Warwick Summer School on Complexity and Inference - Lecture III: Case Studies

Dalia Chakrabarty[1]

[1]University of Warwick,
Department of Statistics

May 15, 2012

Outline of this lecture

- Learning the density of the gravitational mass in a distant galaxy - nonparametric inverse learning given partially sampled, noisy data.
  1. background - motivation for non-parametric approach.
  2. description of underabundance.
  3. invoking domain physics to achieve dimensionality reduction.
  4. model structure.
  5. inference using Metropolis-Hasting.
- Learning the material density and the microscopy correction function, using 2-D image data.
  1. background - harder than usual deprojection problem.
  2. description of ill posedness.
  3. to overcome ill posedness - expanding information, identifying priors, eliciting from literature.
  4. resolution in data guides model structure.
  5. inference - Metropolis within Gibbs.
  6. uniqueness considerations.

Gravitational mass density of a galaxy

Understanding dark matter is one of the fundamental challenges in science today. On the scale of galaxies, this reduces to the estimation of the density of gravitational mass $\rho(x, y, z) > 0,\ \rho : \mathbb{R}^3 \longrightarrow \mathbb{R}_{>0}$ of all matter in the galaxy - dark as well as the luminous mass. Gravitational mass density of luminous matter $\rho_L(x, y, z)$ is achieved by modelling telescopic image data using astronomical modelling. Then dark matter mass density is given by $\rho(x, y, z) - \rho_L(x, y, z)$. So we need to learn $\rho(x, y, z)$.

Gravitational mass density of elliptical galaxies

- $\rho(x, y, z)$ of elliptical galaxies is hard to achieve.

Dark Matter & Galaxies

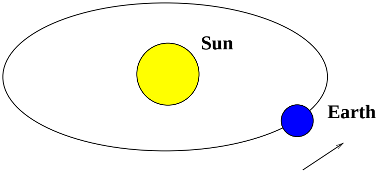
Light distribution from an example “elliptical” galaxy;

Gravitational mass density of elliptical galaxies

Hard to achieve. Data include:

- line-of-sight velocity component of individual galactic particles - noisy, *partially sampled* from the system phase space, can be modelled to estimate $\rho(x, y, z)$ given a model for the phase space probability density function.
- gravitational lensing observations - can be modelled to estimate projections of gravitational mass distribution.
- X-ray measurements - can be modelled under the assumption of hydrostatisc equilibrium and a model for emissivity, using very noisy temperature measurements.

Earth - Sun system



Motion tracks gravitational mass

Earth going around Sun in circular orbit, at distance D from Sun, with circular speed v_C:

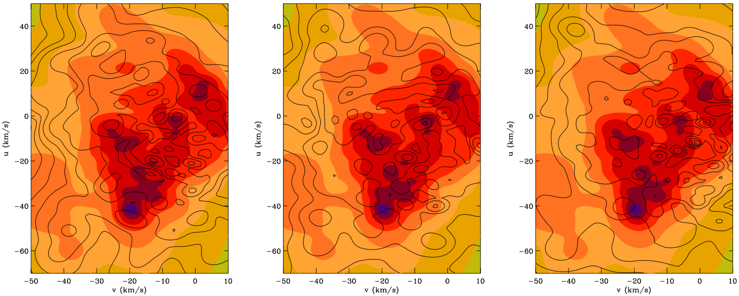
$$\frac{GM_{Sun}}{D^2} = \frac{v_C^2}{D}$$

$$\text{Here } M_{Sun} \equiv M_{enclosed}(D) = \int_0^D \rho(r) 4\pi r^2 dr \qquad (1)$$

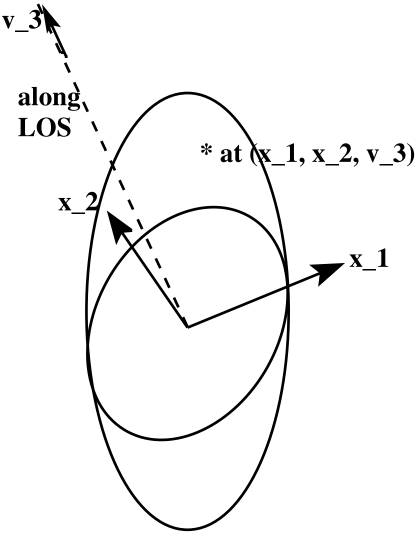
Similarly, motion of galactic particle traces mass enclosed within orbit.

In general it is not possible to express $M(D)$ as function of $|\mathbf{v}|$ of galactic particle (eg. star) because

- only 1 component of **v** is measurable - along LOS.
- particle orbital shape is unknown, i.e. $f(\mathbf{w})$ unknown - need to know $f(\cdot)$.
- strong non-linear effects exist (Chakrabarty 2004, 2007; Chakrabarty & Sideris 2009).



Inside the system



Measurables:

- 1 out of 3 components of **v** of individual resident particles of an astrophysical system, viz. galaxy - v_3.
- where on the image this particle is - x_1, x_2.
- δv_3 - measurement error in v_3.
- measured with instruments mounted on telescopes.
- 1-time measurements.
- no pattern in spatial sampling of v_3 data.

