

**UNIVERSITÀ DEGLI STUDI DI PADOVA**

**Dipartimento di Fisica e Astronomia “Galileo Galilei”**

**Corso di Laurea in Fisica**

**Tesi di Laurea**

**Information Field Theory for Cosmological**

**Perturbations**

Relatore:

**Prof. Sabino Matarrese**

Laureando:

**Leonardo Rebeschini**

**1125253**

**Anno Accademico 2017/2018**



*"Una profonda consapevole  
ignoranza è il preludio di ogni  
avanzamento della scienza".*

James Clerk Maxwell



# Contents

Contents	i
<b>1 Introduction</b>	<b>1</b>
<b>2 Functional Calculus</b>	<b>3</b>
2.1 Introduction to Functionals . . . . .	3
2.2 Functional Calculus . . . . .	5
2.3 Gaussian Functional integral . . . . .	8
2.4 Probability Density Functional . . . . .	10
<b>3 Information Field Theory</b>	<b>13</b>
3.1 Information Theory . . . . .	13
3.2 Bayesian Inference . . . . .	13
3.3 Signal Inference Problem . . . . .	14
3.4 Field Inference . . . . .	14
3.5 Inference on Physical Fields . . . . .	15
3.6 Signal and Noise . . . . .	16
3.7 Hamiltonian Formalism . . . . .	19
3.8 Free Theory . . . . .	20
3.9 Interacting Theory . . . . .	22
3.10 Extrema Principle and MAP . . . . .	22
<b>4 IFT for Cosmological Perturbations</b>	<b>25</b>
4.1 Non-Gaussian CMB Fluctuations via $f_{nl}$ Theory . . . . .	25
4.1.1 Data Model . . . . .	25
4.1.2 Moments Calculation . . . . .	26
4.1.3 CMB Hamiltonian . . . . .	28
4.1.4 $f_{nl}$ Posterior . . . . .	29
4.1.5 Experimental Results . . . . .	30
<b>A Gaussian Field Correlation Functions</b>	<b>31</b>
<b>B CMB <math>f_{nl}</math></b>	<b>33</b>
<b>C CMB Reconstruction</b>	<b>35</b>
<b>Bibliography</b>	<b>37</b>



# Chapter 1

## Introduction

Fields play a very important role in sciences, they describes the spatial variation of a quantity, like the air temperature, as a function of position. Knowing the configuration of a field can be very helpful. Measurements of fields, however, can never provide the precise field configuration with certainty. Physical fields have an infinite number of degrees of freedom, but the data generated by any measurement device is always finite, providing only a finite number of constraints on the field. Probabilistic inference is used to make statements about physical fields exploiting that they follow known physical laws and exhibit correlations. Such information is hold into the field inference in order to overcome the mismatch of field degrees of freedom to measurement points.

To handle this, an information theory for fields is needed, and that is what information field theory is.

Information field theory is a Bayesian statistical field theory, it uses mathematical physics concepts embedded in statistical mechanics formalism and functional calculus to deal with fields.

The specialty of IFT is to deal with continuous fields, and the associated computational problem should be formulated in a way that the resolution of the field becomes irrelevant, as soon a sufficient high resolution is used in numerical applications.

IFT is not a method, but a language to discuss and derive statistical methods, which are used to get information on fields. How can this information be best retrieved? Is there a generic and simple principle from which optimal data analysis strategies derive? Can an information energy be constructed which – if minimized – provides us with the correct knowledge state given the data and prior information? And if this exists, how can this information ground state be found at least approximatively? Which is the inference solution. An information energy, to be minimized, would be very useful to have, since many of the existing minimization techniques, analytical and numerical, can then be applied to it. A number of such functions to be extremized to solve inference problems were proposed in the literature, like the likelihood, the posterior, the Hamiltonian, and the entropy.

The likelihood is the probability that the data has resulted from some signal. The posterior is the reverse, it is the probability that given the data some signal was the origin of it. Extremizing either of them certainly makes sense, but often ignores the presence of slightly less probable, but much more numerous possibilities in the signal phase space. Those have a much larger entropy and are therefore favored by maximum entropy methods. However, maximum entropy alone can not be the inference determining criterion, since it favors states of complete lack of knowledge, irrespective of the data. Thus some counteracting energy is required. In [13] is argued that the

ideal information energy is provided by the Gibbs free energy, which combines both maximum entropy and maximum a posteriori (MAP) principles.

The versatility of IFT to deal with various measurement problems, to take uncertainties on covariances and responses into account and to provide implementable algorithms have let to a number of concrete IFT applications in astronomy and astrophysics. IFT is already applied to a number of areas which deal with the cosmic microwave background, the cosmic large-scale structure, galactic magnetism, interferometric radio astronomy and Gamma-ray and X-ray astronomy.

In chapter 2 functional calculus is introduced in order to know mathematical tools necessary to deal with statistical field theories. In chapter 3 the inference problem is explained implemented with mathematical physics formalism and in chapter 4 an IFT application is presented, discussing on which is called the  $f_{nl}$  model, a simple model concerning a sensitive topic such CMB non-Gaussianity. Here also an experimental result of  $f_{nl}$  measurement is shown.



# Chapter 2

## Functional Calculus

In order to tackle field inference problems, it is necessary to introduce calculus on quantities present in field theory formalism. These quantities are functionals, and our mathematical tool will be functional calculus, which is an extension of the usual differential calculus and could be seen as a calculus about function that depends on an infinite number of degree of freedom.

### 2.1 Introduction to Functionals

Let's consider  $\mathcal{D}$  as the space of square integrable functions defined at each point of space-time<sup>1</sup>, and assume that it is equipped with an orthonormal basis  $[\phi_n]$ :

$$\begin{aligned} i) \int_{\mathcal{D}} dx \phi_n(x) \phi_m(x) &= \delta_{nm} && \text{orthonormality} \\ ii) \sum_n \phi_n(x) \phi_n(y) &= \delta(x-y) && \text{completeness} \end{aligned}$$

so that we can expand any  $q$  as  $q(x) = \sum_n q_n \phi_n(x)$ , with  $q_n \equiv \int_{\mathcal{D}} dx q(x) \phi_n(x)$  the expansion coefficients.

A functional<sup>2</sup> is a mapping from a space of functions into real or complex numbers:

$$\begin{aligned} \mathcal{F} : D &\rightarrow \mathbb{R} \\ q &\rightarrow \mathcal{F}[q] \end{aligned}$$

We will use linear and bilinear functionals, here an example:<sup>3</sup>

$$\begin{aligned} \mathcal{F}_1[q] &\equiv \int_{\mathcal{D}} dx f(x) q(x) \equiv (f, q) \\ \mathcal{F}_2[q] &\equiv \int_{\mathcal{D}_x} \int_{\mathcal{D}_y} dx dy q(x) K(x, y) q(y) \equiv (q, K, q) \end{aligned}$$

where  $f(x)$  and  $K(x, y)$  are given functions.

<sup>1</sup>E.g.  $\mathcal{L}_2(\mathbb{R}^4)$  with  $x \equiv (x_0, x_1, x_2, x_3)$ ,  $q(x)$ .

<sup>2</sup>We shall say that a quantity  $z$  is a functional of the function  $x(t)$  in the interval  $(a, b)$  when it depends on all the values taken by  $x(t)$  when  $t$  varies in the interval  $(a, b)$ ; or, alternatively, when a law is given by which to every function  $x(t)$  defined within  $(a, b)$  (the independent variable within a certain functional field) there can be made to correspond one and only one quantity  $z$ , perfectly determined, and we shall write  $z = \mathcal{F}[x(t)]_a^b$ .

<sup>3</sup> $(\cdot, \cdot)$  is the scalar product referred to the domain space of his input functions, e.g. if  $a, b \in \mathbb{C}^n$   $(a, b) = \sum_{i=1}^n a_i^* b_i$ . If  $a, b \in \mathcal{L}_2(\mathbb{R})$  then  $(a, b) \equiv \int_D a(x) b(x) dx$ . For this part we assuming function to be real in order to simplify the notation.

### Functional Representations

For many purpose it could be useful to represent functionals as functions of an infinite number of variables. Let's consider the following two ways:

i) replace the argument of  $\mathcal{F}(q)$  with its expansion in an orthonormal basis

$$\mathcal{F}[q] \rightarrow \mathcal{F}\left[\sum_n q_n \phi_n\right] = \hat{\mathcal{F}}(q_1, q_2, \dots)$$

Thus, for a given set of  $[\phi_n]$ ,  $\mathcal{F}[q]$  is uniquely determined by the infinite set of expansion coefficient  $[q_n]$ . For example

$$\mathcal{F}_1[q] \rightarrow \hat{\mathcal{F}}_1(q_1, q_2, \dots) = \sum_n q_n f_n \quad , \quad f_n \equiv \int dx \phi_n(x) f(x) = (\phi_n, f)$$

$$\mathcal{F}_2[q] \rightarrow \hat{\mathcal{F}}_2(q_1, q_2, \dots) = \sum_{n,m} q_n q_m K_{nm} \quad , \quad K_{nm} \equiv \int \int dx dy \phi_n(x) \phi_m(y) K(x, y) = (\phi_n, K, \phi_m)$$

ii) divide the domain of variation of  $q$  [e.g.  $\mathbb{R}^4$ ] into infinitely many small cubes [or "hypercubes"] of volume  $\tau$ , in each cube choose an internal point  $x^{(i)}$ . The function  $q(x)$  will be specified approximately by giving the value of it at each  $x^{(i)}$  as follow:  $q(x^{(i)}) \equiv q_i$ . This procedure is called the "pixelization" of space domain. In the limit  $\tau \rightarrow 0$  this specification becomes exact. When the argument of  $\mathcal{F}[q]$  is replaced by the infinite dimensional array  $(\dots, q_i, q_{i+1}, \dots)$  it becomes an ordinary function of the infinite set of variables  $\{q_i\}$ :

$$\mathcal{F}[q] \rightarrow \hat{\mathcal{F}}(\dots, q_i, q_{i+1}, \dots)$$

For example, for  $\tau \rightarrow 0$ :

$$\mathcal{F}_1[q] \rightarrow \hat{\mathcal{F}}_1(\dots, q_i, q_{i+1}, \dots) = \sum_i q_i f_i \quad , \quad f_i \equiv f(x^{(i)}) \tau$$

$$\mathcal{F}_2[q] \rightarrow \hat{\mathcal{F}}_2(\dots, q_i, q_{i+1}, \dots) = \sum_{i,j} q_i q_j K_{i,j} \quad , \quad K_{ij} \equiv K(x^{(i)}, y^{(j)}) \tau^2$$

A functional  $\mathcal{F}[q]$  may sometimes be represented by a series of functionals having the general form:

$$\mathcal{F}[q] = \sum_{n=0}^{\infty} \frac{1}{n!} \int dx_1 \int dx_2 \cdots \int dx_n f(x_1, \dots, x_n) q(x_1) \cdots q(x_n)$$

Examples (exponential and gaussian):

$$\mathcal{F}[q] = e^{(f,q)} \equiv \exp\left(\int dx f(x) q(x)\right) = \sum_{n=0}^{\infty} \frac{1}{n!} \left[ \int dx f(x) q(x) \right]^n$$

$$\mathcal{F}[q] = e^{(q,K,q)} \equiv \exp\left(\int dx dy q(x) K(x, y) q(y)\right) = \sum_{n=0}^{\infty} \frac{1}{n!} \left[ \int \int dx dy q(x) K(x, y) q(y) \right]^n$$

## 2.2 Functional Calculus

### Functional Differentiation

Many of the functions which appear in physics can be differentiated with respect to their arguments. It would be useful to extend the idea of differentiation to functionals as well.

