Automatic physical inference with information maximising neural networks

Tom Charnock

Guilhem Lavaux & Ben Wandelt
Alessandro Manzotti

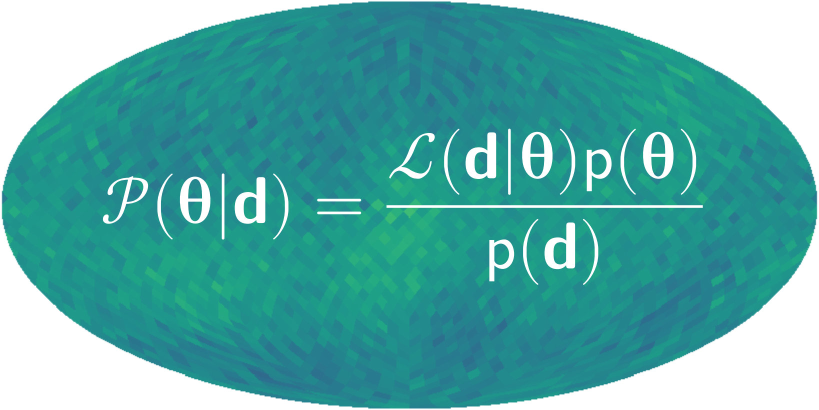
Institut d'Astrophysique de Paris

Physical Review D **97**, 083004

DOI 10.5281/zenodo.1175196

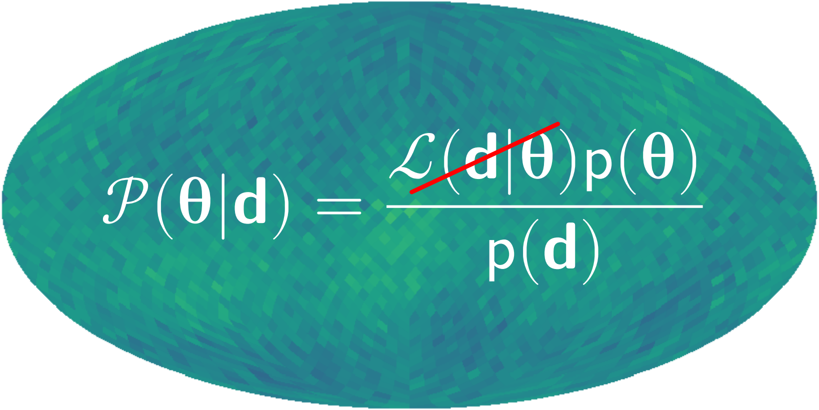
github:information_maximiser

Physical inference



$$\mathcal{P}(\boldsymbol{\theta}|\mathbf{d}) = \frac{\mathcal{L}(\mathbf{d}|\boldsymbol{\theta})\mathrm{p}(\boldsymbol{\theta})}{\mathrm{p}(\mathbf{d})}$$

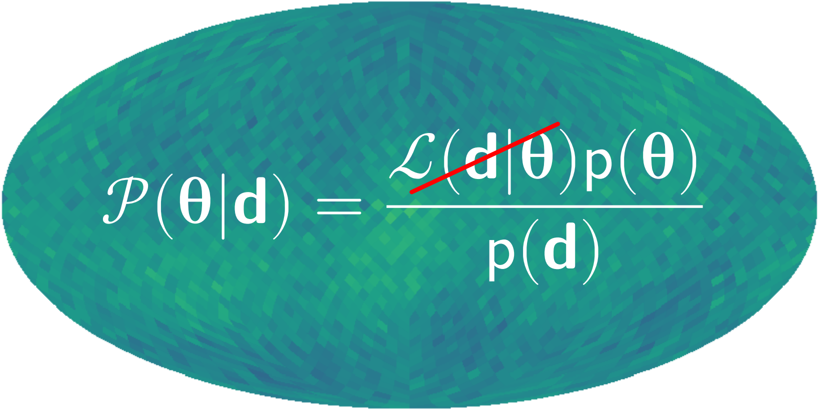
Approximate Bayesian computation



$$\mathcal{P}(\boldsymbol{\theta}|\mathbf{d}) = \frac{\cancel{\mathcal{L}(\mathbf{d}|\boldsymbol{\theta})}\,\mathrm{p}(\boldsymbol{\theta})}{\mathrm{p}(\mathbf{d})}$$

