

Bayesian reconstruction of the cosmological large-scale structure: methodology, inverse algorithms and numerical optimization

F. S. Kitaura* and T. A. Enßlin

Max-Planck Institut für Astrophysik, D-85748 Garching, Germany

1 February 2008

ABSTRACT

We address the inverse problem of cosmic large-scale structure reconstruction from a Bayesian perspective. For a linear data model, a number of known and novel reconstruction schemes, which differ in terms of the underlying signal prior, data likelihood, and numerical inverse extra-regularization schemes are derived and classified. The Bayesian methodology presented in this paper tries to unify and extend the following methods: Wiener-filtering, Tikhonov regularization, Ridge regression, Maximum Entropy, and inverse regularization techniques. The inverse techniques considered here are the asymptotic regularization, the Jacobi, Steepest Descent, Newton-Raphson, Landweber-Fridman, and both linear and non-linear Krylov methods based on Fletcher-Reeves, Polak-Ribière, and Hestenes-Stiefel Conjugate Gradients. The structures of the up-to-date highest-performing algorithms are presented, based on an operator scheme, which permits one to exploit the power of fast Fourier transforms. Using such an implementation of the generalized Wiener-filter in the novel ARGO-software package, the different numerical schemes are benchmarked with 1-, 2-, and 3-dimensional problems including structured white and Poissonian noise, data windowing and blurring effects. A novel numerical Krylov scheme is shown to be superior in terms of performance and fidelity. These fast inverse methods ultimately will enable the application of sampling techniques to explore complex joint posterior distributions. We outline how the space of the dark-matter density field, the peculiar velocity field, and the power spectrum can jointly be investigated by a Gibbs-sampling process. Such a method can be applied for the redshift distortions correction of the observed galaxies and for time-reversal reconstructions of the initial density field.

Key words: large-scale structure of Universe – galaxies: distances and redshifts – methods: data analysis – methods: statistical – methods: numerical – techniques: image processing

1 INTRODUCTION

According to our current picture of cosmogenesis, the galaxies, galaxy clusters, galaxy filaments, and giant voids forming the cosmic large-scale structure (LSS) are products of gravitational instability, which pulls increasingly more matter onto the tiny primordial seed density fluctuations generated at the very first epoch of inflation. The shape and size of the cosmic matter distribution reflects the initial conditions set during or shortly after Big Bang, as well as the interplay of the gravitational self-attraction of matter and the diluting action of the Hubble expansion of cosmic space. Valuable information about the properties and the origin of the cosmic inventory are encoded in the LSS, however, on small-scales, that information is being erased through dynamical non-linear processes.

Our goal is to extract as much of this information as possible from astronomical measurements, which introduce uncertainties

and, consequently, degeneracies. Therefore, we have to adapt an information-theoretical approach to solve the reconstruction problem of cosmography. The Bayesian framework turns out to be the most general approach as we will discuss later. In this paper we present the novel ARGO¹-software package, which reconstructs the three-dimensional density field from the information provided by galaxy surveys with different Bayesian and inverse methods. Here we focus our study on understanding the Bayesian theoretical background and the required algorithmic aspects. Further extensions of the code in which the power-spectrum and the peculiar velocities can be jointly sampled are outlined. These applications are planned to be shown in further publications in which they will be first tested on mock galaxy catalogues.

The large number of telescopes performing galaxy surveys with increasing depth, sky coverage, and accuracy in position and distance (or redshift) determination provide us with superb data on

* E-mail: kitaura@mpa-garching.mpg.de

¹ Algorithm for the **R**econstruction of **G**alaxy-traced **O**verdensities

the cosmic matter distribution at an exponentially increasing rate. One problem is that the discrete objects these instruments reveal to us, the galaxies, are the result of a complex non-linear evolution of cosmic matter combined with complicated astrophysical processes such as star formation. A translation of the galaxy data into the much better understood large-scale dark matter (DM) distribution, which would be much easier to analyse for imprints of cosmologically interesting effects, is far from trivial. The discrete nature of galaxies introduces certain noise, usually modeled by shot noise. Moreover, the partially understood galaxy-formation process inserts systematic uncertainties. In addition, the limited volume of surveys adds complications beyond the problems of galaxy-distance determination being contaminated by observational and velocity redshift-distortions. All these complications have to be dealt with simultaneously and in a controlled fashion. Since it cannot be assumed that the correct or optimal values for the various degrees of freedom of the problem (bias factors, redshift-corrections, etc.) will be guessed a priori, repeated and iterative data analysis is mandatory in order to achieve a high-fidelity and well-understood cosmic map. For example, a correction of redshift-distortions of the galaxies requires the gravitational potential generated by the matter distribution to be reconstructed.

Repeated generation of cosmic matter maps increases the urge to face another challenge, the scaling of the performance of the underlying map-generation algorithms with the data size. Since the matter-density information displayed at a location on a map may depend on all input data (galaxy positions), any algorithm optimised to information theory scales super-linear². With increasing survey sizes, increasing requirements for spatial resolution and volume coverage, and the need to frequently re-iterate the map-generation step, the algorithm has to scale closely to linear with data size, otherwise its application is strongly limited. Former applications in cosmography suffered from such inconvenient performance-scaling, and an effort has to be made to develop simultaneously high-performance and accurate methods.

The work presented in this paper develops the general methodology of Bayesian reconstruction of the cosmic matter distribution, based on the invaluable pioneering work of many other scientists, which will be discussed below, and extends this work to a series of new applications. Existing and novel map making algorithms are summarized in terms of a classification of their Bayesian likelihood and prior functions. The implementation, optimisation, and comparison of various numerical schemes are addressed in detail. This provides a starting point for a correct information-theory approach to cosmography. Many additional problems, not addressed in this paper, such as the galaxy bias, will also have to be solved before accurate maps of the dark matter distribution in our still mysterious Universe can be generated.

Such an undertaking would be highly rewarded in the short and long run. An accurate map of the cosmic matter distribution would be valuable for a manifold of direct scientific applications. These range from structure-formation analysis, to cosmological parameter estimation via power-spectrum measurements, dark energy studies, galaxy-cluster identification and galaxy-bias studies. Accurate cosmic maps would help to determine weak signals associated with the large-scale structure such as the integrated Sachs-Wolf (ISW) effect, or the extended Sunyaev-Zel'Dovich (SZ) ef-

² A map of galaxy counts can be generated by an algorithm with linear scaling to data size however, it is not an optimal representation of the underlying matter field.

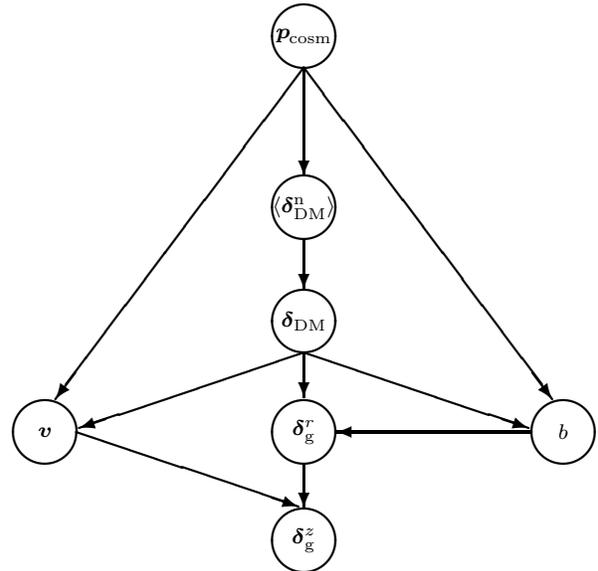


Figure 1. The hierarchical Bayes model for a galaxy distribution in redshift space δ_g^z is represented here in a directed acyclic graph (DAG). The cosmological parameters p_{cosm} govern the rest of the variables. The initial density field coming from e.g. inflationary scenarios can be statistically described by all its moments $\langle \delta_{\text{DM}}^n \rangle$. Here the power spectrum is usually taken, since the initial perturbations are well described by a Gaussian realization of the initial seed fluctuations. The further evolution is described by nearly deterministic processes (given by structure and galaxy formation), which determine the later-time dark matter distribution δ_{DM} with its peculiar velocity field v and the bias function b that relates the galaxy distribution to the dark matter density field. The dark matter distribution δ_{DM} with the bias produces the galaxy distribution in real space δ_g^r . The peculiar velocities v related to the density field through the continuity equation introduce the redshift distortion in δ_g^r finally leading to the galaxy distribution in redshift space δ_g^z .

fect, the detection of which relies on the construction of optimal statistical filters for these signals.

Finally, one could argue that mapping the distribution of matter in the Universe represents a response to mankind's curiosity in its aim to discover *terra incognita* and find an orientation in space and time on cosmological scales and, therefore, should be a goal in itself.

In the remainder of this introduction we give the sources of uncertainties, we present an overview of existent and new Bayesian reconstruction methods, subsequently we briefly describe the algorithmic development presented in this paper, then we summarize non-Bayesian methods and time-reversal reconstruction methods, and in the final part we give a more detailed overview of the structure of this paper.

1.1 Classes of uncertainty

Several classes of uncertainties related to the density-field reconstruction from galaxy surveys demand a statistical approach. Some of the uncertainties are intrinsic to the nature of the underlying signal (the dark matter). Other uncertainties are intrinsic to the nature of the observable (the galaxies). And finally there are uncertainties due to degeneracies which appear through the observation process.

(i) **cosmic variance:** In cosmology it is generally assumed that the structure of the Universe comes from some infinitesimal quantum fluctuations which were frozen out and stretched by an inflationary phase (see Guth 1981; Guth & Pi 1982; Starobinsky

1982; Hawking 1982; Linde 1982; Albrecht & Steinhardt 1982; Bardeen et al. 1983), and later amplified by gravitational instability. According to this picture, the seed fluctuations would have an intrinsic stochastic character and are mainly Gaussian distributed. However, the mechanisms that stretch the quantum fluctuations may also introduce deviations from Gaussianity which would then be imprinted in the seed fluctuations. In general all the moments of the initial fluctuations have to be considered (δ_{DM}^n). Nevertheless, most of the inflationary scenarios predict the density field to be very closely Gaussian distributed and it is generally sufficient to take the second order moment, the two-point correlation function, or the power-spectrum in Fourier-space. We will discuss below how to determine the power-spectrum and techniques to disentangle intrinsic non-Gaussianities within a Bayesian framework. Note that there are alternative models to inflation in which e.g. the seed fluctuations are identified with the topological defects that remain as relics of high-energy phase transitions (Kibble 1976). Accurate reconstructions of the LSS could help to discriminate between the different models.

(ii) **redshift-distortions:** The peculiar motion of galaxies with respect to the Hubble flow of the Universe v introduces uncertainties in their redshift measurement, the so-called redshift-distortions (see e.g. Hamilton 1998, for an introduction to this problem). The measured galaxy overdensities are thus said to be not in real-space δ_{g}^r but in redshift-space δ_{g}^z . In the linear regime, in which galaxies fall into the potential wells on large scales, redshift-distortions cause a squashing of the linear overdensities in radial direction. However, in the non-linear regime, galaxies (e.g. in a galaxy cluster) tend to behave like particles in a gas with randomized motions inside the clusters where the potentials are very high and produce the so-called *finger-of-god* effect, a dispersion along the line of sight. The correction of these distortions is not trivial, since structures erase the information coming from the initial fluctuations after entering the non-linear regime. Consequently, determining the real position of galaxies poses a degenerate problem, which has many possible solutions. Many efforts have been made to correct for these distortions: in the linear regime these efforts start with Kaiser’s pioneering work (see Kaiser 1987) and are followed by the linear redshift-distortions operator (for a detailed derivation see Hamilton 1998). In the non-linear regime, these efforts include a velocity dispersion factor (the *dispersion*-model) corresponding to an exponential pairwise velocity distribution function with no mean streaming (see Ballinger et al. 1996). See Scoccimarro (2004) for an exact relationship between real-space and redshift-space two-point statistics through the pairwise velocity distribution function including all non-linearities. More complex methods of correcting for redshift-distortions were classified by Schmoldt et al. (1999) into iterative methods, that use the redshift-space density to calculate a peculiar velocity field, which can in turn be used to correct the density field distortions (Yahil et al. 1991; Kaiser & Stebbins 1991), and into basis function methods, in which the redshift-space density field is transformed into a combination of angular and radial basis functions from which the radial redshift-distortion is corrected (see e.g. Nusser & Davis 1994; Schmoldt et al. 1999). We propose below a Bayesian method to correct for the linear and non-linear redshift-distortions in a statistical way (see section 2.6).

(iii) **galaxy bias:** The galaxy formation process is a complicated, non-linear and (probably) non-local process. It is known that on large scales the galaxy power-spectrum fits well to the expected DM spectrum predicted from cosmic microwave background (CMB) observations, if some bias factor b between the amplitude of the galaxy and DM fluctuations is assumed. De-

tailed studies show that the bias factor is not universal, but depends on galaxy type, galaxy formation time, redshift, etc. (see e.g. Cooray & Sheth 2002, and references therein). For the purpose of reconstructing the underlying density field, linear biases can easily be tackled within the linear data model described below by including its effects in a selection function. Nevertheless, more complex biases have to be further investigated in a Bayesian framework. Physical processes, which are not perfectly understood within galaxy formation may be treated in a statistical way, encoding the ignorance about certain physical processes in probability distribution functions. Several works study the stochastic nonlinear galaxy biasing (see for example Pen 1998; Dekel & Lahav 1999; Tegmark & Bromley 1999).

(iv) **sampling uncertainties:** The model connecting the continuous dark matter field and the discrete galaxy distribution introduces additional uncertainties. As we have discussed under item (iii) complex physical processes are present in galaxy formation. The usually assumed Poissonian distribution (shot noise) for the galaxy distribution is thus only a crude approximation of the underlying sampling process. Thus the remaining question is what degree of accuracy one wants to achieve in the reconstruction. Nevertheless, further investigation is required in this field. Other sampling uncertainties are not intrinsic to the observable, but come from the mathematical representation we choose. Treating galaxies as counts in cells, for instance, will smooth out the information about their measured position. The resolution level determines again the validity of the method.

(v) **measurement:** The action of measurement introduces uncertainties, either due to the instruments, e.g. blurring by the telescope, or due to the observational strategy, which is included in the noise term, the selection function, and the mask effects. Many of these aspects were already discussed by Zaroubi et al. (1995). We will analyze these issues throughout this paper and propose new solutions to tackle the different problems.

Consequently, extracting the underlying dark matter density field from the luminous matter distribution given by such surveys poses a classical signal reconstruction problem. A Bayesian network depicting the relation of these uncertainties is displayed in fig. (1).

1.2 Bayesian reconstruction methods

Any Bayesian statistical approach requires the definition of a likelihood and a prior. The former is the probability distribution function describing the process generating the observational data. It can be interpreted as a distance measure of the observed data to the underlying signal, as we will discuss below. The prior stands for the distribution function modeling our prior knowledge on the signal to be recovered. Mathematically it can be shown that it regularizes the estimator in the presence of noise (see section 2.5.1). Two kinds of priors have to be distinguished, informative priors, in which the previous physical knowledge about the signal is encoded, and non-informative priors, which try to give objective estimators for the underlying signal based on purely information-theoretical arguments. Here, three non-informative priors are considered: flat priors (see section 2.5.5) with a constant probability distribution function (PDF), entropic priors based on Shannon’s notion of information (see section 2.5.9), and Jeffrey’s prior based on invariant statistical structures under transformation of variables (see section 2.5.8). Finally, a maximization or sampling of the posterior distribution, which is proportional to the product of the likelihood and the prior, has to be done to complete the Bayesian estimation. The

maximization of the posterior is called the maximum a posteriori method (MAP). The maximum likelihood (ML) and maximum entropy method (MEM) are particular cases of the MAP with flat priors and entropic priors, respectively. Complex posterior distribution functions may be sampled iteratively from conditional PDFs in a Markov Chain Monte Carlo fashion (MCMC), see section 2.6. We show how different choices for these distribution functions together with the estimation procedure lead to different reconstruction algorithms, which consequently have distinct application fields (see table 1). A review of existing methods is presented and new applications for the large-scale structure reconstruction, which naturally emerge within the Bayesian formalism, are developed.

In this work we consider Poissonian and Gaussian likelihoods for the galaxy distribution. The former has been previously considered in image restoration especially for deconvolution purposes (see Richardson 1972; Lucy 1974). For example, the Richardson-Lucy algorithm can be derived as the ML of a Poissonian likelihood (see Shepp & Vardi 1982, and appendix B). Here an image can be regarded as photon counts in cells represented by a Poissonian distribution. However, one should notice that this likelihood does not represent the galaxy-formation process. From a pure image reconstruction perspective, it can still be interesting for LSS estimations, because it naturally represents the discrete nature of a galaxy distribution. The Gaussian likelihood allows the incorporation of arbitrary noise structures through the variance. The CMB map-making algorithms, which aim to convert time-ordered data received from satellites into a map of the CMB signal on the sky as a projection on the sphere, usually use this likelihood. In this case, the ML leads to the simple COBE-filter first derived by Janssen & Gulkis (1992). Nevertheless, the complex scanning strategies and foreground removal can add unlimited complexity to these algorithms (e.g. Natoli et al. 2001; Doré et al. 2001; Stompor et al. 2002; Keihänen et al. 2005; Yvon & Mayet 2005).

For the LSS the Gaussian prior arises as the natural informative prior due to the arguments discussed above. We propose a novel algorithm: GAPMAP, which maximizes the posterior with a Gaussian prior and a Poissonian likelihood (see section 2.5.4 and appendix A). In contrast, the Gaussian likelihood with the Gaussian prior leads to the well-known Wiener-filter, which has been used for the LSS reconstruction (see Fisher et al. 1994; Hoffman 1994; Lahav et al. 1994; Lahav 1994; Zaroubi et al. 1995; Fisher et al. 1995; Webster et al. 1997; Zaroubi et al. 1999; Schmoldt et al. 1999; Erdoğan et al. 2004, 2006) and for CMB-mapping (see e.g. Bunn et al. 1994; Tegmark 1997). It is also known to give optimal results in terms of yielding the least square error, see the pioneering work of Rybicki & Press (1992) and Zaroubi et al. (1995). We present in this paper a fast Wiener-filter extra-regularized with Krylov methods as we will see below.

Intrinsic primordial non-Gaussianities can be imprinted in the seed fluctuations depending on the particular theory responsible for the amplification of the fluctuations coming from the early Universe. To find such deviations, non-informative priors, which give non-linear estimates for the underlying signal are required. Entropic priors are well suited here, and have been previously applied for CMB studies. We extend this work for LSS reconstructions and develop the corresponding maximum entropy method for Gaussian and Poissonian likelihoods (see section 2.5.9 and appendix E).

Sampling methods have the advantage of determining the shape of distributions and, thus, leading to a natural estimate of the uncertainty of the estimator. Moreover, the mean can be calculated easily from the sample and is known to give more accurate

results than the maximum in the case of asymmetric PDFs (see e.g. Tanner 1996).

As an example, Hobson & McLachlan (2003) proposed a SZ-cluster detection algorithm using the Metropolis-Hasting algorithm method based on a Poissonian prior distribution, which is designed to find discrete objects. Recently Sutton & Wandelt (2006) developed a reconstruction method for radio-astronomy that samples from the multiplicity function (see eq. 30). Alternative approaches to the maximum likelihood for CMB-mapping algorithms try to jointly reconstruct the CMB-map with its power-spectrum using Gibbs-sampling techniques (Wandelt et al. 2004; Eriksen et al. 2004; O'Dwyer et al. 2005). This approach is especially efficient with respect to other MCMC methods because the transition probability matrix moves the system in each step of the chain. For this special case the importance ratio is always one (see e.g. Neal 1993). This MCMC method requires, however, the complete knowledge of the full conditional PDFs in order to sample from them. Note that the Gaussian prior for the signal simultaneously represents the likelihood for the power-spectrum given the signal, which in this case is an inverse Gamma function for the power-spectrum (see section 2.6.2). This distribution naturally samples the power-spectrum, which strongly deviates from Gaussianity.

With the aim of estimating the power-spectrum in an objective way, non-informative priors are used. Usually a flat prior is taken for the power-spectrum. Alternatively, Jeffrey's prior, for which we give a derivation based on Fisher information (see appendix D), can be used. Alternatively, an entropic prior could also be taken.

Other attempts have been made to estimate the power-spectrum from the LSS based on the distribution of galaxies. A modified Gaussian PDF with a log-normal mean has been used in this approach (see Percival 2005). The same kind of concept, using a modified Gaussian distribution to sample deviations from Gaussianity, has been applied to SZ-cluster detection by Pierpaoli & Anthoine (2005).

In this paper we propose to apply a Gibbs-sampling algorithm to jointly sample the underlying three-dimensional density field with the power-spectrum and the peculiar velocities, which can be used to correct for the redshift-distortions (see 2.6). Applications of this method will be presented in forthcoming publications. Note that the peculiar velocities can also be used to trace the initial density fluctuations back in time as we will discuss below.

1.3 Algorithmic development

In this paper we focus our work on the numerical optimization of inverse techniques to show that a joint estimation of the LSS matter density field and its parameters is feasible (see sections 3 & 4).

The calculation of the reconstructions, either through maximization or through sampling, requires the inversion of certain matrices. For the Wiener-filter, for instance, the reconstruction problem consists in one of its steps on the inversion of the correlation matrix of the data. The methods used in this field so far calculated this matrix and inverted it mainly using the Singular Value Decomposition algorithm that scales as $\mathcal{O}(n^3)$ for a $n \times n$ matrix (see e.g. Zaroubi et al. 1995). However, this approach seems to be hopeless in light of the overwhelming amounts of data coming from different surveys and the possibility of combining them. We made special effort to implement an algorithm in which the involved matrices would not need to be stored taking advantage of an operator formalism, which we worked out here for different reconstruction methods (see table 3 and section 3.3). Such a formalism also allows fast iterative numerical methods that speed the inverse step up to a

scaling of $\mathcal{O}(n \log_2 n)$ thus reducing the main operations to fast Fourier transforms (FFTs). Some of these numerical schemes have been used in CMB-mapping algorithms, but were lacking a detailed comparison of the efficiency of the different methods. Such a comparison is presented here. We derive the different inverse methods in a unified way starting with a Bayesian motivation for iterative schemes (see appendix F) and following with a general formulation of the asymptotic regularization from which the Jacobi, the Steepest Descent, and the Krylov methods are derived. Moreover, non-linear inverse methods are discussed, like the Newton-Raphson, the Lanweber-Fridman and the non-linear Krylov methods. Preconditioning (see appendix G) was taken into account in all the derivations and the importance of such a treatment is tested in section (4) (see fig. 5). In addition, a previously not discussed Krylov method is derived (see formula 94, section 3 and appendix H) and its superior efficiency is demonstrated (see section 4).

1.4 Non-Bayesian reconstruction methods

Let us mention here that there are alternative reconstruction methods which recover the underlying density field based on the observed radial peculiar velocity of galaxies, such as the widely known POTENT-code (Bertschinger & Dekel 1989; Bertschinger et al. 1990; Bertschinger & Dekel 1991). Kaiser & Stebbins (1991) propose a maximum probability technique to reconstruct the density field from peculiar velocities.

Other works are focused on reconstructing the peculiar velocities from density fields (see e.g. Branchini & Plionis 1996; Branchini et al. 1996; Kudlicki et al. 2000; Mohayaee et al. 2004; Mohayaee & Tully 2005). For a review see (Zaroubi 2002a) and references therein.

In addition, several reconstruction techniques, which we do not discuss here are based on geometrical arguments. These techniques include Voronoi tessellations (see e.g. Icke & van de Weygaert 1991; Ebeling & Wiedenmann 1993; Zaninetti 1995; Doroshkevich et al. 1997; Meurs & Wilkinson 1999; Kim et al. 2000; Ramella et al. 2001; Panko & Flin 2004; Zaninetti 2006), Delaunay tessellations (see e.g. Bernardeau & van de Weygaert 1996; Schaap & van de Weygaert 2000; van de Weygaert & Schaap 2001), *friends-of-friends* algorithms (see e.g. Botzler et al. 2004) or *cloud-in-cell* interpolation schemes (see e.g. Gottlöber et al. 2002).

A widely known reconstruction method in various fields is the Pixon method (see e.g. Puetter & Pina 1993). Unlike Bayesian methods, this method does not assign explicit prior probabilities to image models. Instead, it restricts them by seeking minimum complexity. The Pixon method minimizes complexity by smoothing the image model locally as much as the data allow, thus reducing the number of independent patches, or Pixon elements, in the image. For a recent application in astrophysics see e.g. Eke (2001).

1.5 Time-reversal reconstruction of the initial density field

The reconstruction of the initial density fluctuations is closely related to the reconstruction of the large-scale density field at the observed epoch focused on in this paper. However, we believe that fruitful contributions to the field of initial density fluctuations, could be extracted from the work presented here. An accurate overdensity field at the observed epoch and the information about the peculiar velocities could be useful to perform such a time reversal reconstruction. Let us thus briefly review the reconstruction schemes developed in this neighbouring area of cosmology.

The initial density field is of major interest because it represents the origin of the Universe and many theories can be tested with such information. As a direct application, constrained N-body simulations can be done by taking the reconstructed field as the initial conditions to study structure formation by later comparing the results with the observations (see e.g. Ganon & Hoffman 1993; Sheth 1995; Bistolas & Hoffman 1998; Mathis et al. 2002).

As we have discussed above, the large-scale structure contains information about the seed perturbations and its dynamical evolution is well approximated in the linear regime. Following this idea, Weinberg (1992) proposes to reconstruct the seed fluctuations through the Gaussianization of the observed density field, based on the approximation that the rank order of the initial density field smoothed over scales of a few Mpc is preserved under non-linear gravitational evolution and further assuming the initial field to be Gaussian distributed. This method can be regarded as an Eulerian Gaussian mapping scheme.

Other methods run gravity backward in time taking the position and peculiar velocities of objects at a certain redshift. Here, different schemes have been proposed: a huge class relies on Lagrangian dynamical schemes; another class is based on the minimal action principle; and another class is based on optimal mass transportation schemes have been applied for the initial density field reconstruction.

Lagrangian dynamical schemes mainly use the Zel'Dovich approximation (Zel'Dovich 1970) in which the comoving trajectories of the particles are straight lines. In this formalism, the variable under consideration is the displacement of a particle. Several reconstruction schemes are based on this approximation, including the Zel'Dovich-Bernoulli equation derived by Nusser et al. (1991), the Zel'Dovich-continuity equation presented in Gramann (1993) or the path interchange Zel'Dovich approximation scheme (PIZA) used by Croft & Gaztanaga (1997), among others (see for example Dekel et al. 1990; Nusser & Dekel 1992; Narayanan & Weinberg 1998; Valentine et al. 2000). Several of these methods are compared in Narayanan & Croft (1999). More recently, it was proposed to determine the inverse Lagrangian map (defined as the transformation of the present (Eulerian) positions to the respective initial (Lagrangian) positions) by minimising a quadratic cost-function, which searches the optimal mass-transport solution of the Monge-Ampère-Kantorovich problem (Frisch et al. 2002; Mohayaee et al. 2003; Brenier et al. 2003; Mohayaee et al. 2006)

The minimal action principle method was pioneered by Peebles (1989, 1990). One of its first applications was presented in Shaya et al. (1995). Here the gravitational instability problem is treated as a two-point boundary problem and the trajectories of the mass particles are solved by minimizing the action integral. This method was extended by Goldberg & Spergel (2000a,b); Goldberg (2001a,b).

1.6 Structure of the paper

This paper is structured as follows: in section (2) we state the problem of signal reconstruction, then we define the data model. Subsequently, we introduce a general statistical perspective within a Bayesian framework from which different solutions to the reconstruction problem are presented, including Wiener-filtering, the COBE-filter, a novel GAPMAP algorithm with a Poissonian likelihood and a Gaussian prior, Jeffrey's prior and the Maximum Entropy method (MEM). Markov Chain Monte Carlo methods (MCMC) that sample the global probability distribution function of the signal and all underlying parameters are presented as the ideal

approach to achieve a full Bayesian solution of the reconstruction problem. In the numerical method section (3), different iterative inverse schemes which have been implemented in ARGO are presented, including a very efficient novel scheme. The operator formalism is worked out for four novel algorithms in large-scale structure reconstruction. The efficiency of the different inverse schemes is tested with the Wiener-filter under different reconstruction cases with synthetic data, including structured noise, blurring, selection function effects, and windowing in section (4). Particular detailed derivations are presented in the appendix.

2 BAYESIAN APPROACH TO SIGNAL RECONSTRUCTION

The reconstruction of a signal (here: DM distribution) given a set of measurements (here: galaxy catalogues) is usually a highly degenerate problem, as we have discussed above, where the signal is undersampled and modified by systematic and intrinsic errors due to the nature of the observable. This is indeed the situation that we are facing, since most of the galaxy redshift surveys have partial sky coverage and the discrete nature of galaxies introduces shot noise.

An expression for the data as a function of the real signal has to be modeled in a first step. The reconstruction problem is classically seen as the inverse of this functional dependence. The solution to this problem is far from being trivial and essential issues, like solution existence, solution uniqueness, and instability of the solving process, have to be considered. Regarding the solution existence, there will be no model that exactly fits the data, since the mathematical model of the physics of the system is approximate and the data contain noise. That forces us to look for optimal solutions, rather than exact solutions. We will have to deal especially with the last two points mentioned above, uniqueness and stability, because an infinite set of possible solutions can fit the data and because of the ill-conditioned character of the system we are treating. A regularization method that stabilizes the inverse process by imposing additional constraints will be required. We show below how the Bayesian framework permits us to do a regularization in a *natural* way and furthermore to jointly estimate the signal and its parameters. The calculation of the Bayesian estimators will require extra-regularization techniques, which will be presented in section (3). We will start posing the inverse problem by defining the model of the data.

2.1 Data model

The galaxy formation process is also known to be a complicated, non-linear and probably non-local process, as mentioned in the introduction. Thus attempts to invert the galaxy distribution into the original DM distribution may appear naive, if not hopeless. However, it is known that on large scales the galaxy power-spectrum fits well to the expected DM spectrum predicted from CMB observations, if some bias factor between the amplitude of the galaxy and the DM fluctuations is assumed. Detailed studies show that the bias factor is not universal, but depends on galaxy type, galaxy formation time, redshift, etc. Currently, a large effort is being made to understand such dependencies. For the time being, we assume that at some point bias models derived from first principles will be available from which we can derive our data model connecting the signal (DM distribution) to our observable (galaxy counts). Such a model will be complex, non-linear and non-local. The techniques to treat such problems are partially already existing; some still have to

be developed. However, we are confident that such techniques will rely on a foundation of methods solving the much simpler linear inversion problems. Since even those are not fully developed in the field of LSS reconstructions we concentrate here on a linear data model. Nevertheless, a few excursions, to non-linear methods are done in preparation of forthcoming work. We will show how the linear model can also be used to account for non-linearities coming from structure formation (see section 2.6).

2.1.1 Linear data model

The general linear reconstruction problem formally can be written as the inverse problem of recovering the signal \mathbf{s} from the observations \mathbf{d} related in the following way

$$d(\mathbf{x}) = \int d\mathbf{y} R(\mathbf{x}, \mathbf{y}) s^\epsilon(\mathbf{y}), \quad (1)$$

where R represents the kernel of the Fredholm integral equation of the first kind defined by (1), with noise on the signal s being expressed by the superscript ϵ . Discretizing eq. (1) and assuming additive noise, we can formulate the signal degradation model by

$$\mathbf{d} = \mathbf{R}\mathbf{s} + \boldsymbol{\epsilon}. \quad (2)$$

where the $m \times 1$ vector \mathbf{d} represents the data points resulting from the measurements (here: galaxy counts), the statistical noise and the underlying signal are a $m \times 1$ vector $\boldsymbol{\epsilon}$, and a $n \times 1$ vector \mathbf{s} respectively. The object that operates on the signal is \mathbf{R} a $m \times n$ matrix which commonly describes blurring effects caused by the atmosphere, the point-spread function (PSF) of the telescope or the response function of the detectors of the instrument.

