Probabilistic inverse imaging methods in seismology and cosmology

Proximal MCMC and Trans-dimensional Trees

Auggie Marignier[1,2,3], Ana Ferreira[3], Thomas Kitching[2] and Jason McEwen[2]

[1]RSES, ANU [2]MSSL, UCL [3]ES, UCL

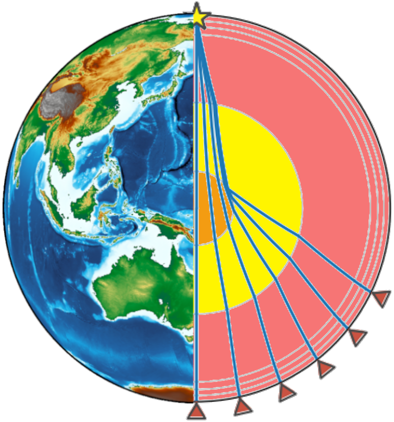
20th July 2023

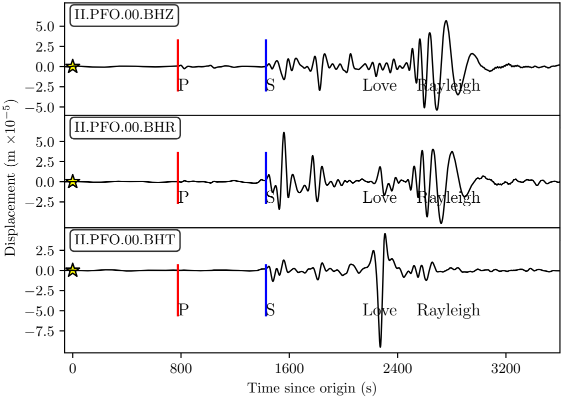
Talk outline

1. Seismology and Cosmology
2. Intro to inverse imaging
3. Bayesian imaging
4. Wavelets and Sparsity
5. Proximal MCMC on the sphere (seismology)
6. Trans-dimensional trees (cosmology)

Seismology

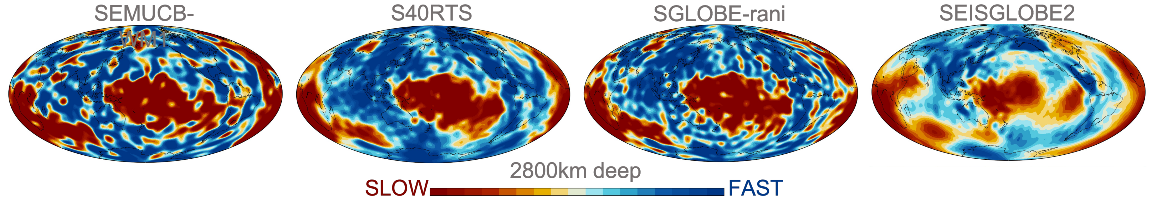
The study of earthquakes and the propagation of seismic waves through the Earth



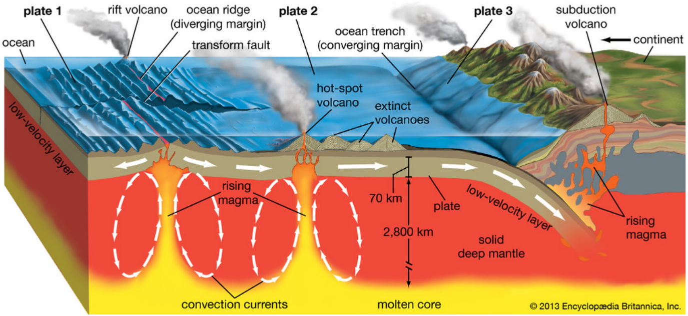


Global Seismic Tomography

Seismic tomography maps the internal structures of the Earth from measurements of seismic waves



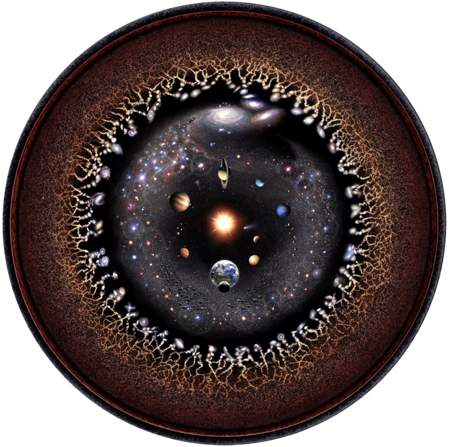
Why?



Some tectonics

Cosmology

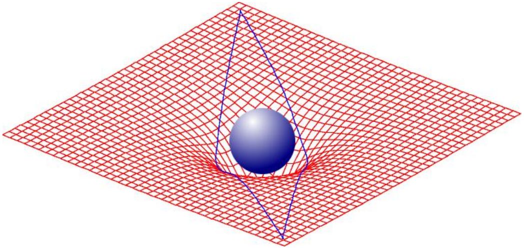
The study of the observable universe, its origins, structures, dynamics and fate. . .



Credit: Pablo Carlos Budassi

Gravitational Lensing

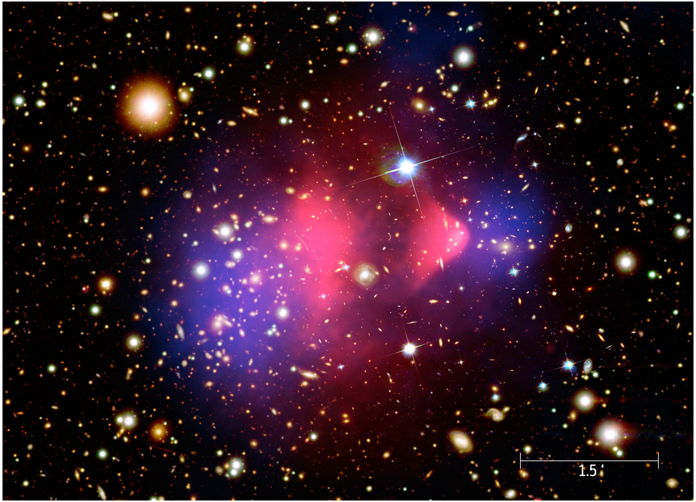
Weak lensing maps the density distribution of the universe from measurements of distorted images



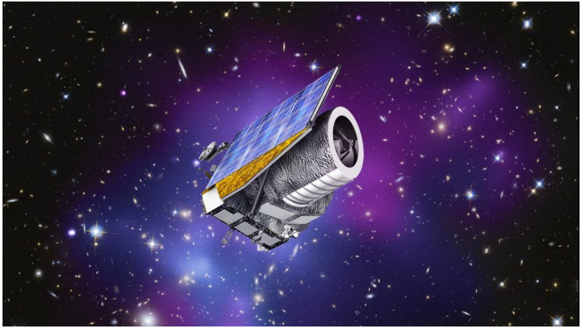
Credit: Mattias Bartelmann

Why?