 Physical Review D **97**, 083004, github:information_maximiser, DOI 10.5281/zenodo.1175196

Approximate Bayesian computation

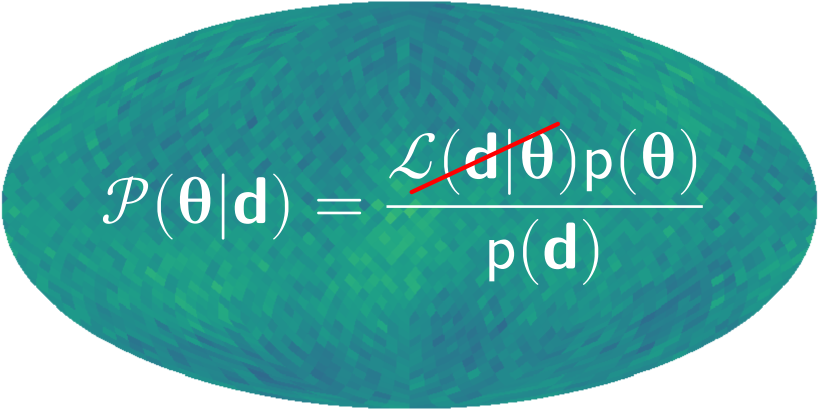


$$\mathcal{P}(\boldsymbol{\theta}|\mathbf{d}) = \frac{\cancel{\mathcal{L}(\mathbf{d}|\boldsymbol{\theta})}\,\mathrm{p}(\boldsymbol{\theta})}{\mathrm{p}(\mathbf{d})}$$

Simulate data at different $\boldsymbol{\theta}$ drawn from $\mathrm{p}(\boldsymbol{\theta})$

Physical Review D **97**, 083004, github:information_maximiser, DOI 10.5281/zenodo.1175196

Approximate Bayesian computation



$$\mathcal{P}(\boldsymbol{\theta}|\mathbf{d}) = \frac{\cancel{\mathcal{L}(\mathbf{d}|\boldsymbol{\theta})}\,\mathrm{p}(\boldsymbol{\theta})}{\mathrm{p}(\mathbf{d})}$$

Simulate data at different $\boldsymbol{\theta}$ drawn from $\mathrm{p}(\boldsymbol{\theta})$

Accept or reject $\boldsymbol{\theta}$ using similarity of simulation and real data

Physical Review D **97**, 083004, github:information_maximiser, DOI 10.5281/zenodo.1175196

Compressed summary statistics

Reduce dimensionality of data

Unlikely for simulation to hit real data in high dimensions

Use summary statistics such as power spectrum, C_ℓ, ect.

Compressed summary statistics

Reduce dimensionality of data

Unlikely for simulation to hit real data in high dimensions

Use summary statistics such as power spectrum, C_ℓ, ect.

Inadequately compressed summary statistics

Upcoming surveys will observe huge amounts of raw data

Even the summary statistics will be $O(10^4)$ (LSST, EUCLID, etc.)

Huge amount of computing power needed

Still nearly impossible to do ABC

Massively optimised parameter estimation and data (MOPED) compression

Fisher information

Amount of information data $\mathbf{d}$ contains about parameters $\boldsymbol{\theta}$

$$\mathbf{F}_{\alpha\beta}(\boldsymbol{\theta}) = -\left\langle \frac{\partial^2 \ln \mathcal{L}(\mathbf{d}|\boldsymbol{\theta})}{\partial\theta_\alpha \partial\theta_\beta} \right\rangle \Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{\text{fid}}}$$

Assuming a Gaussian likelihood with covariance independent of $\boldsymbol{\theta}$

$$-2\ln\mathcal{L}(\mathbf{d}|\boldsymbol{\theta}) = (\mathbf{d}-\boldsymbol{\mu}(\boldsymbol{\theta}))^{\mathsf{T}}\,\mathbf{C}^{-1}\,(\mathbf{d}-\boldsymbol{\mu}(\boldsymbol{\theta})) + \ln|2\pi\mathbf{C}|$$

we find

$$\mathbf{F}_{\alpha\beta}(\boldsymbol{\theta}) = \mathrm{Tr}\left[\boldsymbol{\mu},_{\alpha}^{\mathsf{T}}(\boldsymbol{\theta})\mathbf{C}^{-1}\boldsymbol{\mu},_{\beta}(\boldsymbol{\theta})\right]$$