Let us denote the physical observation process encoded in the \mathbf{R} -matrix as \mathbf{R}_P . We are interested in the selection function of the survey f_S with the corresponding masks f_M , which can also be included in \mathbf{R} . One has to be careful with the data model defined in eq. 2. As several authors point out, there is a correlation between the underlying signal \mathbf{s} and the level of shot noise produced by the discrete distribution of galaxies (see e.g. Seljak 1998). Since, by definition, additive noise assumes no correlation with the signal – otherwise we would have signal content in the noise – we define the effective noise $\boldsymbol{\epsilon}$ as the product of a structure function f_{SF} , which could be correlated with the signal, with a random noise component (ϵ_N) that is uncorrelated with the signal. Given the above definitions, the effective noise $\boldsymbol{\epsilon}$ is uncorrelated with the signal. We may then rewrite eq. (2) in continuous representation as

$$d(\mathbf{x}) = \int d\mathbf{y} R_P(\mathbf{x}, \mathbf{y}) f_S(\mathbf{y}) f_M(\mathbf{y}) s(\mathbf{y}) + f_{SF}(s(\mathbf{x})) \epsilon_N(\mathbf{x}), \quad (3)$$

where $R(\mathbf{x}, \mathbf{y}) = R_P(\mathbf{x}, \mathbf{y}) f_S(\mathbf{y}) f_M(\mathbf{y})$ and $\epsilon(\mathbf{x}) = f_{SF}(s(\mathbf{x})) \epsilon_N(\mathbf{x})$. In practice, we will assume white noise (i.e. constant noise in Fourier space), $\epsilon_N = \epsilon_{WN}$. However, none of the presented techniques in this paper depend on this simplification. Some of the previous studies of large-scale structure reconstruction also included the inverse of the linear redshift-distortions operator as a matrix multiplying \mathbf{R} (see e.g. Lahav et al. 1994). Such an operator cannot easily be found for the non-linear regime. Earlier works try to correct the non-linear redshift-distortions with an additional factor in the power-spectrum analogous to Kaiser's factor (see Kaiser 1987; Ballinger et al. 1996; Erdođdu et al. 2004). Here, we propose a Bayesian solution to the signal reconstruction problem as it will be discussed later.

In most cases, the signal will be strongly underconstrained due to undersampling, i.e. $n \gg m$, which is nearly unavoidable due to

partial sky coverage of surveys. The linear equation (eq. 2) to be inverted is a rank-deficient system. Such systems are characterized by nonuniqueness, since the matrix \mathbf{R} has a nontrivial null space. By superposition, any linear combination of the null space models (models \mathbf{s}_0 that satisfy $\mathbf{R}\mathbf{s}_0 = 0$) can be added to a particular solution leading to infinite solutions. Consequently, we cannot discriminate between situations where the solution is truly zero (see for example Aster et al. 2005). As is well known, a direct inversion of eq. (2) ($\mathbf{R}^{-1}\mathbf{d}$) will amplify the statistical noise and lead to an unstable solution (see e.g. Zaroubi et al. 1995). Instead, a regularization method, which often follows several steps, has to be applied. The first step consists of finding an expression for an estimator of the signal \mathbf{s} that approximately satisfies the data model (eq. 2) and copes with the noise. Further regularization methods are usually required in a second step to actually calculate the estimator. This happens whenever some ill-posed linear or non-linear operators have to be inverted. We shall distinguish between noise regularization and inverse regularization according to the first and the second step, respectively. As Zaroubi et al. (1995) pointed out, using a mean variance estimator alone does not completely solve the inverse problem. Therefore, they proposed the singular value decomposition algorithm (SVD) to extra-regularize these problems. However, this method requires one to calculate the correlation matrix of the data implying a slow algorithm, scaling as $\mathcal{O}(n^3)$, and needs large storage facilities. We will show that a Bayesian approach is a *natural* regularizer for the noise, which then can be regularized further for the inverse purpose with efficient methods that scale as $\mathcal{O}(n \log_2 n)$ (see section 3). Let us address the problem of signal reconstruction from a statistical inference perspective.

2.2 Inversion via statistical estimator

In parametric modeling it is assumed that observational data have been generated by random processes with probability density distributions, depending on the model parameters (see for example Robert 2001). Statistical analysis in this context is essentially an inverse method, which aims at retrieving the causes (here reduced to the parameters of the probabilistic generating mechanism) from the effects (here summarized by the observations).

Traditionally, one tries to find a way where the available information is optimally used and a unique estimator is selected from an infinite set of solutions. One of the classical approaches consists of minimizing the variance of the residuals, which is the variance of the discrepancy between the estimator and the set of possible realizations consistent with the data (see Rybicki & Press 1992). This conjecture is reasonable because the least deviation from the set of *true* signals is searched. The estimator obtained in this way is called the least squares quadratic (LSQ) estimator. However, a transparent statement of the statistical assumptions is missing in this method, contrary to the Bayesian approach used in this work as will be shown below. Moreover, Bayesian statistics allows sampling the PDF of the system under consideration in a natural way. Strictly speaking, one does not look for a unique estimator in this framework. Nevertheless, a summary of the PDF can be given by the mean of the sample (see section 2.6).

The most general approach to determine an estimator, however, should be based on the global (joint) PDF over all relevant quantities, like the signal \mathbf{s} and all model parameters \mathbf{p} , without neglecting any possible dependences. Let us assume that $P(\mathbf{s}, \mathbf{p} | \mathbf{d})$, the joint PDF of the system under consideration, depends on the signal \mathbf{s} and a series of additional parameters \mathbf{p} , given the observations \mathbf{d} . One solution would then be to calculate the expectation of

the signal over the joint PDF space

$$E_{\text{joint}}(\mathbf{s}) \equiv \int d\mathbf{s} d\mathbf{p} [P(\mathbf{s}, \mathbf{p} | \mathbf{d}) \mathbf{s}] \equiv \langle \mathbf{s} \rangle_{(\mathbf{s}, \mathbf{p} | \mathbf{d})}, \quad (4)$$

where we have introduced the ensemble average $\langle \cdot \rangle_{(\mathbf{s}, \mathbf{p} | \mathbf{d})}$ with the subscript representing the PDF over which the integral is done $P(\mathbf{s}, \mathbf{p} | \mathbf{d}) \rightarrow (\mathbf{s}, \mathbf{p} | \mathbf{d})^3$. Expression (4) can consequently be read as the ensemble average over all possible signals and parameters. The joint PDF is unfortunately quite hard to calculate directly, and the integral in eq. (4) is computationally too expensive for realistic cases as it involves many parameters and a large amount of data. To disentangle the uncertainties in parameter and signal spaces, let us apply the product rule of statistics⁴ to eq. (4)

$$\begin{aligned} E_{\text{joint}}(\mathbf{s}) &= \int d\mathbf{p} P(\mathbf{p} | \mathbf{d}) \left[\int d\mathbf{s} [P(\mathbf{s} | \mathbf{p}, \mathbf{d}) \mathbf{s}] \right] \\ &= E_{\mathbf{p}} [E_{\mathbf{s}}(\mathbf{s} | \mathbf{p}, \mathbf{d}) | \mathbf{d}] = \langle \langle \mathbf{s} \rangle_{(\mathbf{s} | \mathbf{p}, \mathbf{d})} \rangle_{(\mathbf{p} | \mathbf{d})}. \end{aligned} \quad (5)$$

This means that the expectation of the signal \mathbf{s} corresponds to the average of the conditional mean of \mathbf{s} over the marginal distribution of \mathbf{p} (see for example Gelman et al. 2004), where the conditional mean is given by

$$E_{\text{cond}}(\mathbf{s}) = E_{\mathbf{s}}(\mathbf{s} | \mathbf{p}, \mathbf{d}) = \int d\mathbf{s} [P(\mathbf{s} | \mathbf{p}, \mathbf{d}) \mathbf{s}] = \langle \mathbf{s} \rangle_{(\mathbf{s} | \mathbf{p}, \mathbf{d})}. \quad (6)$$

Traditionally, the conditional PDF has been used to determine the estimator of the signal assuming that all the parameters are known (e.g. Zaroubi et al. 1995).

As the reconstruction step of the density field is computationally expensive, a joint estimation of the parameters is out of scope. Therefore, the reduced approach of basing the estimators on conditional PDFs provides a computationally more feasible way to tackle problems of this kind. In particular, we will demonstrate that an operator formalism allows efficient sampling of the conditional PDFs, enabling us to sample the joint PDF in a Bayesian framework.

2.3 Bayesian approach

Given a data model, one can usually find an expression for the sampling distribution, i.e. the probability of obtaining the data given the signal and some additional parameters \mathbf{p} , $P(\mathbf{d} | \mathbf{s}, \mathbf{p})$. This is much less difficult than a direct calculation of the posterior $P(\mathbf{s} | \mathbf{d}, \mathbf{p})$. We need an expression which relates both the sampling and the posterior distribution given by Bayes theorem. The derivation of Bayes theorem is straightforward from the joint PDF of the signal and the data, using the product rule and the fact that the joint PDF is invariant under permutations of its arguments⁵. Bayes theorem can be expressed by the following equation

$$P(\mathbf{s} | \mathbf{d}, \mathbf{p}, I) = \frac{P(\mathbf{d} | \mathbf{s}, \mathbf{p}, I)P(\mathbf{s} | \mathbf{p}, I)}{P(\mathbf{d} | \mathbf{p}, I)}, \quad (7)$$

³ Sometimes, however, the ensemble angles will denote the estimator of some signal or parameter in a more general sense, like the maximum likelihood or the maximum a posteriori (see sections 2.4 and 2.5, respectively). Note that a bracket formalism could be introduced at this point, in which eq. (4) would be represented in the following way: $\langle \mathbf{s} | \mathbf{s} | \mathbf{p}, \mathbf{d} \rangle$.

⁴ $P(\mathbf{s}, \mathbf{p} | \mathbf{d}) = P(\mathbf{s} | \mathbf{p}, \mathbf{d})P(\mathbf{p} | \mathbf{d})$

⁵

$$\begin{aligned} P(\mathbf{s}, \mathbf{d}, \mathbf{p}, I) &= P(\mathbf{s} | \mathbf{d}, \mathbf{p}, I)P(\mathbf{d} | \mathbf{p}, I) = \\ P(\mathbf{d}, \mathbf{s}, \mathbf{p}, I) &= P(\mathbf{d} | \mathbf{s}, \mathbf{p}, I)P(\mathbf{s} | \mathbf{p}, I) \end{aligned}$$

where $P(s | \mathbf{p}, I)$ represents the prior knowledge about the signal, as it models the signal before any observations occur. The PDF given by $P(\mathbf{d} | \mathbf{p}, I)$ stands for the so-called evidence that is treated as the normalization of the posterior

$$P(\mathbf{d} | \mathbf{p}, I) = \int d\mathbf{s} P(\mathbf{d} | \mathbf{s}, \mathbf{p}, I) P(\mathbf{s} | \mathbf{p}, I). \quad (8)$$

It is worth mentioning that all the probabilities are conditional to the underlying physical picture, or prior information I . This has to be explicitly considered in case of model comparisons. In the following sections, we will present the steps for completing a Bayesian analysis, starting with the likelihood, then discussing the importance of the prior, and finishing with sampling through the joint signal and parameter space. Note that different choices for these three components (likelihood, prior, and sampling) lead to different classes of reconstruction algorithms. An overview of the different reconstruction scheme implementations based on this classification can be found in table (1).

2.4 The likelihood

The likelihood function is formally any function of the parameters θ proportional to the sample density

$$\mathcal{L}(\theta | \mathbf{d}) \propto P(\mathbf{d} | \theta). \quad (9)$$

Many inference approaches are based on the likelihood function, justified by the likelihood principle, which states that the information obtained by an observation \mathbf{d} about θ is entirely contained in the likelihood function $\mathcal{L}(\theta | \mathbf{d})$. To be specific, if d_1 and d_2 are two observations depending on the same parameter θ such that there exists a constant c satisfying $\mathcal{L}_1(\theta | d_1) = c\mathcal{L}_2(\theta | d_2)$ for every θ , d_1 and d_2 then bring the same information about θ and must hence lead to identical inferences.

Maximum likelihood (ML) methods, for example, rely on the likelihood principle with an estimator of the parameters given by

$$(\theta)_{\text{ML}} = \arg \sup_{\theta} \mathcal{L}(\theta | \mathbf{d}), \quad (10)$$

i.e., the value of θ that maximizes the probability density at \mathbf{d} . Bayesian methods take also advantage of the likelihood principle incorporating the decision-related requirement of the inferential problem through the definition of a prior distribution (see section 2.5). The definition of the likelihood is the first step in a Bayesian framework to determine the posterior distribution (see eq. 7). In using galaxy redshift surveys to trace the matter distribution, we have to deal with the discrete nature of the data sample. Thus the likelihood may be derived here for Poissonian statistics.

2.4.1 Poissonian likelihood

The likelihood of our galaxy distribution may be approximately represented by a Poissonian distribution (the real statistics should describe the much more complex galaxy formation process). Under the assumption of independent and identically distributed (*iid*) observations, this yields

$$\begin{aligned} \mathcal{L}(\mathbf{s} | \mathbf{d}, \mathbf{p}) &\propto \\ P(\mathbf{d} | \mathbf{s}, \mathbf{p}) &= \prod_{i=1}^m \exp(-[(\mathbf{R}\mathbf{s}')_i + c_i]) \frac{[(\mathbf{R}\mathbf{s}')_i + c_i]^{(d'_i + c_i)}}{(d'_i + c_i)!}, \end{aligned} \quad (11)$$

where d'_i are the galaxy counts per cell i and the real, positive signal of the expectation value of the number of galaxies is given by $s'_i = \bar{n}_g(1 + bs_i)$, with $s_i = \delta_{\rho i} = \frac{\rho_i - \bar{\rho}}{\bar{\rho}}$ the DM overdensity,

our target signal. The quantity n_g stands for the mean number of galaxies, $\bar{\rho}$ represents the mean density and b the bias factor. All these quantities are redshift-dependent. The additional parameters \mathbf{p} in this case would be represented by some background c_i and would enter into the operator \mathbf{R} that modifies the signal \mathbf{s} .

For a similar application in astronomy see Lahav & Gull (1989) and Robinson (1991). If d'_i is not converted to an integer, a Gamma function may be used instead of the factorial, $(d'_i + c_i)! \rightarrow \Gamma(d'_i + c_i + 1)$.

2.4.2 Gaussian likelihood

When the number of counts is large the Poisson distribution can be approximated by the normal distribution. In that case, the likelihood can be given by a Gaussian distributed noise

$$\begin{aligned} \mathcal{L}(\mathbf{s} | \mathbf{d}, \mathbf{p}) &\propto \\ P(\mathbf{d} | \mathbf{s}, \mathbf{p}) &= \frac{1}{[(2\pi)^m \det(\mathbf{N})]^{1/2}} \exp\left(-\frac{1}{2} \boldsymbol{\epsilon}^\dagger \mathbf{N}^{-1} \boldsymbol{\epsilon}\right) \\ &\propto \exp\left[-\frac{1}{2} \chi^2(\mathbf{s})\right], \end{aligned} \quad (12)$$

where $\mathbf{N} = \langle \boldsymbol{\epsilon} \boldsymbol{\epsilon}^\dagger \rangle_{(\boldsymbol{\epsilon} | \mathbf{p})}$ is the covariance matrix of the noise $\boldsymbol{\epsilon} \equiv \mathbf{d} - \mathbf{R}\mathbf{s}$, and

$$\chi^2(\mathbf{s}) = (\mathbf{d} - \mathbf{R}\mathbf{s})^\dagger \mathbf{N}^{-1} (\mathbf{d} - \mathbf{R}\mathbf{s}). \quad (13)$$

The parameters \mathbf{p} determine the structure of the noise $\boldsymbol{\epsilon}$ (and therefore the structure of the covariance matrix \mathbf{N}), and also enter into the operator \mathbf{R} . We give different expressions for the noise covariance matrix \mathbf{N} in section (3.3).

Note that χ^2 coincides with the square of the Mahalanobis distance⁶ between \mathbf{d} and $\mathbf{R}\mathbf{s}$, and also coincides with the squared \mathbf{N}^{-1} -norm of the error

$$\chi^2(\mathbf{s}) = D_{\text{Mah}}^2(\mathbf{d}, \mathbf{R}\mathbf{s})_{\mathbf{N}^{-1}} = \|\boldsymbol{\epsilon}\|_{\mathbf{N}^{-1}}^2. \quad (14)$$

In this case, the ML will correspond to the least squares of the error. It will minimize the $\chi^2(\mathbf{s})$ and hence minimize the Mahalanobis distance between the data and the noise-free data model. Therefore, the ML is equivalent to searching the estimator that fits the data better without constraining the model for the signal. Let us study the prior that precisely sets constraints on the signal \mathbf{s} .

2.5 The prior

A second step in Bayesian analysis is to specify the prior distribution for the signal, which contains the prior knowledge about the signal before the measurements were carried out. For little informative data it can strongly affect the posterior distribution and thus modify any inference based on it. For this reason, frequentists criticize Bayesian methods as being subjective. Other definitions of probability, like the frequentist, however, can be shown in most of the situations to be particular cases of the Bayesian approach (see e.g. Tanner 1996), implying the use of an implicit prior. The advantage of defining the prior knowledge about the system under consideration is that the interpretation of the results is straightforward, especially because assumptions flowing into the inference procedure are clearly stated. Once the prior is defined, we can obtain the

⁶ We introduce here a generalized definition of the Mahalanobis distance as: $D_{\text{Mah}}^2(\mathbf{x}, \mathbf{y})_{\mathbf{M}} = (\mathbf{x} - \mathbf{y})^\dagger \mathbf{M} (\mathbf{x} - \mathbf{y})$, with \mathbf{x} and \mathbf{y} being two vectors in the N -dimensional space and \mathbf{M} a $N \times N$ matrix.

Likelihood	Prior	Non-informative priors		Informative priors (MAP)	
		Flat (ML)	Entropic (MEM)	Gaussian	Poissonian
Gaussian				WIENER (Tikhonov, Ridge)	
–Radio			Sutton & Wandelt (2006) [#]		
–CMB	COBE: Janssen & Gulkis (1992) Tegmark (1997,1997b) ROMA: Natoli et al. (2001) MAPCUMBA: Doré et al. (2001) MAXIMA: Stompor et al. (2002) MAGIC [#] : Wandelt et al. (2004) MIRAGE: Yvon & Mayet (2005) MADAM: Keihänen et al. (2005)		Maisinger et al. (1997) Hobson et al. (1998)	Bunn & Sugiyama (1995) Tegmark (1997,1997b)	Hobson & McLachlan (2003) [#]
–LSS				MAGIC [#] : Wandelt et al. (2004) Eriksen et al. (2004) [#] O’Dwyer et al. (2005) [#] Larson et al. (2007) [#] Fisher et al. (1994) Hoffman (1994) Lahav et al. (1994), Lahav (1994) Zaroubi et al. (1995) Fisher et al. (1995) Webster et al. (1997) Zaroubi et al. (1999) Schmoldt et al. (1999) Erdoğdu et al. (2004,2006) ARGO: WIENER**[#]	
			ARGO: MEMG* (section 2.5.9 and appendix E)		(sections 2.5.3, 2.6, 4 and appendix I)
Poissonian		Richardson (1972) Lucy (1974)	ARGO: MEMP* (section 2.5.9 and appendix E)	ARGO: GAPMAP* (section 2.5.4 and appendix A)	
Inverse Gamma					
–CMB		MAGIC [#] : Wandelt et al. (2004) Eriksen et al. (2004) [#] O’Dwyer et al. (2005) [#] Larson et al. (2007) [#]			
–LSS		ARGO*[#] (section 2.6.2)			
Modified Gaussian					
–CMB		Pierpaoli & Anthoine (2005) [#]			
–LSS		Percival (2005) [#]			

*developed and presented in this paper; **developed, tested and presented in this paper; [#]able to sample PDFs

We have left out the reconstruction methods that are focused on the cosmological initial conditions, since they address a different problem and, in general, cannot be classified in terms of the PDFs listed in this table. Neither can other reconstruction algorithms based on geometrical arguments, like Voronoi, Delaunay tessellations, *friends-of-friends* schemes or *cloud-in-cell* interpolation schemes, be classified here.

Table 1. Classification of reconstruction methods in astrophysics based on the prior (columns) and likelihood (rows). Note that most of the reconstruction algorithms in other research areas, such as tomography, where Tikhonov-regularization is widely used, or the algebraic reconstruction technique (ART), which is based on the asymptotic regularization, fall into the class of Wiener-filtering schemes as we show in section (2.5.1) and appendix F. The differences in the ML CMB-map-making algorithms reside mainly in the modeling of the complex noise structure that arises due to the scanning strategies of the satellites and in the various foreground removal methods. The LSS Wiener-filtering methods on the other hand present improvements in the redshift distortions treatment, or are based on the different input data, either galaxy-positions or peculiar velocities. The discrete object detection (Hobson & McLachlan 2003) algorithm was developed to find Snyayev-Zeldovich clusters. This is also the case for the modified Gaussian by Pierpaoli et al. (2005). The reconstruction of the power-spectrum is also listed here. In CMB the joint map and power-spectrum estimation is done by MAGIC. Percival (2005) samples the power-spectrum with a modified Gaussian likelihood given by a log-normal mean. We propose to follow the steps done in CMB and sample the density field and the power-spectrum jointly (see section 2.6.2). This paper covers three new areas in LSS (GAPMAP, MEMG, MEMP) and presents four novel algorithms with which reconstructions can be done very fast.

maximum a posteriori (MAP) estimator, by maximizing the posterior distribution, which is proportional to the likelihood multiplied by the prior,

$$\langle \boldsymbol{\theta} \rangle_{\text{MAP}} = \arg \sup_{\boldsymbol{\theta}} P(\boldsymbol{\theta} | \mathbf{d}). \quad (15)$$

Note that there is a crucial difference to the maximum likelihood estimator (eq. 10) due to the incorporation of the prior information.

2.5.1 Bayes and regularization methods: the prior as a regularizer

Looking at the log-probabilities, we see that the MAP estimator maximizes the following quantity using Bayes theorem ($\log P(\boldsymbol{\theta} | \mathbf{d}) \propto \log(P(\mathbf{d} | \boldsymbol{\theta})P(\boldsymbol{\theta}))$)

$$Q = \log P(\mathbf{d} | \boldsymbol{\theta}) + \log P(\boldsymbol{\theta}). \quad (16)$$

If we assume that the error is Gaussian distributed, (which is a fair assumption if there is no prior information about the noise), and we parametrize the prior of the parameter, say the signal \mathbf{s} , we can rewrite eq. (16) as ($2Q \rightarrow Q$)

$$Q = -\chi^2(\mathbf{s}) + \alpha f_{\text{p}}(\mathbf{s}), \quad (17)$$

where we absorbed the factor 2 in the Lagrangian multiplier α , and f_{p} represents the penalty function that obliges the estimator to fulfil some constraint on the parameter \mathbf{s} , to the detriment of the $\chi^2(\mathbf{s})$ that strongly relies on the data. If we further assume that $\mathbf{N}^{-1} = \mathbf{I}$ (say we have white noise), the Mahalanobis distance reduces to the Euclidean distance ($D_{\text{Mah}}^2(\mathbf{d}, \mathbf{R}\mathbf{s})|_{\mathbf{N}^{-1}=\mathbf{I}} = D_{\text{Euc}}^2(\mathbf{d}, \mathbf{R}\mathbf{s})$), and the quantity one wants to minimize reads

$$\|\boldsymbol{\epsilon}\|^2 + \alpha f_{\text{p}}(\mathbf{s}), \quad (18)$$

where we have absorbed the minus sign in α . Expression (18) is equivalent to least squares with a regularization term, and belongs to Ridge-regression problems (Hoerl 1962; Hoerl & Kennard 1970). Assuming that the penalty function takes the following form $f_{\text{p}}(\mathbf{s}) = \|\mathbf{s}\|^2$, we can write expression (18) as

$$\|\boldsymbol{\epsilon}\|^2 + \alpha \|\mathbf{s}\|^2, \quad (19)$$

which then becomes the Tikhonov regularization method (Tikhonov 1963). The parameter α is called the regularization parameter. These methods lead to linear filters and are essentially identical to Wiener-filtering (Foster 1961), which will be presented in the next section. Note that Tikhonov regularization is equivalent to MAP of a Gaussian likelihood with noise covariance matrix $\mathbf{N} = \mathbf{I}$ and Gaussian prior, with signal covariance matrix $\mathbf{S} = \alpha^{-1}\mathbf{I}$. Nevertheless, the penalty function f_{p} in general can be a non-linear function of the parameter to be estimated (say the signal \mathbf{s}) leading to non-linear estimators. We will introduce MEM as such an example. Tikhonov regularization can also be generalized to non-linear problems by introducing a non-linear kernel operator $\mathbf{R}(\mathbf{s})$.

Summarizing the exposed theory of signal reconstruction, we might interpret the likelihood as some distance measure between the data and the noise-free model of the data, and the prior as some constraint that tightens the estimator to the model of the signal. We have shown here that the classical methods of signal reconstruction, like the Tikhonov regularization, are particular cases of the Bayesian approach. The inclusion of a prior can be regarded as a *natural* regularization, in the sense that the regularization term is provided by a (physical) model of the *true* signal. In appendix F, we discuss the relation between other regularization methods and

the Bayesian approach. In the following subsections we introduce different priors that are relevant for large-scale structure reconstruction and are implemented in ARGO.

2.5.2 Gaussian prior

The distribution of the primordial density field should be very close to Gaussianity according to most of the inflationary scenarios (Guth 1981; Linde 1982; Albrecht & Steinhardt 1982). In fact, the measurements of the CMB show very small deviations from Gaussianity (see e.g. Komatsu et al. 2003). Non-Gaussianities in the matter distribution arose mainly from non-linear gravitational collapse. The non-linear regime of structure formation is responsible for the strong radial redshift-distortions, the *finger-of-god* effect, limiting the accuracy of reconstructions. Previous attempts to correct for these distortions have modified the power-spectrum by introducing a lorentzian factor (see e.g. Ballinger et al. 1996; Erdođdu et al. 2004). In section (2.6) we propose an alternative way to do this in a Bayesian framework, where peculiar velocities are sampled together with the three dimensional map of the matter distribution. For the underlying DM density fluctuation we will assume a Gaussian prior. This is a crude approximation for the density field at the present epoch of the Universe, especially on small-scales. It is, however, a valid description on large-scales and allows to incorporate non-linear corrections in a MCMC fashion, as will be discussed in section (2.6). Following Bardeen et al. (1986) we may thus write the PDF of the signal as a multivariate Gaussian distribution

$$P(\mathbf{s} | \mathbf{p}) = \frac{1}{[(2\pi)^n \det(\mathbf{S})]^{1/2}} \exp\left(-\frac{1}{2}\mathbf{s}^\dagger \mathbf{S}^{-1} \mathbf{s}\right), \quad (20)$$

with \mathbf{S} being the covariance matrix of the signal ($\mathbf{S} = \mathbf{S}(\mathbf{p}) = \langle \mathbf{s}\mathbf{s}^\dagger \rangle_{(\mathbf{s}|\mathbf{p})}$). This formula emphasizes the high dimensional character of the problem (n dimensions of the signal reconstruction, with n being typically between 10^3 and 10^9).

2.5.3 Gaussian prior and Gaussian likelihood: the Wiener-filter

The Gaussian prior together with the Gaussian likelihood lead to the Wiener Filter, completing the square for the signal in the exponent of the posterior distribution, as derived by Zaroubi et al. (1995),

$$\begin{aligned} P(\mathbf{s} | \mathbf{d}, \mathbf{p}) & \propto \exp\left(-\frac{1}{2}\left[\mathbf{s}^\dagger \mathbf{S}^{-1} \mathbf{s} + (\mathbf{d} - \mathbf{R}\mathbf{s})^\dagger \mathbf{N}^{-1}(\mathbf{d} - \mathbf{R}\mathbf{s})\right]\right) \\ & \propto \exp\left(-\frac{1}{2}\left[(\mathbf{s} - \langle \mathbf{s} \rangle_{\text{WF}})^\dagger (\boldsymbol{\sigma}_{\text{WF}}^2)^{-1}(\mathbf{s} - \langle \mathbf{s} \rangle_{\text{WF}})\right]\right) \end{aligned} \quad (21)$$

where the Wiener-filter used to calculate the estimator from the data $\langle \mathbf{s} \rangle_{\text{WF}} = \mathbf{F}_{\text{WF}} \mathbf{d}$ is given by

$$\mathbf{F}_{\text{WF}} = \mathbf{S}\mathbf{R}^\dagger (\mathbf{R}\mathbf{S}\mathbf{R}^\dagger + \mathbf{N})^{-1}, \quad (22)$$

and the corresponding covariance is

$$\boldsymbol{\sigma}_{\text{WF}}^2 = \langle \mathbf{r}\mathbf{r}^\dagger \rangle_{\text{WF}} = (\mathbf{S}^{-1} + \mathbf{R}^\dagger \mathbf{N}^{-1} \mathbf{R})^{-1}, \quad (23)$$

with $\mathbf{r} = \mathbf{s} - \langle \mathbf{s} \rangle_{\text{WF}}$ being the residual. The following notation can be introduced for the posterior PDF: $P(\mathbf{s} | \mathbf{d}, \mathbf{p}) \propto G(\mathbf{s} - \langle \mathbf{s} \rangle_{\text{WF}}, \boldsymbol{\sigma}_{\text{WF}}^2)$, i.e. given a dataset \mathbf{d} derived from a Gaussian process, the possible signals are Gaussian distributed around the Wiener-filter reconstruction $\langle \mathbf{s} \rangle_{\text{WF}}$ with a covariance $\boldsymbol{\sigma}_{\text{WF}}$. The parameters \mathbf{p} enter the operator \mathbf{R} , including

also the cosmological parameters that determine the signal covariance matrix \mathbf{S} . We will discuss in section (2.6) how to sample \mathbf{S} and to determine cosmological parameters.

A remarkable characteristic of the Wiener-filter is that it suppresses the signal in the presence of a high noise level resulting in the null estimator and gives just the deblurred data when noise is negligible. In this sense it is a biased estimator, since its covariance matrix has less power than the original one. Some attempts have been made to derive an equivalent unbiased estimator (see Zaroubi 2002b). However, one might be especially interested in obtaining a conservative estimator. Sampling the joint PDF will fill the missing modes (see e.g. Wandelt et al. 2004) and in this way complete the signal in regions where it is undersampled or the signal to noise ratio is low. It is interesting to note that the Wiener Filter coincides with the MAP estimator in the case of a Gaussian prior on \mathbf{s} and a Gaussian likelihood ($\langle \mathbf{s} \rangle_{\text{WF}} = \langle \mathbf{s} \rangle_{\text{MAP}}$). This filter can also be obtained as the LSQ estimator ⁷ ($\langle \mathbf{s} \rangle_{\text{WF}} = \langle \mathbf{s} \rangle_{\text{LSQ}}$, for an explicit derivation see Zaroubi et al. 1995, and appendix I). Performing the integral of the conditional PDF (see eq. 6) one obtains the same estimator again, thus $\langle \mathbf{s} \rangle_{\text{WF}} = \langle \mathbf{s} \rangle_{(\mathbf{s}|\mathbf{d},\mathbf{p})}$. This is a very important result, since it permits one to sample the conditional PDF. We propose to exploit this property for the joint estimation of the signal and its power-spectrum as is done in the CMB (see Wandelt et al. (2004) and section 2.6.2).

2.5.4 Gaussian prior and Poissonian likelihood: the GAPMAP estimator

The Gaussian likelihood constitutes a valid approximation when the Poissonian character of the distribution is appropriately modeled in the noise correlation matrix \mathbf{N} . However, one would rather describe a discrete sampling process like a galaxy survey with a Poissonian likelihood. Unfortunately, there is no filter available for such a case. Thus, we present a novel iterative equation for the MAP estimator with a Gaussian prior and a Poissonian likelihood, which we call GAPMAP (see appendix A for a derivation)

$$\mathbf{s}^{j+1} = \mathbf{S}\mathbf{R}^\dagger b\overline{n_g} \left(-\vec{1} + \text{diag} \left(\mathbf{R}\overline{n_g}(\vec{1} + b\mathbf{s}^j) + \mathbf{c} \right)^{-1} (\mathbf{d}' + \mathbf{c}) \right). \quad (24)$$

This scheme will be presented in further detail in a separate paper.

2.5.5 Flat prior

With the aim of deriving objective posterior distributions, non-informative prior distributions are introduced. A non-informative prior would suggest that any value is reasonable. Flat priors where the probability distribution is assumed to be constant $\mathbf{P}(\mathbf{s}) = \text{constant}$ are thus very often applied. Note, however, that these are improper priors, since the integral of these distributions diverges to infinity. In this case, the posterior is proportional to the likelihood. The maximum likelihood solution coincides in this way with the MAP estimator assuming a flat prior ($\langle \mathbf{s} \rangle_{\text{ML}} = \langle \mathbf{s} \rangle_{\text{MAP}|\text{flat}}$).

⁷ Note that in this case, the least squares are referred to the residuals \mathbf{r} , i.e. the difference between the real signal \mathbf{s} and the estimated signal $\langle \mathbf{s} \rangle_{\text{LSQ}}$: $\|\mathbf{r}\|^2 = \|\mathbf{s} - \langle \mathbf{s} \rangle_{\text{LSQ}}\|^2$, where the prior on \mathbf{s} is given in a more implicit way by assuming a linear relation between the estimator and the data and statistical homogeneity.