"Direct empirical proof of the existence of dark matter" (Clowe et al 2004)

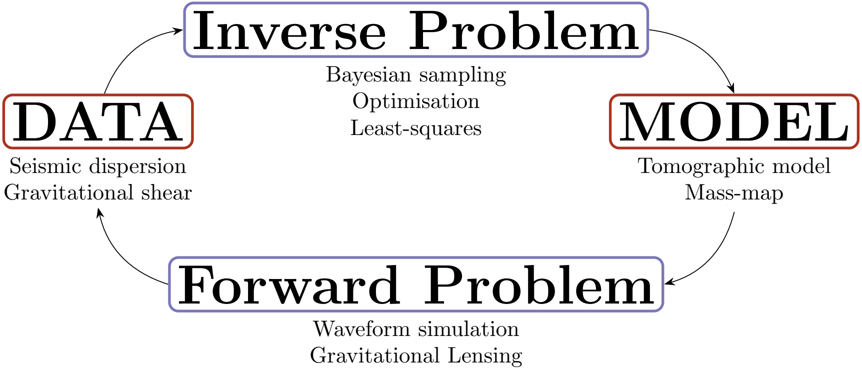


ESA EUCLID - *Exploring the Dark Universe*



VIS (visible light) instrument has 600 000 000 pixels with resolution 0.1''

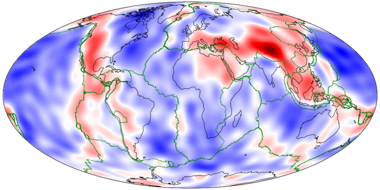
So what’s the connection?

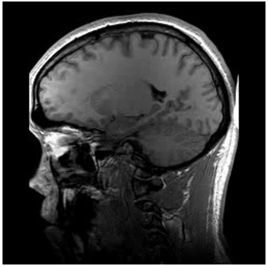


Inverse Imaging

Our aim is to retrieve an image of something we can't see using some other observable data

$$d = G(m)$$

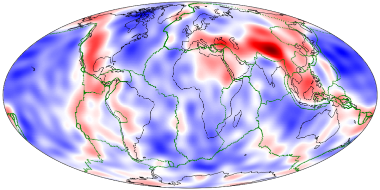


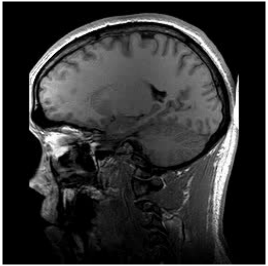


Inverse Imaging

Driven by the requirements of the applications

- Almost always ill-posed
- Non-linear forward models
- Increasing resolution requirements
- Increasing volumes of data
- Uncertainty quantification





Bayesian Inversion

POSTERIOR – what we want

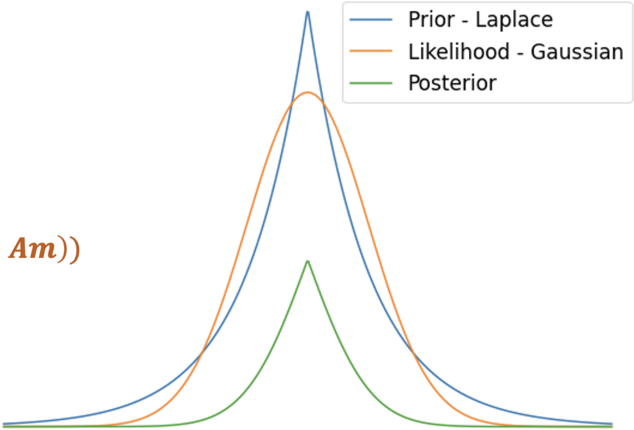
$p(\boldsymbol{m}|\boldsymbol{d}) \propto p(\boldsymbol{d}|\boldsymbol{m})p(\boldsymbol{m})$

LIKELIHOOD – what we have

$p(\boldsymbol{d}|\boldsymbol{m}) \propto \exp(-(\boldsymbol{d} - \boldsymbol{Am})^T \boldsymbol{C}^{-1}(\boldsymbol{d} - \boldsymbol{Am}))$

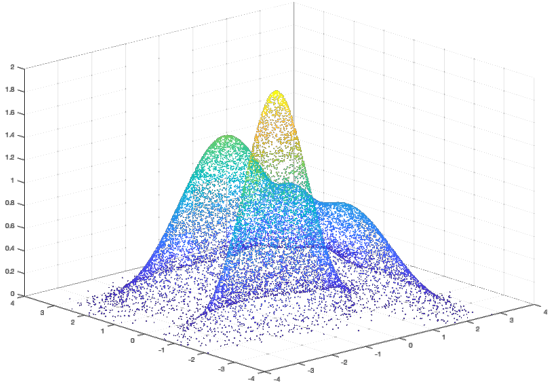
PRIOR – what we think

$p(\boldsymbol{m}) \propto \exp(-\mu|\boldsymbol{m}|)$



Posterior sampling

Repeatedly try points in parameter space and compare predictions with observed data

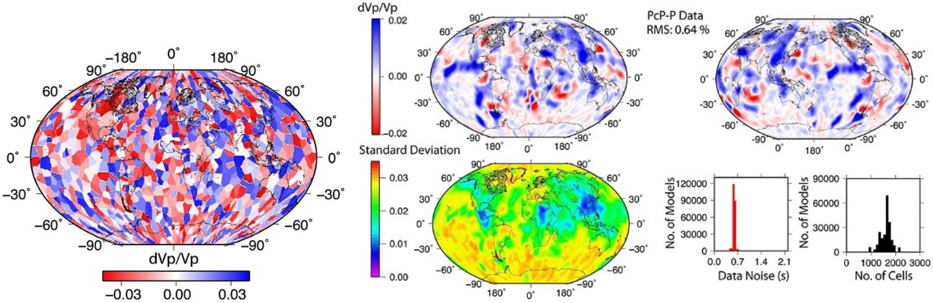


Posterior sampling

Some practical considerations:

- How many parameters?
- How long do predictions take?
- How long do proposals take?
- How many points do you try?