Massively optimised parameter estimation and data (MOPED) compression

Transform the data into parameter space

$$x_\alpha = \mathbf{r}_\alpha^{\mathsf{T}}\mathbf{d}$$

Physical Review D **97**, 083004, github:information_maximiser, DOI 10.5281/zenodo.1175196

Massively optimised parameter estimation and data (MOPED) compression

Transform the data into parameter space

$$x_\alpha = \mathbf{r}_\alpha^\mathrm{T}\mathbf{d}$$

Calculate compression parameters, $\mathbf{r}_\alpha$

Maximise Fisher information whilst ensuring $\mathbf{r}_\alpha$ is orthogonal to $\mathbf{r}_\beta$

$$\mathbf{r}_1 = \frac{\mathbf{C}^{-1}\boldsymbol{\mu}_{,1}}{\sqrt{\boldsymbol{\mu}_{,1}^\mathrm{T}\mathbf{C}^{-1}\boldsymbol{\mu}_{,1}}} \qquad \mathbf{r}_\alpha = \frac{\mathbf{C}^{-1}\boldsymbol{\mu}_{,\alpha} - \sum_{i=1}^{\alpha-1}(\boldsymbol{\mu}_{,\alpha}^\mathrm{T}\mathbf{r}_i)\mathbf{r}_i}{\sqrt{\boldsymbol{\mu}_{,\alpha}^\mathrm{T}\mathbf{C}^{-1}\boldsymbol{\mu}_{,\alpha} - \sum_{i=1}^{\alpha-1}(\boldsymbol{\mu}_{,\alpha}^\mathrm{T}\mathbf{r}_i)^2}}$$

Physical Review D **97**, 083004, github:information_maximiser, DOI 10.5281/zenodo.1175196

Massively optimised parameter estimation and data (MOPED) compression

Transform the data into parameter space

$$x_\alpha = \mathbf{r}_\alpha^T \mathbf{d}$$

Calculate compression parameters, $\mathbf{r}_\alpha$

Maximise Fisher information whilst ensuring $\mathbf{r}_\alpha$ is orthogonal to $\mathbf{r}_\beta$

$$\mathbf{r}_1 = \frac{\mathbf{C}^{-1}\boldsymbol{\mu}_{,1}}{\sqrt{\boldsymbol{\mu}_{,1}^T\mathbf{C}^{-1}\boldsymbol{\mu}_{,1}}} \qquad \mathbf{r}_\alpha = \frac{\mathbf{C}^{-1}\boldsymbol{\mu}_{,\alpha} - \sum_{i=1}^{\alpha-1}(\boldsymbol{\mu}_{,\alpha}^T\mathbf{r}_i)\mathbf{r}_i}{\sqrt{\boldsymbol{\mu}_{,\alpha}^T\mathbf{C}^{-1}\boldsymbol{\mu}_{,\alpha} - \sum_{i=1}^{\alpha-1}(\boldsymbol{\mu}_{,\alpha}^T\mathbf{r}_i)^2}}$$

Lossless compression

Data compressed to number of parameters
Likelihood must be Gaussian

Physical Review D **97**, 083004, github:information_maximiser, DOI 10.5281/zenodo.1175196

Non-linear likelihood-free compression

Might not want to make Gaussian assumption

Information is lost if likelihood is not Gaussian

Non-linear likelihood-free compression

Might not want to make Gaussian assumption

Information is lost if likelihood is not Gaussian

Transform data

Find a non-linear function, $\mathscr{f} : \mathbf{d} \to \mathbf{x}$ which transforms the unknown likelihood into

$$-2\ln\mathcal{L}(\mathbf{x}|\boldsymbol{\theta}) = \left(\mathbf{x} - \boldsymbol{\mu}_{\mathscr{f}}(\boldsymbol{\theta})\right)^{\mathsf{T}} \mathbf{C}_{\mathscr{f}}^{-1} \left(\mathbf{x} - \boldsymbol{\mu}_{\mathscr{f}}(\boldsymbol{\theta})\right)$$