The derivative of a functional  $\mathcal{F}[q]$  ought to tell us how much  $\mathcal{F}[q]$  changes for a given change in  $q$  by a small function  $\eta(x)$ :  $q \rightarrow q + \eta$ . To first order in  $\eta$  the change of  $\mathcal{F}$  will be some linear functional of  $\eta$ :  $\int dy f(y) \eta(y)$ .  $f(y)$  will be called the functional derivative of  $\mathcal{F}[q]$  due to a variation of  $q$  at a generic point  $y$ :

$$\delta\mathcal{F} = \mathcal{F}[q + \eta] - \mathcal{F}[q] \equiv \int \frac{\delta\mathcal{F}}{\delta q(y)} \eta(y) dy + \mathcal{O}(\eta^2) \quad f(y) \equiv \frac{\delta\mathcal{F}}{\delta q(y)} \quad (2.1)$$

We see the analogy with the expression for the change of a function of  $n$  variables, by ordinary calculus:

$$\delta F(x_1, \dots, x_n) = \sum_{j=1}^n \left[ \frac{\partial F}{\partial x_j} \delta x_j + \mathcal{O}(x_j^2) \right]$$

So the variable  $y$  can be seen as a continuous index, which is integrated instead of summed. Note that the dimensions of the functional derivative are not those of the functional divided by the function, as the notation would suggest, but those of the functional divided by the product of the function and its argument. An alternative definition of  $\frac{\delta F}{\delta q}$  comes from the definition of  $\mathcal{F}[q]$  as an ordinary function  $\hat{\mathcal{F}}(\dots, q_i, q_{i+1}, \dots)$ . If the argument is varied as  $q_i \rightarrow q_i + \eta_i$  therefore:

$$\begin{aligned} \delta\mathcal{F} &= \hat{\mathcal{F}}(\dots, q_i + \eta_i, q_{i+1} + \eta_{i+1}, \dots) - \hat{\mathcal{F}}(\dots, q_i, q_{i+1}, \dots) = \sum_i \frac{\partial \hat{\mathcal{F}}}{\partial q_i} \eta_i \\ &= \sum_i \left( \frac{1}{\tau} \frac{\partial \hat{\mathcal{F}}}{\partial q_i} \right) \eta_i \tau = \sum_i f_i \eta_i \tau \end{aligned}$$

in the limit  $\tau \rightarrow 0$ , the sum becomes an integral, and the functional derivative is defined starting from partial derivatives, if the limit exists:

$$\int dy f(y) \eta(y) = \lim_{\tau \rightarrow 0} \lim_{x^{(i)} \rightarrow y} \sum_i f_i \eta_i \tau \quad ; \quad \frac{\delta\mathcal{F}}{\delta q(y)} = \lim_{\tau \rightarrow 0} \lim_{x^{(i)} \rightarrow y} \frac{1}{\tau} \frac{\partial \hat{\mathcal{F}}}{\partial q_i}(x_i)$$

### Chain Rule

A very important rule, which we shall now derive, is the chain rule for functional differentiation. Suppose that  $\mathcal{F}$  is a functional of a function  $g$ , and that this function  $g$  is itself a functional of a function  $f$ .<sup>4</sup> Now suppose that  $f(x)$  changes, to  $f(x) + \eta(x)$  where  $\eta(x)$  is small. Then  $g(y)$  will also change, let us say to  $g(y) + \chi(y)$ , and  $\mathcal{F}$  will change to  $\mathcal{F} + \delta\mathcal{F}$ . From (2.1) these changes are related by

$$\begin{aligned} \delta\mathcal{F} &= \int dy \frac{\delta\mathcal{F}}{\delta g(y)} \chi(y) \\ \chi(y) &= \int dz \frac{\delta g(y)}{\delta f(z)} \eta(z) \end{aligned}$$

<sup>4</sup>Indeed, we have seen that the functional derivative  $\delta\mathcal{F}/\delta f(x_0)$  is both a functional of  $f$  and a function of  $x_0$ .

Combining these two equations and reversing the order of the  $y$ - and  $z$ - integrations gives

$$\delta\mathcal{F} = \int dz \left[ \int dy \frac{\delta\mathcal{F}}{\delta g(y)} \frac{\delta g(y)}{\delta f(z)} \right] \eta(z) \quad (2.2)$$

Comparing (2.1) with (2.2), it follows that the functional derivative of  $\mathcal{F}$  with respect to  $f$  is

$$\frac{\delta\mathcal{F}}{\delta f(z)} = \int dy \frac{\delta\mathcal{F}}{\delta g(y)} \frac{\delta g(y)}{\delta f(z)} \quad (2.3)$$

This is the chain rule for functional differentiation.

### Taylor Series Expansion

If  $\mathcal{F}[q]$  has derivative of all orders then it can be Taylor expanded as

$$\mathcal{F}[q + \delta q] = \mathcal{F}[q] + \sum_{n=1}^{\infty} \frac{1}{n!} \int \cdots \int dx_1 \cdots dx_n \frac{\delta^n \mathcal{F}}{\delta q(x_1) \cdots \delta q(x_n)} \delta q(x_1) \cdots \delta q(x_n)$$

### Linear Transformation

Consider a mapping  $q(x) \rightarrow q'(x)$  of the linear form:  $q(x) = \int dy K(x, y) q'(y)$ , if this transformation possesses an inverse:  $q'(x) = \int dy K^{-1}(x, y) q'(y)$ , using these two expressions together we get:

$$\begin{aligned} q(x) &= \int \int dy dz K(x, y) K^{-1}(y, z) q(z) \\ q'(x) &= \int \int dy dz K^{-1}(x, y) K(y, z) q'(z). \end{aligned}$$

These equations imply:

$$\int dy K(x, y) K^{-1}(y, z) = \int dy K^{-1}(x, y) K(y, z) = \delta(x - z).$$

### Legendre Transformation

Let be  $\mathcal{F}[q]$  a functional, we define the conjugate momentum such that:

$$p(y) = \frac{\delta\mathcal{F}[q]}{\delta q(y)}$$

then define the functional  $\mathcal{G}[p]$  such that:

$$\mathcal{G}[p] = \mathcal{F}[q] - (q, p)$$

this functional  $\mathcal{G}[p]$  is called the "Legendre transform of  $\mathcal{F}[q]$ ". This transformation eliminates  $q(x)$  in favour of  $p(x)$  by the use of its definition, so that  $\mathcal{G}$  is a functional of  $p$  alone.

### Functional Integration

The term "functional integration" refers to the following operation: we consider a set  $\{q\}_n$  of functions, (which may be required to satisfy boundary conditions or some other restriction, e.g.  $q_n \in \mathcal{L}_2(\mathcal{R}) \forall n$ )<sup>5</sup>, work out the value of some functional  $\mathcal{F}$  for every function in the set, and add up all these values. The result is called the functional integral of  $\mathcal{F}$  over the set  $\{q\}_n$ .

This idea is analogous to ordinary integration; the ordinary integral  $\int f(x, y) dx dy$  is the sum of the values taken by  $f$  as its argument moves over a grid of points, times the size  $dx dy$  of each cell of the grid; a functional integral is the sum of the values taken by a functional as its argument moves over some set of functions, times some measure of the 'volume' of function space associated with each function.

Now it is easy in principle to integrate over a single real variable, or a finite number of such variables. But integrating over a set of functions is another matter altogether. In general such a set will be infinite-dimensional, in that an infinite number of real numbers are needed to parameterize its elements. Then this infinite number of parameters must be integrated over, which will involve an infinite number of integrals. This is not obviously well defined, so some care is needed.

Accordingly functional integration may be defined as an integration over infinitely many variables:

$$\int \mathcal{D}q \mathcal{F}[q] \equiv \int \cdots \int \hat{\mathcal{F}}(q_1, q_2, \dots, q_n, \dots) \prod_{i=1}^{\infty} dq_i.$$

Assuming that the integral on the right hand side exists. The notation " $\int \mathcal{D}q$ " means "integrate over all possible values of  $q(x)$ ".<sup>6</sup>

### Change of Integration Variable

Suppose we replace the function  $q(x)$  in a functional integral with a new function  $q'(x)$  which is the result of a linear and symmetric transformation:

$$q(x) = \int dy K(x, y) q'(y) \quad , \quad K(x, y) = K(y, x).$$

We get the new integral

$$\int \mathcal{D}q \mathcal{F}[q] \rightarrow \int \mathcal{D}(K, q') \mathcal{F}[(K, q')]$$

and you can demonstrate that the relation is the following:<sup>7</sup>

$$\int \mathcal{D}q \mathcal{F}[q] = (\mathcal{D}et K) \int \mathcal{D}(K, q') \mathcal{F}[(K, q')].$$

<sup>5</sup>Integrating over all possible functions does not have any obvious physical meaning, and it is not (in general) mathematically defined, the concept of integrating over a restricted set  $\{q\}_n$  of functions, compatible with the physics of the particular problem, is well-defined and useful.

<sup>6</sup>What possible use can functional integration have? In physical terms, the smallest length scales which have been explored are of the order of  $10^{-19}$ m. This is very short, but still a lot bigger than nothing at all. The idea of a continuous function of position can only ever be an approximation to what is known about the physical world.

<sup>7</sup> $\mathcal{D}et K$  is a functional determinant. **Something to explain the functional determinant**

## 2.3 Gaussian Functional integral

We want to calculate a Gaussian functional integrals by analogy (CON UN LIMITE) with ordinary integrals. For a single variable we have the well known result:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\lambda x^2} dx = \frac{1}{\sqrt{\lambda}} \quad \lambda > 0 \quad (2.4)$$

In  $n$  dimensions we have

$$(2\pi)^{-n/2} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} dx_1 dx_2 \cdots dx_n \exp\left(-\frac{1}{2} \sum_{i,j=1}^n A_{ij} x^i x^j\right) = (\det A)^{-1/2}$$

where  $A_{ij}$  are the element of a real and symmetric matrix  $K$  which is required to be positive definite.