2.5.6 Flat prior and Gaussian likelihood: the COBE-filter

In CMB map-making algorithms it is common to use the so-called COBE-filter (see Janssen & Gulkis 1992; Tegmark 1997), which can easily be derived by maximizing the likelihood given in eq. (12)

$$\mathbf{F}_{\text{COBE}} = (\mathbf{R}^\dagger \mathbf{N}^{-1} \mathbf{R})^{-1} \mathbf{R}^\dagger \mathbf{N}^{-1}. \quad (25)$$

This filter has the property that among all unbiased linear estimators (with a noise of zero mean), it leads to the minimum variance (Natoli et al. 2001). Here unbiased means that the statistical mean of the estimator is equal to the *true* signal. This is, however, only fulfilled when the inverse of $\mathbf{R}^\dagger \mathbf{N}^{-1} \mathbf{R}$ exists ⁸. The covariance for the COBE-filter can found to be

$$\sigma_{\text{COBE}}^2 = \langle \mathbf{r}\mathbf{r}^\dagger \rangle_{\text{COBE}} = (\mathbf{R}^\dagger \mathbf{N}^{-1} \mathbf{R})^{-1}. \quad (26)$$

Note that, in general, the following relation holds: $\sigma_{\text{WF}}^2 \leq \sigma_{\text{COBE}}^2$, as a comparison to eq. (23) shows.

Tegmark (1997) claims that several linear filters like the COBE or the Wiener-filter conserve information by comparing the Fisher information matrix corresponding to the filtered signal to the one of the unfiltered time ordered data. This property apparently permits one to perform cosmological parameter estimation from the reconstructed signal after filtering the data. However, linear filters conserve information only if they are invertible, which is not provided for realistic cases as we show in appendix C. A consistent estimation of cosmological parameters has to be done in a full Bayesian framework by estimating the joint PDF of the signal and the parameters, as we will see in section (2.6) (Wandelt et al. 2004).

2.5.7 Flat prior and Poissonian likelihood: the Richardson-Lucy algorithm

A widely used deblurring algorithm in astronomy and medical tomography is the Richardson-Lucy algorithm (Richardson 1972; Lucy 1974), which was shown to be the maximum likelihood solution with a Poissonian likelihood by Shepp & Vardi (1982). We show the derivation in appendix B, as a simplified case with respect to eq. (24). The Richardson-Lucy algorithm cannot prevent serious noise amplifications in the restoration process (see e.g. Carasso 1999). This is a natural consequence when a prior that regularizes the solution is missing. We will show in a forthcoming paper that the Krylov regularization methods we present in this paper (section 3.1.5) are very well suited for deconvolution purposes, as they give faster reconstructions than the Richardson-Lucy algorithm and are stable against noise. A toy application is presented in fig. (8).

2.5.8 Jeffrey's prior

Other non-informative priors have been suggested based on invariant statistical structures under transformation of variables in a Bayesian formalism. Considering a one-to-one transformation in

⁸

$$\begin{aligned} \langle \langle \mathbf{s} \rangle_{\text{COBE}} \rangle_{(\mathbf{d}|\mathbf{s},\mathbf{p})} &= \langle (\mathbf{R}^\dagger \mathbf{N}^{-1} \mathbf{R})^{-1} \mathbf{R}^\dagger \mathbf{N}^{-1} \mathbf{d} \rangle_{(\mathbf{d}|\mathbf{s},\mathbf{p})} \\ &= (\mathbf{R}^\dagger \mathbf{N}^{-1} \mathbf{R})^{-1} \mathbf{R}^\dagger \mathbf{N}^{-1} \langle \mathbf{R}\mathbf{s} + \boldsymbol{\epsilon} \rangle_{(\mathbf{d}|\mathbf{s},\mathbf{p})} \\ &= (\mathbf{R}^\dagger \mathbf{N}^{-1} \mathbf{R})^{-1} \mathbf{R}^\dagger \mathbf{N}^{-1} \mathbf{R}\mathbf{s} \\ &= \mathbf{s}, \text{ if } \mathbf{R}^\dagger \mathbf{N}^{-1} \mathbf{R} \text{ is invertible.} \end{aligned}$$

the one-dimensional case of the parameter: $\phi = f(\theta)$, the equivalence between the respective prior densities is expressed by

$$P(\phi) = P(\theta) \left| \frac{d\theta}{d\phi} \right| = P(\theta) |f'(\theta)|^{-1}. \quad (27)$$

This relation is satisfied by Jeffrey's prior $P(\theta) \propto [J(\theta)]^{1/2}$, where $J(\theta)$ is the Fisher information⁹

$$J(\theta) \equiv \left\langle \left(\frac{\partial \log P(d|\theta)}{\partial \theta} \right)^2 \right\rangle_{(d|\theta)} = - \left\langle \frac{\partial^2 \log P(d|\theta)}{\partial \theta^2} \right\rangle_{(d|\theta)}, \quad (28)$$

and where we have assumed the following regularity condition $\int dd \frac{\partial^2}{\partial \theta^2} P(d|\theta) = 0$. Relation (27) can be proved easily by doing the evaluation $J(\phi) = - \left\langle \frac{\partial^2 \log P(d|\phi)}{\partial \phi^2} \right\rangle_{(d|\phi)} = J(\theta) \left| \frac{d\theta}{d\phi} \right|^2$ (see e.g. Gelman et al. 2004). Note, however, that in the multidimensional case, Jeffrey's prior may lead to incoherences or even paradoxes (see e.g. Berger & Bernardo 1992; Robert 2001). Jeffrey's prior is applied adequately, when not even the order of magnitude of the parameter to be estimated is known a priori. We derive Jeffrey's ignorance prior for the 3-D power-spectrum ($\mathbf{S} = \text{diag}(P_S(\mathbf{k}))$)¹⁰ in appendix D (see section 2.6.2 for an application of this prior).

2.5.9 Entropic prior and Maximum Entropy method

Another approach searches the least informative model compatible with the data using a prior based on Boltzmann's definition of entropy S^E ¹¹ (or equivalently, Shannon's notion of information, see Shannon 1948),

$$P(\mathbf{s} | \mathbf{p}) = \exp(\alpha S^E), \quad (29)$$

and maximizing the resulting posterior distribution, being α some constant, and \mathbf{s} the so-called hidden image (or signal). This inference procedure is called the Maximum Entropy method (MEM) (Jaynes 1963, 1968; Frieden 1972; Gull & Daniell 1978; Gull 1989; Skilling 1989; Maisinger et al. 1997; Hobson et al. 1998). For a review see Narayan & Nityananda (1986). From now on we will represent the underlying signal by \mathbf{s} in the framework of MEM. The MEM can be considered as MAP estimation with an entropic prior.

The particular expression for the entropy depends on the statistical formulation of the non-informative prior. Let us think of a positive signal as a grid with q cells, with each cell i having a certain intensity value s_i , $i = 1, \dots, q$, with an uncertainty on each value given by $\pm \alpha^{-1}$. Then we define some discrete *quanta* n_i on each cell related to the intensity through the uncertainty: $n_i = \alpha s_i$. The signal can be guessed by distributing the n_i *quantas* in the grid. In this way, the image is modeled in this way analogously to the energy configuration space of a thermodynamical system. If we further demand each cell to be *iid*, the number of ways this object can occur is given by the multiplicity

$$W = \frac{N_q!}{n_1! n_2! \dots n_q!}, \quad (30)$$

⁹ The generalization to the multidimensional case leads to the following matrix form: $\mathbf{J}_{ij}(\boldsymbol{\theta}) \equiv \left\langle \frac{\partial \log P(\mathbf{d}|\boldsymbol{\theta})}{\partial \theta_i} \frac{\partial \log P(\mathbf{d}|\boldsymbol{\theta})}{\partial \theta_j} \right\rangle_{(\mathbf{d}|\boldsymbol{\theta})}$ (see appendix C).

¹⁰ Here the autocorrelation matrix \mathbf{S} is represented in k-space. We will discuss this in further detail in section (3.3).

¹¹ Not to be confused by the signal autocorrelation \mathbf{S} .

with N_q being the total amount of *quantas* to be distributed in all cells ($N_q = \sum_i n_i$). The probability of any particular result is then given by the multinomial distribution

$$P(\mathbf{s}' | \mathbf{p}) = W q^{-N_q}. \quad (31)$$

Sutton & Wandelt (2006) propose to sample from the multiplicity function directly to perform reconstructions in radioastronomy. By using Stirling's formula for the factorials ($n! \sim n^n e^{-n}$) we can write

$$\log P(\mathbf{s}' | \mathbf{p}) = -\alpha \sum_i s'_i \log s'_i + \text{const.} \quad (32)$$

Comparing this expression with eq. (29), we recover Shannon's definition of entropy ($S_+^E = \sum_i s'_i \log s'_i$)¹². The expression that is commonly used for the entropy is a generalization of Shannon's formula by Skilling that can be derived based only on consistency arguments within probabilistic information theory for positive and additive distributions (PADs) (Skilling 1989).

This generalization implies the definition of a Lebesgue measure (\mathbf{m}) for the integral of some function of the hidden image to represent the entropy

$$S_+^E(\mathbf{s}' | \mathbf{p}) = \sum_i \left[s'_i - m_i - s'_i \log (s'_i/m_i) \right], \quad (33)$$

here in its discretised form. Skilling's expression for the entropy can also be derived by considering a *team of monkeys* throwing balls at q cells at random with Poissonian expectation μ_i : $P(\mathbf{n}|\boldsymbol{\mu}) = \prod_i \mu_i^{n_i} e^{-\mu_i} / n_i!$, where $n_i = \alpha s_i$ and $\boldsymbol{\mu} = \alpha \mathbf{m}_i$ (Skilling 1989). For a review on further expressions for the entropy see Molina et al. (2001).

The global maximum of S^E over \mathbf{s} in the absence of further constraints is found to be $\mathbf{s}' = \mathbf{m}$. Consequently, \mathbf{m} can also be thought of as a prior model for the image. However, this expression for the entropy will allow reconstructing positive signals only. Zaroubi et al. (1995) propose to define $\mathbf{s}' = \boldsymbol{\rho}$ and $\mathbf{m} = \boldsymbol{\rho}_0$, to avoid the possibility of having a negative distribution for \mathbf{s} .

According to Gull & Skilling (1990) the MEM can be extended to reconstruct distributions, which can be either positive or negative, as in the case of density fluctuations. Such distributions can be described as the difference between two subsidiary positive distributions (PADs)

$$\mathbf{s} = \mathbf{u} - \mathbf{v}, \quad (34)$$

relative to a common model \mathbf{m} ¹³

$$S_{\pm}^E(\mathbf{u}, \mathbf{v} | \mathbf{p}) = \sum_i \left[u_i - 2m_i - u_i \log(u_i/m_i) \right] + \sum_i \left[v_i - 2m_i - v_i \log(v_i/m_i) \right]. \quad (35)$$

One can see from eq. (34) that $\partial S_{\pm}^E / \partial \mathbf{u} = -\partial S_{\pm}^E / \partial \mathbf{v}$, hence yielding

$$\mathbf{u}\mathbf{v} = \mathbf{m}^2. \quad (36)$$

From the relations given by eqs. (34) and (36), it is easy to derive

$$\mathbf{u} = \frac{1}{2}(\mathbf{w} + \mathbf{s}), \quad (37)$$

¹² The “+” symbol in S_+^E denotes that the definition is only valid for positive signals \mathbf{s}' .

¹³ The “±” symbol in S_{\pm}^E denotes that the definition is valid for positive and negative signals \mathbf{s} .

$$\mathbf{v} = \frac{1}{2}(\mathbf{w} - \mathbf{s}), \quad (38)$$

with $w_i = (s_i^2 + 4m_i^2)^{1/2}$. Using these expressions, the total entropy can be rewritten as

$$S_{\pm}^E(\mathbf{s} | \mathbf{p}) = \sum_i \left[w_i - 2m_i - s_i \log \left((w_i + s_i)/2m_i \right) \right]. \quad (39)$$

The Maximum Entropy method gives a non-linear estimator of the underlying signal that one wants to reconstruct. This method is especially interesting to study deviations from Gaussianity (Maisinger et al. 1997; Hobson et al. 1998). It is equivalent to maximize χ^2 with a Lagrangian multiplier, which includes a penalty function given by the entropy. Maximum Entropy in this context searches the hidden image that adds the least additional information to the data.

The quantity we need to maximize is given by

$$Q^E(\mathbf{s} | \mathbf{p}) = \alpha S^E(\mathbf{s} | \mathbf{p}) + \log \mathcal{L}(\mathbf{s} | \mathbf{d}, \mathbf{p}), \quad (40)$$

where the $\log \mathcal{L}$ is given by eq. (13) or eq. (A1). The equation we want to solve is

$$\nabla Q^E(\mathbf{s} | \mathbf{p}) = 0. \quad (41)$$

In section (3.2), different iterative algorithms to solve this non-linear problem will be discussed. The required expressions for the gradient of Q^E and its curvature for positive and positive/negative expressions of the entropy (eqs. 33 and 39) and for both Gaussian and Poissonian likelihoods are presented in appendix E.

Note that in the limit of low density fluctuations, i.e. in the linear regime, the expression of the entropy reduces to the quadratic entropy (eventually with an offset of the origin of \mathbf{s}), $S^E(\mathbf{s} | \mathbf{p}) \simeq -\sum_i s_i^2/2m_i$. This expression is very similar to a Gaussian prior for the signal with a variance given by m . In that case Maximum Entropy leads to the Wiener-filter.

2.6 Markov Chain Monte Carlo: sampling the joint PDF

The drawback of the maximization methods hitherto mentioned, is that they find a unique estimator that is most probably subject to the chosen values for the required parameters. As already mentioned, the complete characterization of a system is contained in the joint PDF in the product space of possible signals and parameters. Thus, it would be desirable to sample from this PDF to find the region of highest confidence for our estimator. This is possible using Markov Chain Monte Carlo (MCMC). The importance of sampling from the joint PDF and the viability of doing that with MCMCs has already been discussed in other contexts in astronomy (Hobson & McLachlan 2003; Jewell et al. 2004; Wandelt et al. 2004). With the MCMC method, the whole system can be moved in its configuration space by updating all variables successively in a Monte Carlo fashion, until the system relaxes (*burns-in*) and reaches the highest density region.

The expectation of the i -th parameter (θ_i) can be calculated by the so-called ergodic average, which is given by the mean of the sample

$$\langle \theta_i \rangle_{(\boldsymbol{\theta} | \mathbf{d})} \simeq \frac{1}{N_b} \sum_{t=0}^{N_b-1} \theta_i^t, \quad (42)$$

with N_b being the size of the sample drawn once the Markov Chain has *burned-in*. In general, the mean estimator is more reliable than the maximum of the distribution, especially in cases with deviations from Gaussianity (see e.g. Gelman et al. 2004). The MCMC

method permits one to approximately solve the integral in eq. (4) through expression (42).

2.6.1 Gibbs sampling

The most straightforward MCMC method is the Gibbs sampler (Geman & Geman 1984), also known as the *heatbath* algorithm. The Gibbs algorithm samples from the joint PDF by repeatedly replacing each component with a value drawn from its distribution conditional on the current values of all other components. This process can be seen as a Markov Chain with transition probabilities π_k for $k = 1, \dots, n$,

$$\pi_k(\boldsymbol{\theta}, \boldsymbol{\theta}') = P(\theta'_k | \{\theta_i : i \neq k\}) \cdot \prod_{i \neq k} \delta_K(\theta_i, \theta'_i), \quad (43)$$

where $\{\theta_i : i \neq k\} = (\theta_1, \dots, \theta_{k-1}, \theta_{k+1}, \dots, \theta_n)$ (see e.g. Neal 1993) and δ_K is the Kronecker delta-function. The Gibbs sampler starts with some initial values $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \dots, \theta_n^{(0)})$ and obtains new updates $\boldsymbol{\theta}^{(j)} = (\theta_1^{(j)}, \dots, \theta_n^{(j)})$ from the previous step $\boldsymbol{\theta}^{(j-1)}$ through successive generation of values

$$\begin{aligned} \theta_1^{(j)} &\sim P(\theta_1 | \{\theta_i^{(j-1)} : i \neq 1\}) \\ \theta_2^{(j)} &\sim P(\theta_2 | \theta_1^{(j)}, \{\theta_i^{(j-1)} : i > 2\}) \\ &\vdots \\ \theta_n^{(j)} &\sim P(\theta_n | \{\theta_i^{(j)} : i \neq n\}) \end{aligned} \quad (44)$$

In this way a random walk on the vector $\boldsymbol{\theta}$ is performed by making subsequent steps in low-dimensional subspaces, which span the full product space. This is similar to individual collisions of particles in a mechanical system that drives a many-body system to an equilibrium distribution for all degrees of freedom. We are especially interested in this sampling method because of its efficiency and simplicity in contrast to other algorithms, which include acceptance and rejection rules like the Metropolis-Hastings algorithm (Metropolis et al. 1953; Hastings 1970). See Wandelt et al. (2004) for applications in CMB-mapping and power-spectrum estimation.

The MCMC method can be applied to perform simultaneously the reconstruction of the density field and the estimation of other parameters, such as the power-spectrum, the peculiar velocities, the bias, or the cosmological parameters (see fig. 1). Here we outline the procedure to apply this method to power-spectrum estimation and redshift-distortion corrections, which can also be used in a joint algorithm. The applications of this method will be presented in forthcoming publications. Note that a higher degree of complexity can be achieved by going beyond linear perturbation theory or considering higher moments of the density field.

2.6.2 Joint signal and power-spectrum estimation

The joint PDF considered here is given by the joint PDF of the signal and the power-spectrum $P(\mathbf{s}, \mathbf{S} | \mathbf{d})$. For the initial guess either an expression for the power-spectrum can be applied (see e.g. Efstathiou et al. 1992; Peacock & Dodds 1994; Smith et al. 1998; Eisenstein & Hu 1999), or the power-spectrum of the CMB can be taken and calculated for the required redshifts with some transfer functions (see e.g. Eisenstein & Hu 1999). Then the following sampling processes are iterated until the chain *burns-in*

$$\mathbf{s}^{(j+1)} \sim P(\mathbf{s} | \mathbf{S}^{(j)}, \mathbf{d}), \quad (45)$$

$$\mathbf{S}^{(j+1)} \sim P(\mathbf{S} | \mathbf{s}^{(j+1)}), \quad (46)$$

where we sample the DM signal with the following PDF (see section 2.5.2)

$$P(\mathbf{s} | \mathbf{S}^{(j)}, \mathbf{d}) \propto G\left(\mathbf{s} - \mathbf{F}_{\text{WF}}(\mathbf{S}^{(j)})\mathbf{d}, \sigma_{\text{WF}}^2(\mathbf{S}^{(j)})\right). \quad (47)$$

By adding a vector with the correct covariance to the Wiener-filtered signal (see Wandelt et al. 2004) we can generate estimates of the signal with the required mean and covariance. Given Bayes theorem we can write

$$\mathbf{s}^{(j)} = \langle \mathbf{s}^{(j)} \rangle_{\text{WF}} + \mathbf{y}_{\sigma_{\text{WF}}}^{(j)}. \quad (48)$$

To generate $\mathbf{y}_{\sigma_{\text{WF}}}^{(j)}$ one has to solve the following set of equations, which are generalized here for windowed signals with a response operator \mathbf{R} (note that similar expressions can be found in Eriksen et al. 2004)

$$\mathbf{y}_{\sigma_{\text{WF}}}^{(j)} = \left((\mathbf{S}^{(j)})^{-1} + \mathbf{R}^\dagger \mathbf{N}^{-1} \mathbf{R} \right)^{-1} \left((\mathbf{S}^{(j)})^{-1/2} \mathbf{x}_{\text{G}_1} + \mathbf{R}^\dagger \mathbf{N}^{-1/2} \mathbf{x}_{\text{G}_2} \right), \quad (49)$$

where x_{G_1} and x_{G_2} are two independent Gaussian variates. One can show by direct calculation that $\mathbf{y}_{\sigma_{\text{WF}}}^{(j)}$ has a covariance given by σ_{WF}^2 .

The power-spectrum can be sampled by a inverse gamma function, which we derive here for the case of the 3-D power-spectrum (see Wandelt et al. 2004, for the analogous CMB case)

$$P(\mathbf{S} | \mathbf{s}) \propto P(\mathbf{S})P(\mathbf{s} | \mathbf{S}). \quad (50)$$

Assuming a Gaussian signal \mathbf{s} (see eq. 20) this yields

$$P(P_{\text{S}}(\mathbf{k}) | \mathbf{s}^{(j)}) \propto P(P_{\text{S}}(\mathbf{k})) \prod_{\mathbf{k}} \frac{1}{\sqrt{P_{\text{S}}(\mathbf{k})}} \exp -\frac{|\mathbf{s}^{(j)}(\mathbf{k})|^2}{2P_{\text{S}}(\mathbf{k})}, \quad (51)$$

with $\mathbf{S} = \text{diag}(P_{\text{S}}(\mathbf{k}))$. The prior $P(P_{\text{S}}(\mathbf{k}))$ can be chosen to be flat ($P(P_{\text{S}}(\mathbf{k})) = \text{const}$) or instead Jeffrey's prior can be used ($P(P_{\text{S}}(\mathbf{k})) = P_{\text{S}}(\mathbf{k})^{-1}$), see section (2.5.8) and appendix D. Note that the likelihood for the power-spectrum given by eq. (51) is clearly non-Gaussian.

2.6.3 Joint signal and peculiar velocities estimation: redshift-distortions correction

The measured redshift of a galaxy, or its so-called recession velocity can be expressed by Hubble's law, that describes the bulk flow of the Universe. However, the peculiar velocity of the galaxies along the line of sight introduces so-called redshift-distortions. This has to be considered in order to correct Hubble's law. Hence, a galaxy's redshift-distance cz (conveniently expressed in velocity units) is given by its true distance $r = H_0 d$ plus its peculiar velocity \mathbf{v} along the line of sight \mathbf{r}_u

$$cz = H_0 d + \mathbf{r}_u \cdot \mathbf{v}. \quad (52)$$

We propose to sample the peculiar velocities in a MCMC fashion, analogous to the case of the power-spectrum

$$\mathbf{s}^{(j+1)} \sim P(\mathbf{s} | \mathbf{v}^{(j)}, \mathbf{S}, \mathbf{d}), \quad (53)$$

$$\mathbf{v}^{(j+1)} \sim P(\mathbf{v} | \mathbf{s}^{(j+1)}), \quad (54)$$

In each step where we sample the peculiar velocity, the redshift-distortion can be corrected using eq. (52)

$$r^{(j+1)} = cz - \mathbf{r}_u \cdot \mathbf{v}^{(j+1)} \quad (55)$$

For low overdensities, linear theory gives good estimates of the peculiar velocity

$$\mathbf{v}_{\text{LT}} = -\beta \nabla \nabla^{-2} \delta_\rho, \quad (56)$$

where $\nabla \nabla^{-2}$ is the inverse Laplacian, and β is the linear growth rate at the present day (f_0) divided by the bias factor b : $\beta = f_0/b$. For high overdensities, the structures tend to be virialized. This means that the galaxies will have random velocities as in a Boltzmann gas, introducing dispersions in the line of sight, the so-called finger-of-god effect. We propose to sample the peculiar velocities from a PDF with a mean $\langle \mathbf{v} \rangle_{\text{M}}$ given by the linear theory \mathbf{v}_{LT} and a velocity dispersion σ_v depending on the local value of the overdensity,

$$P(\mathbf{v} | \mathbf{s}^{(j)}) \propto G\left(\mathbf{v} - \langle \mathbf{v} \rangle_{\text{M}}(\mathbf{s}^{(j)}), \sigma_v^2(\mathbf{s}^{(j)})\right). \quad (57)$$

Such a method would provided linear and non-linear corrections to the redshift-distortions. Note that with a similar approach, non-Gaussianities due to the Lagrangian nature of structure formation could also be addressed.

3 NUMERICAL METHOD

In order to efficiently sample the joint PDF, as it is required in MCMC methods (see section 2.6), fast inverse algorithms need to be considered to regularize the solution. General iterative inverse methods scale as $\mathcal{O}(n^3)$ since they imply matrix multiplications of a $n \times n$ matrix in an iterative fashion (at most n -steps until convergence). This makes the study of the joint PDFs as presented in section (2.6), at a first glance, unfeasible. However, a proper formulation of the problem in an operator formalism allows treating the matrices as operators that have to be neither calculated nor stored. Within this operator formalism, the inversion methods we present here sped up to a scaling of $\mathcal{O}(n \log_2 n)$. We start with a general formulation of iterative methods and subsequently present the different schemes that we have implemented in ARGO. Since a preconditioning treatment can dramatically enhance the performance of iterative schemes (see our numerical experiments in section 4), we pay special attention to this point in the derivation of the different schemes.

3.1 Iterative inverse and regularization methods: a unified formulation of different linear methods

Let us consider a region D in the n -dimensional Euclidean space E_n and denote $L_2(D)$ the Hilbert space of all complex measurable square integrable functions $\int_D d^n z |g|^2(z) < \infty$ with inner product¹⁴

$$\langle g | s \rangle = \int_D d^n z \overline{g(z)} s(z), \quad (58)$$

and norm of $g \in L_2(D)$

$$\|g\| = \langle g | g \rangle^{1/2}. \quad (59)$$

Let Ψ be a subspace of the Hilbert space $L_2(D)$ with the conditions that every element $\psi \in \Psi$ must satisfy being smoothness, limit behaviour at the boundary D , etc. Let us now consider the linear operator \mathbf{A} , defined on the linear manifold Ψ , and suppose that \mathbf{A} is

¹⁴ Here a Dirac type notation is introduced. It should not be confused with the ensemble average notation, which does not have a balk in-between.

a positive definite, i.e. $\langle \mathbf{A}\psi|\psi \rangle \geq 0$ ¹⁵ for all $\psi \in \Psi$. The kind of inverse problem we are interested in belongs to the stationary problems of the form

$$\mathbf{A}\psi = \mathbf{f}, \quad (60)$$

since, for example, for the COBE-filter we have to invert $\mathbf{A}\langle s \rangle_{\text{COBE}} = \mathbf{R}^\dagger \mathbf{N}^{-1} \mathbf{d}$, with $\psi = \langle s \rangle_{\text{COBE}}$, $\mathbf{A} = \mathbf{R}^\dagger \mathbf{N}^{-1} \mathbf{R}$ and $\mathbf{f} = \mathbf{R}^\dagger \mathbf{N}^{-1} \mathbf{d}$, and for the Wiener-filtering we have $\psi = (\mathbf{S}\mathbf{R}^\dagger)^{-1} \langle s \rangle_{\text{WF}}$, $\mathbf{A} = (\mathbf{R}^\dagger \mathbf{S}\mathbf{R} + \mathbf{N})$ and $\mathbf{f} = \mathbf{d}$. Eq. (60) has the same structure as eq. (2), but without a noise term. Hence, a regularization method is again required.

3.1.1 Minimization of the quadratic form

Another way of approaching the linear inverse problem is the minimization of a quadratic form given by

$$Q_{\mathbf{A}}(\psi) = \frac{1}{2} \langle \mathbf{A}\psi|\psi \rangle - \langle \mathbf{f}|\psi \rangle + c. \quad (61)$$

The gradient of $Q_{\mathbf{A}}$ leads to

$$\frac{dQ_{\mathbf{A}}}{d\psi}(\psi) \equiv Q'_{\mathbf{A}}(\psi) = \mathbf{A}\psi - \mathbf{f}, \quad (62)$$

assuming that the operator \mathbf{A} is self-adjoint. Setting the gradient to zero, one obtains eq. (60). The surface defined by a quadratic form with a positive definite matrix \mathbf{A} is shaped like a paraboloid bowl (see e.g. Shewchuk 1994). This ensures the existence of a unique minimum or, equivalently, the convergence of appropriate algorithms.

3.1.2 Solution of the non-stationary problem: asymptotic regularization

Here, a unified framework for the regularization methods that we have implemented in ARGO is given based on the asymptotic regularization. Nevertheless, an original Bayesian motivation to the asymptotic solution is presented in appendix F.

The stationary problem (eq. 60) can be replaced by a non-stationary equation, which relaxes to the equilibrium solution

$$\frac{\partial \psi}{\partial t} + \mathbf{A}\psi = \mathbf{f}. \quad (63)$$

We seek solutions of the form

$$\psi = \sum_l \psi_l \mathbf{u}_l, \quad (64)$$

with a spectrum for the operator \mathbf{A}

$$\mathbf{A}\mathbf{u}_l = \lambda_l \mathbf{u}_l. \quad (65)$$

Expanding \mathbf{f} in this basis, yields

$$\mathbf{f} = \sum_l f_l \mathbf{u}_l. \quad (66)$$

Then we get the following relations for the Fourier coefficients in the stationary case

$$\lambda_l \psi_l = f_l, \quad (67)$$

¹⁵ This expression can be written in matrix notation as $\psi^\dagger \mathbf{A}\psi \geq 0$, where ψ^\dagger is the conjugate and transpose of the vector ψ .

and for the non-stationary case

$$\frac{\partial \psi_l(t)}{\partial t} + \lambda_l \psi_l(t) = f_l, \quad \psi_l(0) = 0, \quad (68)$$

which lead to the following solutions

$$\psi = \sum_l \frac{f_l}{\lambda_l} \mathbf{u}_l, \quad (69)$$

and

$$\psi(t) = \sum_l \frac{f_l}{\lambda_l} (1 - e^{-\lambda_l t}) \mathbf{u}_l, \quad (70)$$

for the stationary and non-stationary cases, respectively. Since the spectrum of a positive definite operator \mathbf{A} is real, $\lambda_l > 0$, it follows that $\lim_{t \rightarrow \infty} \psi|_{\text{non-stationary}} = \psi|_{\text{stationary}}$.

The non-stationary problem can be solved using difference methods with respect to t

$$\psi^{j+1} = \psi^j + \tau^j \mathbf{M}^j (\mathbf{f} - \mathbf{A}\psi^j), \quad (71)$$

with $\{\mathbf{M}^j\}$ being a set of non-singular matrices¹⁶ and $\{\tau_j\}$ being a sequence of real parameters. Here we concentrate on a constant, self-adjoint matrix \mathbf{M} . Let us rewrite eq. (71) as

$$\psi^{j+1} = \psi^j + \tau^j \mathbf{M} \xi^j, \quad (72)$$

with the residuals given by

$$\xi^j = \mathbf{f} - \mathbf{A}\psi^j. \quad (73)$$

The error vectors are defined as

$$\eta^j = \psi^j - \psi^*, \quad (74)$$

where $\psi^* = \mathbf{A}^{-1} \mathbf{f}$ is the exact solution. The matrix \mathbf{M} and the real number $\{\tau_j\}$ are chosen to speed up the convergence. \mathbf{M} usually represents the preconditioning of eq. (71) and τ_j can be interpreted as the time step (see appendix G), and is also called relaxation parameter. Here truncation regularization occurs by quitting the iteration loop. Some stopping rules are therefore required. In the case where no noise regularization was conducted in the first step, they crucially define the noise regularization. In the other cases, they mostly determine algorithmic performance and accuracy. At this point we are interested in the regularization for the inverse purpose, since we have already found expressions which regularize the noise (e.g. Wiener-filter, or MEM). However, the results presented in section 4 show that in some cases truncation leads to better results (see discussion in section 4.2.6). In the following subsections, we will show how different iterative schemes are based on the general formula given by eq. (71). It is worth mentioning that other methods that we do not discuss in this paper, like the algebraic reconstruction technique (ART, see Gordon 1974), can also be expressed through this formula.

3.1.3 Jacobi method

The Jacobi iteration method splits the operator \mathbf{A} in two matrices

$$\mathbf{A} = \mathbf{D} + \mathbf{B}, \quad (75)$$

where \mathbf{D} contains the diagonal elements of \mathbf{A} and \mathbf{B} contains the off-diagonal elements. From eq. (60) one follows

$$\psi = \mathbf{D}^{-1} (\mathbf{f} - \mathbf{B}\psi). \quad (76)$$

¹⁶ We implicitly generalized eq. (63) to $\partial \psi(t)/\partial t = \mathbf{M}(t)(\mathbf{f} - \mathbf{A}\psi)$, where the auxiliary matrix \mathbf{M} is chosen to speed up convergence.

Substituting B by $A - D$ one gets the following iteration scheme

$$\psi^{j+1} = \psi^j + D^{-1}(\mathbf{f} - A\psi^j). \quad (77)$$

The Jacobi method turns out to be a particular case of the iteration scheme given by eq. (71) with a preconditioning matrix given by $M = D^{-1}$ and $\tau^j = 1$. This method can, must be optimized by increasing the timestep τ^j by a certain percentage if the solution converges and decreasing the timestep if the solution diverges. An optimal timestep is hard to find, because the spectrum of the operator A has to be known (see appendix G).