Non-linear likelihood-free compression

Might not want to make Gaussian assumption

Information is lost if likelihood is not Gaussian

Transform data

Find a non-linear function, $\mathscr{f} : \mathbf{d} \to \mathbf{x}$ which transforms the unknown likelihood into

$$-2\ln\mathcal{L}(\mathbf{x}|\boldsymbol{\theta}) = \left(\mathbf{x} - \boldsymbol{\mu}_{\mathscr{f}}(\boldsymbol{\theta})\right)^{\mathsf{T}} \mathbf{C}_{\mathscr{f}}^{-1} \left(\mathbf{x} - \boldsymbol{\mu}_{\mathscr{f}}(\boldsymbol{\theta})\right)$$

What is $\mathscr{f}$?

Non-linear likelihood-free compression

Do not want to make Gaussian assumption

Information is lost if likelihood is not Gaussian

Transform data

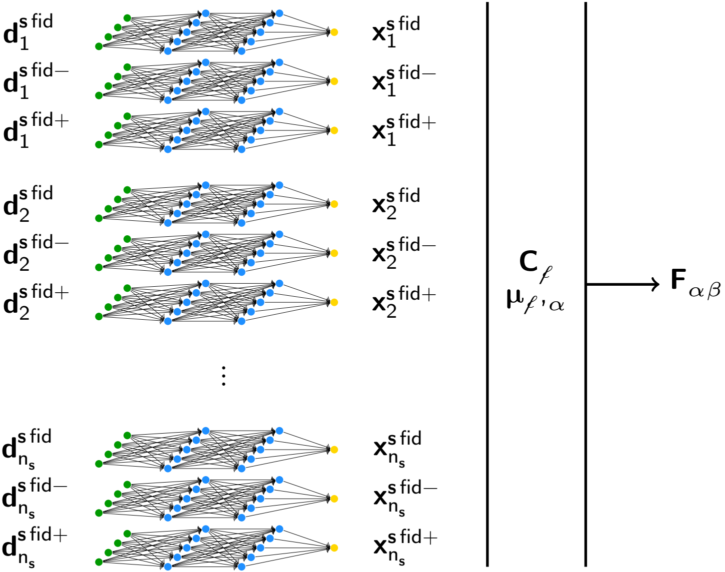
Find a function, $\mathscr{f} : \mathbf{d} \to \mathbf{x}$ which transforms the unknown likelihood into

$$-2\ln\mathcal{L}(\mathbf{x}|\boldsymbol{\theta}) = \left(\mathbf{x} - \boldsymbol{\mu}_{\mathscr{f}}(\boldsymbol{\theta})\right)^{\mathsf{T}} \mathbf{C}_{\mathscr{f}}^{-1} \left(\mathbf{x} - \boldsymbol{\mu}_{\mathscr{f}}(\boldsymbol{\theta})\right)$$

What is $\mathscr{f}$?

A neural network!

Information maximising neural network



Physical Review D **97**, 083004, github:information_maximiser, DOI 10.5281/zenodo.1175196

Automatic physical inference

Make simulations of data, $\{\mathbf{d}_\mathsf{i}^{\mathbf{s}\,\mathsf{fid}}, \mathbf{d}_\mathsf{i}^{\mathbf{s}\,\mathsf{fid}-}, \mathbf{d}_\mathsf{i}^{\mathbf{s}\,\mathsf{fid}+}\}$ at $\boldsymbol{\theta}^{\mathsf{fid}}$

Automatic physical inference

Make simulations of data, $\{\mathbf{d}_{\mathsf{i}}^{\mathbf{s}\,\mathsf{fid}}, \mathbf{d}_{\mathsf{i}}^{\mathbf{s}\,\mathsf{fid}-}, \mathbf{d}_{\mathsf{i}}^{\mathbf{s}\,\mathsf{fid}+}\}$ at $\boldsymbol{\theta}^{\mathsf{fid}}$

