Chapter 4

Physical large-scale structure inference with the BORG algorithm

Contents

> "We are the Borg. Lower your shields and surrender your ships. We will add your biological and technological distinctiveness to our own. Your culture will adapt to service us. Resistance is futile."
> — Star Trek: First Contact (1996)

Abstract

This chapter describes the development and implementation of the BORG algorithm, which aims at physical large-scale structure inference in the linear and mildly non-linear regime. It describes the data model, which jointly accounts for the shape of three-dimensional matter field and its formation history. Based on an efficient implementation of the Hamiltonian Monte Carlo algorithm, BORG samples the joint posterior of the several millions parameters involved, which allows for thorough uncertainty quantification.

This chapter presents BORG (Bayesian Origin Reconstruction from Galaxies), a data assimilation method for probabilistic, physical large-scale structure inference. In section 4.1, the main challenge faced, namely the curse of dimensionality, is discussed. In section 4.2, we describe the latest formulation of BORG data model, initially introduced by Jasche & Wandelt (2013a) and updated by Jasche, Leclercq & Wandelt (2015). Section 4.3 gives considerations about the sampling procedure and the numerical implementation of the algorithm. Finally, in section 4.4, we report on a test of the BORG algorithm using a synthetic catalog of tracers.

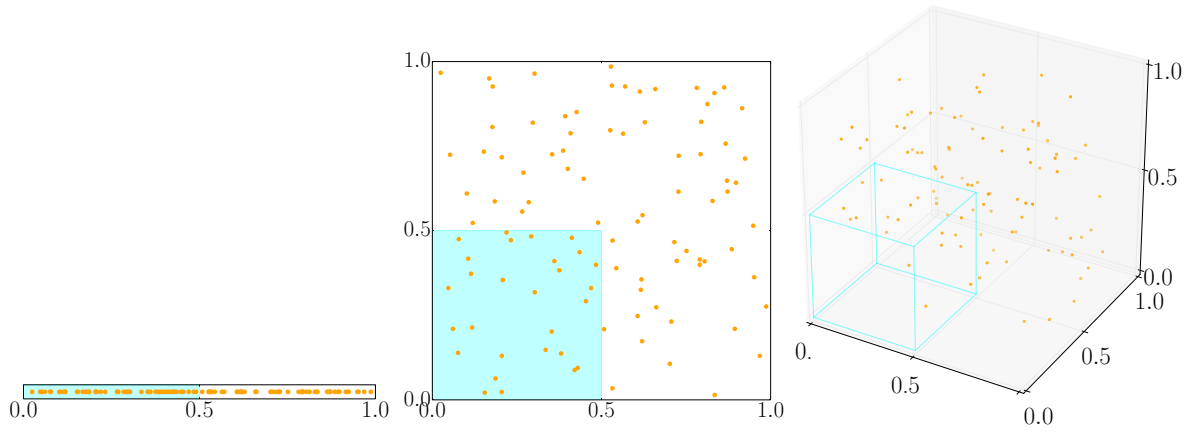

Figure 4.1: Illustration of the curse of dimensionality in one, two and three dimensions. We draw an original sample of 100 random points uniformly distributed in $[0;1]$, then progressively add a second and third coordinate, also uniformly drawn in $[0;1]$. The sparsity of the data (here illustrated by the number of samples in the $\left[0;\frac{1}{2}\right]$ hypercube, in cyan) increases exponentially with the number of dimensions.

Dimension D	$\mathcal{P}_D = 2^{-D}$	Numerical representation
1	2^{-1}	0.5
10	2^{-10}	9.77×10^{-4}
100	2^{-100}	7.89×10^{-31}
1000	2^{-1000}	9.33×10^{-302}
10000	2^{-10000}	0.

Table 4.1: Probability for a sample uniformly drawn in $[0;1]^D$ to be in $\left[0;\frac{1}{2}\right]^D$, as a function of the dimension D. The mathematical result, 2^{-D} (second column) is compared to its double-precision computer representation (third column). For $D \geq 1075$, $\mathcal{P}_D$ is below the minimum positive subnormal double representable.

4.1 The challenge: the curse of dimensionality

Statistical analyses of large-scale structure surveys require to go from the few parameters describing the homogeneous Universe to a point-by-point characterization of the inhomogeneous Universe. The latter description typically involves tens of millions of parameters: the density in each voxel of the discretized survey volume.

"Curse of dimensionality" phenomena (Bellman, 1961) are the significant obstacle in this high-dimensional data analysis problem. They refer to the difficulties caused by the exponential increase in volume associated with adding extra dimensions to a mathematical space. In the following, we discuss the basic aspects of the high-dimensional situation. In particular, we outline three aspects of the curse of dimensionality phenomena.

4.1.1 Sparse sampling

The first and most obvious aspect is the exponential increase of sparsity given a fixed amount of sampling points. Reciprocally, the number of points drawn from a uniform distribution, needed for sampling at a constant density a region in parameter space, increases exponentially with its dimension.

We illustrate this phenomenon in figure 4.1 with 100 points randomly drawn in $[0;1]^D$ for $D = 1, 2, 3$. The number of samples that fall in some fixed region in parameter space exponentially decreases with the dimensionality of the problem. For example, the probability $\mathcal{P}_D$ for a random point to be in the $\left[0;\frac{1}{2}\right]^D$ hyperquadrant (shown in cyan in figure 4.1) is 2^{-D}. Difficulties to represent such probabilities numerically (table 4.1) arise well before $D = 10^7$, as we now discuss.

In standard double-precision binary floating-point format (the IEEE 754 "`binary64`" norm), numbers are represented in base $b = 2$. The bits are laid out as follows (figure 4.2): 1 sign bit, 11 bits for the exponent width, and $p = 52$ bits for the significand precision. The real value assigned by the machine to a set of `binary64` digits

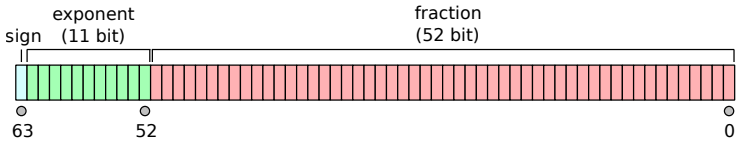

Figure 4.2: Computer representation of double-precision binary floating-point numbers. One bit is used to store the sign, 11 to store the exponent, and 52 bits to store the fractional part. This representation on a finite number of bits implies the existence of both a minimal and a maximal positive representable number.

is

$$(-1)^{\text{sign}} \left(1 + \sum_{i=1}^{52} b_{52-i} 2^{-i}\right) \times 2^{e-1023}, \tag{4.1}$$

where $1 \leq e \leq 2046$ is the "biased exponent" encoded in the 11 exponent bits and b_i are the values of the significand bits.

This representation implies that the maximum relative rounding error when rounding a number to the nearest representable one (the "machine epsilon") is $b^{-(p-1)} = 2^{-52}$. Therefore, the maximum positive double is `max_double` $\equiv (1+(1-2^{-52})) \times 2^{1023} \approx 1.798 \times 10^{308}$ and the minimum positive double is `min_normal_double` $\equiv 2^{-1022} \approx 2.225 \times 10^{-308}$.

In a normal floating-point value, there are no leading zeros in the significand; instead leading zeros are moved to the exponent. By using leading zeros in the significand, it is possible to represent "subnormal numbers", i.e. numbers where this representation would result in an exponent that is too small for the allowed number of bits. The smallest subnormal number representable with the `binary64` norm is `min_subnormal_double` $\equiv 2^{-52} \times 2^{-1022} \approx 4.941 \times 10^{-324}$.

Coming back to the representation of $\mathcal{P}_D$ is a large number of dimensions, the discussion above implies that $\mathcal{P}_D$ is exactly zero, at computer precision, for $D \geq 1075$. More generally, typical probabilities are often below `min_subnormal_double` for $D \gtrsim 1000$, which means that their computer representations as doubles is impossible. Representing such numbers requires more than 64 bits. This number of dimensions is well below that of the problem that we want to tackle, $D \approx 10^7$.