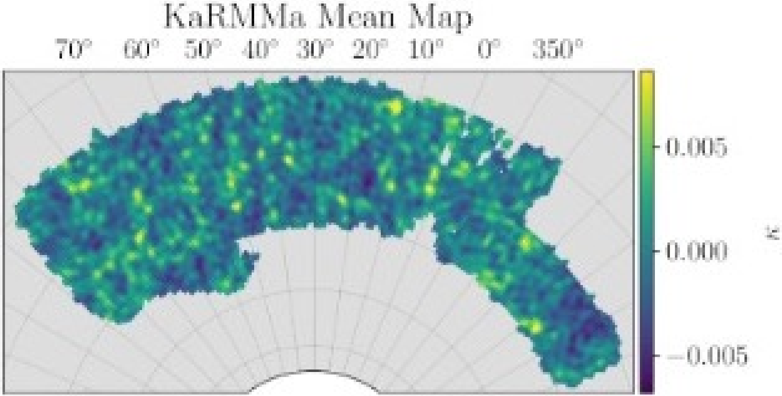
Trans-dimensional MCMC — allow the length of m to vary



Global P wave tomography of Earth's lowermost mantle from partition modelling (Young et al., 2013)

Previous Examples: Cosmology

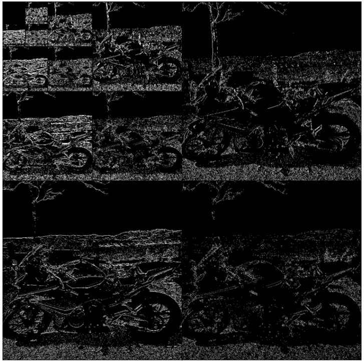
Hamiltonian Monte Carlo — imaging high-dimensional spaces



`KaRMMa` — kappa reconstruction for mass mapping (Fiedorowicz et al., 2022)

Wavelets and Sparsity

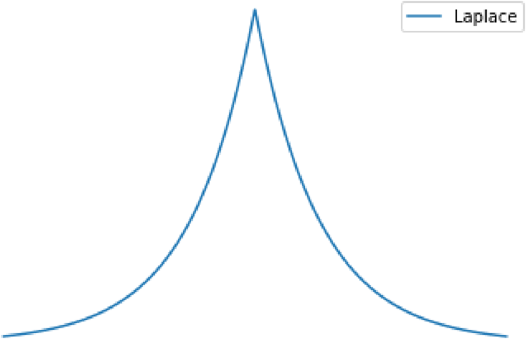
Natural images tend to be sparse in a wavelet basis, so we can use this as prior information



Wavelets and Sparsity

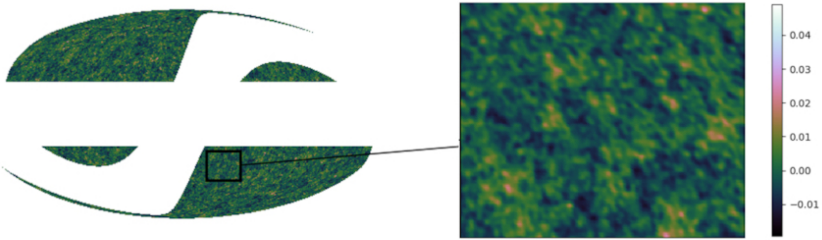
Sparsity is described by the Laplace Distribution

$$p(x) \propto e^{-|x|}$$



Previous Examples: Cosmology

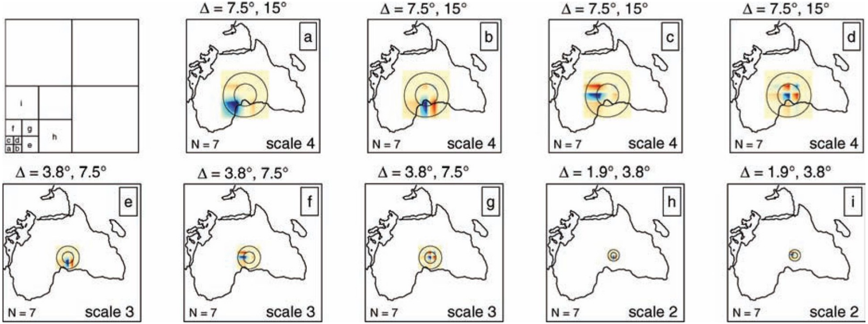
Sparse mass-mapping with proximal convex optimisation



Sparse Bayesian mass-mapping with uncertainties: Full sky observations on the celestial sphere (Price et al., 2020)

Previous Examples: Seismology

Least-squares with sparse regularisation



Solving or resolving global tomographic models with spherical wavelets, and the scale and sparsity of seismic heterogeneity (Simons et al., 2011)

Overall Aim

Advance imaging methods and uncertainty quantification in both seismic and cosmological imaging by transferring ideas from one field to the other

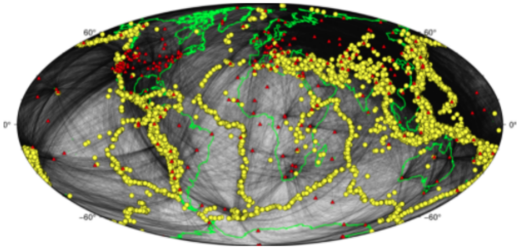
1. Proximal MCMC with wavelet priors (cosmology $\rightarrow$ seismology)
2. Transdimensional MCMC with wavelet priors (seismology $\rightarrow$ cosmology)

Talk outline

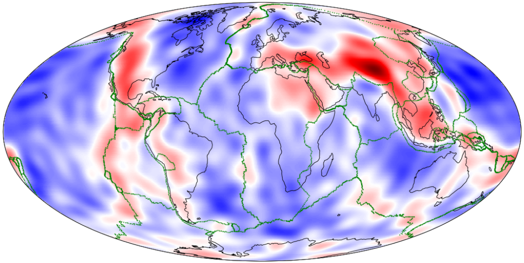
1. Intro to inverse imaging
2. Bayesian imaging
3. Wavelets and Sparsity
4. Proximal MCMC on the sphere (seismology)
5. Trans-dimensional trees (cosmology)

Build global images of surface wave phase speed from surface wave dispersion with full uncertainty quantification

Promote sparsity in a spherical wavelet basis



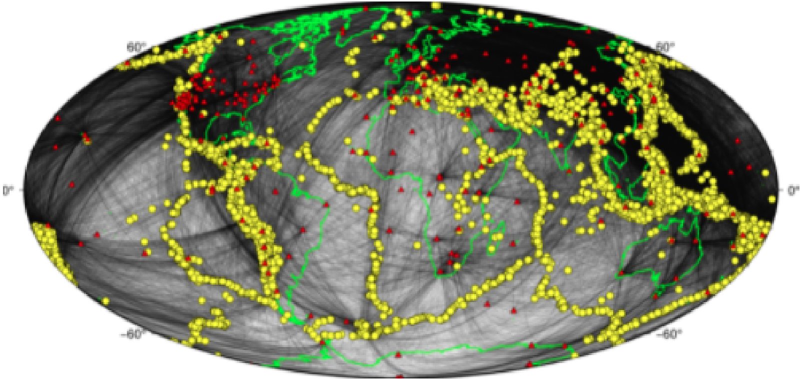
→



The Forward Model

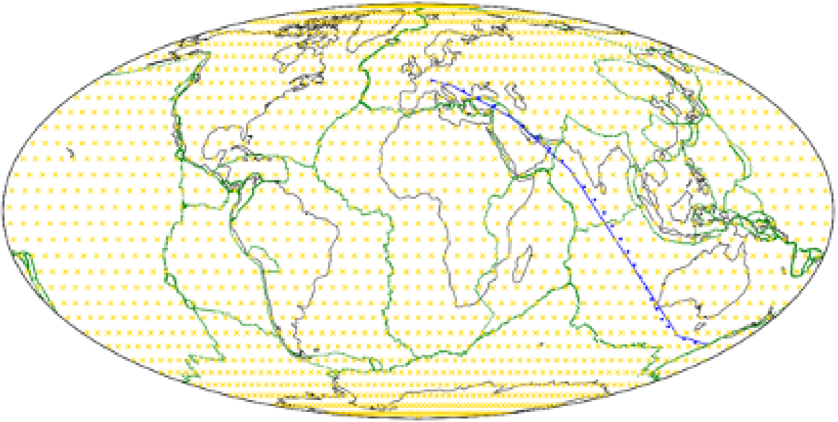
Great circle path integral of the velocity field $c(\theta, \phi)$ for all seismic sources and receivers

$$d_i = \frac{1}{\Delta} \int_{\theta_1^i, \phi_1^i}^{\theta_2^i, \phi_2^i} c(\theta, \phi) ds \quad \text{for } i = 1, \ldots, N_{\text{paths}} \sim \mathcal{O}(10^5)$$