3.1.4 Steepest Descent method

The steepest descent method searches the minimum of the quadratic form by choosing the direction in which Q_A decreases most rapidly. This direction is given by the residual

$$-Q'_A(\psi^j) = \mathbf{f} - A\psi^j = \xi^j. \quad (78)$$

The form of the iteration scheme is thus given by eq. (72), with the length of the step in the direction of the residual given by τ^j . Steepest descent looks for the optimal length which minimizes the quadratic form with respect to τ^j

$$0 = \frac{dQ_A}{d\tau^j}(\psi^{j+1}) = \langle Q'_A(\psi^{j+1}) | \frac{d\psi^{j+1}}{d\tau^j} \rangle = \langle \xi^{j+1} | M\xi^j \rangle. \quad (79)$$

This implies that subsequent searching directions must be orthogonal (say $M = I$). Starting from this condition it is straightforward to derive the expression for τ^j . It is only necessary to use the definition of residual for ξ^{j+1} and substitute ψ^{j+1} from eq. (72).

$$\tau^j = \frac{\langle \xi^j | M\xi^j \rangle}{\langle AM\xi^j | M\xi^j \rangle}. \quad (80)$$

Both the calculation of the factors τ^j and the residuals ξ^j imply applying the operator A , each time on different vectors. It is possible, however, to reduce the operation of A to the same vector for every iteration, but the residuals, must be calculated in a different way. Multiplying both sides of eq. (72) by $-A$ and adding \mathbf{f} , one obtains the following relation for the residuals

$$\xi^{j+1} = \xi^j - \tau^j AM\xi^j. \quad (81)$$

Notice that the vector $AM\xi^j$ already appears in the expression for τ^j , and consequently saves one operation. However, expression (73) has to be periodically used with the feedback of ψ^j , to avoid the accumulation of floating-point roundoff error. The disadvantage of this method is that it ends up searching repeatedly in the same direction. This is especially severe when the quadratic form is highly deformed, which occurs when the matrix A deviates from the unity matrix. We will see, however, that steepest descent competes with any other method when the preconditioning is effective, and thus the stretched shape of the quadratic form is brought close to a spherical symmetric shape. Preconditioning should not imply too many operations; that is the reason why the inverse of the matrix, which contains only the diagonal elements of A , is usually taken for preconditioning. This will work especially fine when the operator A is diagonally dominant, which in our case occurs when nearly full-sky data are available.

3.1.5 Krylov methods: Conjugate Gradients

To make the iteration scheme more efficient, Conjugate Gradients proposes to search each time in a different direction. This

is achieved by imposing A -orthogonality to two different ($i \neq j$) searching vectors μ^i and μ^j

$$\langle \mu^j | \mu^i \rangle_A \equiv \langle A\mu^j | \mu^i \rangle = 0, \quad (82)$$

which are then said to be conjugated. In the preconditioned case, the searching vectors are multiplied by M so that the conjugacy has to be formulated in the following way: $\langle M\mu^j | M\mu^i \rangle_A = 0$ (for $i \neq j$).

The iteration scheme is given by substituting the residuals in eq. (72) by the new searching vectors $\{\mu^j\}$

$$\psi^{j+1} = \psi^j + \tau^j M\mu^j. \quad (83)$$

By subtracting ψ^* we obtain an equation for the errors,

$$\eta^{j+1} = \eta^j + \tau^j M\mu^j. \quad (84)$$

Taking into account the relation between the residuals and the errors

$$\xi^{j+1} = -A\eta^{j+1}, \quad (85)$$

we can derive the recurrent formula for the residuals

$$\xi^{j+1} = -A(\eta^j + \tau^j M\mu^j) = \xi^j - \tau^j AM\mu^j. \quad (86)$$

Here again, expression (73) has to be used periodically with the feedback of ψ^j to avoid the accumulation of floating-point round-off error. The optimal length of the step is found by minimizing the quadratic form

$$0 = \frac{dQ_A}{d\tau^j}(\psi^{j+1}) = -\langle \xi^{j+1} | M\mu^j \rangle = \langle \eta^{j+1} | M\mu^j \rangle_A. \quad (87)$$

Substituting expression (84) in (87) we then obtain

$$\tau^j = -\frac{\langle \eta^j | M\mu^j \rangle_A}{\langle M\mu^j | M\mu^j \rangle_A} = \frac{\langle \xi^j | M\mu^j \rangle}{\langle M\mu^j | M\mu^j \rangle_A}. \quad (88)$$

It can be shown that this formula is equivalent to the following expression

$$\tau^j = \frac{\langle \xi^j | M\xi^j \rangle}{\langle M\mu^j | M\mu^j \rangle_A}, \quad (89)$$

using $\langle \xi^j | M\mu^j \rangle = \langle \xi^j | M\xi^j \rangle$ (see appendix H).

To generate A -orthogonal searching vectors one could think of Gram-Schmidt-conjugation

$$\mu^j = \xi^j + \sum_{k=0}^{j-1} \beta^{jk} \mu^k. \quad (90)$$

Here it was assumed that the residuals $\{\xi^j\}$ form a set of linearly independent vectors (see appendix H). The expression for the factors β^{jk} can be derived by calling A -orthogonality in eq. (90)

$$\begin{aligned} \langle M\mu^j | M\mu^i \rangle_A &= \langle M\xi^j | M\mu^i \rangle_A + \sum_{k=0}^{j-1} \beta^{jk} \langle M\mu^k | M\mu^i \rangle_A \\ 0 &= \langle M\xi^j | M\mu^i \rangle_A + \beta^{ji} \langle M\mu^i | M\mu^i \rangle_A. \end{aligned} \quad (91)$$

One obtains the following formula for the factors

$$\beta^{ji} = -\frac{\langle M\xi^j | M\mu^i \rangle_A}{\langle M\mu^i | M\mu^i \rangle_A}, \quad (92)$$

where $i < j$ according to eq. (90)¹⁷.

¹⁷ Note that the sign of β depends on the definition of the Gram-Schmidt conjugation. An alternative definition with the negation of the residuals would cancel the minus sign in eq. (92). The sign of β can be regarded as a free parameter.

D_m	N_l	$\langle \xi^{j+1} M \xi^{j+1} \rangle$	$\langle \xi^{j+1} M(\xi^{j+1} - \xi^j) \rangle$	$\langle \xi^{j+1} - \xi^j M(\xi^{j+1} - \xi^j) \rangle$ $-\langle \xi^j M \xi^j \rangle$	$-\langle M \xi^{j+1} M \mu^j \rangle_{\mathbf{A}}$
$\langle \xi^j M \xi^j \rangle$		FR	PR	N3/D1	—
$\langle \mu^j M \xi^j \rangle$		N1/D2	N2/D2	N3/D2	—
$-\langle \xi^j M(\xi^{j+1} - \xi^j) \rangle$		N1/D3	N2/D3	N3/D3	—
$-\langle \mu^j M(\xi^{j+1} - \xi^j) \rangle$		N1/D4	HS	N3/D4	—
$-\langle (\xi^{j+1} - \xi^j M(\xi^{j+1} - \xi^j) \rangle$ $-\langle \xi^{j+1} M \xi^{j+1} \rangle$		N1/D5	N2/D5	N3/D5	—
$\langle M \mu^j M \mu^j \rangle_{\mathbf{A}}$		—	—	—	EXP

Table 2. Formulae for the β -factor: $\beta_{lm}^{j+1} = \frac{N_l}{D_m}$. Three of the methods are discussed in the literature: **FR** (Fletcher-Reeves), **PR** (Polak-Ribière, and **HS** (Hestenes-Stiefels). The rest of the formulae are derived in this paper using equivalence relations derived in appendices H1-H3. The **FR** and the **PR** methods are tested against the **EXP** algorithm in section (4).

This method seems to require too much memory, as apparently all previous searching vectors must be stored to calculate the new one. However, only one β -factor remains in the sum in eq. (90), as we show in appendix H3. Hence, Gram-Schmidt orthogonalization can be simplified to the following expression

$$\mu^{j+1} = \xi^{j+1} + \beta^{j+1} \mu^j, \quad (93)$$

where

$$\beta_{\text{EXP}}^{j+1} \equiv \beta^{j+1} \equiv \beta^{j+1,j} = -\frac{\langle M \xi^{j+1} | M \mu^j \rangle_{\mathbf{A}}}{\langle M \mu^j | M \mu^j \rangle_{\mathbf{A}}}, \quad (94)$$

with EXP meaning expensive, since the nominator of β requires an extra \mathbf{A} operation. This additional operation can be saved with alternative methods (see appendix H), like the Fletcher-Reeves method (Fletcher & Reeves 1964)

$$\beta_{\text{FR}}^{j+1} = \frac{\langle \xi^{j+1} | M \xi^{j+1} \rangle}{\langle \xi^j | M \xi^j \rangle}, \quad (95)$$

the Polak-Ribière formula (Polak & Ribière 1969)

$$\beta_{\text{PR}}^{j+1} = \frac{\langle \xi^{j+1} | M(\xi^{j+1} - \xi^j) \rangle}{\langle \xi^j | M \xi^j \rangle}, \quad (96)$$

or the Hestenes-Stiefel expression (Hestenes & Stiefel 1952)

$$\beta_{\text{HS}}^{j+1} = -\frac{\langle \xi^{j+1} | M(\xi^{j+1} - \xi^j) \rangle}{\langle \mu^j | M(\xi^{j+1} - \xi^j) \rangle}. \quad (97)$$

However, β_{EXP} turns out to be a very efficient scheme, which behaves far more stably than the rest (see section 4). Since the β -formulae (eq. 94-97) are mathematically equivalent, one could think of combining them in a single scheme finding numerically different solutions. In such a scheme, one may take advantage of the β_{EXP} -formula only every certain number of iterations, similar, to the update of the residuals (see eqs. 73 and 86). However, this kind of hybrid scheme remains to be thoroughly studied.

Formula (93) shows that new searching vectors are built from a linear combination of the current residual and the previous searching vector. Since the subsequent residuals are given by the linear combination of the previous residual and the \mathbf{A} -operator applied to the searching vector, the manifold where the solution is being searched is spanned by the residuals and the so-called Krylov space. The latter is built by applying the \mathbf{A} operator to the basis vector successively. In this manifold, curved quadratic forms appear to be spherical and thus the searching process becomes more effective. It is possible to derive the Conjugate Gradients method by

minimizing the \mathbf{A} -norm of the error: $\min \|\eta\|_{\mathbf{A}}$ (see e.g. Marchuk 1982). In this sense an optimal solution to the inverse problem can be found even if no unique solution exists. Conjugate Gradients works, even if the operator \mathbf{A} is not a positive definite (for a discussion see e.g. Shewchuk 1994). It can easily be shown that Conjugate Gradients converges at most in n -steps, with n being the number of pixels/vector columns (see e.g. Shewchuk 1994).

3.2 Non-linear inverse methods

Non-linear inverse methods are especially required in reconstruction algorithms that do not assume a gaussian distribution. The iterative method given in eq. (24), which makes use of a Poissonian likelihood, can alternatively be solved with the methods presented in this section. The same applies to the MEM, where zeros of the non-linear eq. (41) have to be found.

The generalization of the regularization methods to non-linear inverse problems is possible with methods like Tikhonov regularization as mentioned in section (2.5) or like asymptotic regularization as will be shown below (a relation between both methods is shown in appendix F). However, the proofs of the convergence properties are different since the spectral theoretical foundation is missing here. We refer the reader to e.g. O’Sullivan (1990).

Let us generalize eq. (60) to non-linear equations of the form

$$\mathbf{A}(\psi) = \mathbf{f}, \quad (98)$$

with \mathbf{A} being a non-linear operator, and solve the non-linear and non-stationary equation given by

$$\frac{\partial \psi}{\partial t} + \mathbf{A}(\psi) = \mathbf{f}, \quad (99)$$

with the forward Euler method. Discretizing the solution yields

$$\psi^{j+1} = \psi^j + \tau^j \mathbf{T}(\psi^j)(\mathbf{f} - \mathbf{A}(\psi^j)), \quad (100)$$

with \mathbf{T} being also a non-linear operator, typically given by $\nabla \mathbf{A}^\dagger$ or $\nabla \mathbf{A}^{-1}$, though more complicated expressions exist (see the Levenberg-Marquardt method or the regularized Gauss-Newton method, Hanke 1997 or Bakushinskii 1992 and Blaschke et al. 1997, respectively).

3.2.1 Newton-Raphson method

One of the most extended non-linear inverse methods is the so-called Newton-Raphson method (for an application in MEMs see

Maisinger et al. 1997; Hobson et al. 1998), which can easily be derived by doing a Taylor expansion of the function under study and truncating it at the first order

$$\psi^{j+1} = \psi^j + (\nabla \mathbf{A}(\psi^j))^{-1}(\mathbf{f} - \mathbf{A}(\psi^j)). \quad (101)$$

This method requires the inverse of the gradient of \mathbf{A} , which for the cases we are interested in is the inverse of a Hessian matrix. Recalling the problem of finding extrema of a function as presented in section (3.1.1) and taking into account eq. (78), the previous equation can be rewritten as

$$\psi^{j+1} = \psi^j - (\nabla \nabla Q_{\mathbf{A}}(\psi^j))^{-1} \nabla Q_{\mathbf{A}}(\psi^j), \quad (102)$$

where $\nabla \nabla Q_{\mathbf{A}} \equiv \partial Q_{\mathbf{A}} / \partial \psi^l \partial \psi^m$ is the Hessian matrix of $Q_{\mathbf{A}}$. For a direct derivation of this equation, we require a Taylor expansion until the second order of $Q_{\mathbf{A}}$, which is where the non-linearity arises. The MEM can be solved (eq. 41) with expression (102) by doing the substitutions: $Q_{\mathbf{A}} \rightarrow Q^{\text{E}}$ and $\psi^j \rightarrow \mathbf{s}^j$. Here the quantity Q^{E} is implicitly approximated by its quadratic expansion $Q_{\mathbf{A}}$. Calculating the inverse of the Hessian $(\nabla \nabla Q_{\mathbf{A}}(\psi^j))^{-1}$ implies solving a linear ill-posed problem in each iteration of the scheme (102). Some solutions have been found to regularize this scheme, like the Levenberg-Marquardt method (see Hanke 1997) or the regularized Gauss-Newton method (see e.g. Bakushinskii 1992; Blaschke et al. 1997).

3.2.2 Landweber-Fridman method

Alternative algorithms to the above mentioned Newton-Raphson class of methods do not need to invert the Hessian matrix and can thus simultaneously speed up and stabilize the inversion process. The Landweber-Fridman algorithm belongs to the class of methods based on steepest descent

$$\psi^{j+1} = \psi^j + (\nabla \mathbf{A}(\psi^j))^{\dagger}(\mathbf{f} - \mathbf{A}(\psi^j)). \quad (103)$$

Making the same substitutions as for eq. (102), we obtain

$$\psi^{j+1} = \psi^j - (\nabla \nabla Q_{\mathbf{A}}(\psi^j))^{\dagger} \nabla Q_{\mathbf{A}}(\psi^j). \quad (104)$$

Here just the adjoint of the Hessian must be taken $(\nabla \nabla Q_{\mathbf{A}}(\psi^j))^{\dagger}$. For a convergence analysis of this method see Hanke et al. (1995).

3.2.3 Non-linear Krylov methods

Another class of methods that do not require one to invert the Hessian matrix are the Krylov-based methods, which we have exposed in the previous section. The difference with respect to the linear case mainly resides in the calculation of the residuals ξ^j and the step size τ^j . The residuals are updated now by the negation of the gradient of the quadratic form that approximates the function under consideration $\xi^j = -\nabla Q_{\mathbf{A}}(\psi^j)$ (see eq. 78). The step size is given by

$$\tau^j = -\frac{\langle \nabla Q_{\mathbf{A}}(\psi^j) | \mathbf{M} \mu^j \rangle}{\langle \mathbf{M} \mu^j | \mathbf{M} \mu^j \rangle_{\nabla \nabla Q_{\mathbf{A}}(\psi^j)}}. \quad (105)$$

The derivation of this expression (see appendix H4) is based on the second order Taylor expansion of $Q_{\mathbf{A}}$. That is why Krylov algorithms which use this formula are called Newton-Krylov methods. There are alternative expressions for the time step τ^j where the Hessian is approximated and does not need to be explicitly calculated, like those using a secant approximation. For various implementations of non-linear Krylov methods see, for example, Shewchuk (1994).

3.3 Operator formalism

The iterative methods presented so far require an operator formalism to become efficient. In this formalism, matrices should be represented in such a way that their action can be expressed as simple operations, like sums and multiplications. In order to achieve this, one has to carefully choose the adequate representation, in which the individual matrix components are diagonal, though the whole matrix may not be. In this section, we present the different operators under consideration (see table 3) in k-space and real-space and discuss their optimal representation. In this way, we can take advantage of the fast Fourier-transform methods (FFTs) that scale as $n \log_2 n$, with n being the length of the arrays, and which ultimately determine the speed of the algorithm.

3.3.1 Fourier-transform definitions and dimensionality of the problem

Let us introduce the following definitions of the N_{D} -dimensional forward and inverse Fourier-transforms

$$\hat{x}(\mathbf{k}) \equiv \text{FT}[x(\mathbf{r})] \equiv \int d^{N_{\text{D}}} \mathbf{r} \exp(i\mathbf{k} \cdot \mathbf{r}) x(\mathbf{r}), \quad (106)$$

and

$$x(\mathbf{r}) \equiv \text{IFT}[\hat{x}(\mathbf{k})] \equiv \int \frac{d^{N_{\text{D}}} \mathbf{k}}{(2\pi)^{N_{\text{D}}}} \exp(-i\mathbf{k} \cdot \mathbf{r}) \hat{x}(\mathbf{k}), \quad (107)$$

respectively.

In general, the reconstruction problem has three spatial dimensions ($N_{\text{D}} = 3$), with the corresponding discrete array lengths for the real-space and k-space vectors given by $\mathbf{r} = (r_x, r_y, r_z)$ and $\mathbf{k} = (k_x, k_y, k_z)$. Each component has the following range: $r_x = \frac{L_x}{n_x} [0, n_x - 1]$, $r_y = \frac{L_y}{n_y} [0, n_y - 1]$, $r_z = \frac{L_z}{n_z} [0, n_z - 1]$ and $k_x = \frac{2\pi}{L_x} [0, n_x - 1]$, $k_y = \frac{2\pi}{L_y} [0, n_y - 1]$, $k_z = \frac{2\pi}{L_z} [0, n_z - 1]$, where the volume of the Universe under consideration is given by $V = L_x \times L_y \times L_z$ in $[(\text{Mpc}/h)^3]$, and the box containing that volume is divided into $n = n_x \times n_y \times n_z$ cells, with n being the length of the array x . In the following, we will treat the operators as being continuous. However, the discrete implementation can be derived in a straightforward way (for a discussion on the relation between discrete and continuous representations see Martel 2005). Note that the methods presented here can be applied in arbitrary dimensions. The number of dimensions N_{D} is thus kept as a free parameter.

In our convention, vectors defined in real-space have plain notation (x) and in k-space they are denoted with hats (\hat{x}). Matrices, however, have two hats in k-space. We represent convolutions with circles “ \circ ” and multiplications with dots “ \cdot ”. Due to the convolution theorem, where convolutions are shown to be multiplications in the counter space, we can either omit hats if they are present or include them if they are not, and replace circles with dots and vice versa “ $\cdot \leftrightarrow \circ$ ” to change from one representation to the other. All the numerical iterative inversion schemes (see section 3) of the different reconstruction algorithms (section 2) require only a small number of basic operators, listed in table (3). To show how the operators listed in table (3) can efficiently be applied we derive their action on an arbitrary vector.

3.3.2 Data model: the response operator

Let us first remember the data model given in eq. (3), and suppose that the operator \mathbf{R}_{P} is given by a convolution in real-space with

$$\begin{aligned}
 \hat{\mathbf{R}}\mathbf{S}\hat{\mathbf{R}}^\dagger\{\hat{\mathbf{x}}\}(\mathbf{k}) &= \int \frac{d^{N_D}\mathbf{k}'}{(2\pi)^{N_D}} \langle \hat{\alpha}(\mathbf{k}) \overline{\hat{\alpha}(\mathbf{k}')} \rangle_{(\mathbf{s}, \boldsymbol{\epsilon}|\mathbf{p})} \{\hat{x}(\mathbf{k}')\} \\
 &= \hat{f}_B(\mathbf{k}) \int \frac{d^{N_D}\mathbf{q}}{(2\pi)^{N_D}} \hat{f}_{SM}(\mathbf{k}-\mathbf{q}) \int \frac{d^{N_D}\mathbf{q}'}{(2\pi)^{N_D}} P_S(\mathbf{q}') (2\pi)^{N_D} \delta_D(\mathbf{q}-\mathbf{q}') \int \frac{d^{N_D}\mathbf{k}'}{(2\pi)^{N_D}} \overline{\hat{f}_{SM}(\mathbf{k}'-\mathbf{q}') \hat{f}_B(\mathbf{k}')} \{\hat{x}(\mathbf{k}')\} \\
 &= \hat{f}_B(\mathbf{k}) \int \frac{d^{N_D}\mathbf{q}}{(2\pi)^{N_D}} \hat{f}_{SM}(\mathbf{k}-\mathbf{q}) P_S(\mathbf{q}) \int \frac{d^{N_D}\mathbf{k}'}{(2\pi)^{N_D}} \overline{\hat{f}_{SM}(\mathbf{k}'-\mathbf{q}) \hat{f}_B(\mathbf{k}')} \cdot \{\hat{x}(\mathbf{k}')\} \\
 &= \hat{f}_B(\mathbf{k}) \int \frac{d^{N_D}\mathbf{q}}{(2\pi)^{N_D}} \hat{f}_{SM}(\mathbf{k}-\mathbf{q}) P_S(\mathbf{q}) \int \frac{d^{N_D}\mathbf{k}'}{(2\pi)^{N_D}} \hat{f}_{SM}(\mathbf{q}-\mathbf{k}') \underbrace{\overline{\hat{f}_B(\mathbf{k}')}}_{\hat{f}_B \cdot \{\hat{x}\}} \{\hat{x}(\mathbf{k}')\} \\
 &\quad \underbrace{\hspace{10em}}_{\hat{f}_{SM} \circ \overline{\hat{f}_B \cdot \{\hat{x}\}}} \\
 &\quad \underbrace{\hspace{10em}}_{P_S \cdot [\hat{f}_{SM} \circ \overline{\hat{f}_B \cdot \{\hat{x}\}}]} \\
 &\quad \underbrace{\hspace{10em}}_{\hat{f}_{SM} \circ [P_S \cdot \overline{\hat{f}_{SM} \circ \overline{\hat{f}_B \cdot \{\hat{x}\}}]}} \\
 &\quad \underbrace{\hspace{10em}}_{\hat{f}_B \cdot [\hat{f}_{SM} \circ [P_S \cdot \overline{\hat{f}_{SM} \circ \overline{\hat{f}_B \cdot \{\hat{x}\}}]}}]
 \end{aligned}$$

$$\begin{aligned}
 \hat{\mathbf{R}}^\dagger \hat{N}_N^{-1} \mathbf{R}\{\hat{\mathbf{x}}\}(\mathbf{k}) &= \overline{\hat{f}_B(\mathbf{k})} \int \frac{d^{N_D}\mathbf{q}}{(2\pi)^{N_D}} \overline{\hat{f}_{SM}(\mathbf{k}-\mathbf{q})} \int \frac{d^{N_D}\mathbf{q}'}{(2\pi)^{N_D}} P_N^{-1}(\mathbf{q}') (2\pi)^{N_D} \delta_D(\mathbf{q}-\mathbf{q}') \int \frac{d^{N_D}\mathbf{k}'}{(2\pi)^{N_D}} \hat{f}_{SM}(\mathbf{k}'-\mathbf{q}') \hat{f}_B(\mathbf{k}') \{\hat{x}(\mathbf{k}')\} \\
 &= \overline{\hat{f}_B(\mathbf{k})} \int \frac{d^{N_D}\mathbf{q}}{(2\pi)^{N_D}} \overline{\hat{f}_{SM}(\mathbf{k}-\mathbf{q})} P_N^{-1}(\mathbf{q}) \int \frac{d^{N_D}\mathbf{k}'}{(2\pi)^{N_D}} \hat{f}_{SM}(\mathbf{k}'-\mathbf{q}) \hat{f}_B(\mathbf{k}') \{\hat{x}(\mathbf{k}')\} \\
 &= \overline{\hat{f}_B(\mathbf{k})} \int \frac{d^{N_D}\mathbf{q}}{(2\pi)^{N_D}} \overline{\hat{f}_{SM}(\mathbf{k}-\mathbf{q})} P_N^{-1}(\mathbf{q}) \int \frac{d^{N_D}\mathbf{k}'}{(2\pi)^{N_D}} \overline{\hat{f}_{SM}(\mathbf{q}-\mathbf{k}')} \underbrace{\hat{f}_B(\mathbf{k}') \{\hat{x}(\mathbf{k}')\}}_{\hat{f}_B \cdot \{\hat{x}\}} \\
 &\quad \underbrace{\hspace{10em}}_{\overline{\hat{f}_{SM} \circ \hat{f}_B \cdot \{\hat{x}\}}} \\
 &\quad \underbrace{\hspace{10em}}_{P_N \cdot [\overline{\hat{f}_{SM} \circ \hat{f}_B \cdot \{\hat{x}\}}]} \\
 &\quad \underbrace{\hspace{10em}}_{\overline{\hat{f}_{SM} \circ [P_N^{-1} \cdot \overline{\hat{f}_{SM} \circ \hat{f}_B \cdot \{\hat{x}\}}]}} \\
 &\quad \underbrace{\hspace{10em}}_{\overline{\hat{f}_B} \cdot [\overline{\hat{f}_{SM} \circ [P_N^{-1} \cdot \overline{\hat{f}_{SM} \circ \hat{f}_B \cdot \{\hat{x}\}}]}}]
 \end{aligned}$$

$$\begin{aligned}
 \hat{\mathbf{R}}^\dagger \hat{N}_{WN}^{-1} \mathbf{R}\{\hat{\mathbf{x}}\}(\mathbf{k}) &= \overline{\hat{f}_B(\mathbf{k})} \int \frac{d^{N_D}\mathbf{q}}{(2\pi)^{N_D}} \overline{\hat{f}_{SM}(\mathbf{k}-\mathbf{q})} \int \frac{d^{N_D}\mathbf{q}'}{(2\pi)^{N_D}} N_{WN}^{-1}(\mathbf{q}-\mathbf{q}') \int \frac{d^{N_D}\mathbf{k}'}{(2\pi)^{N_D}} \hat{f}_{SM}(\mathbf{k}'-\mathbf{q}) \hat{f}_B(\mathbf{k}') \{\hat{x}(\mathbf{k}')\} \\
 &= \overline{\hat{f}_B(\mathbf{k})} \int \frac{d^{N_D}\mathbf{q}}{(2\pi)^{N_D}} \overline{\hat{f}_{SM}(\mathbf{k}-\mathbf{q})} \int \frac{d^{N_D}\mathbf{q}'}{(2\pi)^{N_D}} N_{WN}^{-1}(\mathbf{q}-\mathbf{q}') \int \frac{d^{N_D}\mathbf{k}'}{(2\pi)^{N_D}} \overline{\hat{f}_{SM}(\mathbf{q}-\mathbf{k}')} \underbrace{\hat{f}_B(\mathbf{k}') \{\hat{x}(\mathbf{k}')\}}_{\hat{f}_B \cdot \{\hat{x}\}} \\
 &\quad \underbrace{\hspace{10em}}_{\overline{\hat{f}_{SM} \circ \hat{f}_B \cdot \{\hat{x}\}}} \\
 &\quad \underbrace{\hspace{10em}}_{N_{WN}^{-1} \circ [\overline{\hat{f}_{SM} \circ \hat{f}_B \cdot \{\hat{x}\}}]} \\
 &\quad \underbrace{\hspace{10em}}_{\overline{\hat{f}_{SM} \circ [N_{WN}^{-1} \circ \overline{\hat{f}_{SM} \circ \hat{f}_B \cdot \{\hat{x}\}}]}} \\
 &\quad \underbrace{\hspace{10em}}_{\overline{\hat{f}_B} \cdot [\overline{\hat{f}_{SM} \circ [N_{WN}^{-1} \circ \overline{\hat{f}_{SM} \circ \hat{f}_B \cdot \{\hat{x}\}}]}}]
 \end{aligned}$$

Figure 2. Here the action on an arbitrary vector $\hat{\mathbf{x}}$ of the most complex operators that appear in table (3) is shown. The upper one is required for Wiener-filtering and represents the signal term in the covariance matrix of the data. The middle and lower ones stand for the inverse of the ML variance (eq. 26) and are required for the COBE-filter, the MEMG and for sampling purposes with the Wiener-filter. The equations have to be read from right to left. The braces show the order in which the operations have to be done from top to bottom. One has to be very careful with the correct conjugation of the different functions. Note that, contrary to naive expectations, the conjugation of the first selection function \hat{f}_{SM} to be applied in the upper operation disappears and appears in the middle and lower ones, though initially absent.

	\mathbf{R}	\mathbf{R}^\dagger	\mathbf{S}	\mathbf{S}^{-1}	$\mathbf{S}^{-1/2}$	\mathbf{N}	\mathbf{N}^{-1}	$\mathbf{S}\mathbf{R}^\dagger$	$\hat{\mathbf{R}}^\dagger\mathbf{N}^{-1}$	$\mathbf{R}^\dagger\mathbf{N}^{-1/2}$	$\mathbf{R}^\dagger\mathbf{N}^{-1}\mathbf{R}$	$\mathbf{R}\mathbf{S}\mathbf{R}^\dagger$
COBE							X		X		X	
WIENER	X	X	X	X [#]	X [#]	X	X [#]	X		X [#]	X [#]	X
GAPMAP	X	X	X					X				
MEMG	X	X					X		X		X	
MEMP	X	X										

additional operators required for sampling processes (see eq. 49)

Table 3. Operators in columns needed for the different estimators in rows, the COBE-filter (25), the Wiener-filter (22), the GAPMAP estimator (24), and the MEMs (sections 2.5.9 & 3.2, and appendix E). Note that the trivial diagonal matrices have been left out of this table. The first two estimators are linear estimators, whereas the rest are non-linear. MEMG and MEMP stand for the Maximum Entropy method with a Gaussian likelihood and with a Poissonian likelihood, respectively. Note that some of the operators have to be further inverted either directly, like $(\mathbf{R}^\dagger\mathbf{N}\mathbf{R})^{-1}$ for the COBE-filter, or in combination with other operators, like $(\mathbf{R}\mathbf{S}\mathbf{R}^\dagger + \mathbf{N})^{-1}$ for the Wiener-filter. The methods presented in section (3) show how to do this implicitly by applying the operators in an iterative fashion.

some blurring function f_B

$$d(\mathbf{r}) \equiv \int d^{N_D} \mathbf{r}' f_B(\mathbf{r} - \mathbf{r}') f_S(\mathbf{r}') f_M(\mathbf{r}') s(\mathbf{r}') + f_{SF}(\mathbf{r}) \epsilon_N(\mathbf{r}). \quad (108)$$

The operator \mathbf{R} acting on an arbitrary vector $\{x\}$ is thus given by

$$\mathbf{R}\{x\}(\mathbf{r}) \equiv \int d^{N_D} \mathbf{r}' f_B(\mathbf{r} - \mathbf{r}') f_S(\mathbf{r}') f_M(\mathbf{r}') \{x(\mathbf{r}')\}. \quad (109)$$

The selection function and the masks should conveniently be multiplied in real-space to save convolutions

$$f_{SM}(\mathbf{r}) \equiv f_S(\mathbf{r}) f_M(\mathbf{r}). \quad (110)$$

Accordingly, the same operation as in eq. (109) leads to

$$\hat{\mathbf{R}}\{\hat{x}\}(\mathbf{k}) = \hat{f}_B(\mathbf{k}) \underbrace{\int \frac{d^{N_D} \mathbf{q}}{(2\pi)^{N_D}} \hat{f}_{SM}(\mathbf{k} - \mathbf{q}) \{ \hat{x}(\mathbf{q}) \}}_{\hat{f}_{SM} \circ \{ \hat{x} \}} \quad (111)$$

$$\hat{f}_B \cdot [\hat{f}_{SM} \circ \{ \hat{x} \}],$$

in k-space. Here we have introduced the operator notation in which the equations have to be read from right to left. The braces show the sequence in which the subsequent operations have to be performed in the algorithm. The analogous operation for the adjoint \mathbf{R}^\dagger yields

$$\hat{\mathbf{R}}^\dagger \{\hat{x}\}(\mathbf{k}) = \overline{\hat{f}_B} \cdot \overline{[\hat{f}_{SM} \circ \{ \hat{x} \}]}(\mathbf{k}). \quad (112)$$

3.3.3 Covariance matrix of the data

The data model consists of two terms

$$\alpha(\mathbf{r}) = \int d^{N_D} \mathbf{r}' f_B(\mathbf{r} - \mathbf{r}') f_{SM}(\mathbf{r}') s(\mathbf{r}'), \quad (113)$$

and

$$\epsilon(\mathbf{r}) = f_{SF}(\mathbf{r}) \epsilon_N(\mathbf{r}). \quad (114)$$

The same quantities in k-space are given by

$$\hat{\alpha}(\mathbf{k}) = \hat{f}_B(\mathbf{k}) \int \frac{d^{N_D} \mathbf{q}}{(2\pi)^{N_D}} \hat{f}_{SM}(\mathbf{k} - \mathbf{q}) \hat{s}(\mathbf{q}), \quad (115)$$

and

$$\hat{\epsilon}(\mathbf{k}) = \int \frac{d^{N_D} \mathbf{q}}{(2\pi)^{N_D}} \hat{f}_{SF}(\mathbf{k} - \mathbf{q}) \hat{\epsilon}_N(\mathbf{q}). \quad (116)$$

Consequently, the covariance matrix of the data is given by the following sum

$$\langle \hat{d}(\mathbf{k}) \overline{\hat{d}(\mathbf{k}')} \rangle_{(s, \epsilon | \mathbf{p})} = \langle \hat{\alpha}(\mathbf{k}) \overline{\hat{\alpha}(\mathbf{k}')} \rangle_{(s, \epsilon | \mathbf{p})} + \langle \hat{\epsilon}(\mathbf{k}) \overline{\hat{\epsilon}(\mathbf{k}')} \rangle_{(s, \epsilon | \mathbf{p})}, \quad (117)$$

where we have assumed that the noise is uncorrelated to the signal, which is consistent with our data model. Even though the structure function may be correlated with the signal $\langle \hat{s}(\mathbf{k}) \overline{\hat{f}_{SF}(\mathbf{k}')} \rangle_{(s, f_{SF} | \mathbf{p})} \neq 0$, the random noise part is not $\langle \hat{s}(\mathbf{k}) \overline{\hat{\epsilon}_N(\mathbf{k}')} \rangle_{(s, \epsilon | \mathbf{p})} = 0$. We will calculate the different terms of the data covariance matrix and other related operators in the next sections.