4.1.2 Shape of high-dimensional pdfs

Generally, high-dimensional functions can have more complex features than low-dimensional functions (there is more "space" for that), and hence can be harder to characterize.

Since it is not possible to store arbitrarily small positive numbers, numerical representations of high-dimensional pdfs will tend to have narrow support and very peaked features. This can also cause difficulties, as pdfs have to be normalized to unity: if the support is sufficiently small, the value of the pdf at its peaks can easily be above the maximum double `max_double`, which will cause computer crashes.

4.1.3 Algorithms in high dimensions

It is important to note that curse of dimensionality phenomena are generally not an intrinsic problem of high-dimensional problems, but a joint problem of the data set and the algorithm used. In particular, a dramatic increase of computational time (both to get one sample and to reach the required number of samples) is common. The curse of dimensionality often means that the number of samples available is small compared to the dimension of the space, which can lead to issues such as overfitting the data or getting poor classification or clustering when searching for specific patterns (Verleysen & François, 2005).

For most MCMC algorithms, the slow convergence, due a high rejection rate, is the most significant obstacle. In particular, for many interesting problems (typically non-linear and where components are not independently distributed), traditional sampling techniques that perform a random walk in parameter space, like the Metropolis-Hastings algorithm (see section 3.4.2) will unequivocally fail in $D \approx 10^7$.[1] However, gradients

[1] At least, unless the proposal distribution approximates extremely well the target distribution – which would imply to have already solved the problem!

Code	Density field model	Response operator	Multi-survey	$P(k)$	Photo-z	Galaxy bias model	b	$\tilde{N}$	RSD
ARES	Gaussian (J+10a)	(J+10a)	(JW13b)	(J+10a)		linear (J+10a); M-dep., linear (JW13b)	sampled (JW13b)	(JW13b)	(J+in prep.)
HADES	Log-normal (JK10)	(JK10)	(J+in prep.)	(J+in prep.)	(JW12)	linear (JK10)		(J+in prep.)	
BORG	2LPT (JW13a)	(JW13a)	(JLW15)			linear (JW13a); M-dep., power-law (JLW15)	calibrated with ARES (LJ16); sampled (J+in prep.)	(JLW15)	

Table 4.2: Current status of Bayesian large-scale structure analysis codes ARES, HADES and BORG. Green cells correspond to features implemented in the data model and tested, as reported in the corresponding papers. Blue cells correspond to features which will be described in upcoming publications. The column correspond respectively to: the model used to describe the prior density field; treatment of the survey response operator (survey mask and selection effects); treatment of multiple, independent surveys (or sub-samples of the same survey); power spectrum sampling; photometric redshifts sampling; galaxy bias model (M-dep. stands for luminosity-dependent bias); treatment of bias parameters; sampling of noise levels; treatment of peculiar velocities and redshift-space distortions. The references are J+10a = Jasche *et al.* (2010a); JK10 = Jasche & Kitaura (2010); JW12 = Jasche & Wandelt (2012); JW13a = Jasche & Wandelt (2013a); JW13b = Jasche & Wandelt (2013b); JLW15 = Jasche, Leclercq & Wandelt (2015); LJ16 = Lavaux & Jasche (2016).

of pdfs carry capital information, as they indicate the direction to high-density regions, permitting fast travel through a large volume in parameter space.

One way forward is to reduce the dimensionality of the problem, which is actually an entire research field. For example, principal component analysis converts a set of correlated variables to a set of linearly uncorrelated "principal components". Unfortunately, due to the highly non-linear and complex physics involved in structure formation (see chapter 1), no obvious reduction of the problem size exists in our case. Under the assumption of an initial grf with independent density amplitudes in Fourier space, we cannot make any further dimension reduction, and we have to deal with all $D \approx 10^7$ dimensions. Dimensionality can only be reduced by coarsening the resolution and discarding information.

As we will demonstrate in the rest of this chapter, Hamiltonian Monte Carlo (see section 3.4.3) beats the curse of dimensionality for the problem of physical large-scale structure inference. In particular, the approximate conservation of the Hamiltonian enables us to keep a high acceptance rate, and the use of gradients of the posterior pdf ($\partial\psi(\mathbf{\theta})/\partial\mathbf{\theta}$ in Hamilton's equations) allows efficient search for high density of probability regions.

4.2 The BORG data model

In this section, we discuss the BORG data model, i.e. the set of assumptions concerning the generation of observed large-scale structure data. In other words, we write down a probabilistic data-generating process.

This model was initially introduced by Jasche & Wandelt (2013a). In Jasche, Leclercq & Wandelt (2015), we updated the data model and modified to the original formulation of the BORG sampling scheme to introduce the improvements presented in Jasche & Wandelt (2013b). These improvements permit to account for luminosity-dependent galaxy bias and to perform automatic noise level calibration.

BORG is the successor of ARES (Algorithm for REconstruction and Sampling, Jasche *et al.*, 2010a; Jasche & Wandelt, 2013b) and HADES (HAmiltonian Density Estimation and Sampling Jasche & Kitaura, 2010; Jasche & Wandelt, 2012). In table 4.2, we summarize the different aspects covered by the ARES, HADES, and BORG data models. Contrary to ARES and HADES, which use phenomenological models to describe the density field, BORG involves a physical structure formation model (see table 4.2). LSS observations are merged with actual dynamics. Therefore, even if it is the least advanced algorithm in terms of the aspects covered by the data

model, its physical modeling is the most sophisticated.

In the following, x labels one of the D voxels of the discretized domain, δ^{i} and δ^{f} are realizations of the initial (at $a = 10^{-3}$) and final (at $a = 1$) density contrast, respectively, expressed as D-dimensional vectors. For improved clarity, we use colors in equations to distinguish the different quantities that are involved in the data model.

4.2.1 The physical density prior

In contrast to earlier algorithms (see table 4.2) BORG includes a physical density prior i.e. involves a model for structure formation. This makes the prior (expressed in terms of the final density contrast) highly non-Gaussian and non-linear. Writing down this prior is the subject of the present section.

4.2.1.1 The initial Gaussian prior

As discussed in the introduction and in chapter 1, it is commonly admitted that the density contrast early in the matter era obeys Gaussian statistics. Consistently with the discussion of section 3.2.2, this is the prior that we adopt.

Explicitly, in Fourier space, the prior for the initial density contrast is a multivariate Gaussian process with zero mean and diagonal covariance matrix $\hat{S}$ (see equation (1.14)):

$$\mathcal{P}(\hat{\delta}^{\mathrm{i}}|\hat{S}) = \frac{1}{\sqrt{\left|2\pi\hat{S}\right|}} \exp\left(-\frac{1}{2}\sum_{k,k'} \hat{\delta}^{\mathrm{i}}_{k} \hat{S}^{-1}_{kk'} \hat{\delta}^{\mathrm{i}}_{k'}\right). \tag{4.2}$$

where we explicitly noted by a hat the Fourier-space quantities.

The elements in matrix $\hat{S}$ are fixed parameters in BORG. They characterize the variance of the initial density field and therefore contain a cosmological dependence. We further assume that the covariance matrix $\hat{S}$ is diagonal in Fourier space (this is assuming statistical homogeneity of the initial density contrast, as seen in section 1.2.4.1). The diagonal coefficients are $\sqrt{P(k)/(2\pi)^{3/2}}$, where $P(k)$ are the initial power spectra coefficients for the adopted fiducial cosmological parameters. They are chosen to follow the prescription of Eisenstein & Hu (1998, 1999), including baryonic wiggles.

Alternatively, using the configuration space representation yields

$$\mathcal{P}(\delta^{\mathrm{i}}|S) = \frac{1}{\sqrt{|2\pi S|}} \exp\left(-\frac{1}{2}\sum_{x,x'} \delta^{\mathrm{i}}_{x} S^{-1}_{xx'} \delta^{\mathrm{i}}_{x'}\right). \tag{4.3}$$

4.2.1.2 Translating to the final density field

Following Jasche & Wandelt (2013a), we now show that the problem of physical inference of final density fields can be recast into the problem of inferring the corresponding initial conditions, given the structure formation model.

