

# Cancelling out systematic uncertainties

Jorge Noreña,<sup>1\*</sup> Licia Verde,<sup>1\*</sup> Raul Jimenez,<sup>1</sup> Carlos Peña-Garay<sup>2</sup>  
and Cesar Gomez<sup>3</sup>

<sup>1</sup>*ICREA & Instituto de Ciencias del Cosmos (ICC), University of Barcelona (IEEC-UB), Martí i Franques 1, 08028 Barcelona, Spain*

<sup>2</sup>*Instituto de Física Corpuscular (CSIC-UVEG), València, Spain*

<sup>3</sup>*Physics Department and Instituto de Física Teórica UAM/CSIC, Cantoblanco, 28049 Madrid, Spain*

Accepted 2011 September 2. Received 2011 August 31; in original form 2011 July 4

## ABSTRACT

We present a method to minimize, or even cancel out, the nuisance parameters affecting a measurement. Our approach is general and can be applied to any experiment or observation where systematic errors are a concern e.g. are larger than statistical errors. We compare it with the Bayesian technique used to deal with nuisance parameters: marginalization, and show how the method compares and improves by avoiding biases. We illustrate the method with several examples taken from the astrophysics and cosmology world: baryonic acoustic oscillations (BAOs), cosmic clocks, Type Ia supernova (SNIa) luminosity distance, neutrino oscillations and dark matter detection. By applying the method we not only recover some known results but also find some interesting new ones. For BAO experiments we show how to combine radial and angular BAO measurements in order to completely eliminate the dependence on the sound horizon at radiation drag. In the case of exploiting SNIa as standard candles we show how the uncertainty in the luminosity distance by a second parameter modelled as a metallicity dependence can be eliminated or greatly reduced. When using cosmic clocks to measure the expansion rate of the universe, we demonstrate how a particular combination of observables nearly removes the metallicity dependence of the galaxy on determining differential ages, thus removing the age–metallicity degeneracy in stellar populations. We hope that these findings will be useful in future surveys to obtain robust constraints on the dark energy equation of state.

**Key words:** methods: statistical – cosmology: theory.

## 1 INTRODUCTION

It is often the case that measurements in an experiment are hampered by systematic uncertainties or poorly known quantities that bias or increase the error on the experimental quantity we wish to measure. The usual way to deal with these quantities which we will refer to as ‘nuisance parameters’ is to ‘marginalize’ over them, i.e. in the Bayesian framework, to integrate the likelihood over the full range allowed by the parameter space of the nuisance parameters. This has the inconvenience that it makes the procedure explicitly dependent on the choice of the prior adopted for the nuisance parameters, it is not guaranteed to be unbiased and it is often non-optimal. A trivial case where marginalization will fail at providing the optimal solution is the following: imagine we obtain two measurements  $x$  and  $y$  that depend on two quantities of interest  $\theta_1$ ,  $\theta_2$  and on a nuisance parameter  $n$  in the following form  $x = \theta_1/n$  and  $y = n\theta_2$ .

It is obvious that the best way to remove the nuisance parameter  $n$  is to form the product  $r = xy$ . This reduces the number of ‘data’ but removes completely the dependence of the best-fitting estimate on the nuisance parameter thus removing the bias of the estimate. In real life it will not always be the case that one can completely eliminate the nuisance parameter because the functional form of the observables might not allow one to do so, or because there are less observables than nuisance parameters. It would be of value to have a general prescription that describes whether the observables can be combined in such a way as to completely eliminate one or more nuisance parameters or, if exact cancellation is not possible, which combination could minimize their impact on the final measurement.

Today, mitigation of the effect of systematic uncertainties is a crucial issue. This is especially true in cosmology, where, in the era of precision cosmology brought about by the avalanche of data of the last decade, statistical errors keep shrinking, and the ultimate error floor is often imposed by systematics or nuisance parameters. In the literature, however, there is no generic prescription to address this issue besides marginalization. Still, results obtained by