Measurement errors in v_3

Astronomers, have suggested that the error distribution is Gaussian with variance that is due to the astronomical apparataus, i.e.

$$\epsilon_{v_3} \sim \mathcal{N}(0, \sigma_0^2)$$

where σ_0 is given.

A generic modelling strategy for autonomous, hamiltonian dynamical sytems is being motivated here. So we want to leave incorporation of measurement uncertainties generic. Thus. our model works even when noise distribution is non-Gaussian.

Statement of the problem

$\mathbf{D} = \xi[\rho(x, y, z)]$, where $\rho(x, y, z)$ is unknown and sought. Data is unknown functional of unknown gravitational mass density. Here $\mathbf{D} = \{X_1^{(k)}, X_2^{(k)}, V_3^{(k)}\}_{k=1}^{N_{data}}$

1. Learning a function given measurement vectors? Discretise and learn model parameter vector instead of system function. Seek $\boldsymbol{\rho} = (\rho_1, \rho_2, \ldots, \rho_{N_{rad}})^T$.
2. Comparing model to data? Choose a likelihood function - Gaussian perhaps - of the distance between data **D** and function of $\boldsymbol{\rho}$ that compares with vector $(X_i^{(k)}, X_2^{(k)}, V_3^{(k)})^T$, $= 1, \ldots, N_{data}$. $g(\boldsymbol{\rho}) \longrightarrow$ discrete phase space coordinates? Not possible! Discard this formulation.
3. Consider projecting phase space probability density function - with $\rho(x, y, z)$ embedded in its structure - onto subspace of observables.
4. How to link phase space *pdf* and $\rho(x, y, z)$? Physics!

Galaxies as Mechanical Systems

- Evolution - through iterated applications of *dynamical rule* $g(\mathbf{x}, \mathbf{v}, t)$, $\mathbf{X} \in \mathbb{R}^3$, $\mathbf{V} = \dot{\mathbf{X}}$.
- Evolution of $\mathbf{w_0}$ is deterministic - equations of motion under deterministic (but unknown) gravitational potential $\Phi(\mathbf{x}, \mathbf{v}, t)$
  $g : \mathbb{R}^{3n} \times \mathbb{R}^{3n} \times \mathbb{R}_{>0} \longrightarrow \mathbb{R}$:
  Equations of motion $\ddot{\mathbf{x}} = g(\mathbf{x}, \mathbf{v}, t)$.
- **First model assumptions**:*autonomous Hamiltonian system*.
- $g(\cdot) = -\nabla\Phi(\mathbf{x})$
- Gravitational potential $\nabla^2\Phi(\mathbf{x}) = -4\pi G\rho(\mathbf{x})$, Poisson equation.

Phase space *pdf*

Galactic phase space $\mathcal{W}$ is the space of all possible states - V and X of all galactic particles; W $= (\mathbf{X}, \mathbf{V})^T$. Thus, $\mathcal{W} \subseteq \mathbb{R}^6$.

Phase space *pdf* is $f(\mathbf{x}, \mathbf{v})$, where $f : \mathbb{R}^6 \longrightarrow \mathbb{R}_{\geq 0}$.

Given data D, want to learn

- dynamical rule (functional of $\rho(\mathbf{x})$) that determines evolution of $\{\mathbf{w_0}\}$,
- by embedding $\rho(\mathbf{x})$ in hidden phase space density $f(\mathbf{x}, \mathbf{v})$.
- So we also need to learn $f(\mathbf{x}, \mathbf{v})$.

- Evolution of $f(\mathbf{x}, \mathbf{v})$ is given by Collisionless Boltzmann Equation - *difference with non-stationary inverse problems, though connection possible*.

$$\frac{df(\mathbf{x}, \mathbf{v})}{dt} = 0 \qquad (2)$$

Jeans th; f is a stationary solution of the CBE if it depends on phase space coordinates via integrals of motion. Inversely any function of integrals of motion, is a solution of the CBE.

Invoke domain physics to perform model structuring

- **Second model assumption about topology of phase space**.
- Simplest to assume that phase space is isotropic, so that $f(\mathbf{x}, \mathbf{v})$ is an isotropic function.
- $f(\mathbf{x}, \mathbf{v})$ is an isotropic function if $f(\mathbf{x}, \mathbf{v}) = f(\mathbf{Qx}, \mathbf{Qv})$, for any orthogonal transformation matrix $\mathbf{Q}$ (Truesdell, Noll & Antman 2004; Wang 1969).
- Then, the set of invariants (with respect to $\mathbf{Q}$) of this isotropic scalar function is $\Upsilon_Q = \{\mathbf{x}\cdot\mathbf{x}, \mathbf{v}\cdot\mathbf{v}, \mathbf{x}\cdot\mathbf{v}\}$.
- Then *pdf* of isotropic phase space admits representation $f(\Upsilon_Q) \equiv f(\mathbf{x}\cdot\mathbf{x}, \mathbf{v}\cdot\mathbf{v}, \mathbf{x}\cdot\mathbf{v})$.