As seen before, it is straightforward to express a prior in the initial conditions, $\mathcal{P}(\delta^{\mathrm{i}})$. Given this, we can obtain a prior distribution for the final density contrast at scale factor a by using the standard formula for conditional probabilities:

$$\mathcal{P}(\delta^{\mathrm{f}}) = \int \mathcal{P}(\delta^{\mathrm{f}}, \delta^{\mathrm{i}})\, \mathrm{d}\delta^{\mathrm{i}} \tag{4.4}$$

$$= \int \mathcal{P}(\delta^{\mathrm{f}}|\delta^{\mathrm{i}})\, \mathcal{P}(\delta^{\mathrm{i}})\, \mathrm{d}\delta^{\mathrm{i}}. \tag{4.5}$$

For a deterministic model of structure formation $\delta^{\mathrm{i}} \mapsto \mathcal{G}\left(\delta^{\mathrm{i}}, a\right)$, the conditional probability is given by Dirac delta distributions:

$$\mathcal{P}(\delta^{\mathrm{f}}|\delta^{\mathrm{i}}) = \prod_{x} \delta_{\mathrm{D}}\left(\delta^{\mathrm{f}}_{x} - \left[\mathcal{G}(\delta^{\mathrm{i}}, a)\right]_{x}\right). \tag{4.6}$$

Therefore, given a model $\mathcal{G}$ for structure formation, a prior distribution for the late-time density field can be obtained by a two-step sampling process:

1. drawing an initial condition realization from the prior $\mathcal{P}(\delta^{\rm i})$;
2. propagating the initial state forward in time with $\mathcal{G}$ (this step is entirely deterministic).

This process amounts to drawing samples from the joint prior distribution of initial and final conditions:

$$\mathcal{P}(\delta^{\rm f}, \delta^{\rm i}) = \mathcal{P}(\delta^{\rm i}) \prod_x \delta_{\rm D}\left(\delta^{\rm f}_x - \left[\mathcal{G}(\delta^{\rm i}, a)\right]_x\right). \tag{4.7}$$

Marginalization over initial density realizations then yields samples of the non-Gaussian prior for final density fields. In practice, as initial conditions are also interesting for a variety of cosmological applications, we do not discard them and we always store them, whenever we draw a sample from the prior.

4.2.1.3 The structure formation model

Ideally, the structure formation model should be fully non-linear gravity. For reasons of computational feasibility, in BORG, $\mathcal{G}$ is obtained from second-order Lagrangian perturbation theory and the cloud-in-cell scheme. More specifically, the initial density field is populated by dark matter particles that are evolved according to the equations for 2LPT displacements given in section 1.5.3. In the final state, these particles are assigned to the grid using a CiC scheme, yielding the final density contrast $\delta^{\rm f}$. The reader is referred to appendix B for details on the numerical implementation of 2LPT and CiC.

Using equations (4.3) and (4.7), the joint physical prior for initial and late-time density fields is found to be

$$\mathcal{P}(\delta^{\rm f}, \delta^{\rm i}|S) = \frac{1}{\sqrt{|2\pi S|}} \exp\left(-\frac{1}{2}\sum_{x,x'} \delta^{\rm i}_x S^{-1}_{xx'} \delta^{\rm i}_{x'}\right) \prod_x \delta_{\rm D}\left(\delta^{\rm f}_x - \left[\mathcal{G}(\delta^{\rm i}, a)\right]_x\right). \tag{4.8}$$

Note that the first part (corresponding to the initial conditions) is more easily handled in Fourier space, while the second part (corresponding to the propagation from initial to final conditions) involves density fields in configuration space.

4.2.2 The large-scale structure likelihood

This section discusses the BORG likelihood, $\mathcal{P}(d|\delta^{\rm i})$. The data d used by BORG are galaxy (or matter tracer) number counts in each voxel of the discretized domain. To compute it, the position of galaxies is translated from spherical to Cartesian coordinates using the following coordinate transform:

$$x = d_{\rm com}(z)\cos(\lambda)\cos(\eta), \tag{4.9}$$
$$y = d_{\rm com}(z)\cos(\lambda)\sin(\eta), \tag{4.10}$$
$$z = d_{\rm com}(z)\sin(\lambda), \tag{4.11}$$

with λ being the declination, η the right ascension and $d_{\rm com}(z)$ the radial comoving distance to redshift z for the fiducial cosmology. Galaxies are then binned using the Nearest Grid Point (NGP) assignment scheme to get voxel-wise galaxy number counts.

4.2.2.1 Splitting the galaxy distribution

In order to account for the luminosity-dependence of selection effects and galaxy biases, we split the data into several bins of absolute magnitude. In the following, ℓ labels one of these bins, and N^ℓ is the data set containing the number counts of galaxies in the luminosity bin ℓ and in voxel x, N^ℓ_x.

BORG treats different magnitude bins as independent data sets. Each of them is assigned a likelihood function, $\mathcal{P}(N^\ell|\delta^{\rm i})$. Since it is fair to assume that galaxies in different luminosity bins are independent and identically distributed, once the density field is given, the final likelihood of the total data set $d = \{N^\ell\}$ is obtained by multiplying these likelihood functions,

$$\mathcal{P}(d|\delta^{\rm i}) = \prod_\ell \mathcal{P}(N^\ell|\delta^{\rm i}). \tag{4.12}$$

4.2.2.2 The galaxy distribution as an inhomogeneous Poisson process

Galaxies are tracers of the mass distribution. The statistical uncertainty due to the discrete nature of their distribution is often modeled as a Poisson process (Layzer, 1956; Peebles, 1980; Martínez & Saar, 2002). Before BORG, Poissonian likelihoods have been successfully applied to perform reconstructions of the matter density by Kitaura, Jasche & Metcalf (2010); Jasche & Kitaura (2010); Jasche *et al.* (2010a). Adopting this picture, we write

$$\mathcal{P}(N^\ell|\lambda(\delta^{\rm i})) = \prod_x \frac{\exp\left(-\lambda^\ell_x(\delta^{\rm i})\right)\left(\lambda^\ell_x(\delta^{\rm i})\right)^{N^\ell_x}}{N^\ell_x!}. \tag{4.13}$$

The Poisson intensity field, $\lambda^\ell(\delta^{\rm i})$, characterizes the expected number of galaxies in voxel x given the initial density contrast $\delta^{\rm i}$. As it depends on the position, it is an *inhomogeneous* Poisson process.

Real galaxy samples can have a sub- or super-Poissonian behavior (i.e. be under- or over-dispersed), depending on local and non-local properties (Mo & White, 1996; Somerville *et al.*, 2001; Casas-Miranda *et al.*, 2002). These effects are neglected here, but in the context of large-scale structure reconstructions, deviations from Poissonity have been introduced in the likelihood by Kitaura (2012); Ata, Kitaura & Müller (2015).

4.2.2.3 The Poisson intensity field

The expected number of galaxies in a voxel depends – of course – on the underlying large-scale structure, but also on galaxy bias, redshift-space distortions, dynamical processes along the observer's backwards lightcone, selection effects, and instrumental noise. All these effects should in principle be taken into account in the Poisson intensity field. In the following, we detail, step by step, how to go from $\delta^{\rm i}$ to $\lambda(\delta^{\rm i})$ in the BORG likelihood.

1. **Structure formation.** The first step is to translate initial to evolved dark matter overdensity:

$$\delta^{\rm i} \mapsto \mathcal{G}(\delta^{\rm i}, a). \tag{4.14}$$

   As discussed before, for this step BORG relies on 2LPT instead of fully non-linear gravitational dynamics, meaning that there exists some degree of approximation in the inference process. Accurate quantification this level of approximation is unfortunately not currently possible, as it would require the fully non-linear inference process for reference, which so far is not computationally tractable.

2. **Lightcone effects.** Along with step 1, we could account for lightcone effects so that the distant structures are less evolved than the closest ones. This is exploiting the dependence of $\mathcal{G}$ on a to build the dark matter density on the lightcone. For simplicity, this is not currently implemented in BORG; rather, we run 2LPT up to $a = 1$ everywhere. In the following we simplify the notations and we write $\delta^{\rm f} \equiv \mathcal{G}(\delta^{\rm i}) \equiv \mathcal{G}(\delta^{\rm i}, a = 1)$.