\*E-mail: jorge.norena@icc.ub.edu (JN); liciaverde@icc.ub.edu (LV)

marginalization are often not independent of the systematic effects as we will show in concrete cases. The state of the art in mitigating the effect of systematic uncertainties is represented by some specific examples worked out and applied only to specific cases. For example Song & Percival (2009) noted that two quantities related to the growth of cosmological structures that can be measured from galaxy redshift surveys are  $\beta = f/b$  and  $\sigma_{8,\text{gal}} = b\sigma_{8,\text{m}}$ . Here  $f = \text{dln } \delta/\text{dln } a$  is the logarithmic growth rate of structures and depends on key cosmological parameters that we want to measure,  $b$  is the galaxy bias and our nuisance parameter as it is poorly known,  $\sigma_{8,\text{gal},\text{m}}$  denotes the rms fluctuations of a sphere of  $8 \text{ Mpc } h^{-1}$  for the galaxies and dark matter, respectively, and cosmological information through the growth factor is enclosed in  $\sigma_{8,\text{m}}$ . Similarly to our example above, Song & Percival (2009) suggested to use the combination  $\beta\sigma_{8,\text{g}}$  to remove the bias. In reality, when looking in detail at the equations governing redshift-space distortions in galaxy redshift surveys, one finds that the full redshift-space power spectrum depends on different combinations of  $f$ ,  $b$  and  $\sigma_{8,\text{m}}$ , namely  $f\sigma_{8,\text{m}}$ ,  $fb\sigma_{8,\text{m}}^2$ ,  $b\sigma_{8,\text{m}}$ ; thus the data themselves may allow one to separate the various parameters. The use of ratios of observables to cancel systematics has been widespread. The list of examples also includes observables with unequal energy ranges so that the ratio of observables cancels the dependences on systematic errors in neutrino detectors (Villante, Fiorentini & Lisi 1999) or on the theoretical distribution of dark matter particles (Fox, Kribs & Tait 2011a; Fox, Liu & Weiner 2011b); there are also specific methods that apply in some particular models of dark energy (March et. al. 2011).

Another well-known example is the standard procedure to minimize the effect of noise bias in cosmic microwave background (CMB) angular power spectra. The angular power spectrum computed from the autocorrelation of a map produced by a given detector is given by  $C_\ell^{\text{meas}} = C_\ell^{\text{true}} + C_\ell^{\text{noise}}$ , where  $C_\ell^{\text{noise}}$  denotes the power spectrum of the detector noise and is called ‘noise bias’: a poor knowledge of the noise bias will bias the estimate of the CMB power spectrum. On the other hand the cross power spectrum for maps produced by different, uncorrelated, detectors does not have noise bias.

In any other context a new prescription must be worked out from scratch using detailed knowledge and intuition of the specific problem.

The aim of this paper is to provide a general algorithm that minimizes, in an unbiased way, the impact of systematic uncertainties if they can be somewhat parametrized and poorly known quantities on experimental measurements. The method is completely general and can be applied to any experiment where the nuisance parameters can be characterized (at least approximately). As described here our approach is of straightforward use in cases where statistical errors are negligible compared to systematic errors. Further development is required in order to find the right balance between statistical and systematic errors when this assumption does not hold. After illustrating the approach and deriving the general prescription, we present a suite of applications of different levels of complexity. Though our approach is general we will consider for definiteness cases for which the relevant quantities can be modelled as power laws or linear functions. In order to illustrate the method, we concentrate to some problems in astronomy and cosmology and show how the method reduces the impact of systematics. In particular we address the following problems. (1) Measurements of the expansion history of the universe using baryonic acoustic oscillations (BAO). Here we show how different measurement of the BAO scale at different redshifts can be optimally combined to cancel out uncer-

ainties in the sound horizon scale. (2) Solar neutrinos and the solar metallicity problem. We discuss how to better minimize the impact of uncertainties due to the solar abundances and due to theoretical inputs like opacities and diffusion. (3) Cosmic clocks and how to minimize the influence of systematic uncertainties in the metallicity of galaxies while estimating their age. (4) Type Ia supernova (SNIa) surveys to measure the expansion history of the universe and how to minimize the dependence of the Hubble diagram on a second parameter. We also recover some well-known results such as the cancellation of the noise bias in the measurement of the CMB angular power spectra and the cancellation of astrophysical inputs in dark matter observations through the use of observables with adjusted energy ranges.