Train IMNN to maximise Fisher information

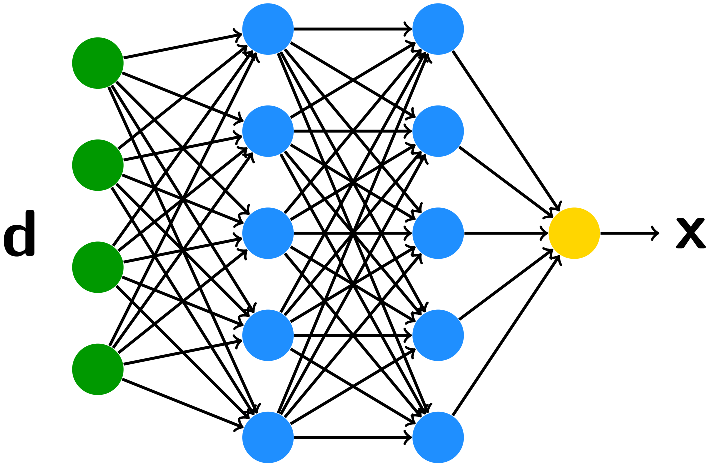
Automatic physical inference

Make simulations of data, $\{\mathbf{d}_\mathsf{i}^{\mathsf{s}\,\mathsf{fid}}, \mathbf{d}_\mathsf{i}^{\mathsf{s}\,\mathsf{fid}-}, \mathbf{d}_\mathsf{i}^{\mathsf{s}\,\mathsf{fid}+}\}$ at $\boldsymbol{\theta}^\mathsf{fid}$

Train IMNN to maximise Fisher information

Compress real data

Automatic physical inference



Automatic physical inference

Make simulations of data, $\{\mathbf{d}_i^{\mathsf{s}\,\mathsf{fid}}, \mathbf{d}_i^{\mathsf{s}\,\mathsf{fid}-}, \mathbf{d}_i^{\mathsf{s}\,\mathsf{fid}+}\}$ at $\boldsymbol{\theta}^{\mathsf{fid}}$

Train IMNN to maximise Fisher information

Compress real data

Draw $\boldsymbol{\theta}$ from $\mathsf{p}(\boldsymbol{\theta})$ and create simulation

Automatic physical inference

Make simulations of data, $\{\mathbf{d}_i^{\mathrm{s\,fid}}, \mathbf{d}_i^{\mathrm{s\,fid}-}, \mathbf{d}_i^{\mathrm{s\,fid}+}\}$ at $\boldsymbol{\theta}^{\mathrm{fid}}$

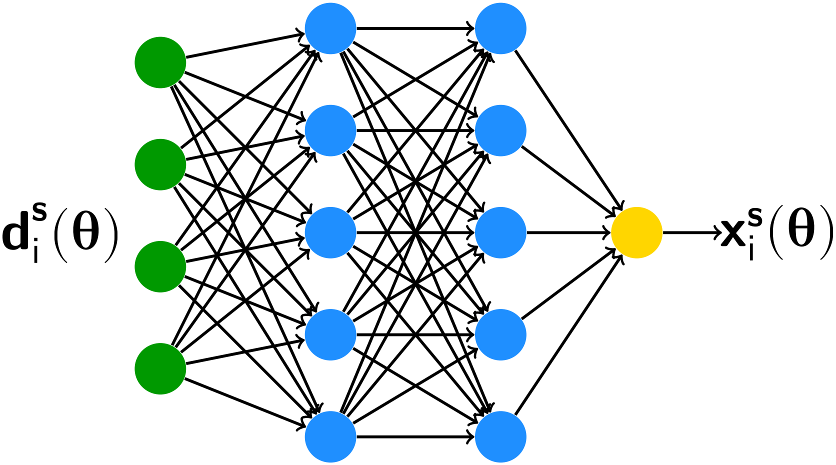
Train IMNN to maximise Fisher information

Compress real data

Draw $\boldsymbol{\theta}$ from $p(\boldsymbol{\theta})$ and create simulation

Compress simulation

Automatic physical inference



 Physical Review D **97**, 083004, github:information_maximiser, DOI 10.5281/zenodo.1175196

Automatic physical inference

Make simulations of data, $\{\mathbf{d}_i^{\mathbf{s}\,\mathrm{fid}}, \mathbf{d}_i^{\mathbf{s}\,\mathrm{fid}-}, \mathbf{d}_i^{\mathbf{s}\,\mathrm{fid}+}\}$ at $\boldsymbol{\theta}^{\mathrm{fid}}$