The Forward Model in Pixel Space

Discretise the path along the surface of the sphere and evaluate the integral numerically

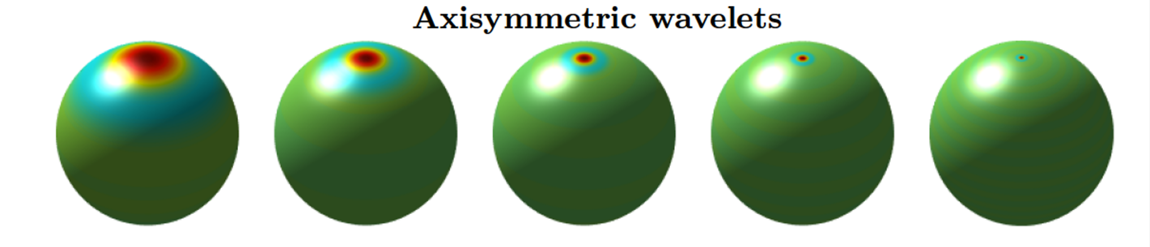


Less accurate but much faster than a common harmonic formulation!

Spherical wavelet basis

Parameter space is the space of wavelet coefficients

At bandlimit $L = 28$, this gives over 4000 parameters to sample



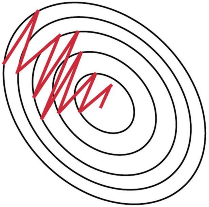
Proximal MCMC

As the number of parameters to sample increases, more complex algorithms are needed to **efficiently navigate the parameter space**

Proximal MCMC

As the number of parameters to sample increases, more complex algorithms are needed to **efficiently navigate the parameter space**

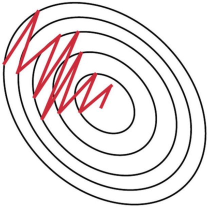
A common choice is to use the **gradient of the posterior** distribution to move towards the global maximum

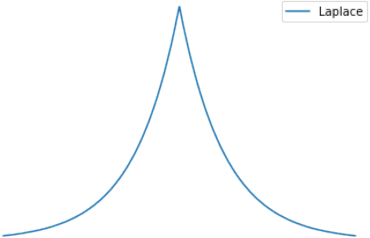


Proximal MCMC

As the number of parameters to sample increases, more complex algorithms are needed to **efficiently navigate the parameter space**

A common choice is to use the **gradient of the posterior** distribution to move towards the global maximum





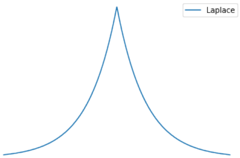
But here the posterior is not differentiable...

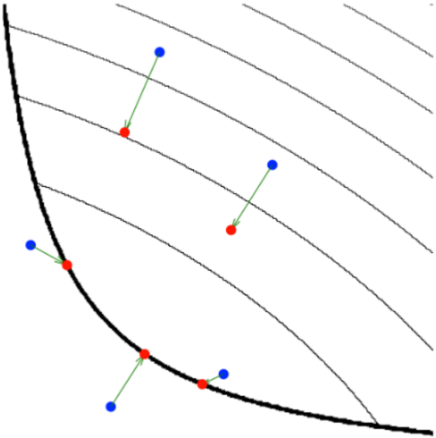
Proximity mapping vs Gradient

$$\mathbf{prox}_{\lambda f}(v) = \underset{x}{\text{argmin}} \left(f(x) + \frac{1}{2\lambda} \|x - v\|_2^2 \right)$$

proximal mapping $\sim$ gradient step in a smoothed function (MY-envelope)

Can be applied to non-smooth functions





Unadjusted Langevin Algorithm

A gradient-based MCMC sampler

To get the next chain sample you need

$$m^{(n+1)} =$$

Unadjusted Langevin Algorithm

A gradient-based MCMC sampler

To get the next chain sample you need the current sample,

$$m^{(n+1)} = m^{(n)}$$

Unadjusted Langevin Algorithm

A gradient-based MCMC sampler

To get the next chain sample you need the current sample, the gradient of the posterior,

$$m^{(n+1)} = m^{(n)} + \frac{\delta}{2}\nabla \log\left(p\left(m^{(n)}|d\right)\right)$$

Unadjusted Langevin Algorithm

A gradient-based MCMC sampler

To get the next chain sample you need the current sample, the gradient of the posterior, and some randomness

$$m^{(n+1)} = m^{(n)} + \frac{\delta}{2}\nabla \log\left(p\left(m^{(n)}|d\right)\right) + \sqrt{\delta}w^{(n)}$$

Moreau-Yosida Unadjusted Langevin Algorithm

A proximal MCMC sampler

To get the next chain sample you need

$$m^{(n+1)} =$$

Moreau-Yosida Unadjusted Langevin Algorithm

A proximal MCMC sampler

To get the next chain sample you need the current sample,

$$m^{(n+1)} = m^{(n)}$$

Moreau-Yosida Unadjusted Langevin Algorithm

A proximal MCMC sampler

To get the next chain sample you need the current sample, the gradient of the likelihood,

$$m^{(n+1)} = m^{(n)} + \delta \nabla p \left(d | m^{(n)} \right)$$

Moreau-Yosida Unadjusted Langevin Algorithm

A proximal MCMC sampler

To get the next chain sample you need the current sample, the gradient of the likelihood, the prox of the prior,

$$m^{(n+1)} = m^{(n)} + \delta \nabla p\left(d \middle| m^{(n)}\right) + \frac{\delta}{\lambda} \mathrm{prox}_{\lambda \|\cdot\|_1}\left(m^{(n)}\right)$$