3.3.4 Covariance matrix of the data: the signal term

Here it becomes necessary to choose the Fourier representation, since it is there that the signal-autocorrelation matrix appears to be diagonal in the form of a power spectrum (eq. 118). Taking into account statistical homogeneity for the signal s

$$\langle \hat{s}(\mathbf{k}) \overline{\hat{s}(\mathbf{k}')} \rangle_{(s | \mathbf{p})} = (2\pi)^{N_D} \delta_D(\mathbf{k} - \mathbf{k}') P_S(\mathbf{k}'), \quad (118)$$

with δ_D being the Dirac-delta function, we can derive the expression for the signal covariance matrix term

$$\begin{aligned} \hat{\mathbf{R}}\mathbf{S}\mathbf{R}^\dagger(\mathbf{k}, \mathbf{k}') &= \langle \hat{\alpha}(\mathbf{k}) \overline{\hat{\alpha}(\mathbf{k}')} \rangle_{(s | \mathbf{p})} \quad (119) \\ &= \hat{f}_B(\mathbf{k}) \int \frac{d^{N_D} \mathbf{q}}{(2\pi)^{N_D}} \hat{f}_{SM}(\mathbf{k} - \mathbf{q}) P_S(\mathbf{q}) \overline{\hat{f}_{SM}(\mathbf{k}' - \mathbf{q})} \hat{f}_B(\mathbf{k}') \\ &= \hat{f}_B(\mathbf{k}) \int \frac{d^{N_D} \mathbf{q}}{(2\pi)^{N_D}} \hat{f}_{SM}(\mathbf{k} - \mathbf{q}) P_S(\mathbf{q}) \hat{f}_{SM}(\mathbf{q} - \mathbf{k}') \overline{\hat{f}_B(\mathbf{k}')}, \end{aligned}$$

For its action on a vector (see fig. 2), we get

$$\hat{\mathbf{R}}\mathbf{S}\mathbf{R}^\dagger\{\hat{x}\}(\mathbf{k}) = \hat{f}_B \cdot [\hat{f}_{SM} \circ [P_S \cdot [\hat{f}_{SM} \circ [\overline{\hat{f}_B \cdot \{ \hat{x} \}]}]]](\mathbf{k}), \quad (120)$$

and consequently

$$\begin{aligned} \hat{\mathbf{S}}\mathbf{R}^\dagger\{\hat{\mathbf{x}}\}(\mathbf{k}) &= \int \frac{d^{N_D}\mathbf{k}'}{(2\pi)^{N_D}} \langle \hat{s}(\mathbf{k})\overline{\hat{d}(\mathbf{k}')} \rangle_{(\mathbf{s}|\mathbf{p})} \{\hat{\mathbf{x}}(\mathbf{k}')\} \\ &= P_S(\mathbf{k}) \int \frac{d^{N_D}\mathbf{k}'}{(2\pi)^{N_D}} \underbrace{\hat{f}_{SM}(\mathbf{k}-\mathbf{k}') \overbrace{\hat{f}_B(\mathbf{k}') \cdot \{\hat{\mathbf{x}}(\mathbf{k}')\}}^{\hat{f}_B \cdot \{\hat{\mathbf{x}}\}}}_{\hat{f}_{SM} \circ [\hat{f}_B \cdot \{\hat{\mathbf{x}}\}]} \\ &= P_S \cdot [\hat{f}_{SM} \circ [\hat{f}_B \cdot \{\hat{\mathbf{x}}\}]]. \end{aligned} \quad (121)$$

The inverse of the signal-autocorrelation matrix can be solved trivially in Fourier-space: $\hat{\mathbf{S}}^{-1} = \text{diag}(P_S(\mathbf{k})^{-1})$. Hence, the inverse square root yields $\hat{\mathbf{S}}^{-1/2} = \text{diag}(P_S(\mathbf{k})^{-1/2})$.

3.3.5 Covariance matrix of the data: the noise term

We assume, analogous to the case of the signal, statistical homogeneity for ϵ_N

$$\langle \hat{\epsilon}_N(\mathbf{k})\overline{\hat{\epsilon}_N(\mathbf{k}')} \rangle_{(\epsilon|\mathbf{p})} = (2\pi)^{N_D} \delta_D(\mathbf{k}-\mathbf{k}') P_N(\mathbf{k}'), \quad (122)$$

and then derive the expression for the noise covariance matrix

$$\begin{aligned} \hat{\mathbf{N}}(\mathbf{k}, \mathbf{k}') &= \langle \hat{\epsilon}(\mathbf{k})\overline{\hat{\epsilon}(\mathbf{k}')} \rangle_{(\epsilon|\mathbf{p})} \\ &= \int \frac{d^{N_D}\mathbf{q}}{(2\pi)^{N_D}} \hat{f}_{SF}(\mathbf{k}-\mathbf{q}) P_N(\mathbf{q}) \hat{f}_{SF}(\mathbf{q}-\mathbf{k}'). \end{aligned} \quad (123)$$

Its action on a vector yields

$$\begin{aligned} \hat{\mathbf{N}}\{\hat{\mathbf{x}}\}(\mathbf{k}) &= \int \frac{d^{N_D}\mathbf{k}'}{(2\pi)^{N_D}} \langle \hat{\epsilon}(\mathbf{k})\overline{\hat{\epsilon}(\mathbf{k}')} \rangle_{(\epsilon|\mathbf{p})} \{\hat{\mathbf{x}}(\mathbf{k}')\} \\ &= \int \frac{d^{N_D}\mathbf{q}}{(2\pi)^{N_D}} \hat{f}_{SF}(\mathbf{k}-\mathbf{q}) P_N(\mathbf{q}) \underbrace{\int \frac{d^{N_D}\mathbf{k}'}{(2\pi)^{N_D}} \hat{f}_{SF}(\mathbf{q}-\mathbf{k}') \{\hat{\mathbf{x}}(\mathbf{k}')\}}_{\hat{f}_{SF} \circ \{\hat{\mathbf{x}}\}} \\ &= \underbrace{\hat{f}_{SF} \circ [P_N \cdot [\hat{f}_{SF} \circ \{\hat{\mathbf{x}}\}]]}_{\hat{f}_{SF} \circ [P_N \cdot [\hat{f}_{SF} \circ \{\hat{\mathbf{x}}\}]}, \end{aligned} \quad (124)$$

In the case where there is no structure function, the noise autocorrelation reduces to

$$\hat{\mathbf{N}}_N(\mathbf{k}, \mathbf{k}') = (2\pi)^{N_D} \delta_D(\mathbf{k}-\mathbf{k}') P_N(\mathbf{k}'). \quad (125)$$

Then, its action is given by

$$\hat{\mathbf{N}}_N\{\hat{\mathbf{x}}\}(\mathbf{k}) = P_N \cdot \{\hat{\mathbf{x}}\}(\mathbf{k}). \quad (126)$$

The corresponding inverse operation is

$$\hat{\mathbf{N}}_N^{-1}\{\hat{\mathbf{x}}\}(\mathbf{k}) = P_N^{-1} \cdot \{\hat{\mathbf{x}}\}(\mathbf{k}). \quad (127)$$

Consequently, we obtain (see fig. 2)

$$\mathbf{R}^\dagger \hat{\mathbf{N}}_N^{-1} \mathbf{R}\{\hat{\mathbf{x}}\}(\mathbf{k}) = \overline{\hat{f}_B} \cdot [\overline{\hat{f}_{SM}} \circ [P_N^{-1} \cdot [\overline{\hat{f}_{SM}} \circ [\hat{f}_B \cdot \{\hat{\mathbf{x}}\}]]]](\mathbf{k}), \quad (128)$$

and

$$\mathbf{R}^\dagger \hat{\mathbf{N}}_N^{-1} \{\mathbf{x}\}(\mathbf{k}) = \overline{\hat{f}_B} \cdot [\overline{\hat{f}_{SF}} \circ [P_N^{-1} \cdot \{\mathbf{x}\}]](\mathbf{k}). \quad (129)$$

The inverse square root of $\hat{\mathbf{N}}_N$ can now be calculated and leads to

$$\hat{\mathbf{N}}_N^{-1/2}(\mathbf{k}) = \text{diag}(P_N^{-1/2}(\mathbf{k})). \quad (130)$$

The operation $\mathbf{R}^\dagger \hat{\mathbf{N}}_N^{-1/2} \{\hat{\mathbf{x}}\}$ can then be obtained by doing the following substitution $\hat{\mathbf{N}}_N^{-1} \rightarrow \hat{\mathbf{N}}_N^{-1/2}$ in eq. (129)

$$\mathbf{R}^\dagger \hat{\mathbf{N}}_N^{-1/2} \{\mathbf{x}\}(\mathbf{k}) = \overline{\hat{f}_B} \cdot [\overline{\hat{f}_{SF}} \circ [P_N^{-1/2} \cdot \{\mathbf{x}\}]](\mathbf{k}). \quad (131)$$

We are especially interested in the case of white noise ($P_N = P_{WN} = \text{const}$) with a structure function (given by the Poissonian shot noise)

$$\hat{\mathbf{N}}_{WN}(\mathbf{k}, \mathbf{k}') = P_{WN} \int \frac{d^{N_D}\mathbf{q}}{(2\pi)^{N_D}} \hat{f}_{SF}(\mathbf{k}-\mathbf{q}) \overline{\hat{f}_{SF}(\mathbf{k}'-\mathbf{q})}. \quad (132)$$

The corresponding action yields

$$\begin{aligned} \hat{\mathbf{N}}_{WN}\{\hat{\mathbf{x}}\}(\mathbf{k}) &= P_{WN} \int \frac{d^{N_D}\mathbf{q}}{(2\pi)^{N_D}} \hat{f}_{SF}(\mathbf{k}-\mathbf{q}) \underbrace{\int \frac{d^{N_D}\mathbf{k}'}{(2\pi)^{N_D}} \hat{f}_{SF}(\mathbf{q}-\mathbf{k}') \{\hat{\mathbf{x}}(\mathbf{k}')\}}_{\hat{f}_{SF} \circ \{\hat{\mathbf{x}}\}} \\ &= \underbrace{P_{WN} \cdot [\hat{f}_{SF} \circ [\hat{f}_{SF} \circ \{\hat{\mathbf{x}}\}]]}_{\hat{f}_{SF} \circ [\hat{f}_{SF} \circ \{\hat{\mathbf{x}}\}]} = P_{WN} \cdot [\hat{f}_{SF}^2 \circ \{\hat{\mathbf{x}}\}] \end{aligned} \quad (133)$$

It can be seen from this equation, that the preferential representation now is in real-space, where \mathbf{N} is diagonal

$$\mathbf{N}_{WN}(\mathbf{r}, \mathbf{r}') = \delta_D(\mathbf{r}-\mathbf{r}') C_{WN} f_{SF}^2(\mathbf{r}'), \quad (134)$$

with $C_{WN} = \text{IFT}[P_{WN}]$ being a constant. The inverse operation yields

$$\mathbf{N}_{WN}^{-1}\{\mathbf{x}\}(\mathbf{r}) = (C_{WN} f_{SF}^2)^{-1} \cdot \{\mathbf{x}\}(\mathbf{r}). \quad (135)$$

Hence, the inverse square root yields

$$\mathbf{N}_{WN}^{-1/2}(\mathbf{r}, \mathbf{r}') = \delta_D(\mathbf{r}-\mathbf{r}') C_{WN}^{-1/2} f_{SF}^{-1}(\mathbf{r}'), \quad (136)$$

and its action in k-space reads

$$\hat{\mathbf{N}}_{WN}^{-1/2}\{\hat{\mathbf{x}}\}(\mathbf{k}) = P_{WN}^{-1/2} \cdot [\hat{f}_{SF}^{-1} \circ \{\hat{\mathbf{x}}\}](\mathbf{k}). \quad (137)$$

Then we get (see fig. 2)

$$\mathbf{R}^\dagger \hat{\mathbf{N}}_{WN}^{-1} \mathbf{R}\{\hat{\mathbf{x}}\}(\mathbf{k}) = \overline{\hat{f}_B} \cdot [\overline{\hat{f}_{SF}} \circ [\hat{\mathbf{N}}_{WN}^{-1} \circ [\overline{\hat{f}_{SF}} \circ [\hat{f}_B \cdot \{\hat{\mathbf{x}}\}]]]](\mathbf{k}), \quad (138)$$

and consequently

$$\mathbf{R}^\dagger \hat{\mathbf{N}}_{WN}^{-1} \{\hat{\mathbf{x}}\}(\mathbf{k}) = \overline{\hat{f}_B} \cdot [\overline{\hat{f}_{SF}} \circ [\hat{\mathbf{N}}_{WN}^{-1} \circ \{\hat{\mathbf{x}}\}]](\mathbf{k}). \quad (139)$$

To calculate $\mathbf{R}^\dagger \hat{\mathbf{N}}_{WN}^{-1/2} \{\hat{\mathbf{x}}\}$ one has to do the following substitution $\hat{\mathbf{N}}_{WN}^{-1} \rightarrow \hat{\mathbf{N}}_{WN}^{-1/2}$ in eq. (139)

$$\mathbf{R}^\dagger \hat{\mathbf{N}}_{WN}^{-1/2} \{\hat{\mathbf{x}}\}(\mathbf{k}) = \overline{\hat{f}_B} \cdot [\overline{\hat{f}_{SF}} \circ [\hat{\mathbf{N}}_{WN}^{-1/2} \circ \{\hat{\mathbf{x}}\}]](\mathbf{k}). \quad (140)$$

In summary, we showed that the action of the different operators on a vector required for the different reconstruction estimators (see table 3) can be calculated in a straightforward way, as an ordered series of products and convolutions. Note that whenever we need to perform a convolution, we change to the counter space representation with FFTs and do multiplications¹⁸ there.

¹⁸ In order to avoid aliasing effects one has to adequately perform *zero-padding* (see e.g. Press et al. 1992).

4 EFFICIENCY AND QUALITY VALIDATION OF THE INVERSE METHODS WITH THE WIENER-FILTER

In this section the Wiener-filter implemented in ARGO is tested with the different linear inverse algorithms presented in the section of numerical methods (3) under several conditions determined by structured noise, blurring, selection function effects and windowing.

The inverse methods that we test here are the Jacobi (J), the Steepest Descent (SD), and several Krylov methods, like the Fletcher-Reeves (FR), the Polak-Ribière (PR), and the *expensive* Conjugate Gradients method (EXP), which requires an additional operation in each iteration (see section 3.1.5 and appendix H3). This scheme has not been previously discussed in the literature and turns out to be very efficient as will be discussed below. Many other Krylov methods (see table 2) can be built from simple equivalence relations, as we show in appendix H. However, only the methods mentioned above are taken into account here, as we consider them to be sufficiently representative. The extra-regularization we propose with these Krylov methods converts the Wiener-filtering in a hybrid Tikhonov-Krylov space regularization method. In addition, we also test the Wiener-filter that uses hermitian redundancy as derived in appendix I. We call the Wiener-filter defined by the mapping equation (16) the conjugated Wiener-filter (CJ), whereas the Wiener-filter defined by eq. (17) has no extra suffix.

With the aim of having full control over the synthetic data, we generate Gaussian random fields¹⁹ with the Peacock & Dodds (1994) formula for the power spectrum. The resulting *real* density field is denoted by $\delta_{\text{real}} \equiv \delta_\rho$, and the reconstruction by $\delta_{\text{rec}} \equiv \psi$. The signals are discretized and arranged as vectors given by $[k + n_z \times (j + n_y \times i)]$, where $i \in [0, n_x - 1]$, $j \in [0, n_y - 1]$, and $k \in [0, n_z - 1]$. The algorithmical part of the reconstruction methods shown in section (3) does not change with the dimensionality, but solely the length of the vectors given by $n = n_x \times n_y \times n_z$ change and thus also the dimension of the involved matrices. The formulation of the matrices is explained in detail in section (3.3). The Fourier transforms must be accordingly called with the dimensions under consideration, which occurs in ARGO by switching between the different FFTs given by FFTW²⁰. In addition, the power spectrum that is used for the reconstruction has to be set up with the corresponding length and the data have to be correctly rearranged to their original dimensions ($[i][j][k] \leftarrow [k + n_z \times (j + n_y \times i)]$) after their manipulation.

4.1 One-dimensional example

We can see in fig. (3) an example of a Gaussian realization in one-dimension (red curve) that can represent a time-line. A structured noise that increases with the distance and with a random noise component was added to the signal. Finally a region was excluded simulating windowing effects. The resulting curve was taken as the input signal (yellow curve). The reconstruction given by ARGO is in blue and green, where the boundary effects were considered in the first case, but not in the second. There the signal was assumed to be zero in the unsampled region. We can see that the blue curve better

¹⁹ We use GARFIELDS: **GA**ussian **R**andom **F**IELDS, a program we developed to generate Gaussian random fields from a given power spectrum. The method can be found in detail in Martel (2005).

²⁰ FFTW is a C subroutine library for computing fast discrete Fourier transforms in one or more dimensions of arbitrary input size and of both real and complex data: <http://www.fftw.org/>

resembles the *real* signal guided by the trend at the boundary. This effect is much larger in multiple dimensions as is shown in fig. (11). In the right plot in fig. (3), two sampling processes are underlying the yellow signal. First, the Gaussian random field that generates the red signal, which is then Poisson sampled thus leading to the yellow data. Again the blue and the green curves represent the reconstructions with and without proper window treatment, respectively. In this case, the blue curve also approaches the *true* signal better.

4.2 Multi-dimensional test cases

ARGO has been implemented such that the global dimension N_D (see section 3.3.1), and even the length in each dimension (n_x, n_y, n_z), can be chosen arbitrarily. Our tests in one-, two- and three dimensions show that the results do not differ qualitatively. The convergence behaviour changes with the length of the arrays ($n = n_x \times n_y \times n_z$) as $n \log_2 n$ fully determined by the FFTs, as we showed in section (3). For the demonstration cases in this paper, we have selected the two-dimensional tests with $128 \times 128 = 16384$ pixels. However, three dimensional tests were also carried out leading to the same conclusions.

4.2.1 Qualitative and quantitative measurement of the quality of the reconstruction

To give a quantitative measurement of the quality of the reconstructions, we define the correlation coefficient r between the reconstructed and the real density field by

$$r \equiv \frac{\sum_i^n \delta_{\rho i} \psi_i}{\sqrt{\sum_i^n \delta_{\rho i}^2} \sqrt{\sum_j^n \psi_j^2}}. \quad (141)$$

This statistical quantity is not very sensitive to the overall distribution and yields good values (close to unity) in some cases even with poor reconstructions (see figure 9). The pixel to pixel plot of the *real* density field against the reconstruction is highly informative because the scatter in the alignment of the pixels around the line of perfect correlation (45° slope) gives a qualitative goodness of the reconstruction. In general, the quality of the recovered density map is better represented by the Euclidean distance between the real and the reconstructed signals. The ensemble average of this quantity can also be regarded as an action or loss function that leads to the Wiener-filter through minimization (see appendix I). Here we introduce the volume-averaged squared Euclidean distance²¹

$$D_{\text{Eucl}}^2(\psi, \delta_\rho) \equiv \frac{1}{V} \int d^{N_D} \mathbf{r} \left[\psi(\mathbf{r}) - \delta_\rho(\mathbf{r}) \right]^2, \quad (142)$$

with $V = L_x \times L_y \times L_z$. We further normalize the Euclidean distance through the following definition

$$D_{\text{Eucl}}^2(\psi, \delta_\rho) \equiv \frac{D_{\text{Eucl}}^2(\psi, \delta_\rho)}{D_{\text{Eucl}}^2(\psi_0, \delta_\rho)}, \quad (143)$$

where ψ_0 is the zero vector. We define the convergence tolerance criterion based on the squared Euclidean distance between subsequent reconstructions

$$\text{tol}_{\text{crit}}^{j+1} \equiv D_{\text{Eucl}}^2(\psi^{j+1}, \psi^j). \quad (144)$$

We prefer this criterion with respect to the squared residuals $\|\xi\|^2$ (see eq. 73) because all the tests show that no further statistical

²¹ Note that $D_{\text{Eucl}}^2(\psi, \delta_\rho) = \frac{1}{V} D_{\text{Eucl}}^2(\psi, \delta_\rho)$.

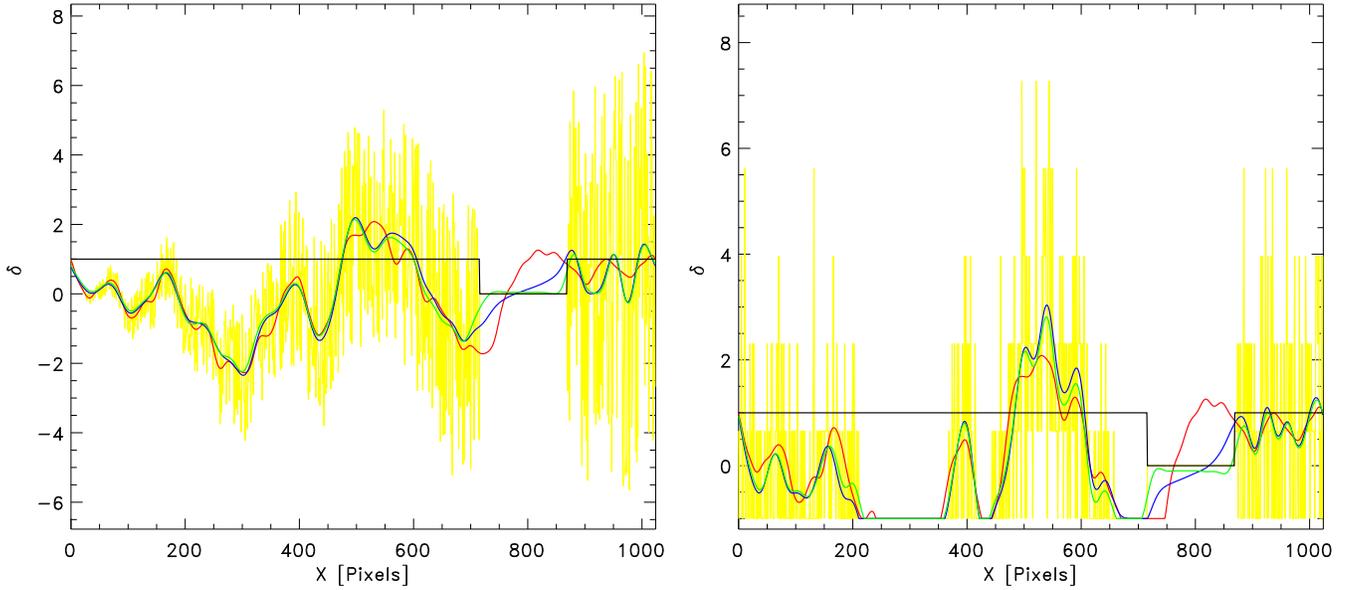


Figure 3. 1-D Reconstruction with structured noise & window: The left plot shows the reconstruction of a one-dimensional noisy signal. The red curve is the *true* underlying signal. The yellow lines represent the measured data in each grid cell. The data are windowed by a function given by the black line. A random noise with a structure function that increases with the distance with respect to the origin has been added to the *true* signal. The green and the blue lines show different reconstructions. In the blue case the windowing is formally treated, whereas in the green case the unseen region is modeled by a mean signal, which is zero in this case. We see that the unsampled region is estimated by the blue curve better than by the green curve, where the edge effects were neglected. The proper treatment of edge-effects gives even better results in the sampled regions close to the borders of the unsampled regions. This improvement can clearly be seen in fig. (11). **Poisson noise:** In the right plot, two sampling processes are underlying the yellow signal. First the Gaussian random field that generates the red signal, which is then Poisson sampled leading to the yellow data. Again, the blue and the green curves represent the reconstructions with and without proper window treatment, respectively.

quality improvement in the reconstructions is reached after $\text{tol}_{\text{crit}}^{j+1}$, as can be inferred from the correlation coefficients r and the normalized squared Euclidean distances $\mathcal{D}_{\text{Eucl}}^2(\psi, \delta_\rho)$.

4.2.2 Numerical performance with and without preconditioning

Here we analyze the convergence behaviour of the different inverse schemes with and without preconditioning. We start by considering a Gaussian random field with some structured noise that increases radially and is modulated by a random noise component. As a preconditioning expression, the diagonal part of the data covariance matrix is chosen, which is given by the sum of

$$\begin{aligned} \hat{R}SR^\dagger(\mathbf{k}, \mathbf{k}) &= \hat{f}_B(\mathbf{k}) \int \frac{d^{N_D} \mathbf{q}}{(2\pi)^{N_D}} \hat{f}_{SM}(\mathbf{k} - \mathbf{q}) P_S(\mathbf{q}) \hat{f}_{SM}(\mathbf{q} - \mathbf{k}) \overline{\hat{f}_B(\mathbf{k})} \\ &= P_B(\mathbf{k}) \underbrace{\int \frac{d^{N_D} \mathbf{q}}{(2\pi)^{N_D}} P_{SM}(\mathbf{k} - \mathbf{q}) P_S(\mathbf{q})}_{P_B \cdot [P_{SM} \circ P_S]}, \end{aligned} \quad (145)$$

and

$$\begin{aligned} \hat{N}(\mathbf{k}, \mathbf{k}) &= \int \frac{d^{N_D} \mathbf{q}}{(2\pi)^{N_D}} \hat{f}_{SF}(\mathbf{k} - \mathbf{q}) P_N(\mathbf{q}) \hat{f}_{SF}(\mathbf{q} - \mathbf{k}) \\ &= \underbrace{\int \frac{d^{N_D} \mathbf{q}}{(2\pi)^{N_D}} P_{SF}(\mathbf{k} - \mathbf{q}) P_N(\mathbf{q})}_{P_{SF} \circ P_N}, \end{aligned} \quad (146)$$

where we have used the following definitions: $P_B \equiv \|\hat{f}_B\|^2$, $P_{SM} \equiv \|\hat{f}_{SM}\|^2$ and $P_{SF} \equiv \|\hat{f}_{SF}\|^2$. We can thus calculate the preconditioning matrix M required for the different schemes (section 3) by just inverting each diagonal component. The results summarized in fig. (5) show important differences between the reconstructions done with (on the left side) and without (on the right side) preconditioning. Some of the methods just speed up, like the various EXP methods or the SD scheme. Others, however, are stabilized and manage to converge to the solution only after preconditioning, like the J, the FR and the CPR methods. Without preconditioning, the latter converges extremely quickly to a wrong solution. This is due to the fact that we did not impose the following stabilization: $\beta_{PR} = \max(\beta_{PR}, 0)$ in this calculation (see Shewchuk 1994, for a discussion). However, our tests show that upon imposing this stabilization the PR-method becomes significantly slower than the rest. On the other hand, the EXP-Krylov methods behave most stably and converge very quickly. In the preconditioned case, we see that all methods converge to the same statistical result, as we can infer from the correlation coefficient r and $\mathcal{D}_{\text{Eucl}}^2(\psi, \delta_\rho)$, except for the PR scheme that yields slightly less optimal results (see the green line in comparison to the rest in panel c). We have tested preconditioning in the rest of the examples and could confirm the results presented in this section. Preconditioning turns out to be necessary to achieve fast algorithms.

4.2.3 Poissonian distribution

In this study case, we investigate the reconstruction of a Gaussian field based on a Poissonian distribution. This model is far from reality, where much more complex processes are known to occur (see

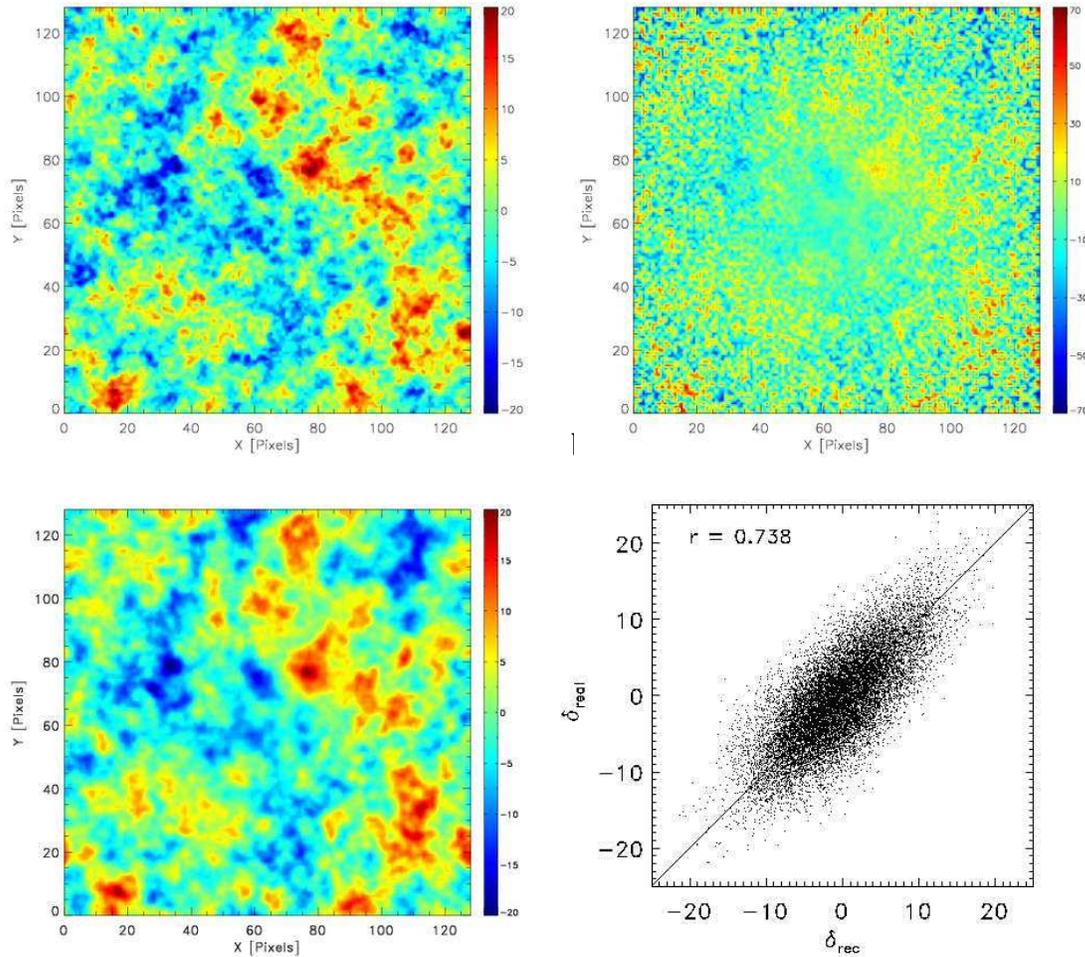


Figure 4. Structured noise treatment: The upper left picture shows the real signal. The upper right picture is the input signal, where some random noise that increases radially was added. Note that the scale of the colourbar changes from a maximum overdensity of 20 to 70. The lower left picture **c** shows the reconstruction. The reconstructions using different numerical methods implemented in ATLAS are indistinguishable. In the lower right image **d**, the real density field is plotted against the reconstructed density field pixel by pixel without any smoothing. The numerical performance of this reconstruction case is shown in the next figure.

discussion in section 2.1). However, we can model a non-Gaussian process in this way and test how good the Wiener-filter reconstruction works under such circumstances. Here the assumed data model does not coincide with the one that has generated the data. However, the Poissonian noise can be modeled in the noise matrix of the Wiener-filtering through the structure function f_S .