Train IMNN to maximise Fisher information

Compress real data

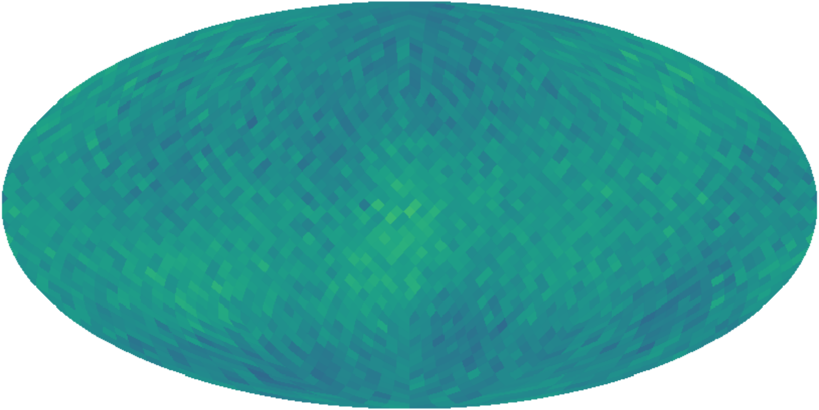
Draw $\boldsymbol{\theta}$ from $\mathrm{p}(\boldsymbol{\theta})$ and create simulation

Compress simulation

Accept or reject proposed sample to build approximate posterior distribution

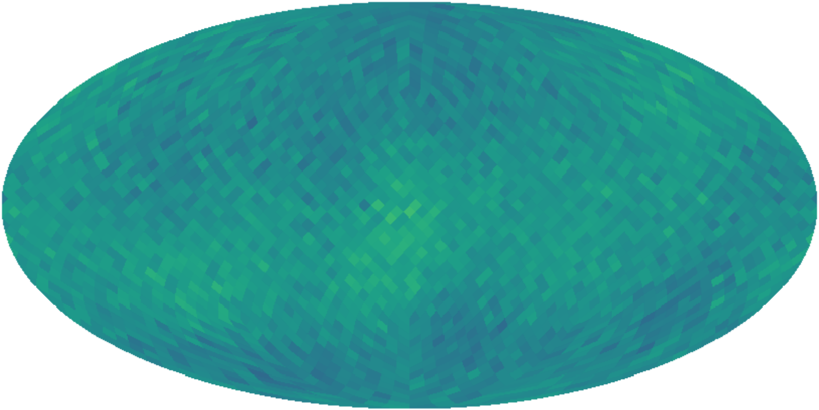
Example time!

CMB polarisation maps



 Physical Review D **97**, 083004, github:information_maximiser, DOI 10.5281/zenodo.1175196

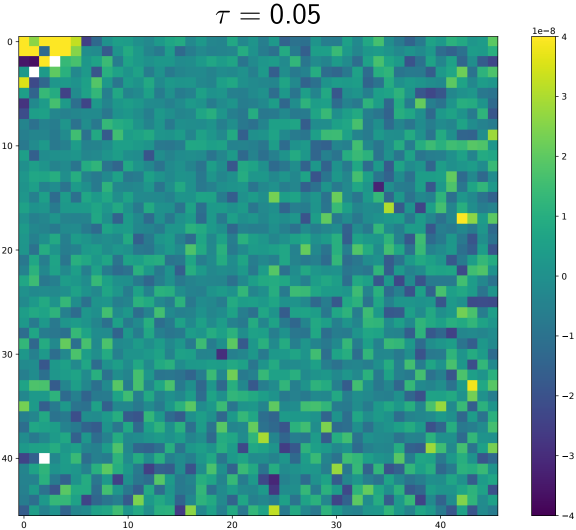
CMB polarisation maps



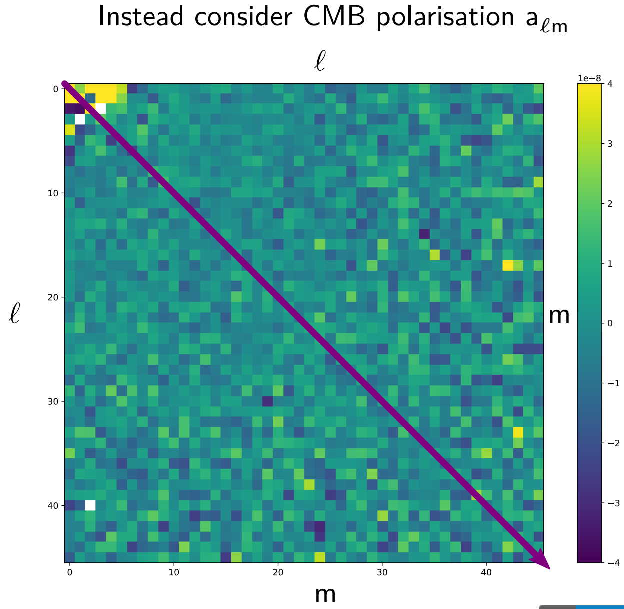
Can we infer the value of τ directly from the maps?

