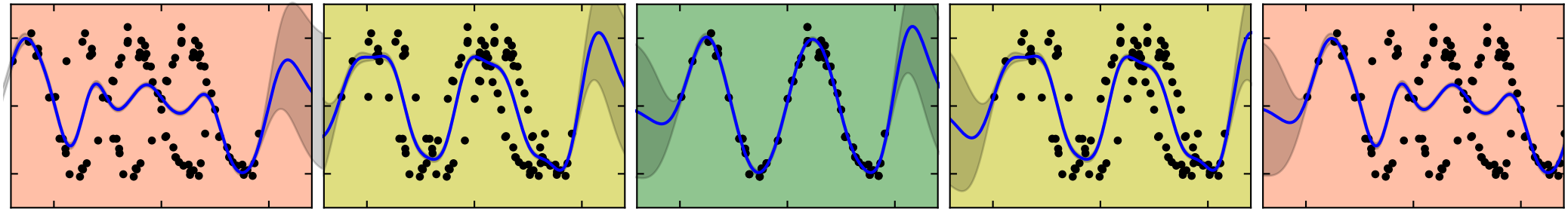
$$\mathrm{ELBO} = \mathbb{E}_{q_\phi(\theta)}\left[\ln p(\theta, x) - \ln q_\phi(\theta)\right]$$

Gradient descent adjusts parameters ϕ of proposal: mean, uncertainty…





We model the correlations between all source parameters and between the lens and the source on large scales. A ten-minute fit (red-yellow-green ellipses) consistently uncovers the simulated lens parameters. The uncertainties are dictated by the image resolution. Including small-scale source layers constrains the lens further.



Varying the lens parameters (mis)aligns multiple projections ⟷ GP marginal likelihood favours perfect alignment (up to noise)

Gaussian processes ⟷ A generative model

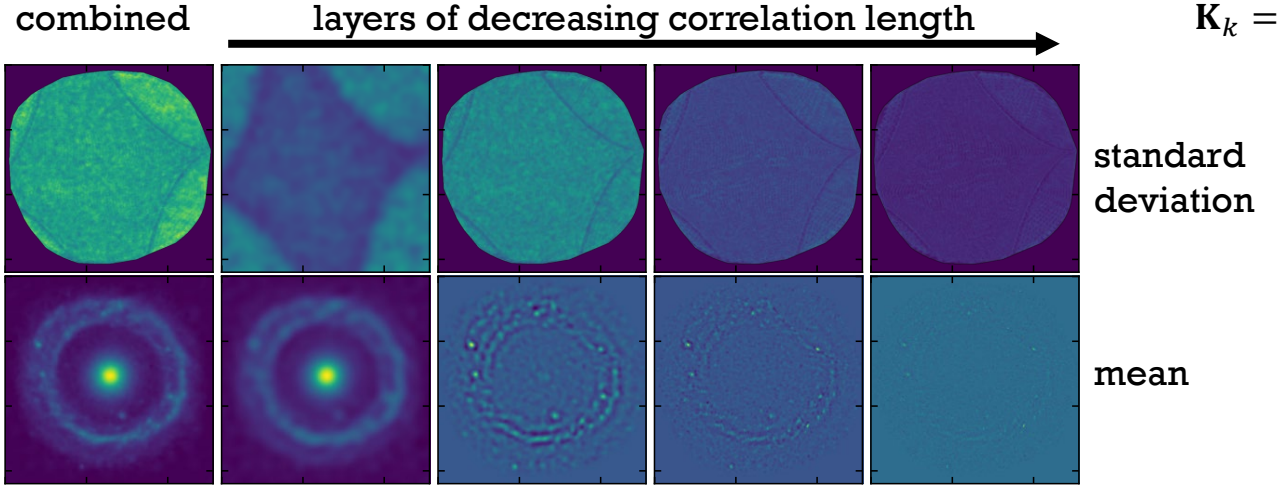
Gaussian processes (GPs) generalise the normal distribution to the space of functions. They achieve two goals. For one, a GP evaluates the amount of variability in the source, which is minimised when multiple projections of the source are well-aligned by the modelled lens (see above). A GP also serves as a posterior for the surface brightness of the source (pictured below), helping disambiguate the effect of lensing substructure and providing an efficient way to sample sources consistent with the observation needed to train substructure inference networks.
To model source features on different scales we use several GP layers with a range of fixed correlation lengthscales.

GPs are non-parametric, providing explicit formulae for the posterior mean and covariance. In order to avoid the necessary expensive matrix inverse and determinant calculation, we re-phrase the GP as a generative model (k labelling layers):

$$\mathbf{data} \sim \mathcal{N}(\textstyle\sum_k \mathbf{f}_k\,, \mathrm{noise}^2), \quad \mathbf{f}_k = \mathbf{T}_k \boldsymbol{\theta}_k, \quad \boldsymbol{\theta}_k \sim \mathcal{N}(0, \alpha_k^2)$$

$\boldsymbol{\theta}_k$ are sets of source parameters inferred variationally;
α_k^2 are the variance hyperparameters of each layer;
$\mathbf{T}_k$ are "transfer matrices" correlating close-by pixels in the source plane. They act as a square-root of the covariance matrix:

$$\mathbf{K}_k = \alpha_k^2 \mathbf{T}_k \mathbf{T}_k^T \quad \leftrightarrow \quad \mathcal{G}(x, 2\Sigma) = \int \mathcal{G}(x, \Sigma)\, \mathcal{G}(x - y, \Sigma)\; dy$$



The variance of smaller-scale layers is automatically suppressed by hyperparameter optimisation.

A fraction of the pixel-pixel covariance is due to an overlap of light-collecting areas in the source plane. We approximate the pixels as Gaussians aligned with the lensed pixel grid, which can be analytically integrated out as in the above equation.