Isotropic phase space *pdf*

Phase space pdf admits representation $f(\Upsilon_Q) \equiv f(\mathbf{x}\cdot\mathbf{x}, \mathbf{v}\cdot\mathbf{v}, \mathbf{x}\cdot\mathbf{v})$.
Recall that f can depend on $\mathbf{x}, \mathbf{v}$ through integrals of motion, aka energy (E), angular momentum (L). Achieved, if phase space is represented as $f(E)$, where $E := \Phi(\sqrt{\mathbf{x}\cdot\mathbf{x}}) + h(\mathbf{v}\cdot\mathbf{v})$.
For the physical interpretation of E as particle energy, $\Phi(|\mathbf{x}|)$ the potential energy, $h(|\mathbf{v}|) \equiv |\mathbf{v}|^2/2$ the kinetic energy.

Other integrals of motion dependent on **x** and **v** as $\Upsilon_Q(\mathbf{x}\cdot\mathbf{x}, \mathbf{v}\cdot\mathbf{v}, \mathbf{x}\cdot\mathbf{v})$? Possible perhaps $\longleftarrow$ more work on the dynamics aspect needs to be input.

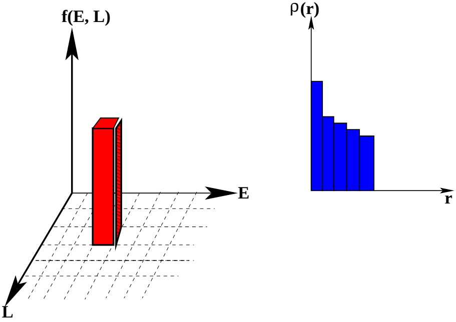
Isotropic phase space implies

- Gravitational potential and thereby grav. mass density depends on $|\mathbf{x}| = x_1^2 + x_2^2 + x_3^2 := r^2$.
- phase space *pdf* is $f(E) = f[\rho(r), v_1^2, v_2^2, v_3^2]$.

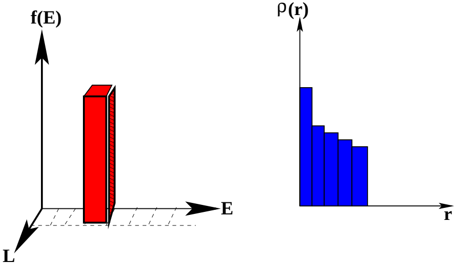
In discretied model,

- discretise range in R - $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_{N_{rad}})^T$, where ρ_i is grav mass density for $r \in [r_{i-1}, r_i)$, $i = 1, \ldots, N_{rad}$.
- discretise range of E - $\mathbf{f} = (f_1, \ldots, f_{N_{eng}})^T$, where f_i is phase space *pdf* for $E \in [E_{j-1}, E_j)$, $j = 1, \ldots, N_{eng}$.

Isotropy $\Longrightarrow$ 1-D mass density histogram.



$f = f(E, L)$ or anisotropic phase space density is akin to a 2-D histogram structure; if $f = f(E)$ is 1-D histogram



Bayesian inverse nonparametric learning of ρ, f

Motion tracks gravitational field implies $\mathbf{D} = \xi(\boldsymbol{\rho})$, where $\boldsymbol{\rho}$ is unknown and sought. **Inverse learning is achieved by attempting the forward problem iteratively**. Likelihood -

projection of $f(E)$ onto the subspace $\mathcal{M} \subset \mathcal{W}$ of these observables, the *pdf* of which is $\nu(x_1, x_2, v_3)$. Assume i.i.d. observable vectors $(x_1^{(k)}, x_2^{(k)}, v_3^{(k)})^T$, $k = 1, \ldots, N_{data}$.

Bayesian inverse nonparametric learning of ρ, f

$$
\begin{aligned}
\mathcal{L} &= \prod_{k=1}^{N_{tot}} \nu(x_1^{(k)}, x_2^{(k)}, v_3^{(k)}) \\
\nu(x_1^{(k)}, x_2^{(k)}, v_3^{(k)}) &= \int_{X_3}\int_{V_1}\int_{V_2} f[E(x_1^{(k)}, x_2^{(k)}, x_3, v_1, v_2, v_3^{(k)})] dx_1 dx_2 dv_3 \\
E(x_1^{(k)}, x_2^{(k)}, x_3, v_1, v_2, v_3^{(k)}) &= \Phi(\sqrt{\{x_1^{(k)}\}^2 + \{x_2^{(k)}\}^2 + \{x_3\}^2}) + \\
&\quad \frac{(\{v_1\}^2 + \{v_2\}^2 + \{v_3^{(k)}\}^2)}{2}.
\end{aligned}
$$