3. **Redshift-space distortions.** At this point, the data model could also include a treatment of redshift-space distortions (see Heavens & Taylor, 1995; Tadros *et al.*, 1999; Percival, Verde & Peacock, 2004; Percival, 2005a; Percival & White, 2009). Though not explicitly included in the present BORG data model, we find empirically that redshift-space distortions are mitigated by the prior preference for homogeneity and isotropy (see chapter 5): BORG interprets deviations from isotropy as noise, and fits an isotropic distribution to the data.

4. **Galaxy bias.** The following step is to get the galaxy density $\rho_{\rm g}$ given the dark matter density ρ. This is making assumptions for physical biasing in galaxy formation. Various LSS inference algorithms assume a linear bias model. In order to be well defined, a Poisson likelihood requires intensities of the inhomogeneous Poisson process to be strictly positive. Since a linear bias model does not guarantee a positive density field and corresponding Poisson intensity, it is not applicable to the present case. For this reason, we assume a phenomenological power-law to account for galaxy biasing:

$$\rho_g \propto \beta\rho^\alpha. \tag{4.15}$$

   In luminosity bin ℓ and in terms of the dark matter overdensity, this is step written

$$\delta^{\rm f} \mapsto \beta^\ell(1+\delta^{\rm f})^{\alpha^\ell} \propto \rho^\ell_g. \tag{4.16}$$

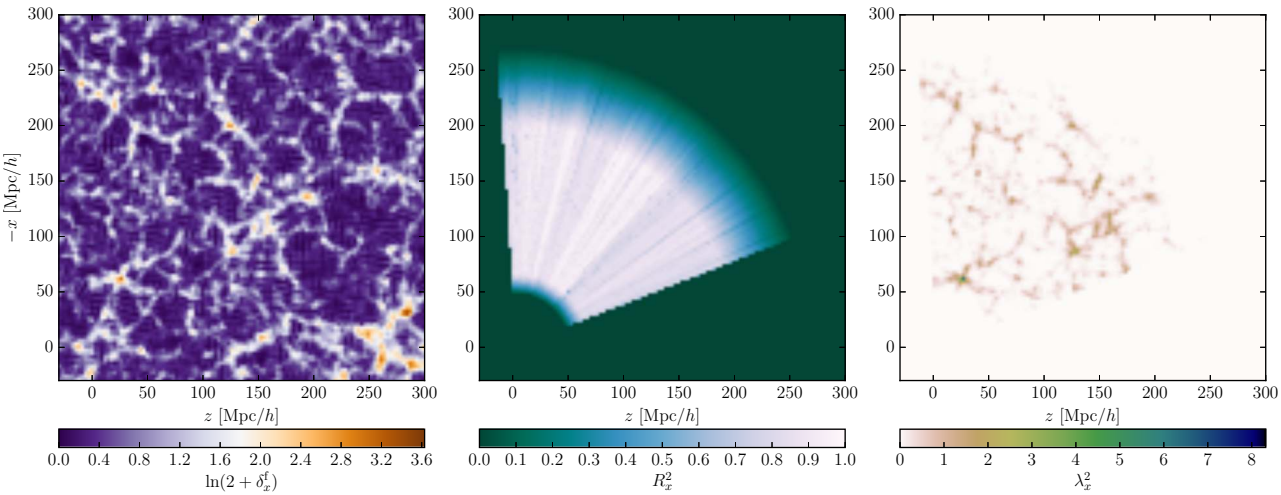

Figure 4.3: Slices through the box used in the BORG SDSS analysis (see chapter 5). *Left panel.* Density in one sample (for clarity, the quantity shown is $\ln(2+\delta_x^{\mathrm{f}})$). *Middle panel.* Survey response operator R_x^2 in the $\ell = 2$ luminosity bin, corresponding to absolute r-band magnitudes in the range $-19.67 < M_{0.1r}^2 < -19.00$. *Right panel.* Poisson intensity field λ_x^2 for this sample and luminosity bin, computed with equation (4.20). The bias and noise parameters are respectively $\alpha^2 = 1.30822$ and $\widetilde{N}^2 = 1.39989$ (see table 5.1).

Note that coefficients α^ℓ and β^ℓ depend on ℓ, which means that the data model accounts for *luminosity-dependent* galaxy biases. Parameters β^ℓ are automatically calibrated during the generation of the Markov Chain (see section 4.3.1). For simplicity, parameters α^ℓ are kept at fixed, fiducial values. In the BORG analysis of the SDSS (chapter 5), these values are determined using a standard model for luminosity-dependent galaxy bias. In their analysis of the 2M++ catalog (Lavaux & Hudson, 2011), Lavaux & Jasche (2016) show that it is possible to calibrate these values with a preliminary ARES inference, for subsequent use in BORG.

5. **Mean number of galaxies.** To get the expected number of galaxies from the unnormalized galaxy density, the quantity $\beta^\ell(1+\delta^{\mathrm{f}})^{\alpha^\ell}$ has to be multiplied by the mean number of galaxies in bin ℓ, $\bar{N}^\ell$. This step is therefore simply:

$$\beta^\ell(1+\delta^{\mathrm{f}})^{\alpha^\ell} \mapsto \bar{N}^\ell\beta^\ell(1+\delta^{\mathrm{f}})^{\alpha^\ell}. \tag{4.17}$$

6. **Observational effects.** The last step is to put in the luminosity-dependent selection effects and the survey mask. For this, we multiply with the linear survey response operator R_x^ℓ, a voxel-wise three-dimensional function that incorporates survey geometries and selection effects:

$$\bar{N}^\ell\beta^\ell(1+\delta_x^{\mathrm{f}})^{\alpha^\ell} \mapsto R_x^\ell\bar{N}^\ell\beta^\ell(1+\delta_x^{\mathrm{f}})^{\alpha^\ell}. \tag{4.18}$$

Eventually, the Poisson intensity field is given by

$$\lambda_x^\ell(\delta^{\mathrm{i}}) = R_x^\ell\bar{N}^\ell\beta^\ell\left(1+\left[\mathcal{G}(\delta^{\mathrm{i}})\right]_x\right)^{\alpha^\ell}. \tag{4.19}$$

We note that $\bar{N}^\ell$ and β^ℓ are degenerate, in the sense that only the product $\bar{N}^\ell\beta^\ell$ matters. We define $\widetilde{N}^\ell \equiv \bar{N}^\ell\beta^\ell$, so that

$$\lambda_x^\ell(\delta^{\mathrm{i}}) = R_x^\ell\widetilde{N}^\ell\left(1+\left[\mathcal{G}(\delta^{\mathrm{i}})\right]_x\right)^{\alpha^\ell}. \tag{4.20}$$

$\widetilde{N}^\ell$ represents the overall noise level in bin ℓ. With the improved BORG data model (Jasche, Leclercq & Wandelt, 2015), we automatically calibrate this parameter (see section 4.3.1). In figure 4.3, we illustrate the construction of the Poisson intensity field for the $\ell = 2$ bin of the SDSS analysis. We show the dark matter density, δ_x^{f}, the survey response operator R_x^2 and the Poisson intensity λ_x^2.

4.2.2.4 The comprehensive large-scale structure likelihood

Noting $d \equiv \{N^\ell\}$ the total data set, i.e. all available galaxy number counts, and $\widetilde{N} \equiv \{\widetilde{N}^\ell\}$ the set of noise parameters in each bin, we obtain the final expression for the LSS likelihood using equations (4.12), (4.13) and (4.20). It reads

$$\mathcal{P}(d|\delta^{\mathrm{i}}, \widetilde{N}) = \prod_{x,\ell} \frac{\exp\left(-R_x^\ell \widetilde{N}^\ell (1 + \left[\mathcal{G}(\delta^{\mathrm{i}})\right]_x)^{\alpha^\ell}\right) \left(R_x^\ell \widetilde{N}^\ell (1 + \left[\mathcal{G}(\delta^{\mathrm{i}})\right]_x)^{\alpha^\ell}\right)^{N_x^\ell}}{N_x^\ell !} \tag{4.21}$$

In this equation, we omitted on the right side of the conditioning bar the sets $\{R_x^\ell\}$ and $\{\alpha^\ell\}$ (one can consider that all probabilities inferred by BORG are conditional on these). However, we now write explicitly $\widetilde{N}$, as this will be of importance later.