Moreau-Yosida Unadjusted Langevin Algorithm

A proximal MCMC sampler

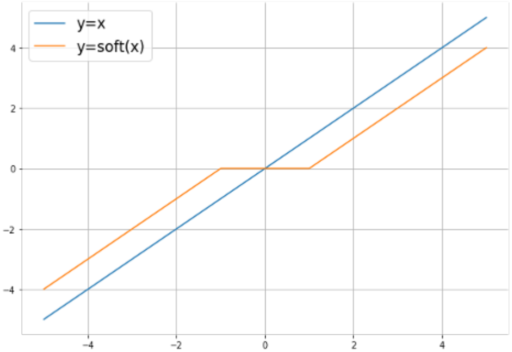
To get the next chain sample you need the current sample, the gradient of the likelihood, the prox of the prior, and some randomness

$$m^{(n+1)} = m^{(n)} + \delta \nabla p\left(d|m^{(n)}\right) + \frac{\delta}{\lambda}\mathrm{prox}_{\lambda\|\cdot\|_1}\left(m^{(n)}\right) + \sqrt{\delta}\,w^{(n)}$$

In general this involves a small convex optimisation problem

In the case of the ℓ_1-norm though, it's very simple

$$f(m) = \mu\|m\|_1 \Rightarrow \mathrm{prox}_{\lambda f}(m) = \mathrm{soft}_{\mu\lambda}(m)$$

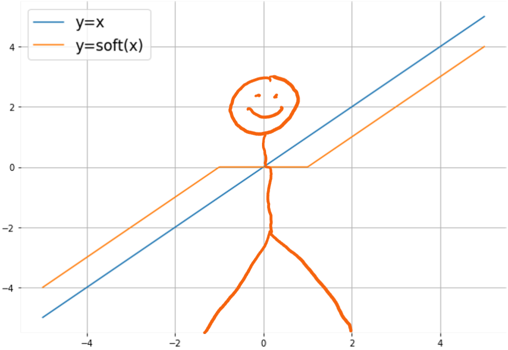


Calculating the prox

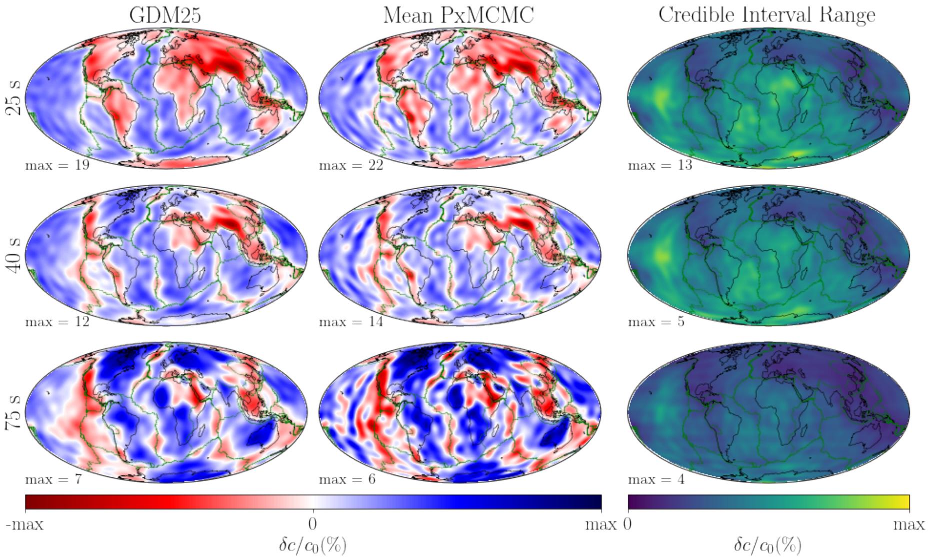
In general this involves a small convex optimisation problem

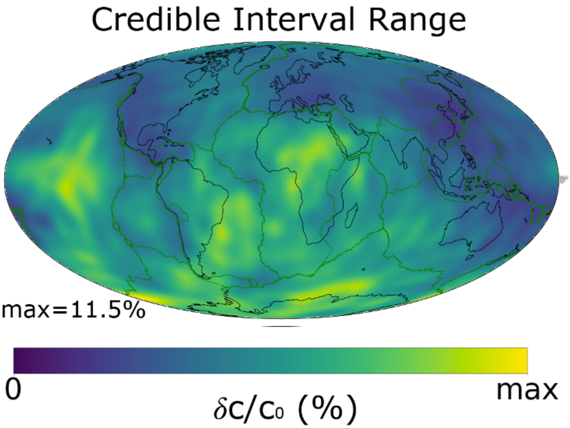
In the case of the ℓ_1-norm though, it's very simple

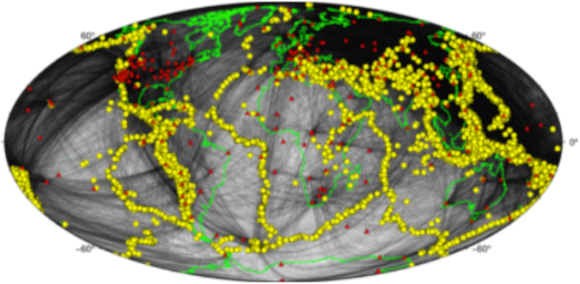
$$f(m) = \mu\|m\|_1 \Rightarrow \text{prox}_{\lambda f}(m) = \text{soft}_{\mu\lambda}(m)$$



Results


GDM25
Mean PxMCMC
Credible Interval Range
25 s
40 s
75 s
max = 19
max = 22
max = 13
max = 12
max = 14
max = 5
max = 7
max = 6
max = 4
-max
0
max
$\delta c / c_0 (\%)$
0
max
$\delta c / c_0 (\%)$


Credible Interval Range
max=11.5%
0
max
δc/c0 (%)



Push to 3D?

Defining 3D spherical wavelets significantly increases the size of the parameter space and computation time of a single iteration.
Initial tests would take well over 2 weeks to converge

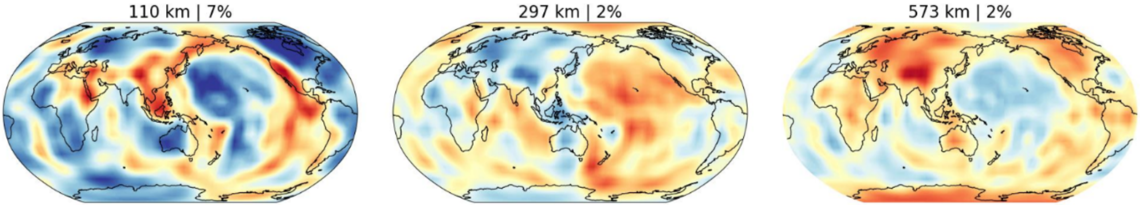


Figure: Attempts using proximal convex optimisation

Conclusions

Proximal MCMC can be used to build 2D spherical images at resolutions expected for seismology ($L < 64$)

Uncertainties are physically reasonable and **aid interpretation**

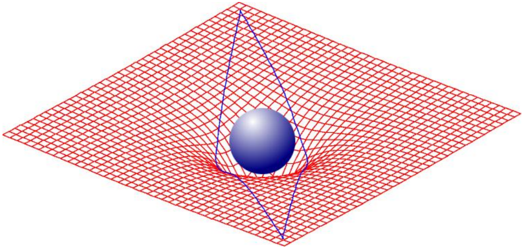
Aiming to build 3D images, but this is computationally very expensive.