Yes, but we're currently improving the network

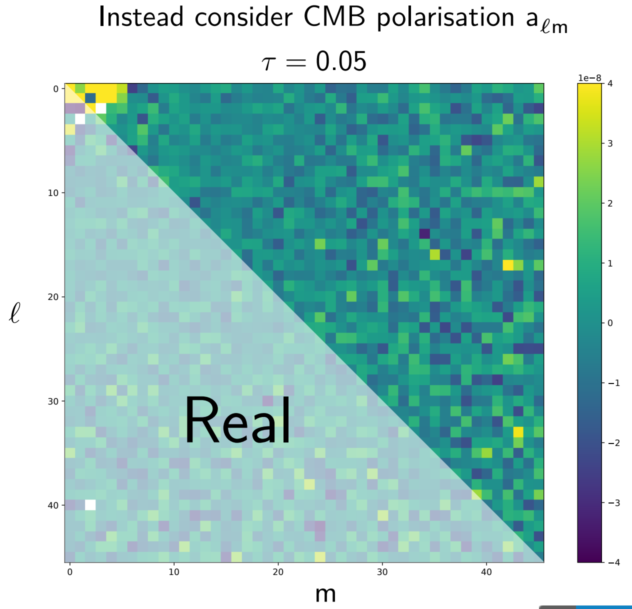
Instead consider CMB polarisation $a_{\ell m}$



Features of the $a_{\ell m}$



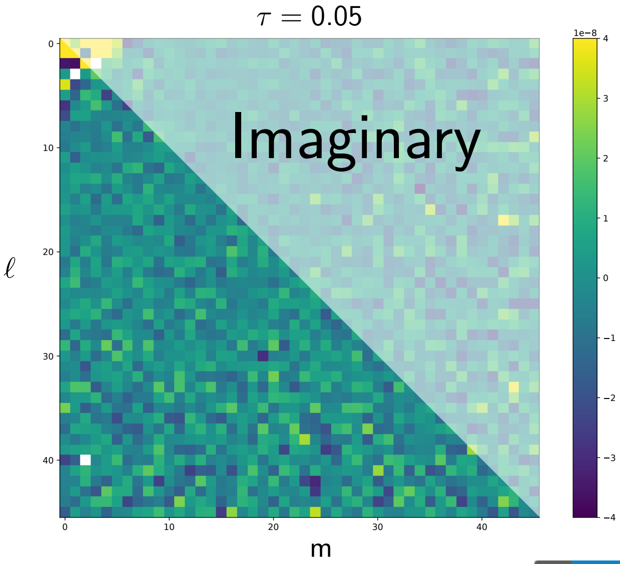
Features of the $a_{\ell m}$



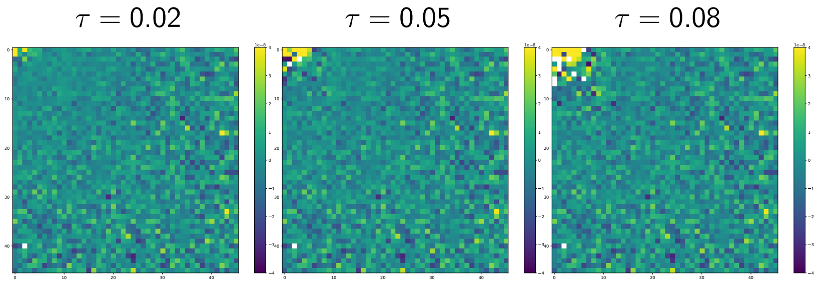
Physical Review D **97**, 083004, github:information_maximiser, DOI 10.5281/zenodo.1175196