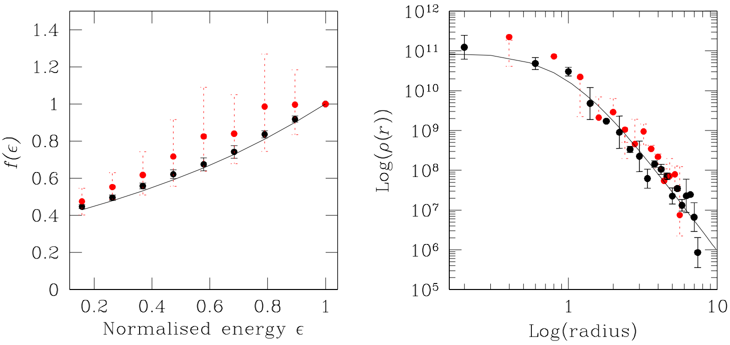
f and ρ are learnt by sampling from the posterior $\Pr[\mathbf{f}, \rho | \{x_1^{(k)}, x_2^{(k)}, v_3^{(k)}\}_{k=1}^{N_{tot}}]$**, using random-walk Metropolis-Hastings.**

Inference

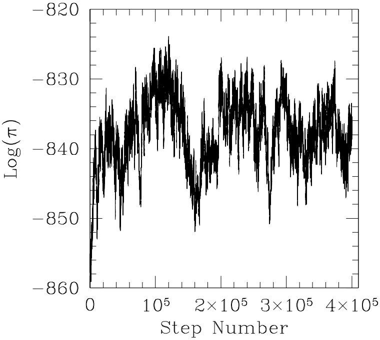
- Positivity for both $\rho(r)$, $f(E)$.
- Monotonic decline for $\rho(r)$. We also choose monotonic declie for $f(E)$ since parametric astronomical models suggest this,but not correct representation. Thus, monotonicity is left modular in the inference scheme.

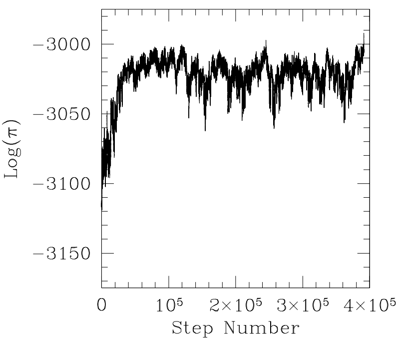
Propose $\tilde{\Delta}_\rho^{(h)} = \Delta_\rho^{(h)} \exp(-\alpha/s)$,
where $\Delta_\rho^{(h)} := \rho_h - \rho_{h+1}$,
$\alpha \sim \mathcal{U}$[-0.5,0.5], s is a constant. $\rho_{N_x+1} := 0$, $\tilde{\rho}_{N_x+1} := 0$.
Then as h is varied from N_x to 1, the proposed h-th component of the unknown gravitational mass density vector is
$\tilde{\rho}_h = \tilde{\rho}_{h+1} + \tilde{\Delta}_\rho^{(h)}$.
Similarly for updating f_j.

Results


1.4
1.2
1
0.8
0.6
0.4
0.2
0
0.2 0.4 0.6 0.8 1
$f(\epsilon)$
Normalised energy ϵ
10^{12}
10^{11}
10^{10}
10^{9}
10^{8}
10^{7}
10^{6}
10^{5}
1 10
$\mathrm{Log}(\rho(r))$
Log(radius)

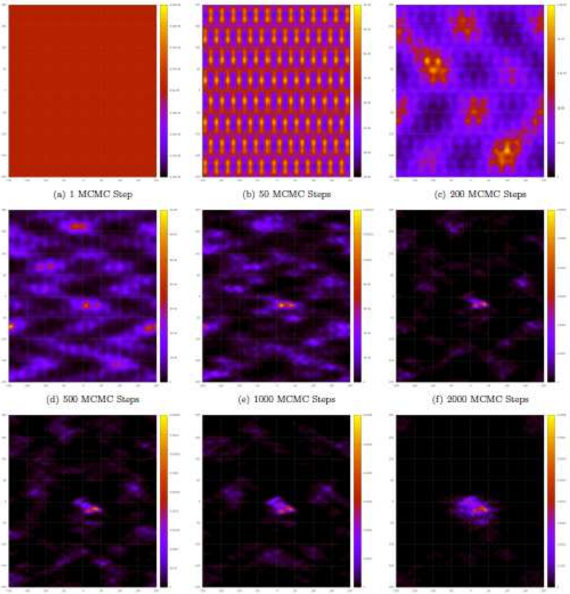
Results





$\longrightarrow$ Test for support in data for the model assumption of isotropic phase space - develop bespoke distribution-free test of hypothesis that work in the context of non-parametric inference.