The results presented in fig. (6) show very good agreement between the reconstruction and the *real* underlying density field (compare panels **a** and **c**). The convergence behaviour and statistical goodness is plotted in the left side of fig. (7), panels **a**, **c** and **e**. There we can see that the FR and PR methods do not converge rapidly (see yellow and green curves in panel **a**). On the contrary, the J, SD, and EXP schemes are very efficient (panel **c**) and lead to very similar results (panels **c** and **e**).

4.2.4 Blurring effects: deconvolution

In this numerical experiment we tested the blurring effects by convolving the density field with a Gaussian. The result is shown in fig. (8), panel **b**. We see how the small structures are smoothed out and only the larger ones prevail. Some noise with a structure function was added to the signal. However, the noise was kept low with

the aim of investigating primarily the blurring effect. The results of the reconstruction that considers only the noise does not change much with respect to the input signal, as can be expected. However, the extra-regularized Wiener-filtering deblurs the image applying eqs. (120) and (121), and yields the figure shown in panel **c**. We see how much of the small scale structure is restored and the peaks become enhanced. The correlation between this reconstruction and the original signal (panel **e**) is significantly better than for the case where the blurring is ignored (panel **f**). We can see in fig. (7) that the deconvolution algorithm is very fast for all the methods except for the FR-scheme. The PR-method is the fastest, but it leads to slightly worse results (see the green curve in panels **c** and **e**). The EXP turns out to be more efficient than the J and SD methods in this case.

4.2.5 Selection function effects

For this case we use a modified data model in which the selection function also affects the noise

$$d = f_S \cdot (s + f_{SF} \cdot \epsilon_{WN}), \quad (147)$$

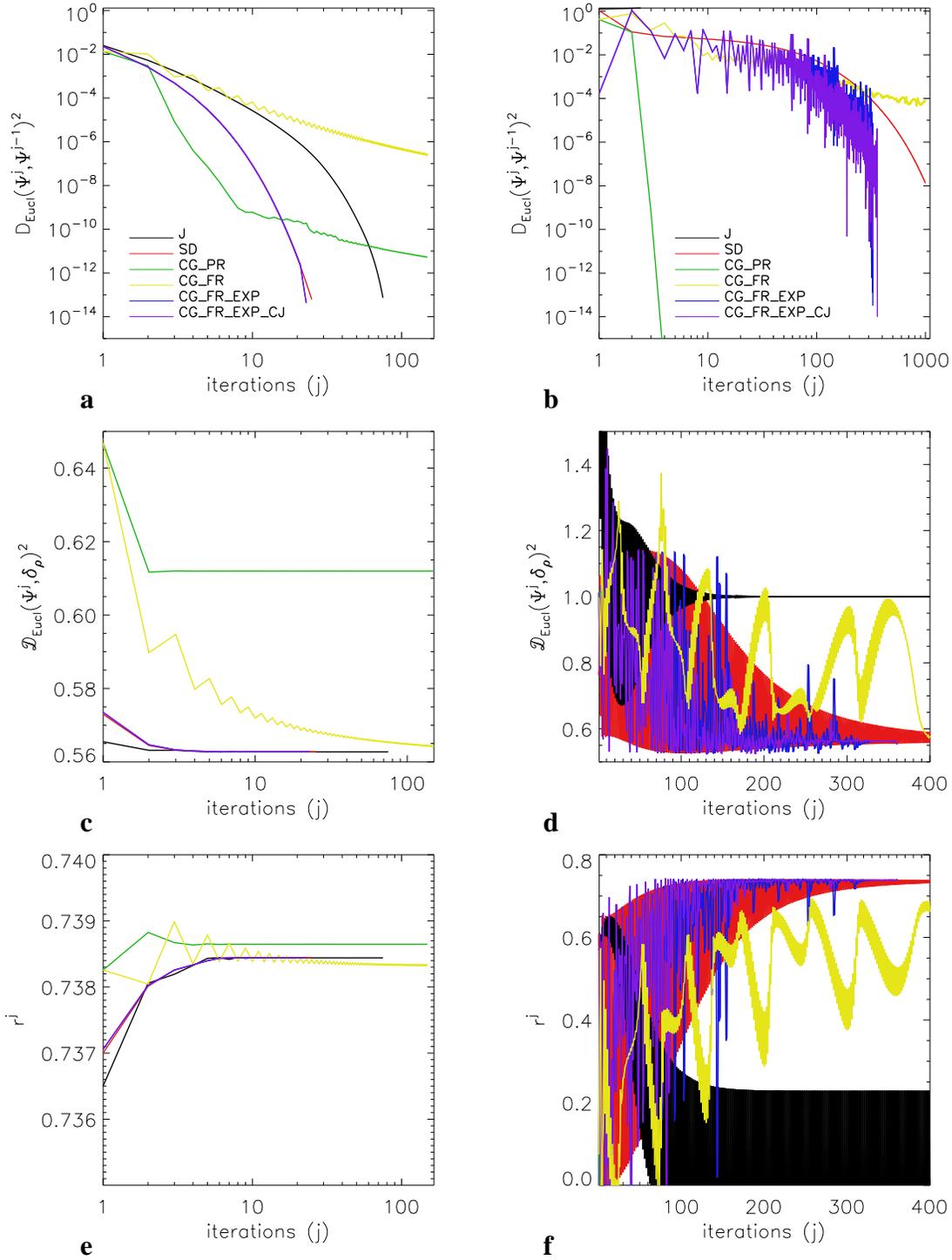


Figure 5. Numerical performance with and without preconditioning: Here the convergence behaviour and the goodness of the reconstructions using different inversion algorithms can be seen. The pictures on the left show the methods using preconditioning, whereas the pictures on the right do not use preconditioning. The upper plots show the squared Euclidean distance between successive reconstructions. The plots in the middle show the normalized Euclidean distance between the different reconstructions and the true signal. The lower plots show the evolution of the statistical correlation coefficient between reconstruction and signal. We see from panel c and panel e that after less than 10 iterations the reconstructions do not significantly improve with most of the inversion algorithms. The different inversion algorithms used are: Jacobi (J), Steepest Descent (SD), Conjugate Gradients (CG), Fletcher Reeves (FR), and Polak Riviere (PR). We also tested a more expensive variant that uses one additional operation of the involved matrix (EXP) and one other variant (CJ), where a degree of freedom in the mapping equation for the Wiener-filter is used.

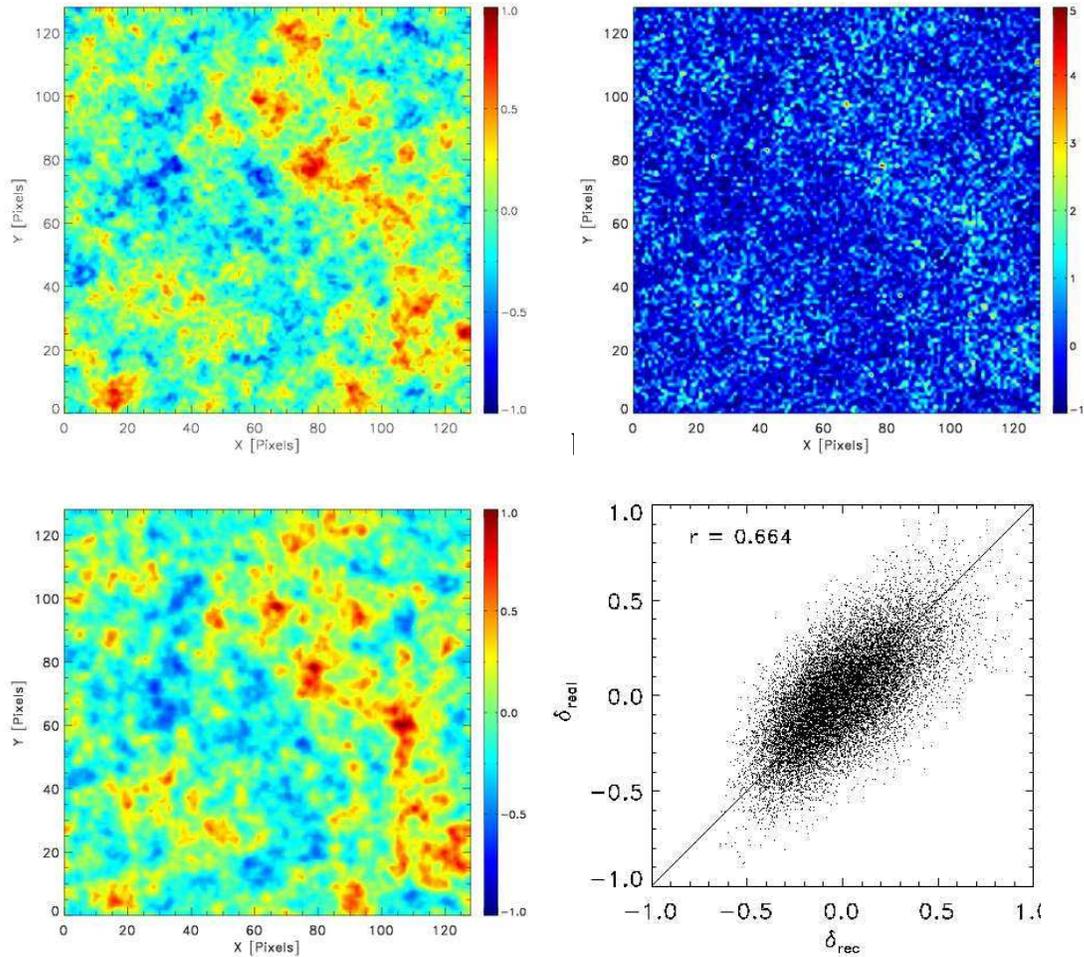


Figure 6. Poissonian noise: Here two stochastic processes are underlying the input signal. First the Gaussian random field that generates the signal in panel **a**, which is then Poisson sampled leading to the signal in panel **b**. The reconstruction in panel **c** is shown to be in good agreement with the underlying signal. The pixel values are correctly distributed as can be seen in panel **d**.

with $f_S \in [0, 1]$, simulating the fading strength of the signal with increasing distance. The results are plotted in fig. (9), where the structure of the signal can be seen to become undistinguishable in radial direction (see panel **b**). Taking only the noise into account leads to very poor reconstructions (see panel **d**). On the contrary, by also considering the selection function effects, the structures are resolved even at contours where only 10 % of the signal plus noise is left (see panel **c**). As can be appreciated in panels **e** and **f** there is an improvement in the correlation between the *real* density field and the reconstructed signal. Panel **e** shows a higher correlation coefficient, but the quality enhancement of the reconstruction can be seen better in the distribution of the density values for each pixel. How the points are correctly spread along the diagonal line can be verified there. The longer Euclidean distance to the *real* density field shows the quantitative difference very clearly, by just comparing the pink curve with the rest (fig. 10 and panel **c**). It is worth mentioning that although the PR test seems to give a comparable result to the calculation that ignores the selection function. The final correlation coefficient in panel **e** shows that the reconstructions actually strongly differ and panel **c** shows that the quality of the recovered signal is notably better for the former experiment.

In addition, we tested the same selection function affecting

only the underlying signal with a model given by

$$d = f_S \cdot s + f_{SF} \cdot \epsilon_{WN}, \quad (148)$$

and obtained the same qualitative results.

4.2.6 Windowing effects

In this section we investigate the mask effects that introduce coupling between different modes in Fourier-space so that the data covariance matrix is no longer diagonal. The input signal is given in panel **b** of fig. (11). The noisy signal from panel **b** in fig. (4) was cut in stripes, to simulate *observed* regions. We compare two reconstructions here, the first one ignores windowing effects given in panel **d** and a second reconstruction employs the proper treatment of the boundary through f_M in the algorithm (see eqs. 120 and 121). The statistical correlation is given in panels **e** and **f**, respectively. Our experiments not only show better results for the latter reconstruction in the unsampled region ($\bar{\Omega}$), represented by the red dots in panels **e** and **f** in fig. (11), but also in the sampled regions (Ω). The global correlation r is significantly improved. Whereas the distribution of the black dots, the values of the densities in the *observed* regions, does not apparently change, the distribution of the unsampled red dots clearly does. These are distributed around

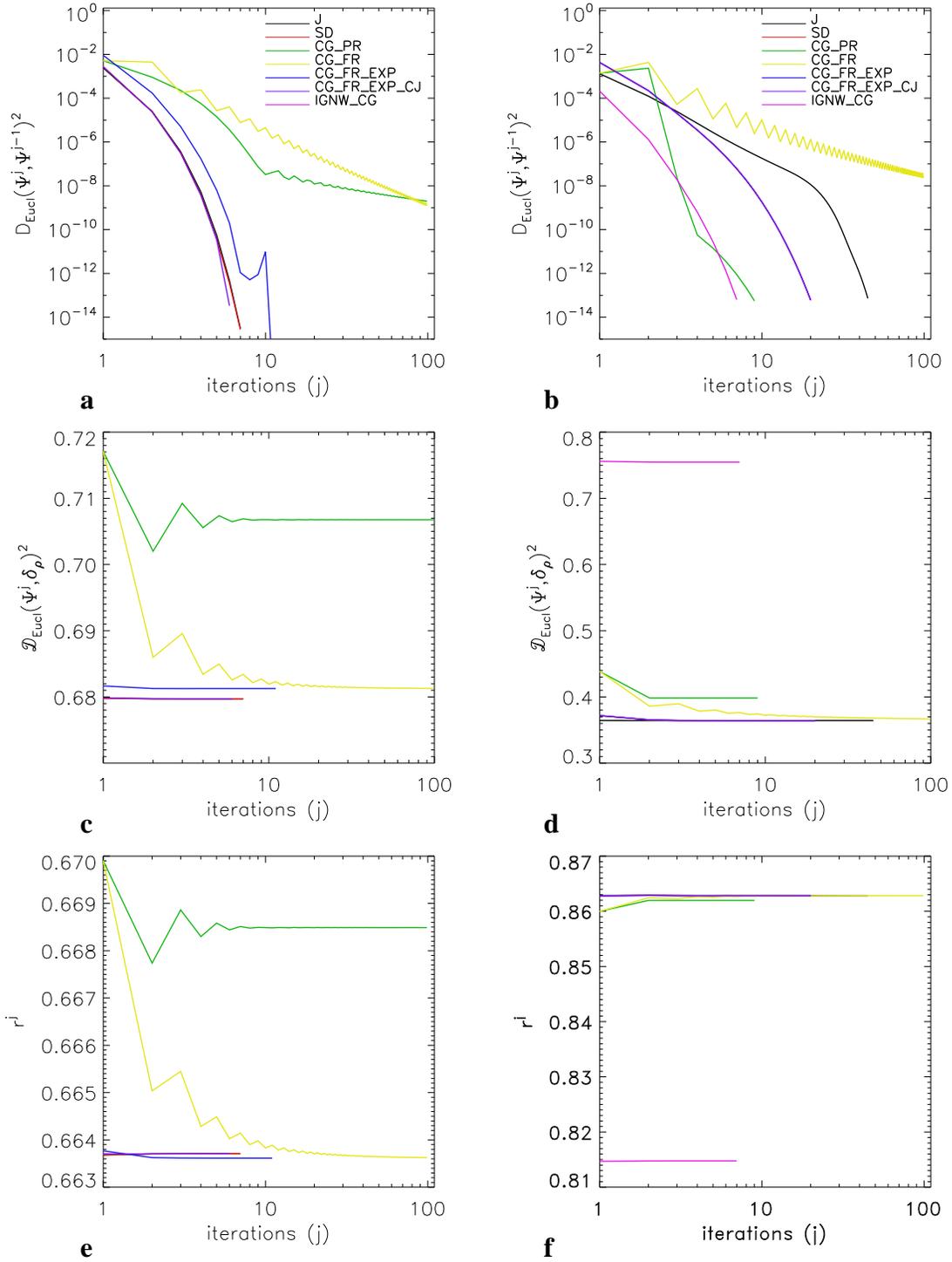


Figure 7. Poissonian noise and numerical performance (panels a, c, e): Here the convergence behaviour and quality of the reconstruction is comparable for the J, SD, EXP methods. The FR and PR schemes do not present a fast convergence (panel a). Nevertheless, the FR scheme (yellow curve) seems to lead to the correct solution (panels c and e). The PR formula, on the contrary, stagnates at reconstructions that have much lower quality compared to the rest of the schemes. **Blurring treatment and numerical performance (panels b, d, f):** In this study case, the EXP algorithm seems to work better than the rest of the schemes. Although the PR formula converges very rapidly (green curve in panel b), it leads to a lower quality reconstruction (panels d and f). The FR scheme converges to the same solution as the J, SD, and EXP algorithms, however, with a slower convergence (yellow curve in panel b). The J and SD methods have an overall good behaviour in this case, but still converge significantly slower than the EXP scheme (their convergence is identical black and red curves are overlotted). The reconstruction considering just the noise is very poor, because the noise is negligible in this case (pink curves).

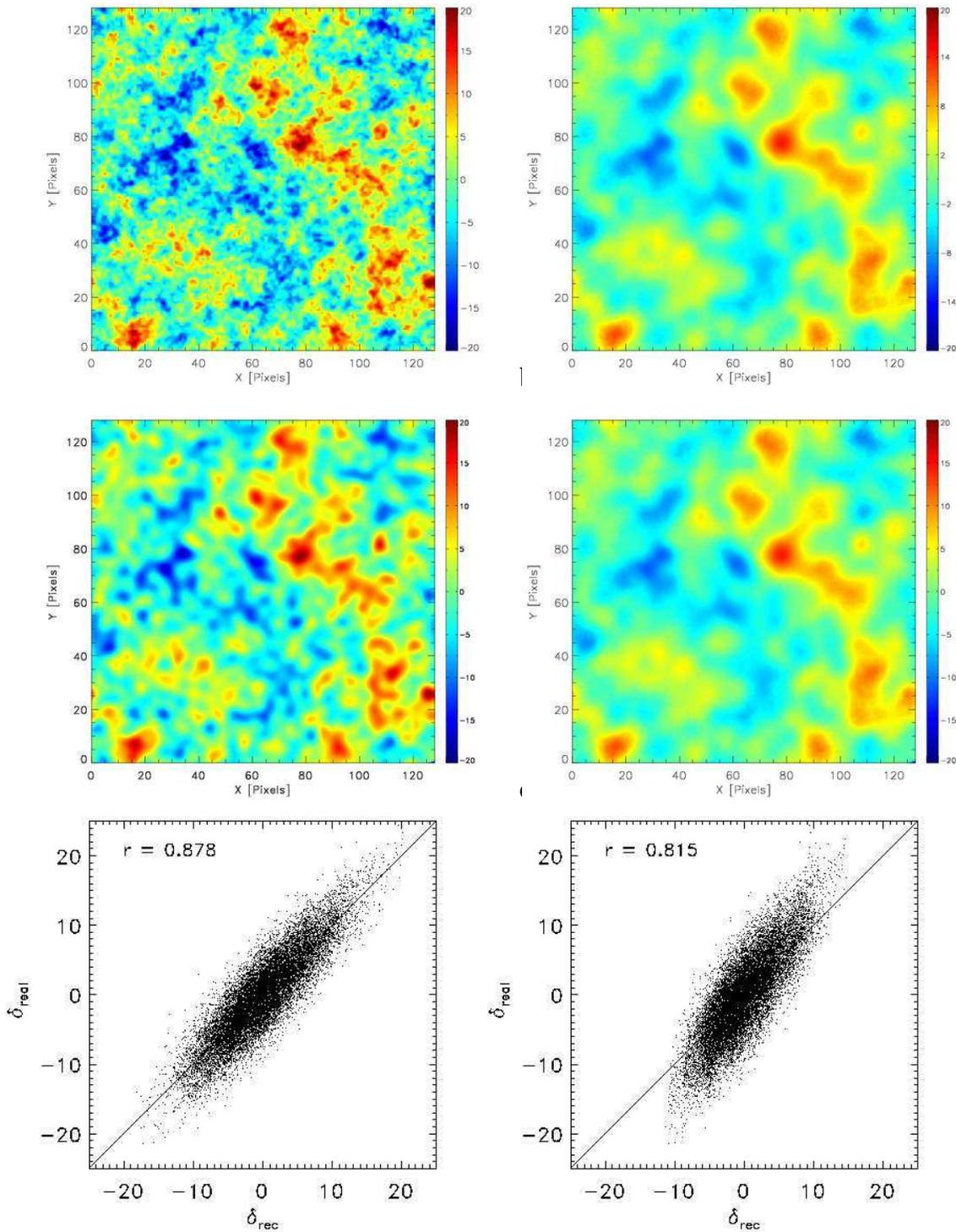


Figure 8. Blurring treatment: Here the signal (panel a) was convolved with a gaussian modeling blurring effects, as shown in panel b. Some low noise with a structure function was added. Panel c shows the deblurred result. Panel d takes only the noise into account. We see in panel f the correlation between the input signal and the *true* signal, because the noise is negligible. The correlation coefficient is thus very high, however, the alignment of the pixels in the plot is not correct. Overdensities and underdensities tend to be underestimated, which is consistent with the blurring effect. The reconstruction given in panel e corrects this effect and consequently a higher correlation coefficient is achieved.

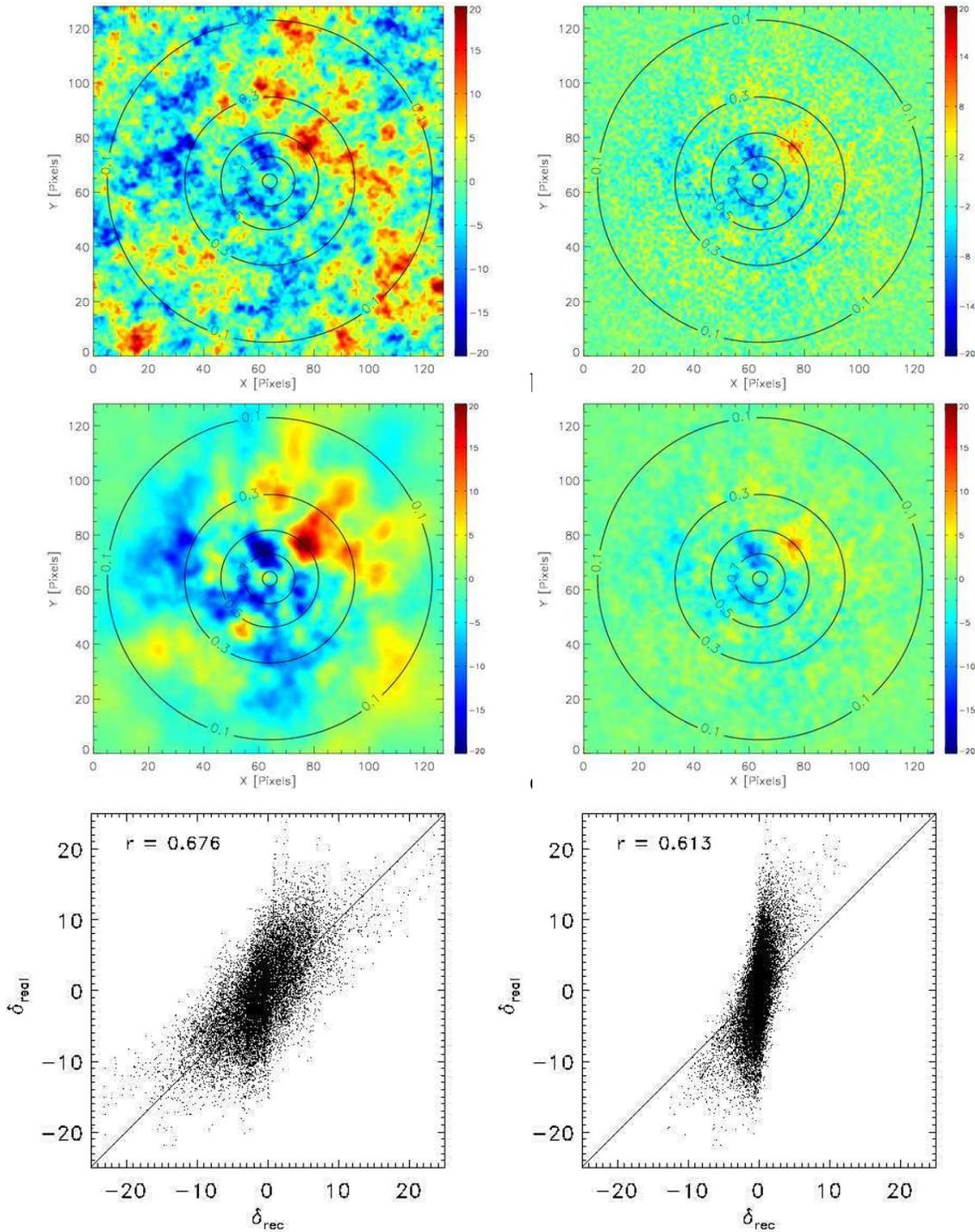


Figure 9. Selection function treatment: Here selection function effects were simulated with a function that takes values between zero and one, decreasing exponentially in radial direction. The contours show different values of this function. Panel **a** shows the real density field. Panel **b** shows the input data, where the true signal was multiplied in real space with the selection function and a radially increasing noise was added. The reconstruction and its correlation with the true signal are represented in panel **c** and **e**, respectively. The reconstruction ignoring selection effects by taking only the noise into account leads to panels **d** and **f**. The reconstruction given in panel **d** is very conservative and smooths the overdensities out due to noise suppression. This leads to a high correlation coefficient, though the individual pixels are clearly not correctly aligned (panel **f**). Panel **e**, on the contrary, shows more structures that are enhanced due to consideration of the selection function effects. This correctly distributes the pixels, as can be seen in panel **e**. The correlation coefficient seems to be significantly better than in panel **f**, however, a better measure of the overall quality of the reconstruction can be seen in next figure.

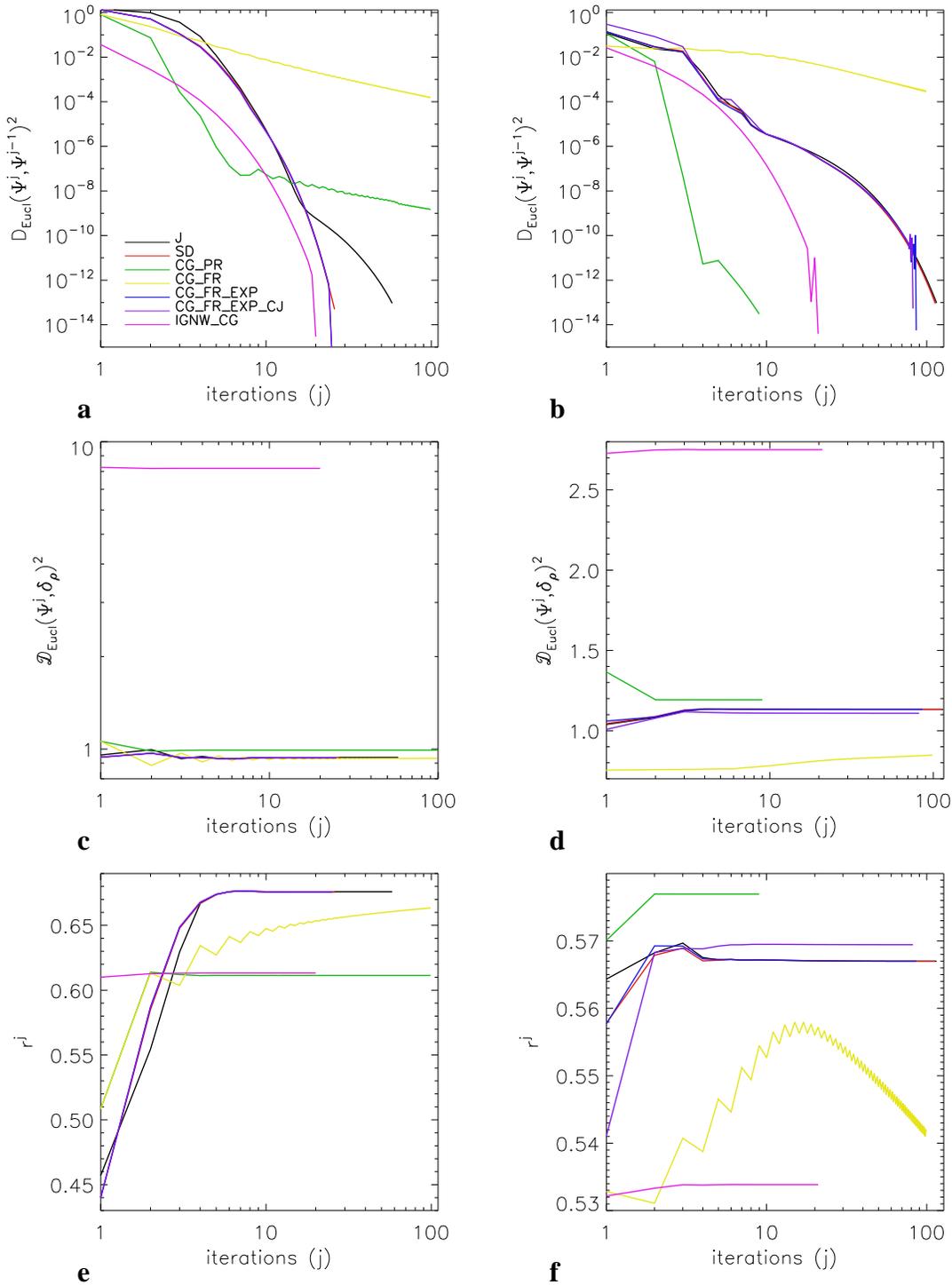


Figure 10. Selection function treatment and numerical performance (panels a, c, e): The same color coding is used as in fig. (5) panel a, except for additional curve (represented in pink) that indicates the reconstruction in which the selection effects are ignored. Panel a shows the squared Euclidean distance between subsequent reconstructions. The squared Euclidean distance between the reconstruction and the true density field is plotted in panel c, showing a huge difference between the reconstruction which takes only the noise into account and ignores the selection function and the rest of the methods. Note that the statistical correlation r is also much better for the case where the selection effects are properly treated (panel e). One concludes from the three plots, that the SD and EXP methods (red, blue and violet curves) clearly converge faster to a more or equally optimal solution in comparison with the rest of the methods. The J scheme shows a significantly slower convergence (black curve in panel a). The PR algorithm stagnates at poorer reconstructions as can be seen from panel c and e. **Windowing treatment and numerical performance (panels b, d, f):** In this case, the PR shows extremely good results: fast convergence (panel b) and a high correlation coefficient (panel d). However, the Euclidean distance is slightly bigger than for the rest of the methods, except for the pink curve (ignoring windowing effects). The FR method is disastrous in this study case and diverges from the solution as can be seen in panel f. The J, SD, and EXP methods show good and stable results. The J and SD algorithms give extremely similar results. Although their convergence behaviour is similar to the EXP schemes, the latter give slightly better results: smaller values for the Euclidean distance and higher values for the correlation coefficient (violet curves in panels d and f, respectively).