4.2.3 The posterior distribution

As usual in Bayesian statistics, the posterior distribution is obtained, up to a normalization constant, by the use of Bayes' formula,

$$\mathcal{P}(\delta^{\mathrm{i}}|d, S, \widetilde{N}) \propto \mathcal{P}(\delta^{\mathrm{i}}|S, \widetilde{N})\, \mathcal{P}(d|\delta^{\mathrm{i}}, S, \widetilde{N}) = \mathcal{P}(\delta^{\mathrm{i}}|S)\, \mathcal{P}(d|\delta^{\mathrm{i}}, \widetilde{N}). \tag{4.22}$$

Substituting equations (4.3) and (4.21) allows to write down the full problem solved by BORG for the density distribution:

$$\mathcal{P}(\delta^{\mathrm{i}}|d, S, \widetilde{N}) \propto \frac{1}{\sqrt{|2\pi S|}} \exp\left(-\frac{1}{2} \sum_{x,x'} \delta_x^{\mathrm{i}} S^{-1}_{xx'} \delta^{\mathrm{i}}_{x'}\right) \prod_{x,\ell} \frac{\exp\left(-R_x^\ell \widetilde{N}^\ell (1 + \left[\mathcal{G}(\delta^{\mathrm{i}})\right]_x)^{\alpha^\ell}\right) \left(R_x^\ell \widetilde{N}^\ell (1 + \left[\mathcal{G}(\delta^{\mathrm{i}})\right]_x)^{\alpha^\ell}\right)^{N_x^\ell}}{N_x^\ell !}. \tag{4.23}$$

It is simpler to express the BORG posterior in terms of the initial conditions, but recall that one gets the final conditions (and in fact the entire LSS history, as demonstrated in chapter 5) automatically and entirely deterministically via the structure formation model $\mathcal{G}$ (see section 4.2.1.2):

$$\mathcal{P}(\delta^{\mathrm{f}}, \delta^{\mathrm{i}}|d, S, \widetilde{N}) = \mathcal{P}(\delta^{\mathrm{i}}|d, S, \widetilde{N}) \prod_x \delta_{\mathrm{D}}\left(\delta_x^{\mathrm{f}} - \left[\mathcal{G}(\delta^{\mathrm{i}})\right]_x\right). \tag{4.24}$$

4.2.4 The Γ-distribution for noise sampling

This section draws from appendix A of Jasche, Leclercq & Wandelt (2015).

We aim at automatically calibrating, during the sampling procedure, the noise level of each luminosity bin, given the data and the current density sample. This requires to write down the conditional probability $\mathcal{P}(\widetilde{N}^\ell|N^\ell, \delta^{\mathrm{f}})$, which we do in this section.

According to Bayes' formula, we can write

$$\mathcal{P}(\widetilde{N}^\ell|N^\ell, \delta^{\mathrm{f}}) \propto \mathcal{P}(\widetilde{N}^\ell)\, \mathcal{P}(N^\ell|\widetilde{N}^\ell, \delta^{\mathrm{f}}), \tag{4.25}$$

where we have assumed the conditional independence $\mathcal{P}(\widetilde{N}^\ell|\delta^{\mathrm{f}}) = \mathcal{P}(\widetilde{N}^\ell)$. In the absence of any further information on the parameter $\widetilde{N}^\ell$, we follow the maximum agnostic approach pursued by Jasche & Wandelt (2013b) by setting the prior distribution $\widetilde{N}^\ell$ constant. By using the Poisson likelihood for $\mathcal{P}(N^\ell|\widetilde{N}^\ell, \delta^{\mathrm{f}})$ (equations (4.13) and (4.20)) into equation (4.25), we obtain the conditional posterior for the noise parameter $\widetilde{N}^\ell$ as:

$$\mathcal{P}(\widetilde{N}^\ell|N^\ell, \delta^{\mathrm{f}}) \propto \exp\left(-\widetilde{N}^\ell A_\ell\right) \times \left(\widetilde{N}^\ell\right)^{B_\ell}, \tag{4.26}$$

where $A_\ell \equiv \sum_x R_x^\ell (1 + \delta_x^{\mathrm{f}})^{\alpha^\ell}$ and $B_\ell \equiv \sum_x N_x^\ell$. By choosing $k_\ell \equiv B_\ell + 1$ and $\theta_\ell \equiv 1/A_\ell$, we yield a properly normalized Γ-distribution for the noise parameter $\widetilde{N}^\ell$, given as:

$$\mathcal{P}(\widetilde{N}^\ell|N^\ell, \delta^{\mathrm{f}}) = \Gamma[k_\ell, \theta_\ell]\left(\widetilde{N}^\ell\right) = \frac{\left(\widetilde{N}^\ell\right)^{k_\ell - 1} \exp\left(-\frac{\widetilde{N}^\ell}{\theta_\ell}\right)}{\theta_\ell^{k_\ell} \Gamma(k_\ell)}. \tag{4.27}$$

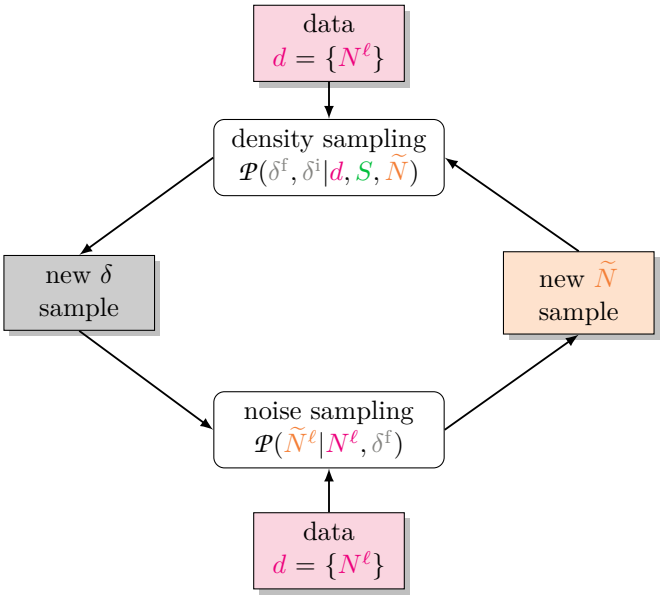

Figure 4.4: Flow chart depicting the multi-step iterative block sampling procedure. In the first step, BORG generates random realizations of the initial and final density fields conditional on the galaxy samples d and on the noise levels $\{\widetilde{N}^\ell\}$. In a subsequent step, the noise parameters $\widetilde{N}^\ell$ are sampled conditional on the previous density realizations.

with shape parameter

$$k_\ell \equiv 1 + \sum_x N^\ell_x, \tag{4.28}$$

and scale parameter

$$\theta_\ell \equiv \frac{1}{\sum_x R^\ell_x (1+\delta^{\rm f}_x)^{\alpha^\ell}}. \tag{4.29}$$

4.3 Sampling procedure and numerical implementation

4.3.1 Calibration of the noise level

This section draws from section 3.2. in Jasche, Leclercq & Wandelt (2015).

Following the approach described in Jasche & Wandelt (2013b), density fields and noise level parameters can be jointly inferred by introducing an additional sampling block to the original implementation of the BORG algorithm. The additional sampling block is designed to provide random samples of the noise parameters $\widetilde{N}^\ell$ given the galaxy data set N^ℓ and the current final density sample $\delta^{\rm f}$.

As indicated by figure 4.4, in a first step, the algorithm infers density fields, then conditionally samples the noise parameters. Iteration of this procedure yields Markovian samples from the joint target distribution.