MCMC performance



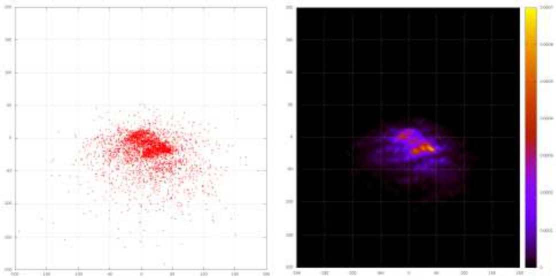
(a) 1 MCMC Step (b) 50 MCMC Steps (c) 200 MCMC Steps

(d) 500 MCMC Steps (e) 1000 MCMC Steps (f) 2000 MCMC Steps

(g) 3000 MCMC Steps (h) 5000 MCMC Steps (i) 10^4 MCMC Steps

Warwick Statistics

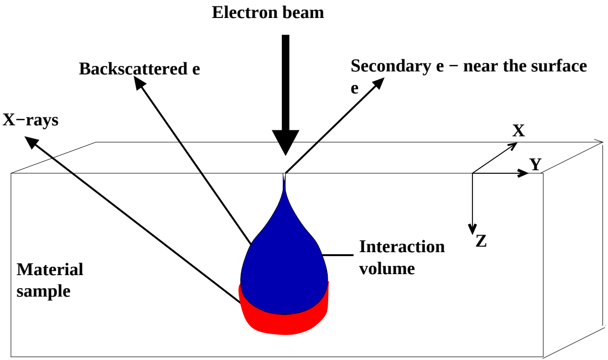
MCMC performance



(a) Samples (b) PDF

From White, Cuevus & Chakrabarty, (under prep)

A demanding inverse problem - motivation



Imaging with Bulk Electron Microscopy - in X-rays, Back Scattered Electrons (BSE), etc., using Scanning Electron Microscope (SEM)

- Electron beam impinges on material, penetrates sample surface.
- Electron irradiation causes interaction of the beam electrons with sample atoms ... interaction bounded by “interaction volume” (morphology dependent on material density, electron beam energy, material characteristics, etc.).
- Results in various kinds of emission (X-rays, secondary electrons, backscattered electrons BSE) - emission can be captured by SEM in 2-D pixels, as image.
- Generated radiation modulated by unknown, material-specific blurring function.
- Image is formed due to projection of generated radiation along the Z-axis, subsequently averaged over lateral extent of “interaction volume”.

A demanding inverse problem

$$I(x_i, y_i) = \langle \mathcal{P}(\rho * \eta) \rangle + \varepsilon \quad \text{or} \tag{4}$$

$$I(x_i, y_i) = \frac{\int_{r=0}^{r_{max}} \int_{\theta=0}^{\theta_{max}} rdrd\theta \int_{z=0}^{z=z_{max}(r,\theta)} \rho(\mathbf{x}) * \eta(\mathbf{x}) dz}{\int_{r=0}^{r_{max}} \int_{\theta=0}^{\theta_{max}} rdrd\theta} + \varepsilon$$

for $i = 1, \ldots, N_{data}$.
$\mathbf{X} = (R, \Theta, Z)^T = (X\ Y\ Z)^T$, with
$x - x_i = r\cos\theta$, $y - y_i = r\sin\theta$
interaction-volume: $r \in [0, r_{max}]$, $\theta \in [0, \theta_{max}]$, $z \in [0, z_{max}]$.

Novelty of problem

Operator $\mathcal{P}$ is projection onto the Z=0 plane, followed by spatial-averaging over the lateral cross-section of the "interaction-volume", resulting in a contractive projection of $\rho \star \eta$ onto the centre of the interaction-volume $\Longrightarrow$ a sequence of inversions of (image) data results in convolution of material density and correction function.

Correction function unknown, material density unknown.

Noise present in image data.

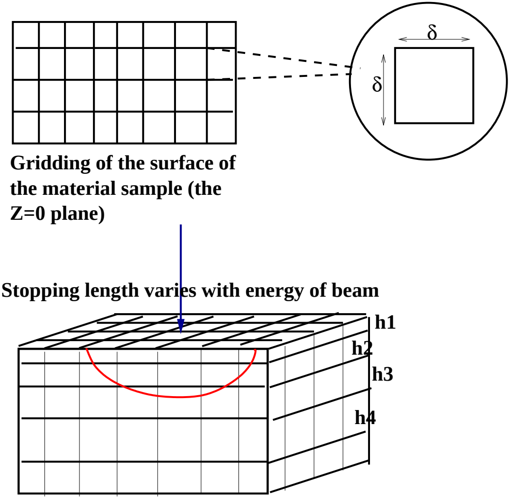
Novelty of solution

Fully discretised model - helps perform forward model iteratively while writing posterior, using likelihood that is function of distance between projected+spatially-averaged $\rho * \eta$.

Expand data space by imaging at multiple beam energies, ϵ_k, $k = 1, \ldots, N_{eng}$

Identify "geometric" priors on $\rho(\mathbf{x})$ + strong priors on $\eta(z)$ via elicitation - helps identifiability.

Uniqueness of solutions for $\rho * \eta$ demonstrated in low noise limit, condition number of problem of learning $\rho * \eta$, in the presence of measurement uncertainties recalled; deviation of uniqueness of solutions for $\rho(\mathbf{x})$, $\eta(z)$ quantified.