Talk outline

1. Intro to inverse imaging
2. Bayesian imaging
3. Wavelets and Sparsity
4. Proximal MCMC on the sphere (seismology)
5. Trans-dimensional trees (cosmology)

Reminder: Gravitational Lensing

Weak lensing maps the density distribution of the universe from measurements of distorted images



Credit: Mattias Bartelmann

Mass-Mapping

Measurements of galaxy shapes $\rightarrow$ shear field $\gamma(\vec{\theta})$

Want convergence field $\kappa(\vec{\theta})$ (related to density)

Mass-Mapping

Measurements of galaxy shapes $\rightarrow$ shear field $\gamma(\vec{\theta})$

Want convergence field $\kappa(\vec{\theta})$ (related to density)

In Fourier space

$$\hat{\gamma}(\vec{k}) = \frac{k_x^2 - k_y^2 + 2ik_x k_y}{k_x^2 + k_y^2}\hat{\kappa}(\vec{k})$$

Measurements of galaxy shapes $\rightarrow$ shear field $\gamma(\vec{\theta})$

Want convergence field $\kappa(\vec{\theta})$ (related to density)

In Fourier space

$$\hat{\gamma}(\vec{k}) = \frac{k_x^2 - k_y^2 + 2ik_x k_y}{k_x^2 + k_y^2}\hat{\kappa}(\vec{k})$$

So our forward model is given by

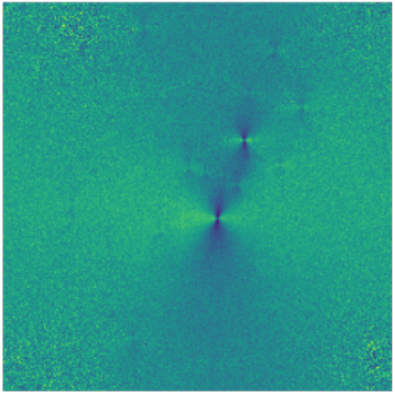
$$\gamma = \mathbf{F}^{-1}\mathbf{D}\mathbf{F}\kappa$$

where $\mathbf{F}$ ($\mathbf{F}^{-1}$) is the (inverse) FFT, and $\mathbf{D}$ is the lensing kernel

Our Aim

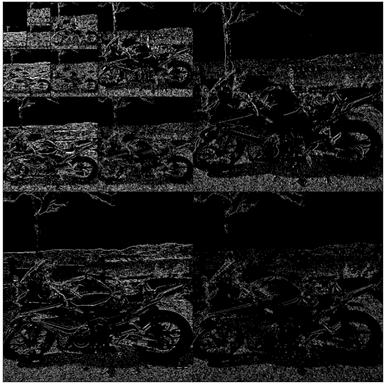
Create mass maps with full uncertainty quantification

Promote sparsity in a wavelet basis



$\rightarrow$

Wavelet Representations of Images



$256 \times 256 \Rightarrow 65,536$ wavelet coefficients $\Rightarrow$ Too many to sample!

Trans-dimensional Bayesian Inversion

Bayes Theorem

$$p(\theta|\mathbf{d}) \propto p(\mathbf{d}|\theta)p(\theta)$$

where θ is a k-dimensional vector of unknown model parameters (wavelet coefficients) where ***k*** **is also unknown**

Trans-dimensional Bayesian Inversion

Bayes Theorem

$$p(\theta|\mathbf{d}) \propto p(\mathbf{d}|\theta)p(\theta)$$

where θ is a k-dimensional vector of unknown model parameters (wavelet coefficients) where $\boldsymbol{k}$ **is also unknown**

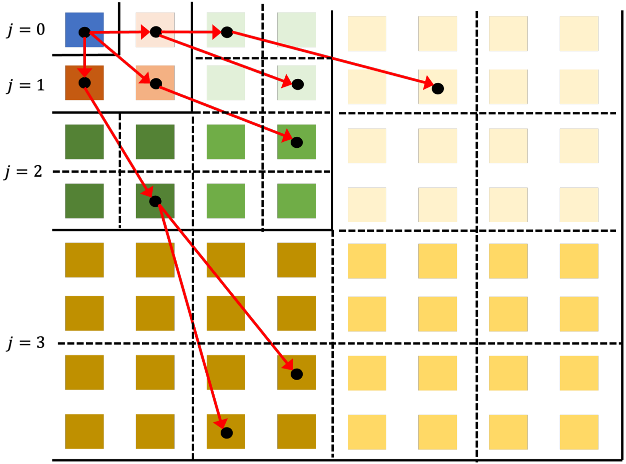
Generalising the common MCMC Metropolis-Hastings acceptance criteria gives

$$\alpha(\theta'|\theta) = \min\left\{1, \frac{p(\theta')p(\mathbf{d}|\theta')q(\theta|\theta')}{p(\theta)p(\mathbf{d}|\theta)q(\theta'|\theta)}|\mathcal{J}|\right\}$$

where $\mathcal{J}$ is the Jacobian matrix of the transformations between parameter spaces

Wavelet Tree Parameterisation

From Hawkins & Sambridge (2015)

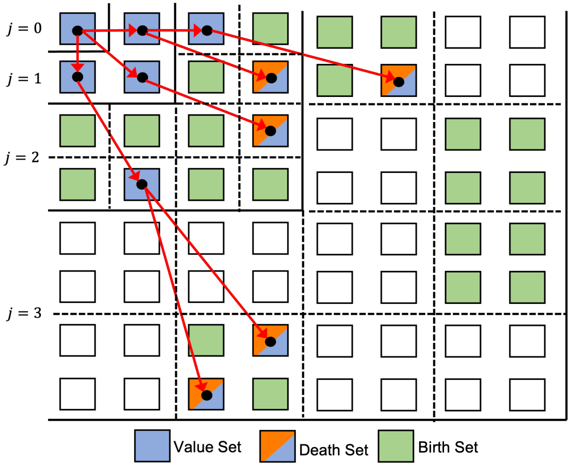


Trans-dimensional Trees

The parameter space can be divided up into three sets

1. The set of k active wavelet coefficients/tree nodes, who's value can change
2. Nodes that could possibly die
3. Nodes that could possibly be born