To prove this result we consider a more general integral:

$$\mathcal{Z}[b] = \mathcal{N} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} dx_1 dx_2 \cdots dx_n \exp\left(-\frac{1}{2} \sum_{i,j=1}^n A_{ij} x^i x^j + i \sum_{j=1}^n b_j x^j\right)$$

with  $\mathcal{N}$  a constant to be specified and  $b_j$  real. Using a vector notation and defining  $y$  and  $z$  as

$$y \equiv iA^{-1}b \quad x = y + z.$$

Using the symmetries of scalar product  $(\cdot, \cdot)$  and our definitions we have

$$\begin{aligned} (y, A, z) + (z, A, y) &= 2(z, A, y) = -2(z, b) \\ (y, A, y) &= (A^{-1}b, A, A^{-1}b) = (A^{-1}b, b) = (b, A^{-1}b) = (b, A^{-1}, b) \\ (b, y) &= -(b, A^{-1}b) = -(b, A^{-1}, b). \end{aligned}$$

We can rewrite the exponent in

$$\begin{aligned} -\frac{1}{2}(x, A, x) + i(b, x) &= \frac{1}{2}[(z, A, z) + (y, A, z) + (z, A, y) + (y, A, y)] + (b, z) + (b, y) \\ &= -\frac{1}{2}[(z, A, z) + (b, A, b)]. \end{aligned}$$

Then the change of variables from  $x$  to  $z$  transform  $\mathcal{Z}[b]$  to

$$\mathcal{Z}[b] = \mathcal{N} e^{-\frac{1}{2}(b, A^{-1}, b)} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} dz_1 dz_2 \cdots dz_n e^{-\frac{1}{2}(z, A, z)}.$$

By an appropriate orthogonal transformation  $z_i = \sum_k l_{ik} \xi_k$  it is possible to transform the quadratic form  $\sum_{j,k} A_{jk} z^j z^k$  to a diagonal form  $\sum_j \lambda_j \xi_j^2$ , with  $\lambda_j$  the eigenvalue of  $A$ . Since the Jacobian of an orthogonal transformation is 1 we have  $\prod_{j=1}^n dz_j = \prod_{j=1}^n d\xi_j$ , and using the  $\exp(\cdot)$  properties the  $n$ -dimensional integral factorize, so we can use the 1-dimensional integral result (2.4).  $\mathcal{Z}[b]$  becomes

$$\mathcal{Z}[b] = \mathcal{N} e^{-\frac{1}{2}(b, A^{-1}, b)} \prod_{j=1}^n \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\lambda_j \xi_j^2} d\xi_j = \mathcal{N} e^{-\frac{1}{2}(b, A^{-1}, b)} \prod_{j=1}^n \left(\frac{2\pi}{\lambda_j}\right)^{(1/2)}$$

Since  $\prod_{j=1}^n \lambda_j \equiv \det(A)$  and if we fix  $N$  such that  $\mathcal{Z}[0] = 1$ , therefore  $\mathcal{N} = (2\pi)^{-n/2}(\det A^{1/2})$ . So we obtain the "generating functional"

$$\mathcal{Z}[b] = \exp\left[-\frac{1}{2}(b, A^{-1}, b)\right].$$

In the  $n \rightarrow \infty$  limit we get from  $\det K \rightarrow \mathcal{D}et K$ , a functional determinant and  $\mathcal{Z}[b]$  gets the same form. But the limit of the previous formula can be just a representation of a functional integral where  $\lim_{n \rightarrow \infty} \prod_{j=1}^n dx_j \rightarrow [\mathcal{D}q]$ . We have<sup>8</sup>

$$\lim_{n \rightarrow \infty} (x, K, x) \rightarrow \int dx dy q(x) K(x, y) q(y).$$

So in this limit the coordinates' vector becomes the function  $x_i \rightarrow q(x)$  while the index  $i$  becomes a continuous index  $i \rightarrow x$ , and

$$\begin{aligned} \lim_{n \rightarrow \infty} i(b, x) &\rightarrow i \int dx J(x) q(x) \\ \lim_{n \rightarrow \infty} (b, K^{-1}, b) &\rightarrow \int dx dy J(x) K^{-1}(x, y) J(y) \end{aligned}$$

the vector  $b$  becomes the function  $J: b_i \rightarrow J(x)$ , which is usually called the "external source", it is a generic (real) functional of  $x$ , which is said to be "linearly coupled" with the "field"  $q(x)$ .  $K^{-1}$  is defined by

$$\int dy K(x, y) K^{-1}(y, z) = \delta(x - z).$$

If we choose  $\mathcal{N}$  to be such that  $\mathcal{Z}[J] = 1$  we can write

$$\begin{aligned} \mathcal{Z}[J] &= \int \mathcal{D}q \exp\left[-\frac{1}{2} \int dx dy q(x) K(x, y) q(y) + i \int dx J(x) q(x)\right] \\ &= \exp\left[-\frac{1}{2} \int dx dy J(x) K^{-1}(x, y) J(y)\right] \\ &= e^{-\frac{1}{2}(J, K^{-1}, J)} \equiv e^{-\frac{1}{2}J^\dagger K^{-1}J}. \end{aligned} \tag{2.5}$$

The functional integral  $\int \mathcal{D}q \mathcal{F}[q]$  can be called, in a somewhat imprecise way, a *path integral*.

---

<sup>8</sup>Bad notation: On the left  $x$  is the configuration vector, summed over his index  $\alpha = 1, \dots, N$ ; on the right  $q$  is the field configuration which is integrated over  $x$ , the coordinate vector.  $x_i \rightarrow q_i \equiv q^{(x_i)}$ . And  $(x, K, x) = \sum_{\alpha=0}^N \sum_{\beta=0}^N x_\alpha K^{\alpha\beta} x_\beta$ ,

## 2.4 Probability Density Functional

The quantity  $[\mathcal{D}q] \exp\left(-\frac{1}{2}(q, k, q)\right)$  which appears in the Gaussian functional integral, if integrated gives  $\mathcal{Z}[0]$ , has all the relevant properties to be a probability (density) functional:

- i) it is semi-positive definite.
- ii) it can be normalized, since  $\mathcal{Z}[0] = 1$ .
- iii) it goes to zero as  $q \rightarrow \pm\infty$ .

We can then formally write:

$$d\mathcal{P}[q] = \mathcal{D}q \mathcal{P}[q] = \mathcal{D}q e^{\left(-\frac{1}{2}(q, k, q)\right)}.$$

This is actually a multivariate Gaussian for an infinite and continuous number of variables. We can then make the following ansatz: the statistical (ensemble) average of a generic functional  $\mathcal{F}[q]$  can be obtained as

$$\langle \mathcal{F}[q] \rangle \equiv \int \mathcal{D}q e^{\left(-\frac{1}{2}(q, K, q)\right)} \mathcal{F}[q] = \int d\mathcal{P}[q] \mathcal{F}[q].$$

If  $q$  is gaussian distributed.<sup>9</sup> We can generalize considering a generic probability density functional  $\mathcal{P}[q]$  which satisfy i), ii) and iii) such that  $q(x)$  is distributed according to it. We define the expectation value of  $\mathcal{F}[q]$  as:

$$\langle \mathcal{F}[q] \rangle \equiv \int \mathcal{D}q \mathcal{P}[q] \mathcal{F}[q].$$

We can also calculate the n-point disconnected correlation functions by repeated functional differentiations of  $\mathcal{Z}[J]$

$$C^{(n)}(x_1, \dots, x_n) \equiv \langle q(x_1), \dots, q(x_n) \rangle = \frac{(-i)^n}{\mathcal{Z}} \frac{\delta^n \mathcal{Z}[J]}{\delta J(x_1) \cdots \delta J(x_n)} \Big|_{J=0}. \quad (2.6)$$

We can also define a set of reduced correlation functions for a random field  $q(x)$  by functional differentiation of  $\log \mathcal{Z}[J]$  :

$$C_c^{(n)}(x_1, \dots, x_n) = (-i)^n \frac{\delta^n \log \mathcal{Z}[J]}{\delta J(x_1) \cdots \delta J(x_n)} \Big|_{J=0}. \quad (2.7)$$

Another case in which is useful the introduction of the external source and functional differentiation is the expectation value of some functional: let  $\mathcal{F}[q]$  be expanded in a functional Taylor series

$$\mathcal{F}[q] = \sum_{n=1}^{\infty} \frac{1}{n!} \int \cdots \int dx_1 \cdots dx_n \frac{\delta^n \mathcal{F}}{\delta q(x_1) \cdots \delta q(x_n)} \Big|_{q=0} \delta q(x_1) \cdots \delta q(x_n).$$

Therefore, using (2.6)

$$\begin{aligned} \langle \mathcal{F}[q] \rangle &= \sum_{n=1}^{\infty} \frac{1}{n!} \int \cdots \int dx_1 \cdots dx_n \frac{\delta^n \mathcal{F}}{\delta q(x_1) \cdots \delta q(x_n)} \Big|_{q=0} C^{(n)}(x_1, \dots, x_n) \\ &= \sum_{n=1}^{\infty} \frac{(-i)^n}{n!} \int \cdots \int dx_1 \cdots dx_n \frac{\delta^n \mathcal{F}}{\delta q(x_1) \cdots \delta q(x_n)} \Big|_{q=0} \left( \frac{\delta}{\delta J(x_1)} \cdots \frac{\delta}{\delta J(x_n)} \right) \mathcal{Z}[J]. \end{aligned}$$

---

<sup>9</sup>Gaussian distributed means that the probability distribution of different configurations of the field  $q$  is a gaussian distributed around the most probable configuration.



Then we can formally write<sup>10</sup>

$$\langle \mathcal{F}[q] \rangle = \mathcal{F}\left[\left(-i\frac{\delta}{\delta J}\right)\right] \mathcal{Z}[J] \Big|_{J=0}. \quad (2.8)$$

In Appendix A there is an example of moments calculation for a Gaussian random field.

---

<sup>10</sup>This expression has meaning only watching at his Taylor expansion.



# Chapter 3

## Information Field Theory

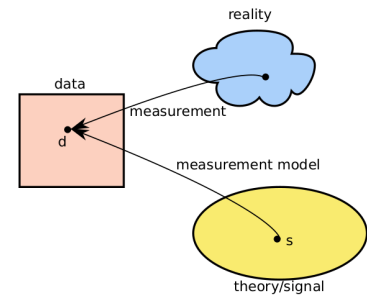
### 3.1 Information Theory

The approach we are going to discuss tells about how we can extract knowledge about physical reality starting from a set of data.

Data are the result of a measurement of some physical quantity and now, in our approach, statistical inference process, the causality between the physical world and the data should be inverted.

Some name convention: the physical state, our system configuration,  $\psi$ ; the signal, the specific aspect of the system that is measured,  $s$ ; the set of data we collect,  $d$ .

The measurement process maps the real state into the data space,  $\psi \rightarrow d[\psi]$ . We just have a model to describe reality and measurement process, a model that has unknown parameters, the signal  $s$ , which shall be determined by the data  $d$ .



### 3.2 Bayesian Inference

We are going to use Bayesian inference that is a statistical inference method based on the Bayesian interpretation of probability.

The system has some background configuration " $\psi$ ", which gives information on signal  $s$  and on measurement process producing data  $d$ :  $s = s[\psi]$  and  $d = d[\psi]$ .

Bayes' theorem states that the posterior probability density  $P(s|d)$ , which summarizes our knowledge on the field after the data is taken, is proportional to the product of the data likelihood  $P(d|s)$  for a given field configuration times its prior probability  $P(s)$ , which summarizes our a priori knowledge on the field, like that smooth configurations are more plausible than rough ones. Bayes' theorem reads<sup>1</sup>:

$$P(s|d) = \frac{P(d, s)}{P(d)} = \frac{P(d|s)P(s)}{P(d)} \quad (3.1)$$

---

<sup>1</sup>In the following  $\psi$  is assumed to be implicitly included among the conditionals of any probability.

$P(s)$  means the probability of the *variable*  $s$  to have the *value*  $s$  given the implicit background configuration information  $\psi$ ,  $P(s_{var} = s_{val}|\psi)$ , with  $s_{var}$  the unknown variable and  $s_{val}$  a concrete value. The joint probability  $P(d, s)$  is decomposed in likelihood and prior<sup>2</sup>. The prior  $P(s)$  summarizes our knowledge on  $s$  before measuring. The likelihood  $P(d|s)$  describes the measurement process. The evidence  $P(d) = \sum_s P(d, s)$  serves as normalization factor<sup>3</sup>:

$$\sum_s P(s|d) = \frac{1}{P(d)} \sum_s P(d, s) = \frac{\sum_s P(d, s)}{\sum_{s'} P(d, s')} = 1$$

### 3.3 Signal Inference Problem

In general, a signal inference problem tries to invert a measurement equation

$$d = f(s, \alpha) \tag{3.2}$$

that describes how the obtained data  $d$  depends on the unknown signal  $s$  and further nuisance parameter  $\alpha$ . Actually, we want to know  $s$  and are not interested in  $\alpha$ , but the latter also influences the measurement outcome. The problem is that the function  $f$  is not necessarily invertible in  $s$ , and that the nuisance parameters may not be known. This problem is particularly severe when the signal is a field, a continuous function  $s(x)$  over a manifold  $\Omega \in x$ .