Features of the $a_{\ell m}$

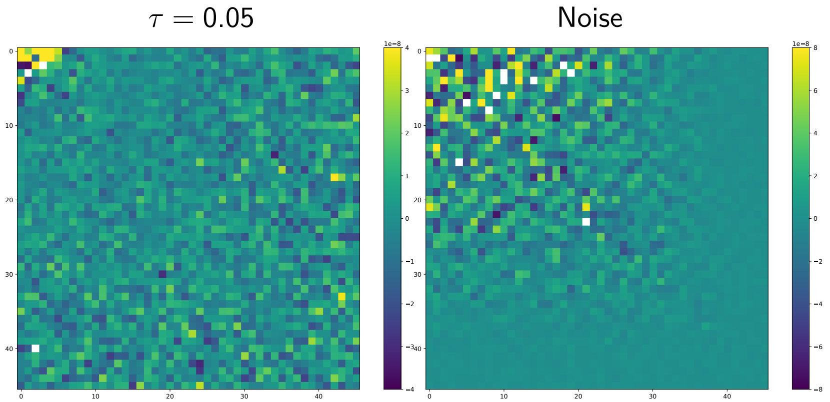
Instead consider CMB polarisation $a_{\ell m}$



Effect of τ on the $a_{\ell m}$



Simulating the $a_{\ell m}$



Automatic physical inference

Make simulations of data, $\{\mathbf{d}_\mathrm{i}^{\mathbf{s}\,\mathrm{fid}}, \mathbf{d}_\mathrm{i}^{\mathbf{s}\,\mathrm{fid}-}, \mathbf{d}_\mathrm{i}^{\mathbf{s}\,\mathrm{fid}+}\}$ at $\tau = 0.055 \pm 0.005$

Train IMNN to maximise Fisher information

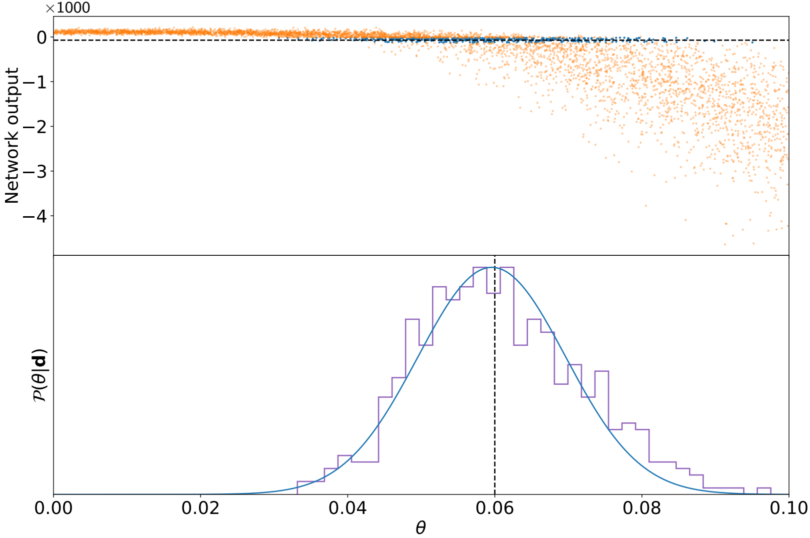
Compress real data ($\tau = 0.06$)

Draw τ from $\mathrm{p}(\tau) = [0.00, 0.10]$ and create simulation

Compress simulation

Accept or reject proposed sample to build approximate posterior distribution

Approximate posterior distribution



What I've just presented

- An optimal compression scheme for data where the likelihood isn't known, but is certainly not Gaussian

What I've just presented

- An optimal compression scheme for data where the likelihood isn't known, but is certainly not Gaussian

Things I haven't mentioned but you should ask me about (now or in the future)

- The networks ability to be robust to poor choice in fiducial parameters
- The relatively small number of simulations to train the network
- The network architecture's effect on the maximum Fisher information obtained
- That you should check out Alessandro's poster on this work
- How much the network loves to pull information out of anything you feed it!

Automatic physical inference with information maximising neural networks

Tom Charnock

Guilhem Lavaux & Ben Wandelt
Alessandro Manzotti

Institut d'Astrophysique de Paris

Physical Review D **97**, 083004

DOI 10.5281/zenodo.1175196

github:information_maximiser