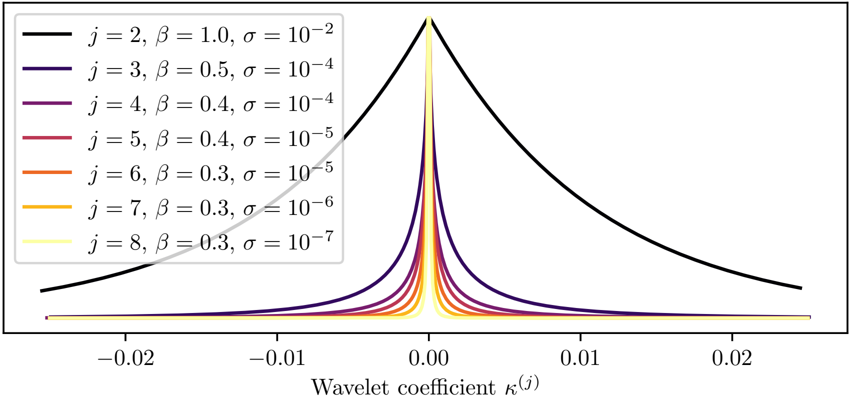
These each have their own proposal distribution $q(\theta|\theta')$



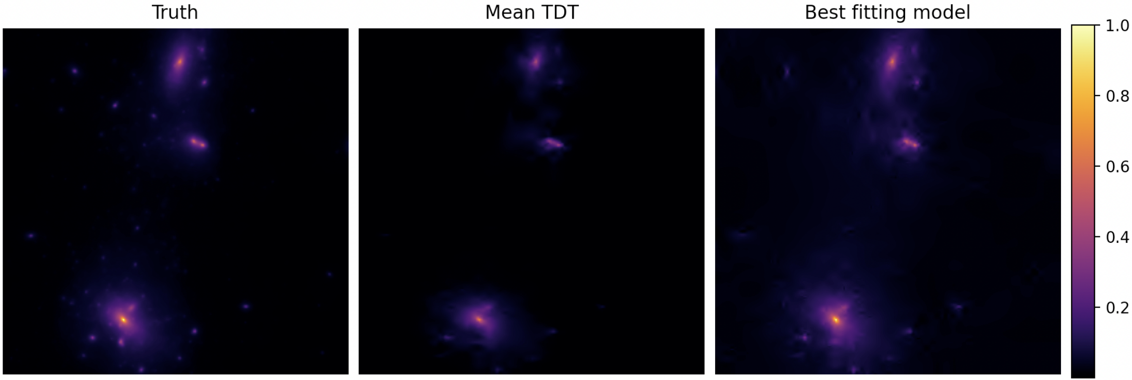
Prior on wavelet coefficients

The Generalised Gaussian Distribution (GGD)

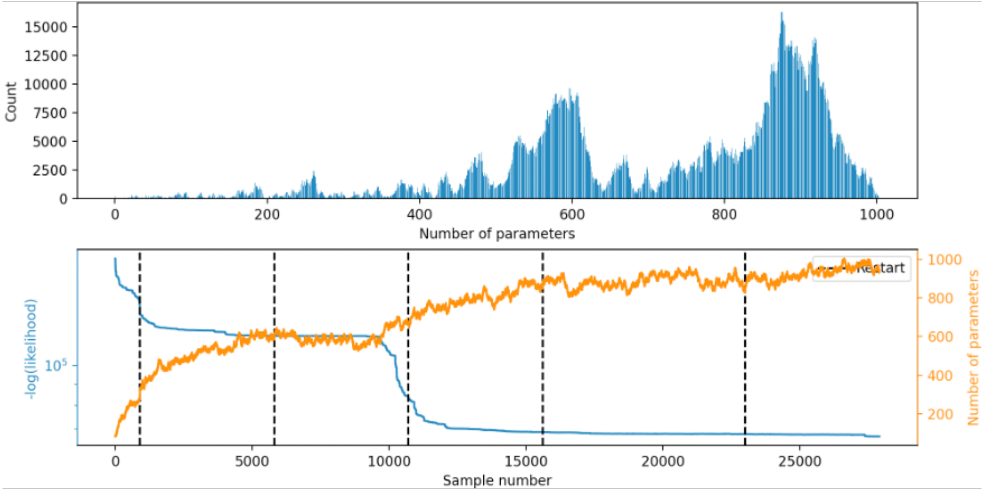
$$f(x|\mu,\sigma,\beta) = \frac{\beta}{2\sigma\Gamma(\beta^{-1})} e^{\left(-\left|\frac{x-\mu}{\sigma}\right|^{\beta}\right)}$$



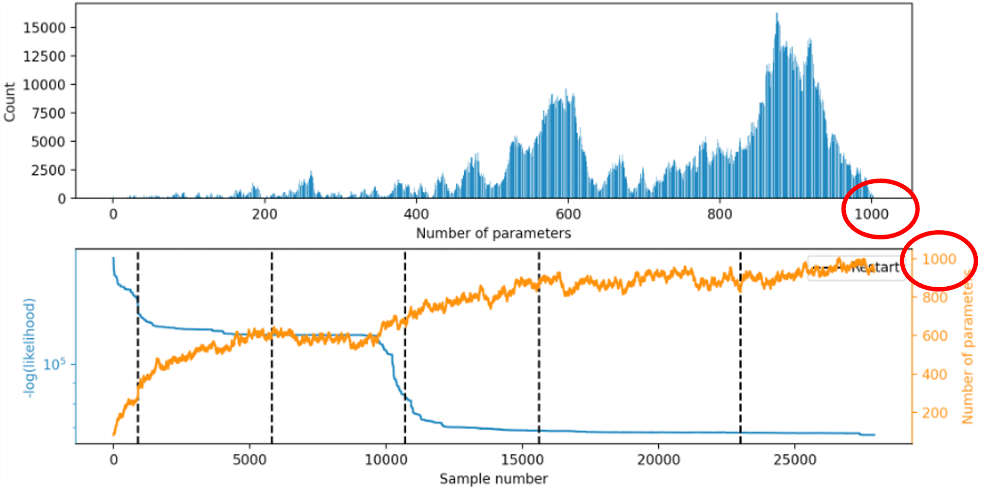
Simple synthetic test



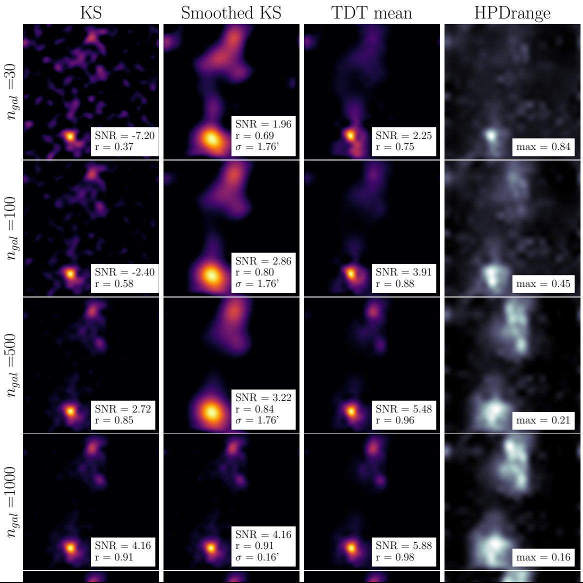
Simple synthetic test



Simple synthetic test



Realistic Noise Levels



Even with the best *Euclid* resolution, most pixels will have infinite noise!