As demonstrated in section 4.2.4, the posterior distributions of noise parameters $\widetilde{N}^\ell$ are Γ-distributions. In the new sampling block, random variates of the Γ-distribution are generated by standard routines provided by the GNU scientific library (Galassi *et al.*, 2003).

4.3.2 Hamiltonian Monte Carlo and equations of motion for the LSS density

Sampling of the posterior distribution for density fields is achieved via Hamiltonian Monte Carlo. As described in section 3.4.3, HMC permits to explore the non-linear posterior by following Hamiltonian dynamics in the high-dimensional parameter space. Omitting normalization constants, the Hamiltonian potential $\psi(\delta^{\rm i})$ can be written as:

$$\psi(\delta^{\rm i}) = -\ln \mathcal{P}(\delta^{\rm i}|d, S, \widetilde{N}) - \ln Z \tag{4.30}$$

$$= \psi_{\rm prior}(\delta^{\rm i}) + \psi_{\rm likelihood}(\delta^{\rm i}), \tag{4.31}$$

with the "prior potential" $\psi_{\rm prior}(\delta^{\rm i})$ given as

$$\psi_{\rm prior}(\delta^{\rm i}) = \frac{1}{2}\sum_{x,x'} \delta^{\rm i}_x S^{-1}_{xx'} \delta^{\rm i}_{x'}, \tag{4.32}$$

and the "likelihood potential" $\psi_{\rm likelihood}(\delta^{\rm i})$ given as

$$\psi_{\rm likelihood}(\delta^{\rm i}) = \sum_{x,\ell} R^\ell_x \widetilde{N}^\ell \left(1 + \left[\mathcal{G}(\delta^{\rm i})\right]_x\right)^{\alpha^\ell} - N^\ell_x \ln\left(R^\ell_x \widetilde{N}^\ell \left(1 + \left[\mathcal{G}(\delta^{\rm i})\right]_x\right)^{\alpha^\ell}\right). \tag{4.33}$$

Given the above definitions of the potential $\psi(\delta^{\rm i})$, one can obtain the required Hamiltonian force (see equation (3.32)) by differentiating with respect to $\delta^{\rm i}_x$:

$$\frac{\partial\psi(\delta^{\rm i})}{\partial\delta^{\rm i}_x} = \frac{\partial\psi_{\rm prior}(\delta^{\rm i})}{\partial\delta^{\rm i}_x} + \frac{\partial\psi_{\rm likelihood}(\delta^{\rm i})}{\partial\delta^{\rm i}_x}. \tag{4.34}$$

The prior term is given by

$$\frac{\partial\psi_{\rm prior}(\delta^{\rm i})}{\partial\delta^{\rm i}_x} = \sum_{x'} S^{-1}_{xx'}\delta^{\rm i}_{x'} \tag{4.35}$$

The likelihood term cannot be obtained trivially. However, the choice of 2LPT and a CiC kernel to model $\mathcal{G}(\delta^{\rm i})$ makes possible to derive this term analytically. This is of crucial importance, because a numerical estimation of gradients is very expensive. A detailed computation can be found in appendix D of Jasche & Wandelt (2013a). The result is

$$\frac{\partial\psi_{\rm likelihood}(\delta^{\rm i})}{\partial\delta^{\rm i}_x} = -D_1 J_x + D_2 \sum_{a>b}\left(\tau^{aabb}_x + \tau^{bbaa}_x - 2\tau^{abab}_x\right), \tag{4.36}$$

where D_1 and D_2 are the first and second-order growth factors at the desired time ($a = 1$), and J_x and τ^{abcd}_x are a vector and a tensor that depend on R^ℓ_x, $\widetilde{N}^\ell$, α^ℓ, N^ℓ_x.

Finally, the equations of motion for the Hamiltonian system can be written as

$$\frac{{\rm d}\delta^{\rm i}_x}{{\rm d}t} = \sum_{x'} M^{-1}_{xx'}\, p_{x'}, \tag{4.37}$$

$$\frac{{\rm d}p_x}{{\rm d}t} = -\sum_{x'} S^{-1}_{xx'}\delta^{\rm i}_{x'} + D_1 J_x(\delta^{\rm i}) - D_2 \sum_{a>b}\left(\tau^{aabb}_x(\delta^{\rm i}) + \tau^{bbaa}_x(\delta^{\rm i}) - 2\tau^{abab}_x(\delta^{\rm i})\right) \tag{4.38}$$

4.3.3 The mass matrix

As mentioned in section 3.4.3, the HMC algorithm possesses a large number of tunable parameters contained in the mass matrix M, whose choice can strongly impact the efficiency of the sampler. As shown in Jasche & Wandelt (2013a, section 5.2 and appendix F), a good approach to obtain suitable masses is to perform a stability analysis of the numerical leapfrog scheme (see section 4.3.4) implemented as integrator. This results in the following prescription:

$$M_{xx'} \equiv S^{-1}_{xx'} - \delta_{\rm K}^{xx'} D_1 \frac{\partial J_x(\delta^{\rm i})}{\partial\delta^{\rm i}_x}(\xi_x), \tag{4.39}$$

where $\delta_{\rm K}$ is a Kronecker delta symbol and ξ_x is assumed to be the mean initial density contrast in high probability regions, i.e. once the sampler has moved beyond the burn-in phase.

Due to the high-dimensionality of the problem, inverting M and storing M^{-1} is computationally impractical. Therefore, a diagonal mass matrix is constructed from equation (4.39).

4.3.4 The leapfrog scheme integrator

For computer implementation, Hamilton's equations, (4.37) and (4.38), must be approximated by discretizing time, using some small stepsize, ε. Several choices of integrator, such as the popular Euler's method, are possible (see section B.5.1).

As discussed in section 3.4.3, it is essential that the adopted scheme respect reversibility and symplecticity, to ensure incompressibility in phase space. Additionally, achieving high acceptance rates require the numerical integration scheme to be very accurate in order to conserve the Hamiltonian. For these reasons, the integrator adopted for implementing BORG is the leapfrog scheme (e.g. Birdsall & Langdon, 1985), which relies on a sequence of "kick–drift–kick" operations that work as follows (see also figure B.3):

$$p_x\left(t+\frac{\varepsilon}{2}\right) = p_x(t) - \frac{\varepsilon}{2}\frac{\partial\psi(\delta^{\mathrm{i}})}{\partial\delta^{\mathrm{i}}_x}\left(\delta^{\mathrm{i}}_x(t)\right), \tag{4.40}$$

$$\delta^{\mathrm{i}}_x(t+\varepsilon) = \delta^{\mathrm{i}}_x(t) + \varepsilon\frac{p_x\left(t+\frac{\varepsilon}{2}\right)}{m_x}, \tag{4.41}$$

$$p_x(t+\varepsilon) = p_x\left(t+\frac{\varepsilon}{2}\right) - \frac{\varepsilon}{2}\frac{\partial\psi(\delta^{\mathrm{i}})}{\partial\delta^{\mathrm{i}}_x}\left(\delta^{\mathrm{i}}_x(t+\varepsilon)\right), \tag{4.42}$$

where m_x is the element of the diagonal mass matrix at position x.

The equations of motion are integrated by making n such steps with a finite step size ε. In order to prevent resonant trajectories, time steps are slightly randomized (ε is randomly drawn from a uniform distribution).

4.4 Testing BORG

Demonstrating of the performance of the BORG algorithm is the subject of sections 6 and 7 in Jasche & Wandelt (2013a). As these results are relevant to set the BORG SDSS analysis on firm statistical grounds, in the following, we briefly report on the original test using mock observations.