### 3.4 Field Inference

A field is a function over some continuous space. The air temperature over Europe, the Earth's magnetic field, the galaxy distribution in the Milky Way, or the dark matter density in the Universe are all fields we want to know as accurately as possible. A physical field has an infinite number of degrees of freedom, as it has a value for each point in space. Knowing a field exactly from a finite number of measurements alone is therefore impossible since it is impossible to determine an infinite number of unknowns from a finite number of constraints, an exact field reconstruction from the data alone is impossible. Additional information is needed.

Additional information might be available from physical law, statistical symmetries, or smoothness properties known to be obeyed by the field. A unique field reconstruction might still be impossible, but the configuration space of possible field realization might be sufficiently constrained to single out a good guess for the field.

The combination of data and additional information is done in an information theoretically correct way by using probabilistic logic. Information field theory (IFT) is therefore information theory applied to fields, Bayesian reasoning with an infinite number of unknowns.

The main difference of IFT to the usual Bayesian inference is that the continuity of the physical space plays a special role. The fact that many physical fields do not exhibit arbitrary roughness due to their causal origins implies that field values at nearby locations are similar, and typically more so the closer the locations are. The consequent exploitation of any knowledge on the field correlation structure permits us to overcome the ill-posedness of the field reconstruction problem.

---

<sup>2</sup> $(d, s)$  means (d "and" s), so the joint probability  $P(d, s) = P(d|s)P(s)$

<sup>3</sup> $s$  e  $s'$  span all possible states.

### 3.5 Inference on Physical Fields

In our attempts to infer the properties of a system we are faced with the problem of how to interpret incomplete, imperfect and noisy data, draw our conclusions based on them, and quantify the uncertainties of our results. [ This is true, from astronomical observation, for using galaxy surveys to map the cosmic Large Scale Structure or for the interpretation of the Cosmic Microwave Background, as well as for many experiments in physical laboratories and compilations of geological, economical, sociological, and biological data about our planet.]

Information theory, which is based on probability theory and the Bayesian interpretation of missing knowledge as a probabilistic uncertainty, provides a suitable framework to handle such problems. Given a model for the Universe or the system under consideration IF permits to describe probabilistically all relevant processes involved in the measurement.<sup>4</sup>

The states of such model, denoted by the state variable  $\psi$ , are identified with the possible configurations of the system, the different physical realities. Each state has a probability  $P(\psi)$  assigned to it, the so-called prior information. This prior contains our knowledge about the system as we model it before any data are taken. For a given cosmological model, the prior may be the probability distribution of the different initial conditions of the Universe. Since our Universe is spatially extended, the state variable will in general contain one or several fields, which are functions over some coordinates  $x$ .

Also the measurement process is described by a data model which defines the so-called likelihood, the probability  $P(d|\psi)$  to obtain a specific data set  $d$  given the physical condition  $\psi$ . For instance, in case the outcome  $d$  of the measurement is deterministic  $P(d|\psi) = \delta(d - d[\psi])$ , where  $d[\psi]$  is the functional dependence of the data on the state. In general the probability distribution function of the data is given in terms of a phase space or path integral over all possible realization of  $\psi$ :<sup>5</sup>

$$P(d) = \int \mathcal{D}\psi P(d|\psi)P(\psi) \quad (3.3)$$

We are usually not interested in the whole state of the system, but only in some specific aspects of it, which we call signal  $s = s[\psi]$ . (parentesi descrittiva che eviterei: ) Since the signal does not contain the full physical state, any physical degree of freedom which is not present in the signal but influences the data will be received as probabilistic uncertainty, or what is called *noise*. The probability distribution function of the signal, its prior  $P(s)$

$$P(s) = \int \mathcal{D}\psi \delta(s - s[\psi])P(\psi) \quad (3.4)$$

is related to the one of the data via the joint probability

$$P(d, s) = \int \mathcal{D}\psi \delta(s - s[\psi])P(d|\psi)P(\psi) \quad (3.5)$$

We can derive signal likelihood  $P(d|s)$  and posterior  $P(s|d)$  from the joint probability:

$$\begin{aligned} P(d, s) &= P(d|s)P(s) = P(s|d)P(d) \\ P(d|s) &= P(d, s)/P(s) \\ P(s|d) &= P(d, s)/P(d) \end{aligned} \quad (3.6)$$

---

<sup>4</sup>The term Bayesian field theory was proposed originally by Lemm for field inference. This term, however, does not follow the convention to name field theories after subjects and not people. We do not talk about Maxwell, Einstein, or Feynman field theory, but about electromagnetic, gravitational, and quantum field theories.

<sup>5</sup>Path integral is defined in the previous chapter.

Before the data are available, the phase space of interest is spanned by the direct product of all possible signals  $s$  and data  $d$ , and all regions with nonzero  $P(d, s)$  are of potential relevance. Once the actual data  $d_{obs}$  have been taken, only a submanifold of this space, as fixed by the data, is of further relevance. The probability function over this subspace is proportional to  $P(d = d_{obs}, s)$ , and needs just to be renormalized by dividing by

$$\begin{aligned} \int \mathcal{D}s P(d_{obs}, s) &= \int \mathcal{D}s \int \mathcal{D}\psi \delta(s - s[\psi]) P(d_{obs}|\psi) P(\psi) \\ &= \int \mathcal{D}\psi P(d_{obs}|\psi) P(\psi) = P(d_{obs}) \end{aligned} \quad (3.7)$$

$P(d_{obs})$  is the unconditioned probability (or evidence) of that data. Thus, we find the resulting information of the data to be the posterior distribution

$$P(s|d_{obs}) = P(d_{obs}, s)/P(d_{obs}) \quad (3.8)$$

This posterior is the fundamental mathematical object from which all our deductions have to be made. It is related via Bayes' theorem (1) to the usually better accessible signal likelihood

$$P(s|d) = P(d|s)P(s)/P(d)$$

which follows from equations (3.6). The normalization term in Bayes's theorem, the evidence  $P(d)$ , is now also fully expressed in terms of the joint probability of data and signal, and (2) becomes

$$P(d) = \int \mathcal{D}s P(d, s)$$

and the underlying physical field  $\psi$  is hidden at this stage in the formalism. The evidence plays a central role in Bayes inference, since it is the likelihood of all the assumed model parameters.

### 3.6 Signal and Noise

If signal and data depend on the same underlying physical properties, there may be correlations between the two, which can be expressed in terms of signal response  $R$  and noise  $n$  of the data as<sup>6</sup>

$$d[s] = R[s] + n_s$$

where the notation denotes the different dependence of response and noise on the signal, in order to highlight that the response should embrace most of the reaction of the data to the signal, whereas the noise should be as independent as possible.

The response is therefore the part of the data which correlates with the signal

$$R[s] \equiv \langle d \rangle_{(d|s)} \equiv \int \mathcal{D}d d P(d|s)$$

<sup>6</sup>The response  $R$  maps the continuous field  $s$  onto a discrete data vector  $d$ . For individual entries of this data vector the equation reads  $d_i = \int_{\Omega} dx R_i(x) s(x) + n_i$ . Here  $i \in \{1, \dots, N_{data}\}$  with  $N_{data} \in \mathbb{N}$  is the index of the data vector's entries.  $\Omega$  is the physical manifold the signal is defined on.

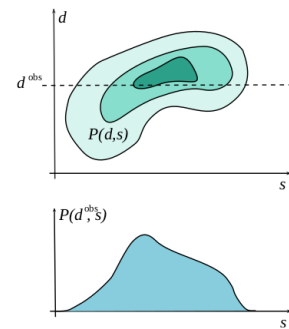


Figure 3.1:  $P(d_{obs})$

and the noise is just defined as the remaining part which does not:

$$n_s \equiv d - R[s] = d - \langle d \rangle_{(d|s)}.$$

Although the noise might depend on the signal, it is per definition linearly uncorrelated to it<sup>7</sup>

$$\langle n_s s^\dagger \rangle_{(d|s)} = (\langle d \rangle_{(d|s)} - R[s]) s^\dagger = 0$$

whereas higher-order correlation might well exist.

These definitions were chosen to be close to the usual language in signal processing and data analysis. They permit one to define signal response and noise for an arbitrary choice of the signal  $s[\psi]$ . No direct causal connection between signal and data is needed in order to have a nontrivial response, since both variables just need to have some couplings to a common subaspect of  $\psi$ .

The above definition of response and noise is however not unique, even for a fixed signal definition, since any data transformation  $d' = T[d]$  can lead to different definitions, as seen from<sup>8</sup>

$$R'[s] \equiv \langle d' \rangle_{(d|s)} = \langle T[d] \rangle_{(d|s)} \neq T[\langle d \rangle_{(d|s)}] = T[R[s]]$$

Thus, the concepts of signal response and therewith defined noise depend on the adopted coordinate system in the data space. This coordinate system can be changed via a data transformation  $T$ , and the transformed data may exhibit a better or worse response to the signal.

Information theory aids in designing a suitable data transformation, so that the signal response is maximal, and the signal noise is minimal, permitting the signal to be best recovered. Then we may aim for an optimal  $T$ , which yields

$$T[d] = \langle s \rangle_{(s|d)} \quad (3.9)$$

We define the posterior average of the signal,  $m_d = \langle s \rangle_{(s|d)}$  to be the map of the signal given the data  $d$  and call  $T$  a map making algorithm if it fulfills (3.9) at least approximately. One may require that the signal response of a map-making algorithm

$$R_T[s] \equiv \langle T[d] \rangle_{(d|s)}$$

is positive definite with respect to signal variations as stated by

$$\frac{\delta R_T[s]}{\delta s} \geq 0$$

This ensures that a map-making algorithm will respond with a non-negative correlation of the map to any signal feature, with respect to the noise ensemble.

The fidelity of a signal reconstruction can be characterized by the quadratic signal uncertainty

$$\sigma_{T,d}^2 = \langle (s - T[d])(s - T[d])^\dagger \rangle_{(s|d)}$$

averaged over typical realizations of signal and noise. Of special interest is the trace of this

$$Tr(\sigma_{T,d}^2) = \int dx \langle |s_x - T_x[d]|^2 \rangle_{(s|d)} \quad (3.10)$$

<sup>7</sup> † means complex conjugation and transposing of a vector or matrix.

<sup>8</sup> Exceptions are some unique relations between signal and state,  $P(\psi|s) = \delta(\psi - \psi[s])$ , and maybe a few other very special cases. In general  $T$  will be a nonlinear operation on the data. For example data dependence of the uncertainty is a common feature of nonlinear inference problems.

since it is the expectation value of the squared distance, in Lebesgue  $L_2$  space, between a signal reconstruction and the underlying signal.

Requesting a map making algorithm to be optimal with respect to eq (3.10) implies

$$T[d] = \langle s \rangle_{(s|d)}$$

and therefore to be optimal in an information theoretical sense according to eq (3.9).

The uncertainty  $\sigma_{T,d}^2$  depends on  $d$  since in Bayesian inference one averages over the posterior, which is conditional to the data.

The Frequentist uncertainty estimate, which is the expected uncertainty of any estimator before the data are obtained, is given by an average over the joint probability function:

$$\sigma_{T,d}^2 = \langle (s - T[d])(s - T[d])^\dagger \rangle_{(d,s)} \quad (3.11)$$

The latter is a good quantity to characterize the overall performance of an estimator, whereas  $Tr(\sigma_{T,d}^2)$  is a more precise indicator of the actual estimator performance for a given data set.