Gridding along Z−axis is non−uniform; #Z−bins = #beam energies

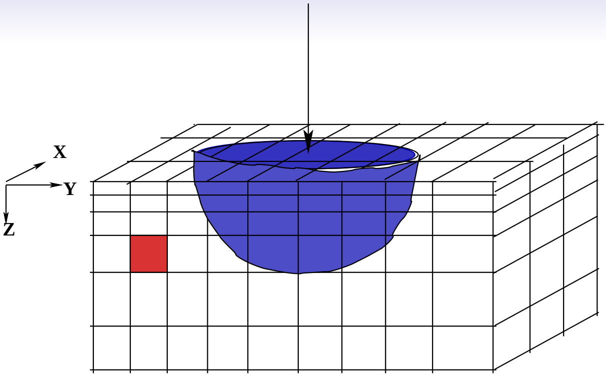


Figure: An example interaction-volume is shown in blue. In this example, there are 9 grid-cells along the Y-axis and only 3 of the 9 grid-cells along the X-axis are shown. In the full 3-D grid with 9 cells along each of the X and Y axes, the example interaction-volume, in blue, is the ik-th one, with $i = 38$ and k=4, and has a radius of $R0^{(4)}$ which by definition of the hemispherical shape, equals $h^{(4)}$. The face on the $x = 0$ plane, of an example voxel that lies between depths $h^{(3)}$ and $h^{(4)}$, is shown in red; this is the ik-th voxel, with i=10, k=4.

Novelty of solution ... cont'd

Fully non-parametric solution for unknown material density function in high-dimensional voxel space - correlation structure amongst components provided through comparison of image data and projection+spatial-averaging of $\rho * \eta$ in discretised model. Density in ik^{th} voxel is $\xi_i^{(k)}$, $i = 1, \ldots, N_{data}$, $k = 1, \ldots, N_{eng}$.

Non-parametric or semi-parametric, discretised solution for unknonwn microscopy correction function $\eta^{(k)}$ sought.

Data

Sequence of N_{eng} 2-D radiation density values recorded in a square spatial array of N_{data} number of pixels, $\{\tilde{I}_i^{(k)}\}_{k=1}^{N_{eng}}, \; i = 1, \ldots, N_{data}$.

Image system at multiple beam energies $\{\epsilon_k\}_{k=1}^{N_{eng}}$, where E is the real-valued discrete energy variable. For $E = \epsilon_k$, the image of the whole material sample is formed by N_{data} number of beam pointings on the surface of the material sample - $\tilde{I}_i^{(k)}$.

Model assumptions

Hemispherical "interaction volume" assumed to enhance simplicity of model $\rightsquigarrow$ could render problem harder than it is. Morphological details of "interaction volume" known as function of material properties from microscopy literature.

Correction function depends only on Z, i.e. independent of beam location $\rightsquigarrow$ aids identifiability of solutions of $\rho(\mathbf{x})$, $\eta(z)$.

Geometric priors

In the forward problem, we check if $\langle \mathcal{P}(\rho * \eta) \rangle = \langle \mathcal{P}(\rho * \eta) \rangle_{-mk}$, where $\langle \mathcal{P}(\cdot) \rangle_{-mk} :=$ spatial-averaging without including $\rho(x_m, y_m, z^{(k)})$. If so, then we set $\rho(x_m, y_m, z^{(k)}) = 0$, i.e. $\xi_m^{(k)} = 0$.
Define

$$\tau_i^{(k)} := \frac{I_i^{(k)}}{I_i^{(k-1)}}, \text{ if } \quad I_i^{(k)} < I_i^{(k-1)},\ I_i^{(k-1)} \neq 0 \qquad (5)$$

$$\tau_i^{(k)} := 1 \quad \text{otherwise}$$

$$\nu_i^{(k)}(\tau_i^{(k)}) = p^{\tau_i^{(k)}}(1-p)^{1-\tau_i^{(k)}}, \qquad (6)$$

where $\nu_i^{(k)}$ is a probability density and the hyperparameter $P \in \mathbb{R},\ 0 \leq p \leq 1$. $p \sim \mathcal{U}[0.9, 0.99]$.

Geometric priors

$$\pi_0(\rho) \propto \exp\left(-\frac{1}{2}\langle\rho, \mathbf{S}^{-1}\rho\rangle\right).$$

We set the $N_{eng} \times N_{eng}$ precision matrix for any $i = 1, \ldots, N_{data}$ to be

$$\begin{aligned} \mathbf{S_i}^{-1} &= \mathbf{H\_iH\_i}^T \\ H\_i_{k\ell} &= \nu_i^{(k)} \cdot \nu_i^{(\ell)} \text{if } k = \ell \\ H\_i_{k\ell} &= 0 \text{if } k \neq \ell. \end{aligned} \tag{7}$$

Here $k, \ell = 1, \ldots, N_{eng}$ and $H\_i_{k\ell}$ is the $k\ell$-th element of $\mathbf{H\_i}$. $\nu_i^{(k)}$ is defined in Equation 6.

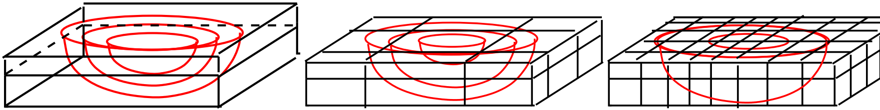
Priors on $\eta(z)$

Non-parametric model for $\eta(z)$:
$\pi(\eta) = \mathcal{N}_F(\eta|\eta 0, S)$**, where** $S = f(\eta 0)|\Psi(0)$**. A uniform prior is assigned to the hyperparameter** $\eta 0$**.**

Semi-parametric model for $\eta(z)$:
Shape of $\eta(z)$ approximated using available 2-parameter function in the literature, with Gaussian priors on the two parameters.