4.4.1 Generating mock observations

The first step is to generate an initial Gaussian random field (see section B.3). This was done on a three-dimensional Cartesian grid of 128^3 voxels covering a comoving cubic box of length 750 Mpc/h with periodic boundary conditions. The Fourier-space covariance matrix includes an Eisenstein & Hu (1998, 1999) cosmological power spectrum with baryonic wiggles. The cosmological parameters are fixed at fiducial values,

$$\Omega_\Lambda = 0.78, \Omega_{\mathrm{m}} = 0.22, \Omega_{\mathrm{b}} = 0.04, \sigma_8 = 0.807, h = 0.702, n_{\mathrm{s}} = 0.961. \tag{4.43}$$

The Gaussian initial conditions are populated by a Lagrangian lattice of 256^3 particles, that are propagated forward in time using the same implementation of second-order Lagrangian perturbation theory as used in BORG. The final density field is constructed from the resultant particle distribution using the cloud-in-cell scheme. Note that it is crucial to use the 2LPT model for structure formation at this point, instead of, for example, a full N-body simulation, in order to demonstrate that BORG correctly infers the input field. Only in this fashion can we demonstrate that the BORG complicated statistical machinery works, and compare the input and output without differences due to additional physics.

An artificial tracer catalog is then generated by simulating an inhomogeneous Poisson process characterized by equations (4.13) and (4.20) (see also figure 4.3 for an illustration). For the purpose of the test run, the problem is simplified to only one luminosity bin ($\ell = 0$), the mean number of galaxies $\bar{N}^0$ is fixed, and the tracers are supposed to be unbiased (which amounts to fixing $\alpha^0 = 1$, $\beta^0 = 1$). However, the survey response operator R^0_x involves a highly-structured survey mask (mimicking the geometry of the Sloan Digital Sky Survey data release 7) and realistic selection functions (based on standard Schechter luminosity functions), in order to demonstrate the possibility of doing large-scale structure inference from real data sets.

4.4.2 Convergence and correlations of the Markov Chain

As mentioned in section 3.4.3, HMC is designed to have the target distribution as its stationary distribution. Therefore, the sampling process provides samples of the posterior distribution (equation (4.23)) after an initial burn-in phase. Jasche & Wandelt (2013a) showed that during this phase, of the order of 600 samples, the power spectrum converges at all scales towards the true power in the initial density field. The absence of any power excess or deficiency demonstrates the correct treatment of the response operator. The analysis also showed that burn-in also manifests itself in the acceptance rate, which has a dip around after 100 samples, then increases and asymptotes at a constant value of around 84%.

Generally, successive samples of the chain will be correlated to previous samples. The correlation length of the chain determines the amount of independent samples that can be drawn from the total chain. Jasche & Wandelt (2013a) estimated the correlation length to about 200 samples and obtained a total of 15,000 samples; which amounts to around 72 independent samples after burn-in.

These statistical tests demonstrate that exploring the large-scale structure posterior is numerically feasible despite the high dimensionality of the problem.

4.4.3 Large-scale structure inference

This section discusses the large-scale structure inferred via the application of BORG to the synthetic data set. Figure 4.5 shows slices through various three-dimensional quantities: the true initial density field, one sample of initial conditions, the posterior mean for the initial density field; the same quantities for final density fields; the posterior standard deviation in the initial and final conditions; and the mock data set.

Comparison of initial and final density fields permits to check the correspondence between structures with growing statistical complexity. Furthermore, comparison of final density fields to the data demonstrates the accuracy of the inference of the underlying dark matter density field. In particular, one can see that the algorithm extrapolates unobserved filaments between clusters, based on the physical picture of structure formation provided by 2LPT. At high redshift or near the survey boundaries, complex structures appear continuous, which proves that the algorithm augments unobserved or poorly constrained regions with statistically correct information, consistently with the structure formation model. Therefore, each individual sample is a physical dark matter realization, to the level of accuracy of 2LPT.

The variation between samples quantifies joint and correlated uncertainties. This is illustrated in figure 4.5 by unobserved regions in the posterior means, where the values in different samples average to cosmic mean density, and by the posterior standard deviations. Therefore, contrary to other reconstruction approaches found in the literature, BORG possesses a demonstrated capability of quantifying uncertainty of inferred maps, locally and globally. These uncertainties can then be propagated to any derived quantity, as we demonstrate for example with cosmic web types in chapter 9.

Finally, Jasche & Wandelt (2013a) demonstrated that the inferred initial density contrast follows Gaussian one-point statistics, that inferred density fields cross-correlate with the true solution as expected (i.e. $R(k) \equiv P_{\delta_{\rm inferred}\times\delta_{\rm true}}/\sqrt{P_{\delta_{\rm inferred}}P_{\delta_{\rm true}}} \to 1$ as $k \to 0$), and that BORG also infers the underlying velocity field in detail.

4.5 Future extensions of BORG

The method described in this chapter forms the basis of a sophisticated, but also extensible, physical large-scale structure inference framework. In particular, natural extensions of the BORG algorithm would enable automatic calibration of bias parameters (the exponents α^ℓ in previous sections) and of the covariance matrix of initial fluctuations (the matrix S). This would allow precise inference of the early-time matter power spectrum from biased catalogs of tracers. As noted in the introduction, this endeavor could yield a vast gain of information for the determination of cosmological parameters, in comparison to state-of-the-art techniques.

Let us consider a set of comoving wavenumbers $\{k_n\}$ and let us denote by $P \equiv \{P(k_n)\}$ the set of corresponding power spectrum coefficients. Since direct sampling from $\mathcal{P}(P|d)$ is impossible, or at least difficult, Jasche *et al.* (2010a) proposed to explore the full multi-dimensional joint posterior of power spectra coefficients and density fluctuations, $\mathcal{P}(\delta^{\rm f}, P|d)$. They employ a two-steps Gibbs sampling scheme, a method previously applied to CMB data analysis (Wandelt, Larson & Lakshminarayanan, 2004; Eriksen *et al.*, 2004; Jewell, Levin & Anderson, 2004):

$$\delta^{\rm f} \curvearrowleft \mathcal{P}(\delta^{\rm f}|P, d), \tag{4.44}$$

$$P \curvearrowleft \mathcal{P}(P|\delta^{\rm f}, d), \tag{4.45}$$

where the arrow denotes a random draw from the pdf on its right. The ARES code is an implementation of this scheme. It assumes the conditional independence $\mathcal{P}(P|\delta^{\rm f}, d) = \mathcal{P}(P|\delta^{\rm f})$, which yields an inverse-Gamma distribution for power spectrum coefficients, and a Gaussian prior for $\delta^{\rm f}$ (i.e. a Wiener posterior for $\mathcal{P}(\delta^{\rm f}|P, d)$; see Jasche *et al.*, 2010a). In Jasche & Wandelt (2013b), updates and improvements of ARES are introduced, in order to account for uncertainties arising from galaxy biases and normalizations of the galaxy density (i.e. noise levels).