### 3.7 Hamiltonian Formalism

We argued that the posterior  $P(s|d)$  (3.8) contains all available information on the signal, Although the posterior might not be easily accessible mathematically, we assume in the following that the prior  $P(s)$  of the signal before the data are taken as well as the likelihood of the data given a signal  $P(d|s)$  are known or at least can be Taylor-Frèchet expanded around some reference field configuration. Introducing the language of statistical field theory, Bayes's theorem permits one to express the posterior as<sup>9</sup>

$$P(s|d) = \frac{P(d, s)}{P(d)} = \frac{P(d|s)P(s)}{P(d)} = \frac{e^{-H[s]}}{\mathcal{Z}} \quad (3.12)$$

having defined the information Hamiltonian

$$H[s] \equiv H_d[s] \equiv -\log[P(d, s)] = -\log[P(d|s)P(s)]$$

and the evidence of the data, as

$$P(d) \equiv \int \mathcal{D}s P(d|s)P(s) = \int \mathcal{D}s e^{-H[s]} \equiv \mathcal{Z}$$

and the partition function  $\mathcal{Z} \equiv \mathcal{Z}_d$ . It is also convenient to include a moment generating function into the definition of the partition function

$$\mathcal{Z}[J] \equiv \int \mathcal{D}s e^{-H[s] + iJ^\dagger s} \quad (3.13)$$

This means that  $P(d) = \mathcal{Z} = \mathcal{Z}[0]$ , but also permits to calculate any moment and cumulants<sup>10</sup> of the signal field  $s$  via Frèchet differentiation as (2.6), (2.7)

$$\begin{aligned} C^{(n)}(x_1, \dots, x_n) &\equiv \langle s(x_1) \cdots s(x_n) \rangle_{(s|d)} = \frac{(-i)^n}{\mathcal{Z}} \frac{\delta^n \mathcal{Z}[J]}{\delta J(x_1) \cdots \delta J(x_n)} \Big|_{J=0} \\ C_c^{(n)}(x_1, \dots, x_n) &\equiv \langle s(x_1) \cdots s(x_n) \rangle_{(s|d)}^c = (-i)^n \frac{\delta^n \mathcal{Z}[J]}{\delta J(x_1) \cdots \delta J(x_n)} \Big|_{J=0} \end{aligned} \quad (3.14)$$

The assumption that  $H[s]$  can be expanded in a Taylor series permits one to write

$$H[s] = \frac{1}{2} s^\dagger D^{-1} s - J^\dagger s + H_0 + \sum_{n=3}^{\infty} \frac{1}{n!} \Lambda_{(n)}^\dagger s^n \quad (3.15)$$

the first three Taylor coefficients have special roles. The constant  $H_0$  is fixed by the joint probability density of signal and data normalization condition.  $J$ , the linear coefficient, is called information source. This term is usually directly and linearly related to the data. The quadratic coefficient,  $D^{-1}$ , defines the information propagator  $D(x, y)$ , which propagates information on the signal at  $y$  to location  $x$ , and thereby, for example, permits one to partially reconstruct the signal at locations where no data were taken. The anharmonic tensors  $\Lambda_{(n)}$  create interactions between the modes of the free, harmonic theory. Since this free theory will be the basis for the full interaction theory, we first investigate the case  $\Lambda_{(n)} = 0$ .

<sup>9</sup>The last term is just a renaming of the numerator and denominator of the first fraction, which highlights the connection to statistical mechanics.

<sup>10</sup>Correlation functions are also called moments, and connected correlation functions are called cumulants.

### 3.8 Free Theory

#### Gaussian Data Model

The simplest data model to be described assumes a signal with Gaussian prior  $P(s)$ , linear response on signal  $R[s]$ , signal independent and Gaussian distributed noise  $P(n|s)$ .

The signal Gaussian prior is

$$P(s) = \mathcal{G}(s, S) \equiv \frac{1}{|2\pi S|^{1/2}} \exp\left(-\frac{1}{2}s^\dagger S^{-1}s\right) \quad (3.16)$$

where  $S = \langle ss^\dagger \rangle$  is the signal covariance..

The signal is assumed to be provided by nature and our measurement device according to a linear data model

$$d = Rs + n \quad (3.17)$$

here the response  $R[s] = Rs$  is linear, and the noise  $n_s = n$  is signal independent<sup>11</sup>. Typically, the data space is discrete, whereas the signal space may be continuous. In that case each  $d_i$  data point is given by

$$d_i = \int dx R_i(x)s(x) + n_i$$

In this simple model we assume, the noise to be signal independent and Gaussian, and therefore distributed as

$$P(n|s) = \mathcal{G}(n, N)$$

where  $N = \langle nn^\dagger \rangle$  is the noise covariance matrix.

Since the noise is the difference of the data to the signal response,  $n = d - Rs$ , the likelihood of the data is given by

$$P(d|s) = P(n = d - Rs) = \mathcal{G}(d - Rs, N). \quad (3.18)$$

Then the Hamiltonian of the Gaussian model is

$$\begin{aligned} H_{\mathcal{G}}[s] &\equiv -\log P(d, s) = -\log (P(d|s)P(s)) \\ &= -\log (\mathcal{G}(d - Rs, N) \mathcal{G}(s, S)) \\ &= H[d|s] + H[s] \end{aligned}$$

then using log properties and collecting the terms with same order in  $s$  it becomes

$$\begin{aligned} H_{\mathcal{G}}[s] &= \frac{1}{2} \left( s^\dagger R^\dagger N^{-1} - s^\dagger j - j^\dagger s + d^\dagger N^{-1} d + \ln |2\pi N| \right) + \frac{1}{2} \left( s^\dagger S^{-1} s + \ln |2\pi S| \right) \\ &= \frac{1}{2} s^\dagger D^{-1} s - j^\dagger s + H_0^{\mathcal{G}} \end{aligned} \quad (3.19)$$

having defined  $D$ , propagator of the free theory,

$$D = [S^{-1} + R^\dagger N^{-1} R]^{-1} \quad (3.20)$$

and  $j$  the information source as<sup>12</sup>

$$j = R^\dagger N^{-1} d \quad (3.21)$$

<sup>11</sup>The linear response matrix  $R$  of our instrument can contain window and selection functions, blurring effects, and even a Fourier transformation of the signal space, if our instrument is an interferometer

<sup>12</sup> $j$  reads:  $j(x) = \sum_{i,k} R_i^*(x) N_{ik}^{-1} d_k$ . in case of discrete data but continuous signal spaces.

depends linearly on the data in a response-over-noise weighted fashion. And the term

$$H_0^{\mathcal{G}} = \frac{1}{2}d^\dagger N^{-1}d + \frac{1}{2}\log(|2\pi S||2\pi N|)$$

contains all  $s$ -independent normalization constants.

Rewriting 3.19, reading  $j^\dagger s = \frac{1}{2}(j^\dagger s + s^\dagger j)$  and inserting the identity  $\mathbb{1} = D D^{-1} = D^{-1}D$ , exploiting that  $D$  is invertible, we get

$$\begin{aligned} H_{\mathcal{G}}[s] &\doteq \frac{1}{2}, [s^\dagger D^{-1}s - j^\dagger D D^{-1}s - s^\dagger D^{-1}Dj] \\ &\doteq \frac{1}{2}, [s^\dagger D^{-1}s - (Dj)^\dagger D^{-1}s - s^\dagger D^{-1}Dj] \end{aligned}$$

Defining  $m \equiv Dj$  and  $m^\dagger \equiv (Dj)^\dagger$ , using the fact that  $D$  is symmetric and real, and adding the  $s$ -independent term  $m^\dagger D^{-1}m$ , we obtain

$$H_{\mathcal{G}}[s] \doteq \frac{1}{2}(s - m)^\dagger D^{-1}(s - m). \quad (3.22)$$

The partition function of the free theory

$$\begin{aligned} \mathcal{Z}_{\mathcal{G}}[J] &= \int \mathcal{D}s e^{-H_{\mathcal{G}}[s] + J^\dagger s} \\ &= \int \mathcal{D}s \exp\left(-\frac{1}{2}s^\dagger D^{-1}s + (J + j)^\dagger s - H_0^{\mathcal{G}}\right) \end{aligned} \quad (3.23)$$

is a Gaussian path integral, which can be calculated exactly, yielding

$$\mathcal{Z}_{\mathcal{G}}[J] = |2\pi D|^{1/2} \exp\left(\frac{1}{2}s^\dagger D^{-1}s + (J + j)^\dagger D(J + j) - H_0^{\mathcal{G}}\right)$$

Then using the explicit partition function one can calculate the expectation of the signal given the data  $\langle s \rangle_d$ , which we call the map  $m_d$  generated by the data

$$\begin{aligned} m_d \equiv \langle s \rangle_d &= \frac{\delta \log \mathcal{Z}_{\mathcal{G}}[J]}{\delta J} \Big|_{J=0} \\ &= Dj \\ &= [S^{-1} + R^\dagger N^{-1}R]^{-1} R^\dagger N^{-1}d \\ &= (F_{WF})d \end{aligned} \quad (3.24)$$

Using (3.20) and (3.21). The last expression shows that the map is given by the data after applying a generalized Wiener filter,  $m_d = (F_{WF})d$ . The propagator  $D(x, y)$  describes how the information on the density field contained in the data at location  $x$  propagates to position  $y$ :<sup>13</sup>

$$m(y) = \int dx D(x, y)j(x) \quad (3.25)$$

The connected autocorrelation of the signal given the data is the propagator itself, according to (A.1)

$$\langle ss^\dagger \rangle_d^c = D = [S^{-1} + R^\dagger N^{-1}R]^{-1} \quad (3.26)$$

<sup>13</sup>In practice, one will use an iterative linear algebra method like the conjugate gradient method to solve numerically the equation  $D^{-1}m = j$  for  $m$  on a computer.

And all higher connected correlation functions are zero. Therefore, the signal given the data is a Gaussian random field around the mean  $m_d$  and with a variance of the residual error  $r = s - m_d$  provided by the propagator itself

$$\langle rr^\dagger \rangle = \langle ss^\dagger \rangle_d - \langle s \rangle_d \langle s^\dagger \rangle_d = \langle ss^\dagger \rangle_d^c = D$$

The posterior is therefor a Gaussian given by

$$P(s|d) = \mathcal{G}(s - m_d, D) \quad (3.27)$$

On the one hand it is the susceptibility of our mean field  $m$  to the force of the information source  $j$ , since  $m = Dj$ , on the other hand it describes the remaining a posteriori uncertainty  $D = \langle (s - m)(s - m)^\dagger \rangle_{(s|d)}$ .

### 3.9 Interacting Theory

If any of the assumptions of our Wiener filter theory scenario is violated, in that the signal response is non-linear, the field or the noise is non-Gaussian, the noise variance depends on the signal, or the noise or signal covariances are unknown and have to be determined from the data itself, the resulting information Hamiltonian will contain anharmonic terms. These terms couple the different eigenmodes of the information propagator and lead to an interacting field theory.

Assuming that the Hamiltonian can be Taylor-Fréchet expanded in the signal fields, the Hamiltonian can be written as  $H = H_{free} + H_{int}$ .<sup>14</sup>

$$\mathcal{H}[s] = \frac{1}{2} s^\dagger D^{-1} s - J^\dagger s + H_0 + \sum_{n=3}^{\infty} \frac{1}{n!} \Lambda_{(n)}^\dagger s^n \quad (3.28)$$

Let us assume that the interaction terms are small, usually this can be achieved by shifting the field values to  $s' = s - s_{cl}$  where  $s_{cl}$  is the minimum of the Hamiltonian, the classical field, or in inference language, the maximum a posteriori estimator. Expanding  $H(d, s') = H(d, s = s_{cl} + s')$  around  $s' = 0$  then often ensures small interaction terms around the origin.