Need to decrease the image resolution to reduce the noise per pixel

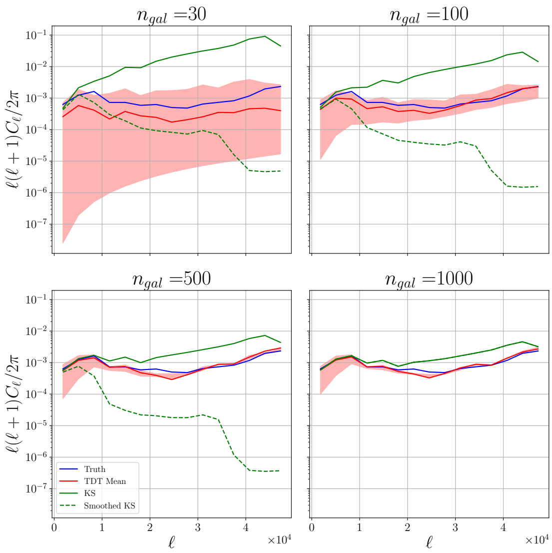
Better image reconstruction than standard methods + uncertainties

Realistic Noise Levels

Even with the best *Euclid* resolution, most pixels will have infinite noise!

Need to decrease the image resolution to reduce the noise per pixel

Better image reconstruction than standard methods + uncertainties



Conclusions

Trans-dimensional approach gives promising results on simulations

By slowly growing the parameter space, it is more efficiently sampled, making this high-dimensional problem computationally tractable

Produces better images than standard approaches even at high-noise levels

Overall Conclusions

Advanced techniques are starting to make probabilistic sampling feasible for imaging problems

Sparsity/compressed sensing is playing a significant role

Looking forward, as resolution demands increase, these efficient samplers and parameterisations will be crucial

Thank you!

Additional Slides

Spherical wavelet transform

Denote the set of spherical harmonic coefficients of some spherical signal by a hat i.e.

$$\hat{\boldsymbol{x}} = \boldsymbol{Y}\boldsymbol{x}$$

Spherical wavelet transform

Denote the set of spherical harmonic coefficients of some spherical signal by a hat i.e.

$$\hat{\boldsymbol{x}} = \boldsymbol{Y}\boldsymbol{x}$$

The spherical wavelet transform $\boldsymbol{\Psi}$ of some spherical signal $\boldsymbol{x}$ is composed of a spherical harmonic transform $\boldsymbol{Y}$ and a harmonic wavelet multiplication $\boldsymbol{W}$

$$\hat{\alpha} = \boldsymbol{W}\hat{\boldsymbol{x}}$$

$$\alpha = \boldsymbol{\Psi}\boldsymbol{x} = \boldsymbol{Y}^{-1}\boldsymbol{W}\hat{\boldsymbol{x}} = \boldsymbol{Y}^{-1}\boldsymbol{W}\boldsymbol{Y}\boldsymbol{x}$$

Spherical wavelet transform

Denote the set of spherical harmonic coefficients of some spherical signal by a hat i.e.

$$\hat{\boldsymbol{x}} = \boldsymbol{Y}\boldsymbol{x}$$

The spherical wavelet transform $\boldsymbol{\Psi}$ of some spherical signal $\boldsymbol{x}$ is composed of a spherical harmonic transform $\boldsymbol{Y}$ and a harmonic wavelet multiplication $\boldsymbol{W}$

$$\hat{\boldsymbol{\alpha}} = \boldsymbol{W}\hat{\boldsymbol{x}}$$

$$\boldsymbol{\alpha} = \boldsymbol{\Psi}\boldsymbol{x} = \boldsymbol{Y}^{-1}\boldsymbol{W}\hat{\boldsymbol{x}} = \boldsymbol{Y}^{-1}\boldsymbol{W}\boldsymbol{Y}\boldsymbol{x}$$

And the inverse spherical wavelet transform is then

$$\boldsymbol{x} = \boldsymbol{\Psi}^{-1}\boldsymbol{x} = \boldsymbol{Y}^{-1}\boldsymbol{W}^{\dagger}\boldsymbol{Y}\boldsymbol{\alpha}$$

The Forward Model in Harmonic Space

The great circle path integral can be calculated by rotating the field $c(\theta, \phi)$ to $c(\theta', \phi')$ such that $(\theta_2, \phi_1) \to (\pi/2, 0)$ and $(\theta_2, \phi_2) \to (\pi/2, \Delta)$ (i.e. the path is now along the equator)

$$\frac{1}{\Delta}\int_{\theta_1,\phi_1}^{\theta_2,\phi_2} c(\theta,\phi)ds = \sum_{\ell}\sum_{m}\left(\frac{-i}{m}\right)\left(Y_{\ell m}\left(\frac{\pi}{2},\Delta\right) - Y_{\ell m}\left(\frac{\pi}{2},0\right)\right)\sum_{n} D^{\ell}_{mn} c_{\ell n}$$

$$Y_{\ell m} = \text{spherical harmonics}$$
$$D^{\ell}_{mn} = \text{Wigner-D matrices}$$
$$c_{\ell n} = \text{spherical harmonic coefficients}$$

The Forward Model in Harmonic Space

As a matrix multiplication, if we sample the harmonic wavelet coefficients $\hat{\alpha}$ we have

$$\boldsymbol{d} = \boldsymbol{\Phi_h} \hat{\boldsymbol{c}} = \boldsymbol{\Phi_h} \boldsymbol{W}^\dagger \hat{\alpha}$$

where $\boldsymbol{\Phi_h} \in \mathbb{C}^{N_{\text{paths}} \times L^2}$ is a generally *dense* matrix representing the path integral

The Forward Model in Harmonic Space

As a matrix multiplication, if we sample the harmonic wavelet coefficients $\hat{\alpha}$ we have

$$\boldsymbol{d} = \boldsymbol{\Phi_h} \hat{\boldsymbol{c}} = \boldsymbol{\Phi_h} \boldsymbol{W}^{\dagger} \hat{\alpha}$$

where $\boldsymbol{\Phi_h} \in \mathbb{C}^{N_{\text{paths}} \times L^2}$ is a generally *dense* matrix representing the path integral

Notice that this avoids computationally expensive spherical harmonic transforms ($\boldsymbol{Y}$)

The Forward Model in Harmonic Space

As a matrix multiplication, if we sample the harmonic wavelet coefficients $\hat{\boldsymbol{\alpha}}$ we have

$$\boldsymbol{d} = \boldsymbol{\Phi_h} \hat{\boldsymbol{c}} = \boldsymbol{\Phi_h} \boldsymbol{W}^\dagger \hat{\boldsymbol{\alpha}}$$

where $\boldsymbol{\Phi_h} \in \mathbb{C}^{N_{\text{paths}} \times L^2}$ is a generally *dense* matrix representing the path integral

Notice that this avoids computationally expensive spherical harmonic transforms ($\boldsymbol{Y}$)

But $\boldsymbol{\Phi_h}$ is so large and dense that its multiplication is much slower than spherical harmonic transforms!