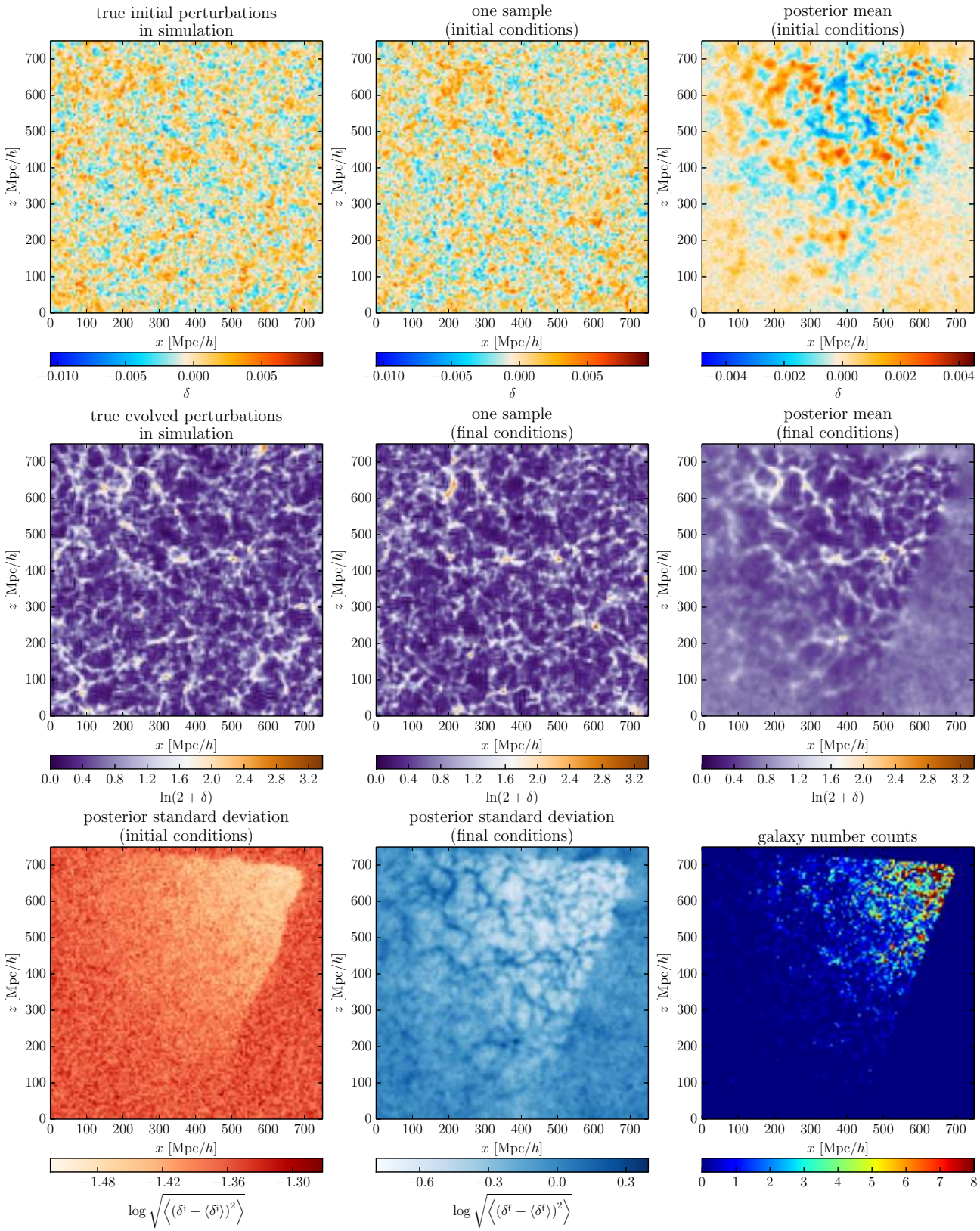

Figure 4.5: Slices through the box used for testing BORG on a synthetic data set. Various quantities (indicated above the panels) are shown. The comparison between panels illustrates the performance of BORG at inferring density fields and demonstrates its capability of quantifying uncertainties. This figure shows results originally obtained by Jasche & Wandelt (2013a), courtesy of Jens Jasche.

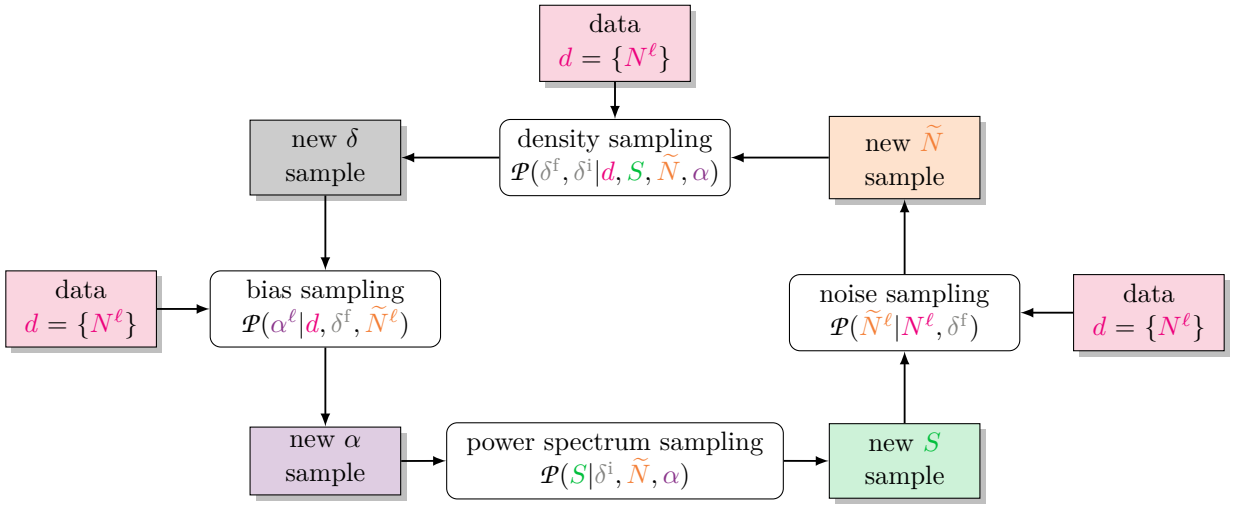

Figure 4.6: Flow chart depicting the multi-step iterative block sampling procedure for a natural extension of the BORG algorithm. In the first step, BORG generates random realizations of initial and final density fields conditional on the galaxy samples d, on the covariance matrix of initial fluctuations, S, on the noise levels $\{\widetilde{N}^\ell\}$ and on the bias parameters $\{\alpha^\ell\}$. In subsequent steps, the bias parameters, the covariance matrix and the noise parameters are sampled conditional on respective previous samples and on the data when necessary. Iterations of this procedure yield samples from the full joint posterior distribution, $\mathcal{P}(\delta^{\rm f}, \delta^{\rm i}, S, \widetilde{N}, \alpha|d)$.

Following these ideas, an extended BORG algorithm should perform iterative block sampling according to the scheme given in figure 4.6 (for reference, see also figure 4.4 for the current BORG algorithm, and figure 1 in Jasche & Wandelt, 2013b, for the ARES algorithm). In comparison to the conditional posterior expressions written down by Jasche *et al.* (2010a) and Jasche & Wandelt (2013b), this procedure would involve the expression of $\mathcal{P}(\alpha^\ell|d, \delta^{\rm f}, \widetilde{N}^\ell)$ in terms of the BORG power-law bias model (instead of the linear bias model of ARES) and of $\mathcal{P}(S|\delta^{\rm i}, \widetilde{N}, \alpha)$ in terms of initial (instead of final) density fields.[2] In ARES, density sampling is by far the most expensive step. It can be done by constructing the Wiener-filtered map (which requires inversions of large matrices, see equations (1.27) and (1.28)) and augmenting missing fluctuations from the prior (Jasche *et al.*, 2010a), by means of HMC (Jasche & Wandelt, 2013b), or by using an auxiliary messenger field, which removes the need for matrix inversion (Jasche & Lavaux, 2015; see also Elsner & Wandelt, 2013). For the BORG data model, involving a structure formation model instead of a Gaussian prior for the galaxy density, HMC is the state-of-the-art technique.

An upcoming improvement of BORG will involve the joint sampling of density $\delta^{\rm i}$, noise levels $\widetilde{N}^\ell$ and bias parameters α^ℓ. Unfortunately, computational time issues mean that joint, physical inference of density and power spectra is still out of reach. Correlation lengths are of the order of 200 samples for BORG density fields (Jasche & Wandelt, 2013a) and 100 samples for ARES power spectrum coefficients (Jasche & Wandelt, 2013b).[3] Preliminary tests indicate that the correlation length for the joint inference process is of the order of a few hundred samples. However, even with a correlation length of 100 samples, accurate characterization of power spectra and corresponding uncertainties require, at least, about 40,000 samples. With the current performance of the BORG sampler (discussed in sections 4.4.2 and 5.2), such a run would take several years on a typical computer. For this reason, this thesis focuses on sampling the matter density field for a fixed power spectrum of primordial fluctuations, rather than sampling this as well. Algorithmic and methodological innovations that would render such a run possible are currently being discussed but will require a considerable additional implementation effort and are outside the scope of this thesis.

[2] As noted in section 4.2.1.1, the Fourier-space representation of S is a diagonal matrix containing the coefficients $\sqrt{P(k)/(2\pi)^{3/2}}$.
[3] See Jasche & Wandelt, 2013b; Jewell *et al.*, 2009, for the discussion of a method designed to reduce the otherwise prohibitively long correlation length of ARES chains.