In this case, it is possible to expand the mean field value, or any other quantity of interest, around its free theory value. Since the terms of such an expansion can become numerous and complex, this is best done diagrammatically.

### 3.10 Extrema Principle and MAP

The Hamiltonian permits one to ask for classical equations derived from an extremal principle. This is justified, on the one hand, as being just the result of a the saddle point approximation of the exponential in the partition function. On the other hand, the extrema principle is equivalent to the MAP estimator, which is quite commonly used for the construction of signal filters.<sup>15</sup>

The classical theory is expected to capture essential features of the field theory. However, if the field fluctuations are able to probe phase-space regions away from the maximum in which the

<sup>14</sup>Where  $\sum_{n=3}^{\infty} \frac{1}{n!} \Lambda_{(n)}^\dagger s^n$  reads  $\sum_{n=3}^{\infty} \frac{1}{n!} \int \cdots \int dx_1 \cdots dx_n \Lambda_{x_1 \cdots x_n}^{(n)} s(x_1) \cdots s(x_n)$ .

<sup>15</sup>An exhaustive introduction into and discussion of the MAP approximation to Gaussian and non-Gaussian signal fields is provided by Lemm [10].

Hamiltonian (or posterior) has a more complex structure, deviations between classical and field theory should become apparent.

Extremizing the free theory Hamiltonian (3.19), respect the signal field  $s$ , around the configuration  $s = m$ ,

$$\left. \frac{\delta H_g[s]}{\delta s} \right|_{s=m} = 0 \implies D^{-1} m - j = 0$$

we get the classical mapping equation  $m = D j$ , which is the same result found using the field theory result (3.24).

It is also possible to measure the sharpness of the maximum of the posterior by calculating the Hessian curvature matrix

$$\mathcal{H}[m] = \left. \frac{\delta^2 H_g[s]}{\delta s^2} \right|_{s=m} = D^{-1}.$$

In the Gaussian MAP approximation, the inverse of the Hessian is identical to the residual covariance

$$\langle r r^\dagger \rangle = \mathcal{H}^{-1}[m] = D$$

which for the pure Gaussian model is of course identical to the exact result, as given by the field theory.



# Chapter 4

## IFT for Cosmological Perturbations

In the previous chapter Information Field Theory was presented as a field theory approach to Bayesian inference, written in Statistical Mechanics formalism. In the Gaussian case achieved results are exact, if we deal with non Gaussian fields we get to interactive terms in the Hamiltonian, this set out of question the possibility of an analytical solution, so the way is to make a perturbative expansion using Feynman rules and solve the problem iteratively. The following chapter treats of an actual use of IFT.

### 4.1 Non-Gaussian CMB Fluctuations via $f_{nl}$ Theory

#### 4.1.1 Data Model

We investigate the so-called  $f_{nl}$  theory of local non-Gaussianities in the CMB temperature fluctuations. This problem has currently a high scientific relevance due to the strongly increasing availability of high fidelity CMB measurements, which permit one to constrain the physical conditions at very early epochs of the Universe.

Over the very uniform CMB sky with a mean temperature  $T_{CMB} = 2,728K$ , small temperature fluctuations on the level of

$$\delta T_{obs}^{\{I,E,B\}}/T_{CMB} \sim 10^{-\{5,6,7\}}$$

are observed in total intensity (I), in polarization E and B mode respectively. The weak B modes are mainly due to the lensing of E modes and some unknown level of gravity waves. We will disregard them in the following. These CMB temperature fluctuations are believed and observed to follow mostly a Gaussian distribution. However, inflation predicts some level of non-Gaussianity<sup>1</sup>.

As a first step, we assume that the data  $d$  are given and we want to reconstruct  $\varphi$  from them. The primordial, as well as some of the secondary CMB temperature fluctuations, are a response to the gravitational potential initially seeded during inflation. Since we are interested in primordial fluctuations, we write

$$d \equiv \delta T_{obs}^{\{I,E\}}/T_{CMB} = R\varphi + n \quad (4.1)$$

---

<sup>1</sup>Some of the secondary anisotropies imprinted by the LSS of the Universe via CMB lensing, the integrated Sachs-Wolfe and the Rees-Sciama effects should also have imprinted non-Gaussian signatures.

where  $\varphi$  is the three-dimensional, primordial gravitational potential, and  $R$  is the response on it of a CMB instrument, observing the induced CMB temperature fluctuations in intensity and E-mode polarization. These are imprinted by a number of effects, like gravitational redshifting, the Doppler effect, and anisotropic Thomson scattering.

The precise form of the response does not matter for a development of these concepts and can be inserted later<sup>2</sup>.

The noise  $n$  subsumes all deviation of the measurement from the signal response due to instrumental and physical effects, which are not linearly correlated with the primordial gravitational potential, such as detector noise, remnants of foreground signals, but also primordial gravitational wave contributions to the CMB fluctuations.

The small level of non-Gaussianity expected in the CMB temperature fluctuations is a consequence of some non Gaussianity in the primordial gravitational potential. Despite the lack of a generic non-Gaussian probability function, many of the inflationary non-Gaussianities seem to be well described by a local process, which taints an initially Gaussian random field,  $\phi \rightarrow P(\phi) = \mathcal{G}(\phi, \Phi)$ <sup>3</sup> with some level of non-Gaussianity.

A well controllable realization of such a tarnishing operation is provided by a slightly nonlinear transformation of  $\phi$  into the primordial gravitational potential  $\varphi$  via

$$\varphi(x) = \phi(x) + f_{nl}(\phi^2(x) - \langle \phi^2(x) \rangle_{(\phi)}) \quad (4.2)$$

for any  $x$ . The parameter  $f_{nl}$  controls the level and nature of non-Gaussianity via its absolute value and sign, respectively.

The data model reads

$$d = R(\phi + f_{nl}(\phi^2 - \hat{\Phi})) + n \quad (4.3)$$

In the following we assume the noise  $n$  to be Gaussian with covariance  $N = \langle nn^\dagger \rangle_{(n)}$ .

### 4.1.2 Moments Calculation

The nonlinearity of the relation between the hidden Gaussian random field  $\phi$  and the observable gravitational potential  $\varphi$  (4.2) imprints non-Gaussianity into the latter. In order to be able to extract the value of the non-Gaussianity parameter  $f$  from any data containing information on  $\varphi$ , we need to know its statistic at least up to the four-point function, the kurtosis, which we derive with IFT methods. In the following sometimes is used the letter  $f$  for convenience instead of  $f_{nl}$ .

In order to do that, it is convenient to define a  $\varphi$ -moment generating function  $\mathcal{Z}[J]$

$$\begin{aligned} \mathcal{Z}[J] &= \int \mathcal{D}\phi P(\phi) e^{J^\dagger \varphi(\phi)} \\ &= \int \mathcal{D}\phi \mathcal{G}(\phi, \Phi) e^{J^\dagger \varphi(\phi)} \\ &= \frac{1}{|2\pi\Phi|^{1/2}} \int \mathcal{D}\phi e^{-\frac{1}{2}\phi^\dagger \Phi^{-1} \phi} e^{(J^\dagger(\phi + f(\phi^2 - \Phi)))} \end{aligned} \quad (4.4)$$

<sup>2</sup>The response, which translates the 3D gravitational field into temperature maps, is well known from CMB theory and can be calculated with publicly available codes.

<sup>3</sup> $\Phi$  is the  $\phi$  covariance  $\langle \phi \phi^\dagger \rangle_{(\phi)}$ .



rewriting  $f\phi^2$  as  $\frac{1}{2}(2\phi^\dagger f\phi)$  and using the general result (2.5) one gets to<sup>4</sup>

$$\mathcal{Z}[J] = \frac{1}{|2\pi(1 - 2\Phi\hat{f}J)|^{1/2}} \exp\left(\frac{1}{2}J^\dagger(\Phi^{-1} - 2\hat{f}J)^{-1}J - (fJ)^\dagger\hat{\Phi}\right) \quad (4.5)$$

Then  $\log \mathcal{Z}[J]$  function results<sup>5</sup>

$$\log \mathcal{Z}[J] = \frac{1}{2}J^\dagger[\Phi^{-1} - 2\hat{f}J]^{-1}J - (fJ)^\dagger\hat{\Phi} - \frac{1}{2}\text{tr}[\log(1 - 2\Phi\hat{f}J)] \quad (4.6)$$

This permits one to calculate via  $J$  derivatives the mean<sup>6</sup>,

$$\begin{aligned} \bar{\varphi}(x) &\equiv \langle \varphi \rangle_{(\phi)} \\ &= \frac{\delta \log \mathcal{Z}[J]}{\delta J_x} \Big|_{J=0} \\ &= J^\dagger[\Phi^{-1} - 2\hat{f}J]^{-1} \Big|_{J=0} - f\hat{\Phi} - \frac{1}{2} \frac{\delta}{\delta J} \left( \text{tr}(\log(1 - 2\Phi\hat{f}J)) \right) \Big|_{J=0} \\ &= -f\hat{\Phi} + \frac{\hat{\Phi}f}{(1 - 2\Phi\hat{f}J)} \Big|_{J=0} \\ &= 0 \end{aligned} \quad (4.7)$$

the covariance,

$$\begin{aligned} C_{xy}^{(\varphi)} &\equiv \langle \varphi_x \varphi_y \rangle_{(\phi)}^c \\ &= \frac{\delta^2 \log \mathcal{Z}[J]}{\delta J_x \delta J_y} \Big|_{J=0} \\ &= \frac{\delta}{\delta J} \left( J^\dagger(\Phi^{-1} - 2\hat{f}J)^{-1} \Big|_{J=0} - f\hat{\Phi} + \frac{\hat{\Phi}f}{(1 - 2\Phi\hat{f}J)} \right) \Big|_{J=0} \\ &= \Phi^{-1} - 2f\Phi^2f \end{aligned} \quad (4.8)$$

the 3-points correlation function, or skewness<sup>7</sup>,

$$\begin{aligned} B_{xyz}^{(\varphi)} &\equiv \langle \varphi_x \varphi_y \varphi_z \rangle_{(\phi)}^c \\ &= 2 \left( \Phi_{xy} f_y \Phi_{yz} + 2 \Pi_{xyz} \right) + 8 \Phi_{xy} f_y \Phi_{yz} f_z \Phi_{zx} f_x \end{aligned} \quad (4.9)$$

and the 4-points correlation function, or kurtosis

$$\begin{aligned} T_{xyzu}^{(\varphi)} &= \Phi_{xy} \Phi_{zu} \Phi_{xz} \Phi_{yu} \Phi_{xu} \Phi_{yz} + \langle \varphi_x \varphi_y \varphi_z \varphi_u \rangle_{(\phi)}^c \\ &= \left( \frac{1}{8} \Phi_{xy} \Phi_{zu} + 2 \Phi_{xy} f_y \Phi_{yz} f_z \Phi_{zu} + \Phi_{xy} f_y \Phi_{yz} f_z \Phi_{zu} f_u \Phi_{ux} f_x \right) + 23 \Pi_{xyzu} \end{aligned} \quad (4.10)$$

of the gravitational potential.

<sup>4</sup>where in the first term  $|1 - 2\Phi\hat{f}J|$  stays for that combination of functional determinant of  $\mathbb{1}$ ,  $\hat{\Phi}$  and  $\hat{f}J$ .