Taking image details on board

- Size of “interaction volume” $R0^{(k)}$ determined by material properties, given ϵ_k.
- (Lateral) size of voxel determined by resolution of imaging technique at hand ($\delta \in \mathbb{R}$).
- 3 models - depend on resolution of image technique.



Spatially-averaged+projected $\rho * \eta$ from all voxels included within the "interaction volume" formed at the i^{th} beam pointing, at the k^{th} beam energy value to get $\tilde{I}_i^{(k)}$

$$
\begin{aligned}
\tilde{I}_i^{(k)} &= \frac{1}{(R0^{(k)})^2} \int_{x=0}^{R0^{(k)}} dx \int_{y=0}^{\sqrt{(R0^{(k)})^2 - x^2}} dy\, f(x,y), \\
f : \mathbb{R} \times \mathbb{R} &\longrightarrow \mathbb{R} \\
f(x,y) &= \sum_{t=0}^{j(x,y)} \xi_{n(x,y)}^{(t)} \phi^{(t)} + \\
& \xi_{n(x,y)}^{(j(x,y)+1)} \phi^{(j(x,y)+1)} \frac{\Delta(x,y)}{(R0^{(j(x,y)+1)} - R0^{(j(x,y))})}
\end{aligned}
\tag{8}
$$

Warwick Statistics

Gaussian likelihood

$$\mathcal{L}(I_i^{(k)}|\rho(x_i, y_i, z_i^{(k)}), \Phi(z^{(k)})) = \frac{1}{\sqrt{2\pi}\sigma_i^{(k)}} \exp\left[-\frac{(I_i^{(k)} - \tilde{I}_i^{(k)})^2}{2(\sigma_i^{(k)})^2}\right], \quad k = 1, \ldots, N_{eng}, \; i = 1, \ldots, N_{data} \tag{9}$$

Posterior

Posterior density, given the image data, is defined using Bayes rule. Assume contribution from all "interaction volume", are conditionally *iid* so that

$$\pi(\xi_1^{(1)},\ldots,\xi_1^{(N_{eng})},\ldots,\xi_{N_{data}}^{(1)},\ldots,\xi_{N_{data}}^{(N_{eng})},\Phi^{(1)},\ldots,\Phi^{(N_{eng})}|\mathrm{data}) \propto \prod_{i=1}^{N_{data}}\prod_{k=1}^{N_{eng}}\mathcal{L}(\tilde{I}_i^{(k)}|\rho(x_i,y_i,z_i^{(k)}),\Phi(z^{(k)})) \pi_0(\xi_1^{(1)},\ldots,\xi_{N_{data}}^{(N_{eng})})\nu_0(\Phi^{(1)},\ldots,\Phi^{(N_{eng})})$$

where $\pi_0(\xi_1^{(1)},\ldots,\xi_{N_{data}}^{(N_{eng})})$ is the joint prior probability density of $\rho(\mathbf{x})$.

Inference

- State space of the fully discretised model is $N_{data} \times N_{eng} + N_{eng}$-dimensional while for the semi-parametric implementation of the correction function, the state-space is $N_{data} \times N_{eng} + 2$-dimensional.
- Sample from high-dimensional posterior using Metropolis-within-Gibbs.
- Adaptive Metropolis-Hastings (Haario et. al 2006).

$$\tilde{\xi}_i^{(k)}|_n \sim \mathcal{N}_F(\mu_i^{(k)}|_n, \varsigma_i^{(k)}|_n) \quad k = 1, \ldots, N_{eng} \tag{10}$$

$$\begin{aligned}
\mu_i^{(k)}|_n &= \xi_i^{(k)}|_{n-1}, \quad \forall\, k = 1 \ldots, N_{eng},\ i = 1, \ldots, N_{data},\ n = 1, \ldots \\
\left(\varsigma_i^{(k)}|_n\right)^2 &= \frac{\sum_{p=n_0}^{n-1}\left(\xi_i^{(k)}|_p\right)^2}{n-n_0} - \left[\frac{\sum_{p=n_0}^{n-1}\left(\xi_i^{(k)}|_p\right)}{n-n_0}\right]^2 \quad \text{if} \quad n \geq n_0 \\
&= T\xi_i^{(k)}|_0 \quad \text{if} \quad n < n_0
\end{aligned}$$

In the small noise limit

In the small noise limit, it can be shown that joint posterior probability density for the material density and blurring function, given the image data, for all beam locations (all i) and all beam energies (all k), reduces to a product of $N_{data} \times N_{eng}$ Dirac measures, with the ik-th measure centred at the solution to the equation $\tilde{I}_i^{(k)} = I_i^{(k)}$. Then $\tilde{\mathbf{I}} := \mathcal{C}[(\rho * \eta)_i^{(k)}]$ implies that the unique matrix $[(\rho * \eta)_i^{(k)}] = \mathbf{C}^{+}\mathbf{I}$.