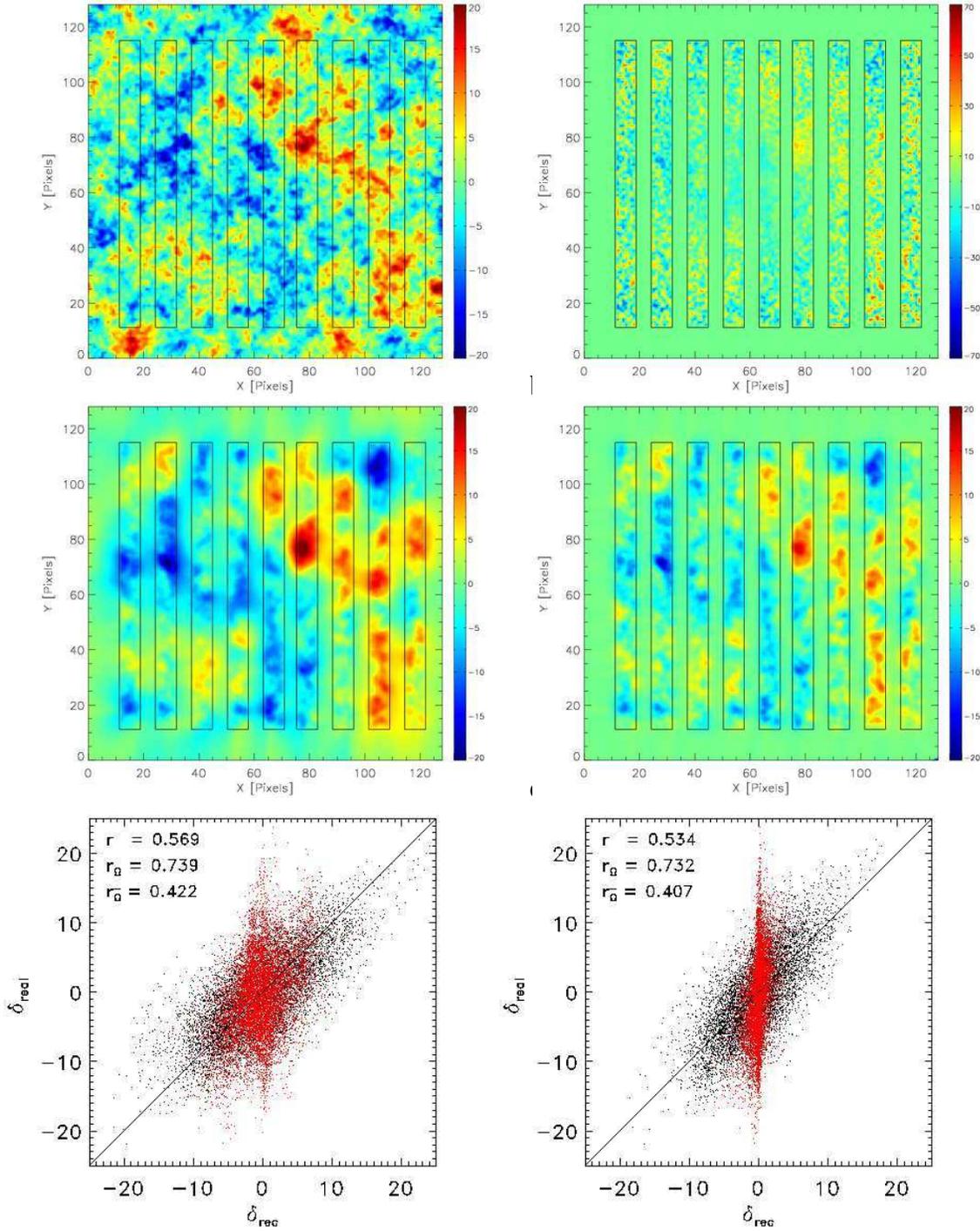


Figure 11. Windowing treatment: Here the edge effects are shown in two dimensions. The true signal was multiplied by a windowing function that is one in the observed region (Ω) and zero in the unknown region ($\bar{\Omega}$). The sampled regions are given by the vertical stripes. In addition, a radially increasing noise was added (see panel **b**). Panel **c** shows the reconstruction handling the edge effects. Panel **d** represents the result taking only the noise into account. We see in panel **c** how the information is propagated into the unsampled regions leading to a closer resemblance of the real signal, whereas the noise is just suppressed in panel **d**. Panels **e** and **f** show the correlation coefficients for the whole reconstructed region, split into the sampled (black dots) and the unsampled regions (red dots). Note that the red dots are strongly aligned around the zero value in panel **f**, whereas they are correctly spread in panel **e**, statistically representing the information propagation process mentioned above.

the zero value for the case where windowing is ignored, because a zero signal is assumed by ARGO in the $\bar{\Omega}$ region. In contrast we see that the red dots are distributed along the diagonal line when edge effects are considered. This is equivalent to a propagation of the information to the unsampled regions or the appropriate interpolation and extrapolation of signals. Looking at the numerical performance in fig. (7) reveals that most of the methods behave very similarly, except for the PR and FR schemes that deviate from the rest. The former converges rapidly to a good solution that has a higher correlation (see green curve in panel f), but a slightly worse Euclidean distance to the *true* signal. The FR on the other hand converges extremely slowly. The correlation coefficient is at a stage where it becomes dramatically worse (see yellow curve in panel f). The smaller Euclidean distance is no measure for the quality in this case, because these low values can be achieved when the reconstruction is very conservative (closer to zero) and has no structure. Notice how many schemes start with better values for that distance measure (see panel d). The EXP methods converge faster and the CJ version leads to even slightly better results (see violet curve in panels d and f).

It is also worth mentioning that the best reconstructions in terms of high correlation coefficients and low Euclidean distances to the underlying signal are achieved only after three iterations for the J, SD, and EXP methods, prior to numerical convergence. We furthermore tested ARGO under extreme noise conditions in which the inversion diverges and produces density values that approach infinity. At early iterations, extremely good reconstructions were produced. These examples underline the regularization character of the inversion schemes under consideration in this publication. However, for the cases we are interested in, where the noise is mainly determined by the discrete sampling of galaxies, no additional stopping rules are required and the inversion algorithms can be run until full convergence.

5 SUMMARY AND CONCLUSIONS

The goal of this work is to exploit the Bayesian formalism to develop methods that reconstruct the underlying dark-matter distribution from the discrete sample of galaxies and their three-dimensional positions provided by galaxy redshift surveys. Such a general Bayesian analysis permits one to innovate methods and push this field forward to develop more accurate reconstruction algorithms.

We show how a series of uncertainties demand a statistical approach (see figure 1 and section 1.1). Some of the uncertainties are intrinsic to the nature of the underlying signal (the dark matter) and have a stochastic character, the cosmic variance. Other uncertainties are intrinsic to the nature of the observable (the galaxies) and lead to a kind of shot noise, galaxy-bias and redshift-distortions. Additional uncertainties, such as windowing, selection function effects and blurring effects, arise due to the observation process. The degeneracies that are produced by such uncertainties require regularization techniques, which should converge to optimal solutions. We discuss the different Bayesian approaches specified through different options for the likelihood and the prior, and see how *natural* regularizations can be performed by the prior-choice (see section 2.5). Moreover, we see how the definition of particular likelihoods and priors define classes of algorithms, each specific to a different problem approach (see table 1).

We develop new algorithms in this Bayesian framework which account for the discrete nature of a galaxy distribution by taking a

Poissonian likelihood. This is done for the case of a Gaussian prior leading to the GAPMAP estimator (see section 2.5.4 and appendix A) and for the case of an entropic prior (see section 2.5.9 and appendix E). The Maximum Entropy method is studied in detail as a non-informative prior, which does not assume a particular pattern for the underlying signal. This can be interesting when searching for intrinsic deviations from Gaussianity (see section 2.5.9 and references therein).

We extend the Wiener-filter (see section 2.5.3 and appendix I) and propose novel algorithms to do a joint estimation of the density field, its power-spectrum, and the peculiar velocities of the galaxies (see section 2.6). We also address the possibility of extending such work to determine cosmological parameters and the bias between galaxies and dark matter.

Such an aim requires a large number of repeated reconstructions, which can be only achieved with highly efficient inverse algorithms. We develop here the necessary numerical schemes in a preconditioned way for linear and non-linear inverse problems (see section 3 and appendix H & G). Such iterative schemes acquire their real power only in an operator formalism, which we derive in detail for different Bayesian methods (see section 3.3). A novel Krylov formula (see section 3.1.5 and appendix H) turns out to be superior in terms of performance and fidelity, as we show in section (4).

The novel ARGO-software package is presented in this paper. Different inverse schemes are tested with the Wiener-filter implemented in ARGO under several conditions determined by structured noise, blurring, selection function effects and windowing (see section 4).

We conclude that fast three-dimensional reconstructions of the large-scale structure scaling as $n \log_2 n$ (with n being the total number of grid cells) can be done with hybrid Wiener-Krylov iterative schemes under an operator formalism, which takes advantage of the speed of FFTs. This opens new horizons of possibilities, such as joint parameter and signal estimation, in the field of large-scale structure reconstruction.

It is our goal to apply such techniques to reconstruct the underlying density field, the power-spectrum and the peculiar velocities from galaxy surveys. Still, different problems, such as galaxy-bias studies, have to be further analysed. However, we are confident that such issues can be tackled from an information-theory approach.

ACKNOWLEDGEMENTS

We acknowledge the valuable and encouraging discussions to Benjamin Wandelt, Simon D. M. White, Jeremy Blaizot and Hans-Martin Adorf. We want to thank especially Jens Jasche, Pilar Esquej and Sheridan Sauer for the careful proof-reading of this work. We are also very grateful to Martin Reinecke and Daniel Sauer for their computational advice. F. Kitaura thanks the International Max Planck Research School for supporting this project, the Faculty for Statistics at Penn State University for the statistics course in 2005, and especially Calyampudi R. Rao and Tom Loredo for very enlightening lectures and conversations.

REFERENCES

- Albrecht A., Steinhardt P. J., 1982, Physical Review Letters, 48, 1220
- Aster R. A., Brochers B., Thurber C. H., 2005, Parameter estimation and inverse problems. Elsevier Academic Press, London
- Bakushinskii A. B., 1992, Compt.Math.Phys., 32, 1353

- Ballinger W. E., Peacock J. A., Heavens A. F., 1996, *MNRAS*, 282, 877
- Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, *ApJ*, 304, 15
- Bardeen J. M., Steinhardt P. J., Turner M. S., 1983, *Phys. Rev. D*, 28, 679
- Berger J. O., Bernardo J. M., 1992, in Berger J. O., Bernardo J. M., Dawid A. P., Smith A. F. M., eds, *Bayesian Statistics 4 On the development of the reference prior method*. Oxford University Press, Oxford, pp 35–60
- Bernardeau F., van de Weygaert R., 1996, *MNRAS*, 279, 693
- Bertschinger E., Dekel A., 1989, *ApJ*, 336, L5
- Bertschinger E., Dekel A., 1991, in Latham D. W., da Costa L. A. N., eds, *ASP Conf. Ser. 15: Large-Scale Structures and Peculiar Motions in the Universe Mapping Large-Scale Flows in Three Dimensions: Method*. pp 67–
- Bertschinger E., Dekel A., Faber S. M., Dressler A., Burstein D., 1990, *ApJ*, 364, 370
- Bistolas V., Hoffman Y., 1998, *ApJ*, 492, 439
- Blaschke B., Neubauer A., Scherzer O., 1997, *IMA Journal of Numerical Analysis*, 17, 421
- Botzler C. S., Snigula J., Bender R., Hopp U., 2004, *MNRAS*, 349, 425
- Branchini E., Plionis M., 1996, *ApJ*, 460, 569
- Branchini E., Plionis M., Sciamia D. W., 1996, *ApJ*, 461, L17+
- Brenier Y., Frisch U., Hénon M., Loeper G., Matarrese S., Mohayaee R., Sobolevskii A., 2003, *MNRAS*, 346, 501
- Bunn E. F., Fisher K. B., Hoffman Y., Lahav O., Silk J., Zaroubi S., 1994, *ApJ*, 432, L75
- Bunn E. F., Sugiyama N., 1995, *ApJ*, 446, 49
- Carasso A. S., 1999, *SIAM Journal on Numerical Analysis*, 36, 1657
- Cooray A., Sheth R., 2002, *Physics Reports*, 372, 1
- Croft R. A. C., Gaztanaga E., 1997, *MNRAS*, 285, 793
- Dekel A., Bertschinger E., Faber S. M., 1990, *ApJ*, 364, 349
- Dekel A., Lahav O., 1999, *ApJ*, 520, 24
- Doré O., Teyssier R., Bouchet F. R., Vibert D., Prunet S., 2001, *Astr.Astrophys.*, 374, 358
- Doroshkevich A. G., Gottlober S., Madsen S., 1997, *Astr.Astrophys.Suppl.*, 123, 495
- Ebeling H., Wiedenmann G., 1993, *Phys. Rev. E*, 47, 704
- Efstathiou G., Bond J. R., White S. D. M., 1992, *MNRAS*, 258, 1P
- Eisenstein D. J., Hu W., 1999, *ApJ*, 511, 5
- Eke V., 2001, *MNRAS*, 324, 108
- Erdogdu P., Lahav O., Huchra J., et al. 2006, *MNRAS*, 373, 45
- Erdogdu P., Lahav O., Zaroubi S., et al. 2004, *MNRAS*, 352, 939
- Eriksen H. K., O'Dwyer I. J., Jewell J. B., Wandelt B. D., Larson D. L., Górski K. M., Levin S., Banday A. J., Lilje P. B., 2004, *Rev.Astrn.Astrophys.*, 155, 227
- Fisher K. B., Lahav O., Hoffman Y., Lynden-Bell D., Zaroubi S., 1995, *MNRAS*, 272, 885
- Fisher K. B., Scharf C. A., Lahav O., 1994, *MNRAS*, 266, 219
- Fletcher R., Reeves C. M., 1964, *Computer Journal*, 7, 163
- Foster M., 1961, *J. SIAM*, 9, 387
- Frieden B. R., 1972, *Journal of the Optical Society of America (1917-1983)*, 62, 511
- Frisch U., Matarrese S., Mohayaee R., Sobolevskii A., 2002, *Nature*, 417, 260
- Ganon G., Hoffman Y., 1993, *ApJ*, 415, L5
- Gelman A., Carlin J. B., Stern H. S., Rubin D., 2004, *Bayesian data analysis*. Chapman & Hall/CRC, Boca Raton, Florida
- Geman S., Geman D., 1984, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721
- Goldberg D. M., 2001a, *ApJ*, 552, 413
- Goldberg D. M., 2001b, *ApJ*, 550, 87
- Goldberg D. M., Spergel D. N., 2000a, *ApJ*, 544, 21
- Goldberg D. M., Spergel D. N., 2000b, in Courteau S., Willick J., eds, *ASP Conf. Ser. 201: Cosmic Flows Workshop Using Perturbative Least Action to Run N-Body Simulations Back in Time*. pp 282–
- Gordon R., 1974, *IEEE Trans NS*, 21, 78
- Gottlöber S., Kerscher M., Kravtsov A. V., Faltenbacher A., Klypin A., Müller V., 2002, *Astr.Astrophys.*, 387, 778
- Gramann M., 1993, *ApJ*, 405, 449
- Gull S. F., 1989, in Skilling J., ed., *Maximum Entropy and Bayesian Methods Developments in Maximum Entropy data analysis*. Kluwer Academic Publishers, Dordrecht, pp 53–71
- Gull S. F., Daniell G. J., 1978, *Nature*, 272, 686
- Gull S. F., Skilling J., 1990, *The MEMSYS5 User's Manual*. Maximum Entropy Data Consultants Ltd, Royston
- Guth A. H., 1981, *Phys. Rev. D*, 23, 347
- Guth A. H., Pi S.-Y., 1982, *Physical Review Letters*, 49, 1110
- Hamilton A. J. S., 1998, *Linear Redshift Distortions: A Review*
- Hanke M., 1997, *Inverse Problems*, 13, 75
- Hanke M., Neubauer A., Scherzer O., 1995, *Numerische Mathematik*, 72, 21
- Hastings W. K., 1970, *Biometrika*, 57, 97
- Hawking S. W., 1982, *Communications in Mathematical Physics*, 87, 395
- Hestenes M. R., Stiefel E., 1952, *Journal of Research of the National Bureau of Standards*, 49, 6
- Hobson M. P., Jones A. W., Lasenby A. N., Bouchet F. R., 1998, *MNRAS*, 300, 1
- Hobson M. P., McLachlan C., 2003, *MNRAS*, 338, 765
- Hoerl A. E., 1962, *Chemical Engineering Progress*, 58, 54
- Hoerl A. E., Kennard R. W., 1970, *Technometrics*, 12, 55
- Hoffman Y., 1994, in Balkowski C., Kraan-Korteweg R. C., eds, *ASP Conf. Ser. 67: Unveiling Large-Scale Structures Behind the Milky Way Wiener Reconstruction of the Large-Scale Structure in the Zone of Avoidance*. pp 185–
- Icke V., van de Weygaert R., 1991, *QJRAS*, 32, 85
- Janssen M. A., Gulkis S., 1992, in Signore M., Dupraz C., eds, *NATO ASIC Proc. 359: The Infrared and Submillimetre Sky after COBE Mapping the sky with the COBE differential microwave radiometers*. pp 391–408
- Jaynes E. T., 1963, in Ford K., ed., *Statistical Physics Information Theory and Statistical Mechanics*. Benjamin, New York, p. 181
- Jaynes E. T., 1968, *IEEE Trans. on Systems Science and Cybernetics*, SSC-4, 227
- Jewell J., Levin S., Anderson C. H., 2004, *ApJ*, 609, 1
- Kaiser N., 1987, *MNRAS*, 227, 1
- Kaiser N., Stebbins A., 1991, in Latham D. W., da Costa L. A. N., eds, *ASP Conf. Ser. 15: Large-Scale Structures and Peculiar Motions in the Universe Reconstructing the Density Field from Peculiar Velocities: A Maximum Probability Technique*. pp 111–
- Keihänen E., Kurki-Suonio H., Poutanen T., 2005, *MNRAS*, 360, 390
- Kibble T. W. B., 1976, *Journal of Physics A Mathematical General*, 9, 1387
- Kim R., Strauss M., Bahcall N., Gunn J. E., Lupton R. H., Vogeley M. S., Schlegel D., the SDSS Collaboration 2000, in Mazure A., Le Fèvre O., Le Brun V., eds, *Clustering at High Redshift Vol. 200 of Astronomical Society of the Pacific Conference Series, Finding Clusters of Galaxies in the Sloan Digital Sky Survey using Voronoi Tessellation*. pp 422–
- Komatsu E., Kogut A., Nolte M. R., Bennett C. L., Halpern M., Hinshaw G., Jarosik N., Limon M., Meyer S. S., Page L., Spergel D. N., Tucker G. S., Verde L., Wollack E., Wright E. L., 2003, *Rev.Astrn.Astrophys.*, 148, 119
- Kudlicki A., Chodorowski M., Plewa T., Różycka M., 2000, *MNRAS*, 316, 464
- Lahav O., 1994, in Balkowski C., Kraan-Korteweg R. C., eds, *ASP Conf. Ser. 67: Unveiling Large-Scale Structures Behind the Milky Way Wiener Reconstruction of All-Sky Spherical Harmonic Maps of the Large-Scale Structure*. pp 171–
- Lahav O., Fisher K. B., Hoffman Y., Scharf C. A., Zaroubi S., 1994, *ApJ*, 423, L93+
- Lahav O., Gull S. F., 1989, *MNRAS*, 240, 753
- Larson D. L., Eriksen H. K., Wandelt B. D., Górski K. M., Huey G., Jewell J. B., O'Dwyer I. J., 2007, *ApJ*, 656, 653
- Linde A. D., 1982, *Physics Letters B*, 108, 389
- Lucy L. B., 1974, *AJ*, 79, 745
- Maisinger K., Hobson M. P., Lasenby A. N., 1997, *MNRAS*, 290, 313
- Marchuk G. I., 1982, *Methods of numerical mathematics*. Springer-Verlag, New York
- Martel H., 2005, *ArXiv Astrophysics e-prints*

- Mathis H., Lemson G., Springel V., Kauffmann G., White S. D. M., Eldar A., Dekel A., 2002, *MNRAS*, 333, 739
- Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., Teller E. T., 1953, *Journal of Chemical Physics*, 21, 1087
- Meurs E. J. A., Wilkinson M. I., 1999, in Giuricin G., Mezzetti M., Salucci P., eds, *Observational Cosmology: The Development of Galaxy Systems Vol. 176 of Astronomical Society of the Pacific Conference Series, Large-scale structure assessed with Voronoi techniques*. pp 333–+
- Mohayaee R., Frisch U., Matarrese S., Sobolevskii A., 2003, *Astr.Astrophys.*, 406, 393
- Mohayaee R., Mathis H., Colombi S., Silk J., 2006, *MNRAS*, 365, 939
- Mohayaee R., Tully B., Frisch U., 2004, *ArXiv Astrophysics e-prints*
- Mohayaee R., Tully R. B., 2005, *ApJ*, 635, L113
- Molina R., Nunez J., Cortijo F. J., Mateos J., 2001, *Signal Processing Magazine, IEEE*, 18, 11
- Narayan R., Nityananda R., 1986, *Ann.Rev.Astrn.Astrophys.*, 24, 127
- Narayanan V. K., Croft R. A. C., 1999, *ApJ*, 515, 471
- Narayanan V. K., Weinberg D. H., 1998, *ApJ*, 508, 440
- Natoli P., de Gasperis G., Gheller C., Vittorio N., 2001, *Astr.Astrophys.*, 372, 346
- Neal R. M., 1993, in *Technical Report CRG-TR-93-1 Probabilistic Inference Using Markov Chain Monte Carlo Methods*. Dept. of Computer Science, University of Toronto
- Nusser A., Davis M., 1994, *ApJ*, 421, L1
- Nusser A., Dekel A., 1992, *ApJ*, 391, 443
- Nusser A., Dekel A., Bertschinger E., Blumenthal G. R., 1991, *ApJ*, 379, 6
- O'Dwyer I. J., Bersanelli M., Childers J., et al. 2005, *.Rev.Astrn.Astrophys.*, 158, 93
- O'Sullivan F., 1990, *SIAM J. Numer. Anal.*, 27, 1635
- Panko E., Flin P., 2004, in Diaferio A., ed., *IAU Colloq. 195: Outskirts of Galaxy Clusters: Intense Life in the Suburbs Application of the Voronoi tessellation technique for galaxy cluster search in the Münster Red Sky Survey*. pp 245–247
- Peacock J. A., Dodds S. J., 1994, *MNRAS*, 267, 1020
- Peebles P. J. E., 1989, *ApJ*, 344, L53
- Peebles P. J. E., 1990, *ApJ*, 362, 1
- Pen U.-L., 1998, *ApJ*, 504, 601
- Percival W. J., 2005, *MNRAS*, 356, 1168
- Pierpaoli E., Anthoine S., 2005, *Advances in Space Research*, 36, 757
- Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, *Numerical recipes in C. The art of scientific computing*. Cambridge: University Press, —c1992, 2nd ed.
- Puetter R. C., Pina R. K., 1993, in Fowler A. M., ed., *Proc. SPIE Vol. 1946, p. 405–416, Infrared Detectors and Instrumentation*, Albert M. Fowler; Ed. The pixon and Bayesian image reconstruction. pp 405–416
- Ramella M., Boschin W., Fadda D., Nonino M., 2001, *Astr.Astrophys.*, 368, 776
- Richardson W. H., 1972, *Journal of the Optical Society of America (1917-1983)*, 62, 55
- Robert C. P., 2001, *The Bayesian choice*. Springer-Verlag, New York
- Robinson D. R. T., 1991, in Grandy W. T., Schick L. H., eds, *Maximum Entropy and Bayesian methods Maximum Entropy with Poissonian statistics*. pp 337–341
- Rybicki G. B., Press W. H., 1992, *ApJ*, 398, 169
- Schaap W. E., van de Weygaert R., 2000, *Astr.Astrophys.*, 363, L29
- Schmoldt I. M., Saar V., Saha P., Branchini E., Efstathiou G. P., Frenk C. S., Keeble O., Maddox S., McMahan R., Oliver S., Rowan-Robinson M., Saunders W., Sutherland W. J., Tadros H., White S. D. M., 1999, *ApJ*, 118, 1146
- Scoccimarro R., 2004, *Phys. Rev. D*, 70, 083007
- Seljak U., 1998, *ApJ*, 503, 492
- Shannon C. E., 1948, *A mathematical theory of communication*. Vol. 27, *Bell System Technical Journal*
- Shaya E. J., Peebles P. J. E., Tully R. B., 1995, *ApJ*, 454, 15
- Shepp L. A., Vardi Y., 1982, *IEEE Trans. Med. Imaging*, 1, 113
- Sheth R. K., 1995, *MNRAS*, 277, 933
- Shewchuk J. R., 1994, *An Introduction to the Conjugate Gradient Method Without the Agonizing Pain*. published in the web
- Skilling J., ed. 1989, *Maximum Entropy and Bayesian methods* : 8 : 1988
- Smith C. C., Klypin A., Gross M. A. K., Primack J. R., Holtzman J., 1998, *MNRAS*, 297, 910
- Starobinsky A. A., 1982, *Physics Letters B*, 117, 175
- Stompor R., Balbi A., Borrill J. D., Ferreira P. G., Hanany S., Jaffe A. H., Lee A. T., Oh S., Rabii B., Richards P. L., Smoot G. F., Winant C. D., Wu J.-H. P., 2002, *Phys. Rev. D*, 65, 022003
- Sutton E. C., Wandelt B. D., 2006, *.Rev.Astrn.Astrophys.*, 162, 401
- Tanner M. A., 1996, *Tools for statistical inference*. Springer-Verlag, New York
- Tegmark M., 1997, *Phys. Rev. D*, 56, 4514
- Tegmark M., 1997, *ApJ*, 480, L87+
- Tegmark M., Bromley B. C., 1999, *The Astrophysical Journal*, 518, L69
- Tikhonov A. N., 1963, *Soviet Math Dokl*, 4, 1035
- Valentine H., Saunders W., Taylor A., 2000, *MNRAS*, 319, L13
- van de Weygaert R., Schaap W., 2001, in Bandy A. J., Zaroubi S., Bartelmann M., eds, *Mining the Sky Tessellation Reconstruction Techniques*. pp 268–+
- Vogeley M. S., Szalay A. S., 1996, *ApJ*, 465, 34
- Wandelt B. D., Larson D. L., Lakshminarayanan A., 2004, *Phys. Rev. D*, 70, 083511
- Webster M., Lahav O., Fisher K., 1997, *MNRAS*, 287, 425
- Weinberg D. H., 1992, *MNRAS*, 254, 315
- Yahil A., Strauss M. A., Davis M., Huchra J. P., 1991, *ApJ*, 372, 380
- Yvon D., Mayet F., 2005, *Astr.Astrophys.*, 436, 729
- Zaninetti L., 1995, *Astr.Astrophys.Suppl.*, 109, 71
- Zaninetti L., 2006, *Chinese Journal of Astronomy and Astrophysics*, 6, 387
- Zaroubi S., 2002a, *ArXiv Astrophysics e-prints*
- Zaroubi S., 2002b, *MNRAS*, 331, 901
- Zaroubi S., Hoffman Y., Dekel A., 1999, *ApJ*, 520, 413
- Zaroubi S., Hoffman Y., Fisher K. B., Lahav O., 1995, *ApJ*, 449, 446
- Zel'Dovich Y. B., 1970, *Astr.Astrophys.*, 5, 84

APPENDIX A: GAPMAP: MAP WITH A GAUSSIAN PRIOR AND A POISSONIAN LIKELIHOOD

Remember $P(\mathbf{s} | \mathbf{d}, \mathbf{p}) \propto \mathcal{L}(\mathbf{d} | \mathbf{s}, \mathbf{p})P(\mathbf{s} | \mathbf{p})$ to be extremized. First we write the log-likelihood taking the logarithm of eq. (12)

$$\log \mathcal{L}(\mathbf{s} | \mathbf{d}, \mathbf{p}) = \sum_i \left[-(\mathbf{R}\mathbf{s}')_i - c_i + (d'_i + c_i) \log \left((\mathbf{R}\mathbf{s}')_i + c_i \right) - \log \left((d'_i + c_i)! \right) \right]. \quad (\text{A1})$$

Then we differentiate with respect to the signal to yield

$$\frac{\partial \log \mathcal{L}(\mathbf{s} | \mathbf{d}, \mathbf{p})}{\partial s_k} = \sum_i \left[R_{ik} b \bar{n}_g \left(-1 + \left(\sum_j R_{ij} s'_j + c_i \right)^{-1} (d'_i + c_i) \right) \right]. \quad (\text{A2})$$

The same exercise for the Gaussian prior leads to

$$\frac{\partial \log P(\mathbf{s} | \mathbf{p})}{\partial s_k} = - \sum_j S_{kj}^{-1} s_j. \quad (\text{A3})$$

Now we demand $0 = \partial \log P(\mathbf{s} | \mathbf{d}, \mathbf{p}) / \partial s_k$ to get an equation for the MAP estimator. After applying \mathbf{S} to the equation we obtain

$$\sum_i \sum_l \left[S_{kl} R_{il} b \bar{n}_g \left(-1 + \left(\sum_m R_{im} \bar{n}_g (1 + b s_m^j) + c_i \right)^{-1} (d'_i + c_i) \right) \right] - s_k^j = 0. \quad (\text{A4})$$

Adding the index $j+1$ and j to \mathbf{s} on lhs and rhs respectively, an iteration scheme is formed

$$s_k^{j+1} = \sum_i \sum_l \left[S_{kl} R_{il} b \bar{n}_g \left(-1 + \left(\sum_m R_{im} \bar{n}_g (1 + b s_m^j) + c_i \right)^{-1} (d'_i + c_i) \right) \right]. \quad (\text{A5})$$

Let us simplify this algorithm for positive signals \mathbf{s}' in matrix notation

$$\mathbf{s}'^{j+1} = \bar{\mathbf{s}}'^2 \mathbf{S} \mathbf{R}^\dagger \left[-\bar{\mathbf{1}} + \text{diag}(\mathbf{R} \mathbf{s}'^j + \mathbf{c})^{-1} (\mathbf{d}' + \mathbf{c}) \right] + \bar{\mathbf{s}}', \quad (\text{A6})$$

where we made following substitutions $b \rightarrow 1$ and $\bar{n}_g \rightarrow \bar{\mathbf{s}}'$, with $\bar{\mathbf{s}}'$ being the average of the positive signal.

APPENDIX B: POISSONIAN MAXIMUM LIKELIHOOD

The context in which the Richardson-Lucy algorithm is applied has positive intensity signals and the kernel \mathbf{R} in eq. (1) is understood as a blurring function that can be expressed mathematically as a convolution with the *true* signal \mathbf{s} . We will further assume no background ($\mathbf{c} = 0$) so that the log-likelihood of eq. (12) can be written as

$$\log \mathcal{L}(\mathbf{s}' | \mathbf{d}', \mathbf{p}) = \sum_i \left[-(\mathbf{R}\mathbf{s}')_i + d'_i \log(\mathbf{R}\mathbf{s}')_i - \log(d'_i!) \right], \quad (\text{B1})$$

differentiating with respect to the signal yields

$$0 = \frac{\partial \log \mathcal{L}(\mathbf{s}' | \mathbf{d}', \mathbf{p})}{\partial s'_k} = \sum_i \left[R_{ik} \left(-1 + (\mathbf{R}\mathbf{s}')_i^{-1} d'_i \right) \right]. \quad (\text{B2})$$

We can multiply this equation with the signal \mathbf{s}' and make an iterative method which coincides with Richardson-Lucy algorithm

$$\mathbf{s}'^{j+1} = \text{diag}(\mathbf{R}^\dagger \text{diag}(\mathbf{R} \mathbf{s}'^j)^{-1} \mathbf{d}') \mathbf{s}'^j, \quad (\text{B3})$$

with $\mathbf{R}^\dagger \bar{\mathbf{1}} = \bar{\mathbf{1}}$ due to the convolution operation.

APPENDIX C: LINEAR FILTERS NEED TO BE INVERTIBLE TO CONSERVE INFORMATION

The Fisher information matrix \mathbf{J} for a Gaussian distribution²² with zero mean and covariance matrix \mathbf{C} calculated by Vogeley & Szalay (1996) has the form

$$\mathbf{J}_{ij} = \frac{1}{2} \text{tr}(\mathbf{G}_i \mathbf{G}_j), \quad (\text{C1})$$

with

$$\mathbf{G}_i = \mathbf{C}^{-1} \mathbf{C}_{,i}, \quad (\text{C2})$$

where the comma notation $\mathbf{C}_{,i}$ stands for the derivative with respect to the parameter θ_i : $d\mathbf{C}/d\theta_i$. Following Tegmark (1997), we calculate the Fisher information matrix \mathbf{J} for the filtered and unfiltered signal. Let us assume a linear filter \mathbf{L} , which provides us with an estimator of the signal

$$\langle \mathbf{s} \rangle_L \equiv \mathbf{L} \mathbf{d}. \quad (\text{C3})$$

The correlation matrix of the estimator yields

$$\mathbf{C}^{\text{est}} = \langle \langle \mathbf{s} \rangle_L \langle \mathbf{s} \rangle_L^\dagger \rangle_{(\mathbf{s}, \epsilon | \mathbf{p})} = \mathbf{L}^\dagger \left(\mathbf{R} \mathbf{S} \mathbf{R}^\dagger + \mathbf{N} \right) \mathbf{L}. \quad (\text{C4})$$

We get then

$$\mathbf{C}_{,i}^{\text{est}} = \mathbf{L}^\dagger \left(\mathbf{R} \mathbf{S}_{,i} \mathbf{R}^\dagger \right) \mathbf{L}, \quad (\text{C5})$$

$$\mathbf{G}_i^{\text{est}} = \tilde{\mathbf{L}} \left(\mathbf{R} \mathbf{S} \mathbf{R}^\dagger + \mathbf{N} \right)^{-1} \tilde{\mathbf{L}}^\dagger \mathbf{L}^\dagger \left(\mathbf{R} \mathbf{S}_{,i} \mathbf{R}^\dagger \right) \mathbf{L}, \quad (\text{C6})$$

where we have denoted the approximate inverse of \mathbf{L} as $\tilde{\mathbf{L}}$. Doing the same for the data yields

$$\mathbf{C}^{\text{data}} = \langle \mathbf{d} \mathbf{d}^\dagger \rangle_{(\mathbf{s}, \epsilon | \mathbf{p})} = \left(\mathbf{R} \mathbf{S} \mathbf{R}^\dagger + \mathbf{N} \right), \quad (\text{C7})$$

$$\mathbf{C}_{,i}^{\text{data}} = \mathbf{R} \mathbf{S}_{,i} \mathbf{R}^\dagger, \quad (\text{C8})$$

$$\mathbf{G}_i^{\text{data}} = \left(\mathbf{R} \mathbf{S} \mathbf{R}^\dagger + \mathbf{N} \right)^{-1} \left(\mathbf{R} \mathbf{S}_{,i} \mathbf{R}^\dagger \right). \quad (\text{C9})$$

If we now insert expression (C6) in the Fisher matrix (C1), we get

$$\mathbf{J}_{ij}^{\text{est}} = \frac{1}{2} \text{tr} \left(\mathbf{G}_i^{\text{est}} \mathbf{G}_j^{\text{est}} \right) = \frac{1}{2} \text{tr} \left(\tilde{\mathbf{L}} \mathbf{C}^{\text{data}-1} \tilde{\mathbf{L}}^\dagger \mathbf{L}^\dagger \mathbf{C}_{,i}^{\text{data}} \mathbf{L} \tilde{\mathbf{L}} \mathbf{C}^{\text{data}-1} \tilde{\mathbf{L}}^\dagger \mathbf{L}^\dagger \mathbf{C}_{,j}^{\text{data}} \mathbf{L} \right) \quad (\text{C10})$$

In general, this will differ from the Fisher matrix of the data. If we assume, however, that the linear operator is invertible ($\exists \mathbf{L}^{-1}$), then eq. (C10) reduces to

$$\mathbf{J}_{ij}^{\text{est}} = \frac{1}{2} \text{tr} \left(\mathbf{L}^{-1} \mathbf{G}_i^{\text{data}} \mathbf{G}_j^{\text{data}} \mathbf{L} \right). \quad (\text{C11})$$

Invoking that the trace of a product of matrices is invariant under cyclic permutations, we see that

$$\mathbf{J}_{ij}^{\text{est}} = \frac{1}{2} \text{tr} \left(\mathbf{G}_i^{\text{data}} \mathbf{G}_j^{\text{data}} \right) = \mathbf{J}_{ij}^{\text{data}}. \quad (\text{C12})$$

This shows the result that any linear invertible filter conserves information, regardless of the parameters that one wants to estimate. However, one should be careful with this statement because linear filters are, in general, not invertible unless the data and signal space have the same dimension, the noise is non-zero for any frequency, and the \mathbf{R} - and \mathbf{S} -matrices are invertible. Usually the data and signal space will differ and the \mathbf{R} -matrix will not be exactly invertible.