<sup>5</sup>Using the following property:  $A$  is a non singular squared matrix  $\log(\det A) = \text{tr}(\log A)$ .

<sup>6</sup>Using the relation:  $\frac{\partial}{\partial B} \text{tr}(AB^T) = A$ .

<sup>7</sup> $\Pi_{\alpha_1, \dots, \alpha_n}$  means  $n! - 1$  index permutations (one term is just written).

### 4.1.3 CMB Hamiltonian

Although we are not interested in the auxiliary field  $\phi$ , it is nevertheless very useful for its marginalization to define its Hamiltonian, which is

$$\begin{aligned} H_f[d, \phi] &= -\log \left( \mathcal{G}(\phi, \Phi) \mathcal{G}(d - R(\phi + f(\phi^2 - \hat{\Phi}), N)) \right) \\ &= \frac{1}{2} \phi^\dagger D^{-1} \phi + H_0 - j^\dagger \phi + \sum_{n=0}^4 \frac{1}{n!} \Lambda_{(n)}^\dagger \phi^n \end{aligned} \quad (4.11)$$

with<sup>89</sup>

$$\begin{aligned} D^{-1} &= \Phi^{-1} + R^\dagger N^{-1} R \equiv \Phi^{-1} + M \\ j &= R^\dagger N^{-1} d \\ \Lambda^{(0)} &= j^\dagger (f \hat{\Phi}) + \frac{1}{2} (f \hat{\Phi})^\dagger M (f \hat{\Phi}) \\ \Lambda^{(1)} &= -(f \hat{\Phi})^\dagger M \\ j' &= j - \Lambda^{(1)\dagger} \\ \Lambda^{(2)} &= -2 \widehat{f j'} \end{aligned} \quad (4.12)$$

and  $H_0$  collects all terms independent of  $\phi$  and  $f$ <sup>10</sup>.

If we consider large scales, dominated by the Sachs-Wolfe effect, a local approximation exists in which the response and the noise covariance are diagonal in the position space, we have

$$N_{xy} = \sigma_n^2(x) \delta(x - y) \quad (4.13)$$

and  $R = -3$  for the total intensity fluctuations, thus

$$M_{xy} = 9 \sigma_n^{-2}(x) \delta(x - y) \quad (4.14)$$

if we restrict the signal space to the last-scattering surface, which is isomorphic to  $S^2$ . This permits one to simplify the Hamiltonian to

$$H_f[d, \phi] = \frac{1}{2} \phi^\dagger D^{-1} \phi + H_0 - j^\dagger \phi + \sum_{n=0}^4 \frac{1}{n!} \lambda_{(n)}^\dagger \phi^n \quad (4.15)$$

with

$$\begin{aligned} D^{-1} &= \Phi^{-1} + 9 (\sigma_n^{-2}) \\ \lambda^{(0)} &= 3 (\hat{\Phi} / \sigma_n^2)^\dagger \left( \frac{3}{2} f^2 \hat{\Phi} - f d \right) \\ \lambda^{(1)} &= -3 (\hat{\Phi} \sigma_n^2)^\dagger f \\ j' &= j - \lambda^{(1)} = 3 (3 \hat{\Phi} f - d) / \sigma_n^2 \\ \lambda^{(2)} &= -2 \widehat{f j'}, \quad \lambda^{(3)} = 54 f / \sigma_n^2, \quad \lambda^{(4)} = 108 f^2 / \sigma_n^2 \end{aligned} \quad (4.16)$$

<sup>8</sup> $\Lambda^{(3)}$  and  $\Lambda^{(4)}$  are more complicated, the whole expressions are in [1].

<sup>9</sup>Here and in the following, the hat on the vector  $fj$  denotes a diagonal matrix,  $\widehat{fj}$ , whose entries are given by  $\widehat{fj}_{xx} = f_x j'_x$ .

<sup>10</sup>Equation (4.12) permits us to consider values of  $f_{nl}$  that vary from location to location. However, we want to concentrate on a single value of  $f_{nl}$ .

The numerical coefficients of the last two terms may look large, however, these coefficients stand in front of terms of typically  $\phi^3 \sim 10^{-15}$ , and  $\phi^4 \sim 10^{-20}$ , which ensures their well behavedness in any diagrammatic expansion series. We define the Wiener-filter reconstruction of the gravitational potential as

$$m_0 = D j. \quad (4.17)$$

#### 4.1.4 $f_{nl}$ Posterior

We aim to determine the PDF for the  $f_{nl}$  parameter, and thus are interested in Since now we are not interested in reconstructing the primordial fluctuations, but to determine the PDF for the  $f_{nl}$  parameter.

Thus are interested in

$$\begin{aligned} P(f|d) &\propto P(d|f)P(f) \\ &\propto \int \mathcal{D}\phi P(d, \phi|f) \\ &= \int \mathcal{D}\phi e^{-H[d, \phi|f]} \end{aligned} \quad (4.18)$$

where we have assumed that  $P(f) = const.$  for simplicity. However, we are not able to perform the path integration, because the Hamiltonian is not quadratic in the field  $\phi$ .

An expansion in Feynman diagrams had therefore been proposed in [1]. Where the approach consist on marginalize the former by calculating the log evidence  $\log P(d|f)$  up to quadratic order in  $f$ :

$$\begin{aligned} \log \mathcal{Z}[d|f] &= \log \int \mathcal{D}\phi P(d, \phi|f) \\ &= \log \int \mathcal{D}\phi e^{-H[d, \phi|f]} \end{aligned} \quad (4.19)$$

Although  $f$  is not known, the the terms in the diagrammatic expansion which are proportional to  $f$  and  $f^2$  can be calculated separately, permitting one to write down the Hamiltonian of  $f$

$$\begin{aligned} H_d[f] &\equiv -\log P(d|f)P(f) \\ &= \tilde{H}_0 + \frac{1}{2} f^\dagger \tilde{D}^{-1} f + \tilde{j}^\dagger f + \mathcal{O}(f^3) \end{aligned} \quad (4.20)$$

where we collected the linear and quadratic coefficients into  $\tilde{j}$  and  $\tilde{D}^{-1}$ . Then the optimal  $f$  estimator to the lowest order is therefore

$$m_f = \langle f \rangle_{((\phi, f|d))} = \tilde{D} \tilde{j} \quad (4.21)$$

and its uncertainty variance is just  $\tilde{D}$ .

Now, the central idea to circumvent this problem is to use a saddle-point approximation by performing a Taylor expansion of the Hamiltonian up to the second order in  $\phi$  around its minimum [11]. This is possible because  $|\phi| \sim \mathcal{O}(10^{-5})$  provides us with a small parameter and  $P(\phi \approx 1)$  is negligibly small. To calculate this expansion, we need the first and second functional derivatives

of  $H(d, \phi|f)$  with respect to  $\phi$ . The minimum  $m$ , in the large-scale approximation<sup>11</sup>, is given by,

$$\begin{aligned} 0 &= \left. \frac{\delta H[d, \phi|f]}{\delta \phi} \right|_{\phi=m} \\ &= (D^{-1} + \lambda^{(2)})m - j + \lambda^{(1)} + \frac{1}{2}\lambda^{(3)}m^2 + \frac{1}{6}\lambda^{(4)}m^3 \end{aligned} \quad (4.22)$$

and the Hessian

$$\begin{aligned} D_{d,f}^{-1} &= \left. \frac{\delta^2 H[d, \phi|f]}{\delta \phi^2} \right|_{\phi=m} \\ &= D^{-1} + \lambda^{(2)} + \widehat{\lambda^{(3)}m} + \frac{1}{2}\widehat{\lambda^{(4)}m^2} \end{aligned} \quad (4.23)$$

In our saddle-point approximation, the Hamiltonian therefore has the form

$$H[d, \phi|f] = H[d, m|f] + \frac{1}{2}(\phi - m)^\dagger D_{d,f}^{-1}(\phi - m) + \mathcal{O}((\phi - m)^3) \quad (4.24)$$

where we ignore third and fourth order terms in  $(\phi - m)$ . Now we are able to compute analytically the  $\phi$  marginalization,

$$\begin{aligned} P(d|f) &\propto \int \mathcal{D}\phi e^{-H[d, \phi|f]} \\ &\approx \int \mathcal{D}(\phi - m) \left| \frac{\delta(\phi - m)}{\delta \phi} \right|^{-1} \exp\left(-H[d, m|f] - \frac{1}{2}(\phi - m)^\dagger D_{d,f}^{-1}(\phi - m)\right) \\ &= |2\pi D_{d,f}|^{1/2} \exp\left(-H[d, m|f]\right) \end{aligned} \quad (4.25)$$

At this point we are able to calculate analytically the maximum a posteriori estimator for  $f_{nl}$ , the  $f_{MAP}$ , by setting

$$\left. \frac{\partial P(f|d)}{\partial f} \right|_{f=f_{MAP}} = 0 \quad (4.26)$$

This yields<sup>12</sup>

$$0 = \frac{1}{2} \text{tr} \left( \frac{\partial}{\partial f} \ln (2\pi D_{d,f})^{-1} \right) + \frac{\partial H[d, m|f]}{\partial m} \frac{\partial m}{\partial f} + \left. \frac{\partial H[d, m|f]}{\partial f} \right|_{f=f_{MAP}} \quad (4.27)$$

solving this equation we get  $f_{MAP}$ .

### 4.1.5 Experimental Results

While standard single-field inflation theories predict small values of  $f_{nl} \ll 1$ , multifield inflation theories predict larger  $f_{nl}$  values up to the order of  $\mathcal{O}(10^2)$ . Therefore, any detection or upper limit of  $f_{nl}$  rules out some inflation models and might enable us to select between the remaining ones. Recent data from the *Planck* collaboration [12] satellite constrain the non-Gaussianity, combining temperature and polarization data, to

$$f_{nl} = 0.8 \pm 5.0 \text{ (68\% C.L.)}. \quad (4.28)$$

Which is compatible with zero.

<sup>11</sup>Where the response and the noise covariance matrix are diagonal.

<sup>12</sup>Note that we have used the implicit function theorem to calculate the partial derivative of the implicitly defined function  $m(f)$  with respect to  $f$ .

## Appendix A

# Gaussian Field Correlation Functions

We can see an example. Consider a Gaussian field  $q(x)$  with kernel  $K(x, y)$ , it has the partition function  $\mathcal{Z}[J] = e^{-\frac{1}{2}(J, K^{-1}, J)}$ . We can try to calculate the 1-point and 2-point correlation functions, finding:

$$\begin{aligned}
 \langle q(x) \rangle &\equiv \left. \frac{\delta \mathcal{Z}[J]}{\delta J(x)} \right|_{J=0} \\
 &= - (K^{-1}, J) \mathcal{Z}[J] \Big|_{J=0} \\
 &= 0 \\
 \langle q(x), q(y) \rangle &\equiv \left. \frac{\delta^2 \mathcal{Z}[J]}{\delta J(x) \delta J(y)} \right|_{J=0} \\
 &= \left( -K^{-1}(x, y) + (K^{-1}, K^{-1} J) \right) \mathcal{Z}[J] \Big|_{J=0} \\
 &= K^{-1}(x, y)
 \end{aligned} \tag{A.1}$$

therefore for a Gaussian variable the 2-point correlation function is the inverse of the kernel  $K(x, y)$ . If we want to calculate the  $C^{(n)}(x_1, \dots, x_n)$  we get <sup>1</sup>.

$$\begin{aligned}
 C^{(2n+1)}(x_1, \dots, x_{2n+1}) &= 0 \\
 C^{(2n)}(x_1, \dots, x_{2n}) &= (2n-1)!! (K^{-1}(x_i, x_{i+1}))^{sym}
 \end{aligned}$$

we see from this example that n-point correlation functions of a Gaussian field are all derivable from the 2-point one, in the sense that correlation functions are decomposable as products of the 2-point function.