Learning ρ and η from the unique $\rho * \eta$ is ill-posed but ratio of unknown to known parameters $\geq$ 0.99 typically $\longrightarrow$ deviations from uniqueness accommodated in 95% highest probability density regions.

Simulation - coarsest resolution

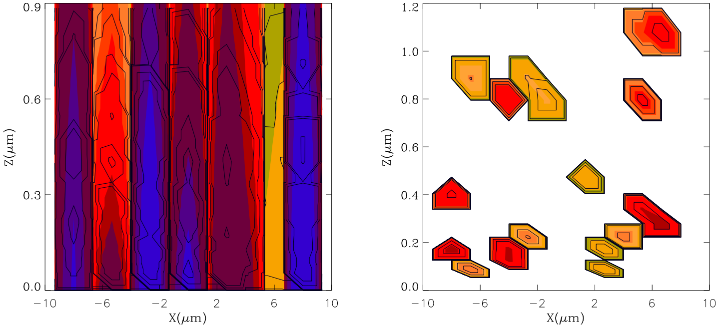
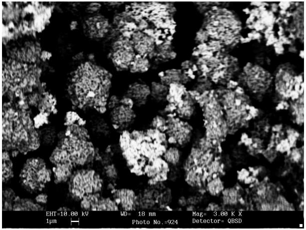
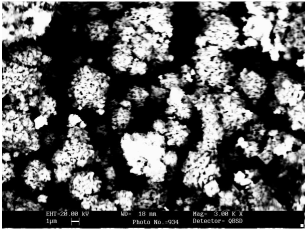
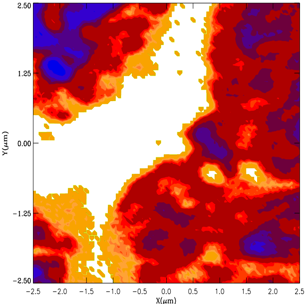


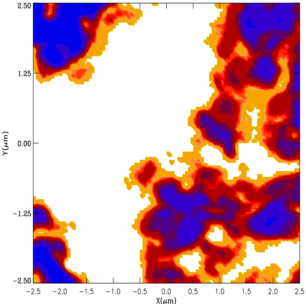
Figure: Simulated images of Copper-Tungsten alloy - sparse (left) and dense (right) density structure.

Application to real image data (11 images) of a nano-composite of Ni-Ag nanoparticles, made by dropcast method, imaged in BackScattered Electrons with an SEM - resolution of about 50 nm.

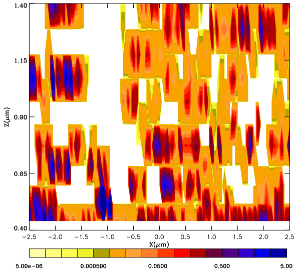


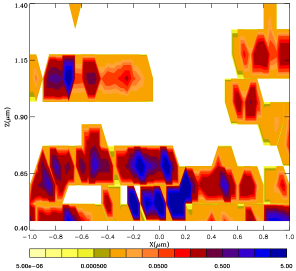


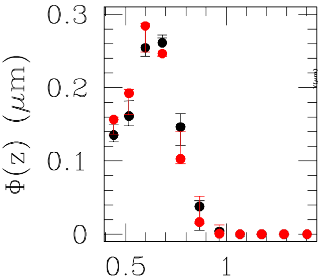


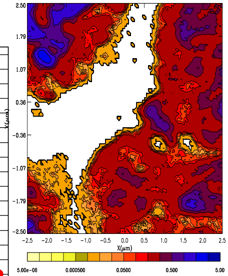


Learnt functions

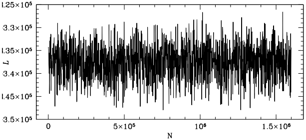


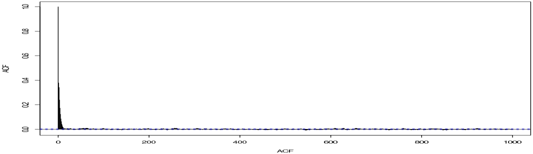


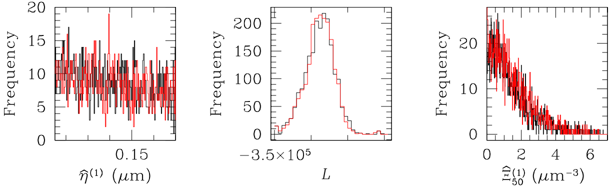




MCMC Diagnostics







Summary

- In complex systems, with many degrees of freedom, non-linearity guides correlation structure amongst parameters, as well as evolution – non-parametric learning.
- Decompose to smaller problems, input domain physics as much as possible.
- Identify details in topology of relevant subspaces if possible - sparsity in the model, morphological details.
- Inference in high-dim state space - MCMC $\longleftarrow$ practice + understanding + luck.
- Test for assumptions in data - develop bespoke non-parametric tests.