²² Here a Gaussian likelihood is assumed, but the result does not rely on the Gaussianity of the data (see e.g. Seljak 1998).

APPENDIX D: JEFFREY'S PRIOR FOR THE 3-DIMENSIONAL POWER SPECTRUM

Let us start by assuming a Gaussian likelihood²³

$$P(\mathbf{s} | P_S(\mathbf{k})) \propto \prod_{\mathbf{k}} \frac{1}{\sqrt{P_S(\mathbf{k})}} \exp -\frac{|s(\mathbf{k})|^2}{2P_S(\mathbf{k})}. \quad (\text{D1})$$

The log-likelihood is then given by

$$\log \left(P(\mathbf{s} | P_S(\mathbf{k})) \right) \propto \sum_{\mathbf{k}} \left[\log \left(P_S(\mathbf{k}) \right) + \frac{|s(\mathbf{k})|^2}{P_S(\mathbf{k})} \right]. \quad (\text{D2})$$

We now need the second derivatives of the log-likelihood with respect to the parameter P_S

$$\frac{\partial^2}{\partial P_S(\mathbf{k})^2} \log \left(P(\mathbf{s} | P_S(\mathbf{k})) \right) \propto \left[-\frac{1}{P_S^2(\mathbf{k})} + \frac{2|s(\mathbf{k})|^2}{P_S^3(\mathbf{k})} \right]. \quad (\text{D3})$$

The next step consists of calculating the Fisher information by performing the integral $\int ds P(\mathbf{s} | P_S(\mathbf{k}))$ on the above quantity, which is equivalent to performing the following ensemble average (see section 2.2)

$$J(P_S(\mathbf{k})) = \left\langle \frac{\partial^2}{\partial P_S(\mathbf{k})^2} \log \left(P(\mathbf{s} | P_S(\mathbf{k})) \right) \right\rangle_{(\mathbf{s}|\mathbf{p})} \propto \frac{1}{P_S^2(\mathbf{k})}, \quad (\text{D4})$$

where we have taken into account that $P_S(\mathbf{k}) = \langle |s(\mathbf{k})|^2 \rangle_{(\mathbf{s}|\mathbf{p})}$. Finally the square-root of the Fisher information leads to Jeffrey's prior

$$P(P_S(\mathbf{k})) = \sqrt{J(P_S(\mathbf{k}))} \propto P_S(\mathbf{k})^{-1}. \quad (\text{D5})$$

Following Wandelt et al. (2004) we can argue in a more intuitive way that $P(P_S(\mathbf{k})) \propto P_S(\mathbf{k})^{-1}$ is a solution to a measure invariant under scale transformations of the form $P(P_S(\mathbf{k}))dP_S(\mathbf{k}) = P(\alpha P_S(\mathbf{k}))\alpha dP_S(\mathbf{k})$ (here we have generalized this result to the 3-dimensional power spectrum).

APPENDIX E: MEM WITH GAUSSIAN AND POISSONIAN LIKELIHOODS

The quantity to maximize is given by

$$Q^E(\mathbf{s} | \mathbf{p}) = \alpha S^E(\mathbf{s} | \mathbf{p}) + \log \mathcal{L}(\mathbf{s} | \mathbf{d}, \mathbf{p}). \quad (\text{E1})$$

After some calculations we see that the gradient of the entropy for PADs is

$$\nabla S_+^E(\mathbf{s}' | \mathbf{p})_i = -\log \left(\frac{s'_i}{m_i} \right), \quad (\text{E2})$$

and for positive and negative distributions

$$\nabla S_{\pm}^E(\mathbf{s} | \mathbf{p})_i = -\log \left(\frac{w_i + s_i}{m_i} \right). \quad (\text{E3})$$

We took into account that $\partial w_i / \partial s_j = s_i / w_i \delta_{ij}$. It is then more straightforward to calculate the S^E curvature for PADs

$$\nabla \nabla S_+^E(\mathbf{s}' | \mathbf{p}) = -\text{diag}(\mathbf{s}')^{-1}, \quad (\text{E4})$$

and for positive and negative distributions,

$$\nabla \nabla S_{\pm}^E(\mathbf{s} | \mathbf{p}) = -\text{diag}(\mathbf{w})^{-1}. \quad (\text{E5})$$

Analogously, we calculate the gradient of the log $\mathcal{L}(\mathbf{s} | \mathbf{d})$ for the

Gaussian case valid for positive (s') and positive and negative signals (s_{\pm})

$$\nabla \log \mathcal{L}_G(\mathbf{s} | \mathbf{d}, \mathbf{p})_i = -\frac{1}{2} \nabla \chi^2(\mathbf{s})_i = -\left(\mathbf{R}^\dagger \mathbf{N}^{-1} (\mathbf{R}\mathbf{s} - \mathbf{d}) \right)_i, \quad (\text{E6})$$

and the corresponding curvature

$$\nabla \nabla \log \mathcal{L}_G(\mathbf{s} | \mathbf{d}, \mathbf{p}) = -\frac{1}{2} \nabla \nabla \chi^2(\mathbf{s}) = -\mathbf{R}^\dagger \mathbf{N}^{-1} \mathbf{R}. \quad (\text{E7})$$

The Poissonian case leads to

$$\begin{aligned} \nabla \log \mathcal{L}_P(\mathbf{s} | \mathbf{d}, \mathbf{p})_i & \quad (\text{E8}) \\ &= b \bar{n}_g \sum_{\mathbf{k}} \left[R_{ki} \left(-1 + \left(\sum_j R_{kj} s'_j + c_k \right)^{-1} (d'_k + c_k) \right) \right] \\ &= b \bar{n}_g \left[\mathbf{R}^\dagger \left(-\mathbf{1} + \text{diag} \left((\mathbf{R}\mathbf{s}' + \mathbf{c})^{-1} (\mathbf{d}' + \mathbf{c}) \right) \right) \right]_i, \end{aligned}$$

and

$$\begin{aligned} \nabla \nabla \log \mathcal{L}_P(\mathbf{s} | \mathbf{d}, \mathbf{p})_{ij} & \quad (\text{E9}) \\ &= -b^2 \bar{n}_g^2 \sum_{\mathbf{k}} \left[R_{ki} \left(\sum_l R_{kl} s'_l + c_k \right)^{-2} R_{kj} (d'_k + c_k) \right] \\ &= -b^2 \bar{n}_g^2 \left[\mathbf{R}^\dagger \left(\text{diag} \left((\mathbf{R}\mathbf{s}' + \mathbf{c})^{-2} \mathbf{R}^\dagger (\mathbf{d}' + \mathbf{c}) \right) \right) \right]_{ij}. \end{aligned}$$

Note that when dealing with overdensity fields one should do the following substitution: $s'_i = \bar{n}_g (1 + bs_i)$ in the last two expressions.

Summing up, we have the following gradient of Q^E for PADs

$$\nabla Q_+^E(\mathbf{s}' | \mathbf{p})_i = -\alpha \log \left(\frac{s'_i}{m_i} \right) + \nabla \log \mathcal{L}(\mathbf{s}' | \mathbf{d}, \mathbf{p})_i, \quad (\text{E10})$$

and for positive and negative distributions

$$\nabla Q_{\pm}^E(\mathbf{s} | \mathbf{p})_i = -\alpha \log \left(\frac{w_i - s_i}{m_i} \right) + \nabla \log \mathcal{L}(\mathbf{s} | \mathbf{d}, \mathbf{p})_i, \quad (\text{E11})$$

and the corresponding curvatures

$$\nabla \nabla Q_+^E(\mathbf{s}' | \mathbf{p}) = -\alpha \text{diag}(\mathbf{s}')^{-1} + \nabla \nabla \log \mathcal{L}(\mathbf{s}' | \mathbf{d}, \mathbf{p}), \quad (\text{E12})$$

$$\nabla \nabla Q_{\pm}^E(\mathbf{s} | \mathbf{p}) = -\alpha \text{diag}(\mathbf{w})^{-1} + \nabla \nabla \log \mathcal{L}(\mathbf{s} | \mathbf{d}, \mathbf{p}). \quad (\text{E13})$$

The corresponding likelihood (Gaussian or Poissonian) has to be inserted in each of the expressions for the gradient or curvature of Q^E . For the choice of an optimal regularization constant α see e.g. Maisinger et al. (1997) and Hobson et al. (1998).

APPENDIX F: BAYES, TIKHONOV, ASYMPTOTIC REGULARIZATION AND LEARNING ALGORITHMS

We want to solve eq. (60) from a Bayesian perspective. Let us assume a Gaussian likelihood with covariance \mathbf{I}

$$\mathcal{L}(\boldsymbol{\psi} | \mathbf{f}, \mathbf{p}) = G(\mathbf{f} - \mathbf{A}\boldsymbol{\psi}, \mathbf{I}), \quad (\text{F1})$$

which is a fair assumption in the absence of noise (eq. (60) is equivalent to eq. (2) without noise, $\epsilon = 0$). Let us further assume a Gaussian prior around a prior solution $\boldsymbol{\psi}^*$ with covariance $\tau \tilde{\mathbf{M}}^{-1}$

$$P(\boldsymbol{\psi} | \mathbf{p}) = G(\boldsymbol{\psi} - \boldsymbol{\psi}^*, \tau \tilde{\mathbf{M}}^{-1}). \quad (\text{F2})$$

We can now calculate the MAP which coincides in this case with the mean of the posterior. Let us look at the quantity given by the log-posterior PDF

$$\|\mathbf{f} - \mathbf{A}\boldsymbol{\psi}\|^2 + \tau \|\boldsymbol{\psi} - \boldsymbol{\psi}^*\|_{\tilde{\mathbf{M}}}^2, \quad (\text{F3})$$

²³ Note that the likelihood for $P_S(\mathbf{k})$ is the prior for \mathbf{s} .

which is a generalization of Tikhonov regularization. Minimizing the negative log-posterior yields the following equation for the Bayesian estimator $\langle\psi\rangle_{\text{B}}$

$$\mathbf{A}^\dagger(\mathbf{A}\langle\psi\rangle_{\text{B}} - \mathbf{f}) + \tau^{-1}\tilde{\mathbf{M}}(\langle\psi\rangle_{\text{B}} - \psi^*) = 0. \quad (\text{F4})$$

If we now choose $\tilde{\mathbf{M}} = \mathbf{A}^\dagger\mathbf{M}^{-1}$ (\mathbf{M} is an invertible matrix) we get

$$\mathbf{A}^\dagger(\mathbf{M}^{-1}(\psi^* - \langle\psi\rangle_{\text{B}}) + \tau(\mathbf{f} - \mathbf{A}\langle\psi\rangle_{\text{B}})) = 0, \quad (\text{F5})$$

This equation will be fulfilled if the following equality holds

$$\langle\psi\rangle_{\text{B}} = \psi^* + \tau\mathbf{M}(\mathbf{f} - \mathbf{A}\langle\psi\rangle_{\text{B}}). \quad (\text{F6})$$

The estimator $\langle\psi\rangle_{\text{B}}$ for the solution to the inverse problem (eq. (60)) is expressed in eq. (F6) as the prior solution ψ^* plus a correction term given by the residual $\mathbf{f} - \mathbf{A}\langle\psi\rangle_{\text{B}}$. Since only the residual based on the prior solution is known, the following substitution must be done on the right-hand-side (rhs) $\langle\psi\rangle_{\text{B}} \rightarrow \psi^*$ leading to

$$\langle\psi\rangle_{\text{B}} \simeq \psi^* + \tau\mathbf{M}(\mathbf{f} - \mathbf{A}\psi^*). \quad (\text{F7})$$

This can be interpreted as an iterative scheme, in which the estimator is the update $j+1$ ($\langle\psi\rangle_{\text{B}} \rightarrow \psi^{j+1}$ on the left-hand-side (lhs)) of the estimator at the previous step j ($\psi^* \rightarrow \psi^j$ on the rhs)

$$\psi^{j+1} = \psi^j + \tau\mathbf{M}(\mathbf{f} - \mathbf{A}\psi^j). \quad (\text{F8})$$

In this way, we have found the general iterative method (eq. 71) derived with the asymptotic regularization in section (3.1.2). From the Bayesian point of view, this scheme could be interpreted as a learning algorithm, in which the estimator of the solution to the inverse problem is calculated from the prior solution and becomes itself the prior solution for the subsequent iteration.

APPENDIX G: PRECONDITIONING

We can enhance the convergence of the iteration methods by multiplying the matrix we want to invert by another matrix that is close to its inverse

$$\mathbf{M}\mathbf{A}\psi = \mathbf{M}\mathbf{f}, \quad (\text{G1})$$

with $\mathbf{M} \sim \mathbf{A}^{-1}$. Let us show this by deriving eq. (71) in a different way. We can invert $\mathbf{M}\mathbf{A}$ using the Neumann expansion for the inverse of an operator

$$\psi = (\mathbf{M}\mathbf{A})^{-1}\mathbf{M}\mathbf{f} = \sum_{i=0}^{\infty}(\mathbf{I} - \mathbf{M}\mathbf{A})^i\mathbf{M}\mathbf{f}. \quad (\text{G2})$$

This iteration scheme will converge if $\|\mathbf{I} - \mathbf{M}\mathbf{A}\| < 1$. Let us introduce the following notation

$$\psi \equiv \sum_{i=0}^{\infty} \psi[i], \quad (\text{G3})$$

$$\psi^j \equiv \sum_{i=0}^j \psi[i], \quad (\text{G4})$$

with

$$\psi[i] \equiv (\mathbf{I} - \mathbf{M}\mathbf{A})^i\mathbf{M}\mathbf{f}. \quad (\text{G5})$$

It follows that

$$\psi[i+1] = (\mathbf{I} - \mathbf{M}\mathbf{A})\psi[i], \quad (\text{G6})$$

and summing over i we get

$$\sum_{i=0}^j \psi[i+1] = \sum_{i=0}^j \psi[i] - \sum_{i=0}^j \mathbf{M}\mathbf{A}\psi[i]. \quad (\text{G7})$$

Manipulating the indices, we see that

$$\sum_{i=0}^j \psi[i+1] = \sum_{i=0}^{j+1} \psi[i] - \psi[0]. \quad (\text{G8})$$

Combining the last two equations we obtain eq. (71)²⁴

$$\psi^{j+1} = \psi^j + \mathbf{M}(\mathbf{f} - \mathbf{A}\psi^j), \quad (\text{G9})$$

with

$$\psi[0] = \psi^0 = \mathbf{M}\mathbf{f}. \quad (\text{G10})$$

The meaning of the preconditioning matrix \mathbf{M} is clear when we look at eq. (G2). There it can be seen that a much more rapid convergence is obtained if $(\mathbf{I} - \mathbf{M}\mathbf{A})$ is close to zero, that is if \mathbf{M} is close to the inverse of \mathbf{A} .

APPENDIX H: KRYLOV METHODS: CONJUGATE GRADIENTS

H1 Orthogonality between the residuals and the searching vectors

Eq. (87) tells us that each error vector η^{j+1} is \mathbf{A} -orthogonal to the previous searching vector $\mathbf{M}\mu^j$. Since all different searching vectors $\mathbf{M}\mu^i$ are \mathbf{A} -orthogonal to each other by construction, and the error vectors are given by the linear combination of the previous error vector and the previous searching vector (eq. (84)), it follows that each error vector η^{j+1} is \mathbf{A} -orthogonal to all previous searching vectors μ^i , i.e. for $i \leq j$,

$$\langle\eta^{j+1}|\mathbf{M}\mu^i\rangle_{\mathbf{A}} = 0. \quad (\text{H1})$$

Using eq. (85) we can write eq. (H1) as

$$\langle\xi^{j+1}|\mathbf{M}\mu^i\rangle = 0, \quad (\text{H2})$$

being $i \leq j$.

Applying the inner product between the searching vectors $\mathbf{M}\mu^i$ and the recurrent formula for the residuals (eq. 86), we get

$$\langle\xi^{j+1}|\mathbf{M}\mu^i\rangle = \langle\xi^j|\mathbf{M}\mu^i\rangle - \tau^j\langle\mathbf{M}\mu^j|\mathbf{M}\mu^i\rangle_{\mathbf{A}}. \quad (\text{H3})$$

For $i \neq j$ this equation reduces to

$$\langle\xi^{j+1}|\mathbf{M}\mu^i\rangle = \langle\xi^j|\mathbf{M}\mu^i\rangle. \quad (\text{H4})$$

From eq. (H2) and eq. (H4) we conclude that for $i < j$,

$$\langle\xi^j|\mathbf{M}\mu^i\rangle = 0. \quad (\text{H5})$$

H2 The set of residuals as a basis of linearly independent vectors

Taking the Gram-Schmidt orthogonalization scheme (eq. 90) and multiplying it with the residuals, we obtain

$$\langle\xi^i|\mathbf{M}\mu^j\rangle = \langle\xi^i|\mathbf{M}\xi^j\rangle + \sum_{k=0}^{j-1} \beta^{kj} \langle\xi^i|\mathbf{M}\mu^k\rangle. \quad (\text{H6})$$

²⁴ The iteration time step τ has been absorbed here in the matrix \mathbf{M} .

Using the result obtained in the appendix H1 (eq. H5), one shows the orthogonality (strictly orthogonal, if $M = I$) between any different residuals (for $i \neq j$)²⁵

$$\langle \xi^i | M \xi^j \rangle = 0. \quad (\text{H7})$$

For $i = j$ by combining (H5) and (H6) we get the relation we used in equation (89)

$$\langle \xi^i | M \mu^i \rangle = \langle \xi^i | M \xi^i \rangle. \quad (\text{H8})$$

H3 Formulae for the β -factor

From the scalar product between eq. (86) and the residual ξ^i

$$\langle \xi^{j+1} | M \xi^i \rangle = \langle \xi^j | M \xi^i \rangle - \tau^j \langle M \mu^j | M \xi^i \rangle_{\mathbf{A}}, \quad (\text{H9})$$

it is clear that the β -factors are all zero except for one. Notice that the denominator in β , given by $\langle M \mu^j | M \xi^i \rangle_{\mathbf{A}}$ cancels out if neither $i = j + 1$ nor $i = j$. The latter is excluded according to the definition of β (see eqs. 90 and 92). Gram-Schmidt orthogonalization thus simplifies to eq. (93), with

$$\beta_{\text{EXP}}^{j+1} = -\frac{\langle M \xi^{j+1} | M \mu^j \rangle_{\mathbf{A}}}{\langle M \mu^j | M \mu^j \rangle_{\mathbf{A}}}. \quad (\text{H10})$$

Other expressions for this factor can be derived by replacing $i = j + 1$ in eq. (H9)

$$\langle M \mu^j | M \xi^{j+1} \rangle_{\mathbf{A}} = -\frac{1}{\tau^j} \langle \xi^{j+1} | M \xi^{j+1} \rangle. \quad (\text{H11})$$

Substituting this expression in eq. (92) and using the formula for τ^j (eq. 89) one obtains the Fletcher-Reeves equation

$$\beta_{\text{FR}}^{j+1} = \frac{\langle \xi^{j+1} | M \xi^{j+1} \rangle}{\langle \xi^j | M \xi^j \rangle}. \quad (\text{H12})$$

Polak-Ribière's formula can now be obtained trivially by taking expression (H7) into account. Let us do an invariant operation by adding $-\langle \xi^{j+1} | M \xi^j \rangle$ to the nominator in Fletcher-Reeves formula

$$\langle \xi^{j+1} | M \xi^{j+1} \rangle - \langle \xi^{j+1} | M \xi^j \rangle = \langle \xi^{j+1} | M (\xi^{j+1} - \xi^j) \rangle, \quad (\text{H13})$$

which immediately leads to Polak-Ribière's expression

$$\beta_{\text{PR}}^{j+1} = \frac{\langle \xi^{j+1} | M (\xi^{j+1} - \xi^j) \rangle}{\langle \xi^j | M \xi^j \rangle}. \quad (\text{H14})$$

In order to get Hestenes-Stiefels formula one has to consider eqs. (H8) and (H5) in the denominator of β_{PR}

$$\langle \xi^j | M \xi^j \rangle = \langle \mu^j | M \xi^j \rangle - \langle \mu^j | M \xi^{j+1} \rangle = \langle \mu^j | M (\xi^j - \xi^{j+1}) \rangle, \quad (\text{H15})$$

resulting in the following expression

$$\beta_{\text{HS}}^{j+1} = -\frac{\langle \xi^{j+1} | M (\xi^{j+1} - \xi^j) \rangle}{\langle \mu^j | M (\xi^{j+1} - \xi^j) \rangle}. \quad (\text{H16})$$

Due to the relations derived in this appendix other equivalent formulae for β (summarized in table 2) can be found, which differ in their numerical behaviour. Note that from the 16 possible schemes presented here, only 3 are discussed in the literature.

²⁵ This result is at first glance only valid for $i < j$. However, with the additional requirement that the matrix M be self-adjoint, the generalization to $i \neq j$ is trivial.

H4 Preconditioned non-linear time step

The function under consideration is expanded until the second order around $\tau^j M \mu^j$ according to eq. (83)

$$Q_{\mathbf{A}}(\psi^j + \tau^j M \mu^j) \simeq Q_{\mathbf{A}}(\psi^j) + \tau^j \langle \nabla Q_{\mathbf{A}}(\psi^j) | M \mu^j \rangle + \frac{\tau^{j2}}{2} \langle M \mu^j | M \mu^j \rangle_{\nabla \nabla Q_{\mathbf{A}}(\psi^j)}. \quad (\text{H17})$$

Then the derivative with respect to the searching vector is done to find the extremum

$$\frac{d}{d\tau^j} Q_{\mathbf{A}}(\psi^j + \tau^j M \mu^j) \simeq \langle \nabla Q_{\mathbf{A}}(\psi^j) | M \mu^j \rangle + \tau^j \langle M \mu^j | M \mu^j \rangle_{\nabla \nabla Q_{\mathbf{A}}(\psi^j)}. \quad (\text{H18})$$

By setting this equation to zero, one finds an expression for the time step

$$\tau^j = -\frac{\langle \nabla Q_{\mathbf{A}}(\psi^j) | M \mu^j \rangle}{\langle M \mu^j | M \mu^j \rangle_{\nabla \nabla Q_{\mathbf{A}}(\psi^j)}}. \quad (\text{H19})$$

Note that the last equation can be rewritten using relation (H8) as

$$\tau^j = -\frac{\langle \nabla Q_{\mathbf{A}}(\psi^j) | M \nabla Q_{\mathbf{A}}(\psi^j) \rangle}{\langle M \mu^j | M \mu^j \rangle_{\nabla \nabla Q_{\mathbf{A}}(\psi^j)}}. \quad (\text{H20})$$

APPENDIX I: THE MAPPING EQUATION FOR THE WIENER-FILTER IN K-SPACE

Following the concept of minimum variance (e.g. Rybicki & Press 1992; Zaroubi et al. 1995), we define an action given by the normalized volume integral of the square of the difference between the reconstruction (ψ) and the ensemble of different possible realizations of the density field ($s = \delta_\rho$)

$$\mathcal{A} = \langle \frac{1}{V} \int d^{N_D} \mathbf{r} [\psi(\mathbf{r}) - s(\mathbf{r})]^2 \rangle_{(s, \epsilon | \mathbf{p})}. \quad (\text{I1})$$

From the statistical point of view, the action \mathcal{A} is the loss function that has to be minimized. Note that this action can be expressed as the ensemble average of the squared Euclidean distance between the real density field s and the reconstruction ψ

$$\mathcal{A} = \frac{1}{V} \langle D_{\text{Eucl}}^2(\psi, s) \rangle_{(s, \epsilon | \mathbf{p})}. \quad (\text{I2})$$

Transforming expression (I1) into Fourier space yields

$$\mathcal{A} = \frac{1}{V} \int \frac{d^{N_D} \mathbf{k}}{(2\pi)^{N_D}} \left[\langle \hat{\psi}(\mathbf{k}) \overline{\hat{\psi}(\mathbf{k})} \rangle_{(s, \epsilon | \mathbf{p})} + \langle \hat{s}(\mathbf{k}) \overline{\hat{s}(\mathbf{k})} \rangle_{(s, \epsilon | \mathbf{p})} - \langle \hat{\psi}(\mathbf{k}) \overline{\hat{s}(\mathbf{k})} \rangle_{(s, \epsilon | \mathbf{p})} - \langle \hat{s}(\mathbf{k}) \overline{\hat{\psi}(\mathbf{k})} \rangle_{(s, \epsilon | \mathbf{p})} \right]. \quad (\text{I3})$$

Assuming a linear relation between the reconstruction ψ and the data d

$$\hat{\psi}(\mathbf{k}) = \int \frac{d^{N_D} \mathbf{k}'}{(2\pi)^{N_D}} \hat{F}_{\text{WF}}(\mathbf{k}, \mathbf{k}') \hat{d}(\mathbf{k}'), \quad (\text{I4})$$

and statistical homogeneity $\langle \langle \hat{s}(\mathbf{k})\hat{s}(\mathbf{k}') \rangle \rangle_{(\mathbf{s}, \boldsymbol{\epsilon}|\mathbf{p})} = (2\pi)^{N_D} \delta_D(\mathbf{k} - \mathbf{k}') P_S(\mathbf{k}')$, yields

$$\begin{aligned} \mathcal{A} = & \frac{1}{V} \int \frac{d^{N_D} \mathbf{k}}{(2\pi)^{N_D}} \int \frac{d^{N_D} \mathbf{k}'}{(2\pi)^{N_D}} \left[\right. \\ & \hat{F}_{WF}(\mathbf{k}, \mathbf{k}') \int \frac{d^{N_D} \mathbf{q}}{(2\pi)^{N_D}} \overline{\hat{F}_{WF}(\mathbf{k}, \mathbf{q}) \langle \hat{d}(\mathbf{k}') \overline{\hat{d}(\mathbf{q})} \rangle}_{(\mathbf{s}, \boldsymbol{\epsilon}|\mathbf{p})} \\ & + (2\pi)^{N_D} \delta_D(\mathbf{k} - \mathbf{k}') \langle \hat{s}(\mathbf{k}') \overline{\hat{s}(\mathbf{k}')} \rangle_{(\mathbf{s}, \boldsymbol{\epsilon}|\mathbf{p})} \\ & - \overline{\hat{F}_{WF}(\mathbf{k}, \mathbf{k}') \langle \hat{d}(\mathbf{k}') \overline{\hat{s}(\mathbf{k}')} \rangle}_{(\mathbf{s}, \boldsymbol{\epsilon}|\mathbf{p})} \\ & \left. - \overline{\hat{F}_{WF}(\mathbf{k}, \mathbf{k}') \langle \hat{s}(\mathbf{k}) \overline{\hat{d}(\mathbf{k}')} \rangle}_{(\mathbf{s}, \boldsymbol{\epsilon}|\mathbf{p})} \right]. \end{aligned} \quad (15)$$

Now the action is minimized with respect to the linear operator, $\frac{\delta \mathcal{A}}{\delta \hat{F}_{WF}} = 0$, to obtain the following mapping equation

$$\int \frac{d^{N_D} \mathbf{q}}{(2\pi)^{N_D}} \hat{F}_{WF}(\mathbf{k}, \mathbf{q}) \langle \hat{d}(\mathbf{q}) \overline{\hat{d}(\mathbf{k}')} \rangle_{(\mathbf{s}, \boldsymbol{\epsilon}|\mathbf{p})} = \langle \hat{s}(\mathbf{k}) \overline{\hat{d}(\mathbf{k}')} \rangle_{(\mathbf{s}, \boldsymbol{\epsilon}|\mathbf{p})}. \quad (16)$$

Note that eq. (16) allows us to substitute \mathbf{k}' by $-\mathbf{k}'$, which is equivalent to the conjugation of $\hat{d}(\mathbf{k}')$ due to the hermitian redundancy of real numbers

$$\int \frac{d^{N_D} \mathbf{q}}{(2\pi)^{N_D}} \hat{F}_{WF}(\mathbf{k}, \mathbf{q}) \langle \hat{d}(\mathbf{q}) \hat{d}(\mathbf{k}') \rangle_{(\mathbf{s}, \boldsymbol{\epsilon}|\mathbf{p})} = \langle \hat{s}(\mathbf{k}) \hat{d}(\mathbf{k}') \rangle_{(\mathbf{s}, \boldsymbol{\epsilon}|\mathbf{p})}. \quad (17)$$

The linear operator one obtains in this way is different, but fulfils the same requirements. We compare both cases in section (4). Let us see how one would apply such a filter. The covariance matrix of the data is given by

$$\langle \hat{d}(\mathbf{k}) \hat{d}(\mathbf{k}') \rangle_{(\mathbf{s}, \boldsymbol{\epsilon}|\mathbf{p})} = \langle \hat{\alpha}(\mathbf{k}) \hat{\alpha}(\mathbf{k}') \rangle_{(\mathbf{s}, \boldsymbol{\epsilon}|\mathbf{p})} + \langle \hat{\epsilon}(\mathbf{k}) \hat{\epsilon}(\mathbf{k}') \rangle_{(\mathbf{s}, \boldsymbol{\epsilon}|\mathbf{p})}, \quad (18)$$

and its action on some vector by

$$\begin{aligned} & \int \frac{d^{N_D} \mathbf{k}'}{(2\pi)^{N_D}} \langle \hat{\alpha}(\mathbf{k}) \hat{\alpha}(\mathbf{k}') \rangle_{(\mathbf{s}, \boldsymbol{\epsilon}|\mathbf{p})} \{ \hat{x}(\mathbf{k}') \} \\ & = \hat{f}_B \cdot \left[\hat{f}_{SM} \circ \left[P_S \cdot \overline{\left[\hat{f}_{SM} \circ \left[\hat{f}_B \cdot \{ \hat{x} \} \right] \right]} \right] \right] (\mathbf{k}), \end{aligned} \quad (19)$$

and

$$\begin{aligned} & \int \frac{d^{N_D} \mathbf{k}'}{(2\pi)^{N_D}} \langle \hat{\epsilon}(\mathbf{k}) \hat{\epsilon}(\mathbf{k}') \rangle_{(\mathbf{s}, \boldsymbol{\epsilon}|\mathbf{p})} \{ \hat{x}(\mathbf{k}') \} \\ & = \hat{f}_{SF} \circ \left[P_N \cdot \overline{\left[\hat{f}_{SF} \circ \{ \hat{x} \} \right]} \right] (\mathbf{k}). \end{aligned} \quad (110)$$

The correlation matrix between the data and the signal applied to that vector yields

$$\begin{aligned} & \int \frac{d^{N_D} \mathbf{k}'}{(2\pi)^{N_D}} \langle \hat{s}(\mathbf{k}) \hat{d}(\mathbf{k}') \rangle_{(\mathbf{s}, \boldsymbol{\epsilon}|\mathbf{p})} \{ \hat{x}(\mathbf{k}') \} \\ & = P_S \cdot \overline{\left[\hat{f}_{SM} \circ \left[\hat{f}_B \cdot \{ \hat{x} \} \right] \right]} (\mathbf{k}). \end{aligned} \quad (111)$$

We see that the difference with respect to the operations derived in section (3.3) resides in the conjugation of certain functions.