---

<sup>1</sup>*sym* means the total symmetrization operation of the term:  $(K^{-1}(x_1, x_2) K^{-1}(x_3, x_4) \dots K^{-1}(x_{2n-1}, x_{2n}))$ . In this case i goes from 0 to  $n-1$  so the symmetrization consist in a sum over  $(2n-1)!!$  terms.



# Appendix B

## CMB $f_{nl}$

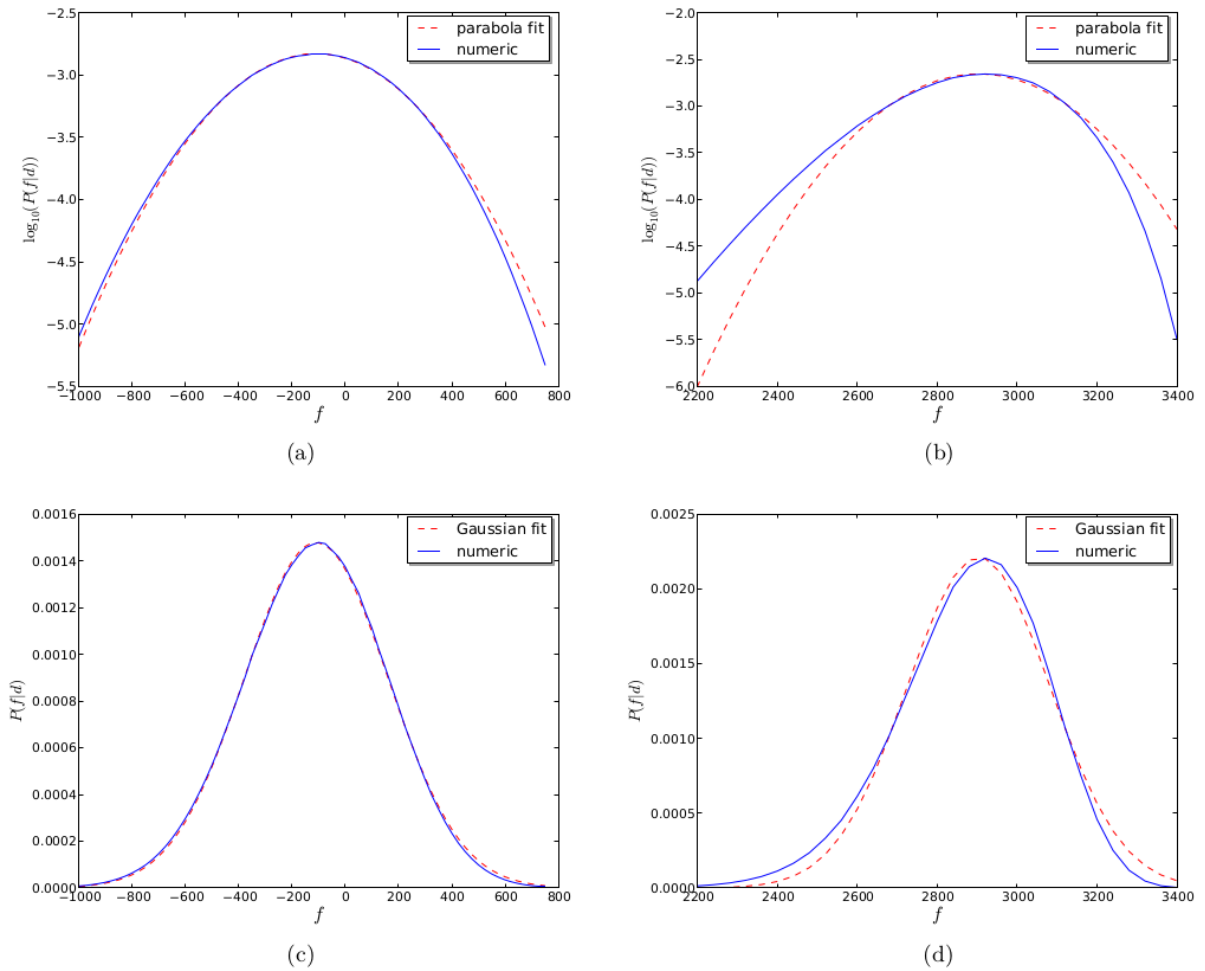


Figure B.1: Normalized posterior distributions for  $f$  in a one-dimensional test case with data generated from  $f_{gen} = 3$  [panels (a), and (c)] and  $f_{gen} = 3000$  [panels (b), and (d)]. The (a), (b) upper [(c), (d) lower] panels show the numerically calculated posterior including a parabola (Gaussian) fit. For  $f_{gen} = 3000$ , the  $PDF$  is negatively skewed and thus significantly non-Gaussian. The fitting curves in the upper panels arise from translating the Gaussian fit of the lower panels into a quadratic function. [11]





## Appendix C

# CMB Reconstruction

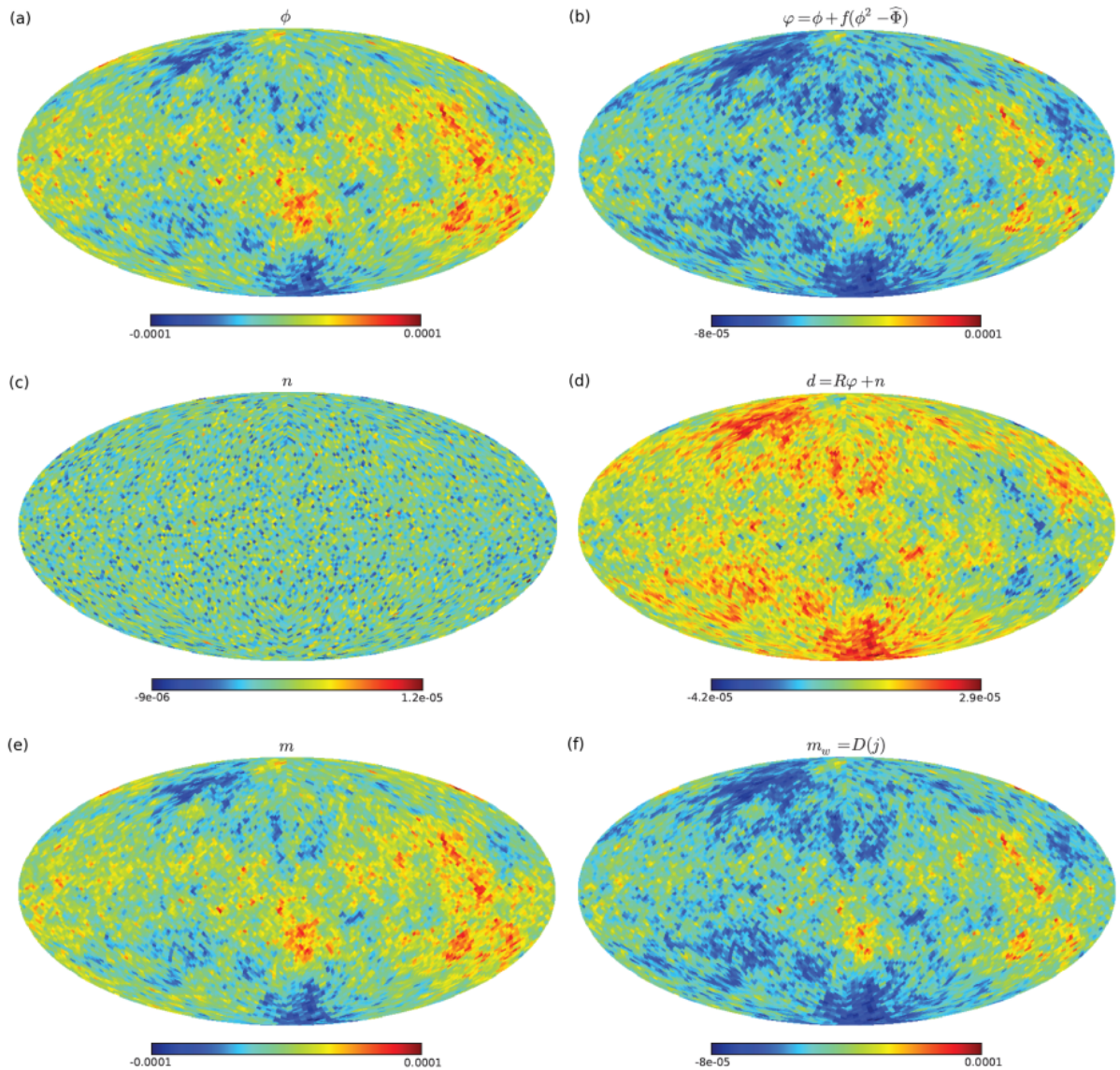


Figure C.1: Reconstruction of the primordial gravitational potential  $\varphi$  or the auxiliary Gaussian field  $\phi$  by using the maximum of the Hamiltonian and by applying a Wiener Filter, respectively. The upper four panels [(a) – (d)] are showing the generation of the mock data  $d$ , whereas the last two panels [(e), (f)] are showing the reconstructions of the original fields. Note the different color codes. [11]

# Bibliography

- [1] Torsten A. Enßlin, Mona Frommert, and Francisco S. Kitaura, (2009), *Information field theory for cosmological perturbation reconstruction and nonlinear signal analysis*, Physical Review D 80, 105005.
- [2] Torsten A. Enßlin, (2018), *Information theory for fields* arXiv:1804.03350v1.
- [3] Torsten A. Enßlin, (2018), Lecture Notes, *Information Theory and Information Field Theory*.
- [4] Torsten Enßlin, (2013), *Information field theory*, Max Planck Institute for Astrophysics, Garching bei München, arXiv:1301.2556v1.
- [5] Fortschritte der Physik, 31, M.H. Zaidi, (1983), 403-418, *Functional Methods*, Department of Physics University of Virginia.
- [6] Ariel Caticha, (2008), *Lecture on probability, entropy, and statistical physics* Department of Physics, University at Albany, arXiv:0808.0012v1.
- [7] Sabino Matarrese, (1990), *An introduction to path integrals and their cosmological applications*, Physics and Astrophysics department "Galileo Galilei", University of Padova.
- [8] Vito Volterra, (1959), *Theory of functionals and of integral and integro-differential equations*, Dover Publications, Inc.
- [9] J.J. Binney, N.J. Dowrick, A.J. Fisher, M.E.J. Newman, (1992), *The Theory of Critical Phenomena, an Introduction to Renormalization Group*, Clarendon Press Oxford.
- [10] Jörg C. Lemm, (2000), *Bayesian Field Theory Nonparametric Approaches to Density Estimation, Regression, Classification, and Inverse Quantum Problems*, Institut für Theoretische Physik I Universität Münster.
- [11] Sebastian Dorn, Niels Oppermann, Rishi Khatri, Marco Selig, and Torsten A. Enßlin, (2018), *Fast and precise way to calculate the posterior for the local non-Gaussianity parameter  $f_{nl}$  from cosmic microwave background observations*, D-85748 Garching, D-80333 München, Germany.
- [12] Planck Collaboration (2015), *Planck 2015 results*, EDP Sciences.
- [13] Torsten A. Enßlin and Cornelius Weig, (2010), *Inference with minimal Gibbs free energy in information field theory*, Garching, Germany, arXiv:1004.2868v3.