## 2 A SYSTEMATIC APPROACH. ANALYSIS AND RECIPE

We consider  $N$  observables  $O_i$ ,  $i = 1, \dots, N$  that depend on  $m$  accurate or interesting quantities  $\mu_i$ ,  $i = 1, \dots, m$  and  $n$  ‘unknown’ or ‘biased’ nuisance quantities  $\nu_i$ ,  $i = 1, \dots, n$ , i.e.  $O_i(\mu_1, \dots, \mu_m, \nu_1, \dots, \nu_n)$ . Our working assumption is that we are ignorant of the mean values of the nuisance parameters  $\nu_j$  and their errors  $\langle \nu_j \nu_l \rangle$  and  $\langle \mu_j \nu_l \rangle$ . Our goal is to find combinations of the observable quantities  $f_k(O_1, \dots, O_N)$ ,  $k = 1, \dots, M$  such that they are not affected by our ignorance of the nuisance parameters. This requirement implies that (a) the variance of  $f_k$  is independent of the variance and covariance of the  $\nu_i$ , guaranteed by the lack of explicit dependence of  $f_k$  on  $\nu_i$  and (b) the mean value of  $f_k$  should be independent of the mean value of the  $\nu_i$ , guaranteed by the null total derivative of  $f_k$  with respect to the  $\nu_i$ . These conditions can be written explicitly as a set of first-order partial differential equations:

$$f_k = f_k(O_1, \dots, O_N), \quad (1)$$

$$\frac{df_k}{d\nu_i} = \sum_{j=1}^N \frac{\partial f_k}{\partial O_j} \frac{\partial O_j}{\partial \nu_i} = 0 \quad \text{for } i = 1, \dots, n. \quad (2)$$

Note that even if the nuisance parameters have a distribution which is very different from a Gaussian, the condition that  $f_k$  have no explicit dependence on  $\nu_i$  guarantees that the correlation functions involving an arbitrary number of solutions  $f_k$ , i.e.  $\langle f_1^{\alpha_1} \dots f_M^{\alpha_M} \rangle$  will be independent of any correlation functions involving the nuisance parameters  $\nu_i$ . If the distribution of the nuisance parameters is assumed to be Gaussian, it is enough to require that the partial derivative of  $f_k$  with respect to  $\nu_i$  vanishes,  $\partial f_k / \partial \nu_i = 0$  for  $i = 1, \dots, n$ , in order to guarantee that the correlation functions of  $f_k$  are independent of correlations involving the nuisance parameters. Furthermore, in the Gaussian case the correlations between the nuisance parameters and the solutions will vanish  $\langle \nu_i f_k \rangle = 0$  due to the vanishing total derivative, but in the general non-Gaussian case they are different from zero  $\langle \nu_i f_k \rangle \neq 0$ .

A natural interpretation of the recipe defined by equation (2) is the *renormalization group* equation (Wilson & Kogut 1974). In this framework, the nuisance *unknown* quantities  $\nu_i$  are interpreted as setting the renormalization group scale. Physical quantities  $f_k(\nu_i, O_j(\nu))$  are then defined as those functionals of the observables that are invariant under arbitrary changes of the values of the nuisance, i.e. as the ones satisfying the generalized renormalization group equation:

$$\left( \frac{\partial}{\partial \nu_i} + \beta_{i,j} \frac{\partial}{\partial O_j} \right) f_k(\nu_i, O_j(\nu)) = 0, \quad (3)$$

with  $\beta_{i,j} = \partial O_j / \partial v_i$  the corresponding  $\beta$  functions. In other words, by means of this recipe we identify the physics that is invariant under arbitrary rescalings of the nuisance quantities. This nuisance-independent physics is completely characterized by the *nuisance scaling dimensions*  $\beta_{i,j}$  of the observables. In the limit case where nuisance can be totally washed out this recipe will unravel the underlying responsible *scale invariance*.

For definiteness, let us show the solutions for the case of  $N$  observable quantities  $O_i$  which are modelled by power laws of the  $n$  nuisance parameters  $v_j$ :

$$O_i - \hat{O}_i = g(\boldsymbol{\mu}) \prod_{j=1}^n (v_j - \hat{v}_j)^{a_{ij}}, \quad (4)$$

where  $g$  is some function of  $\boldsymbol{\mu}$ , a vector containing all other quantities on which the observables depend,  $a_{ij}$  is known exponents and  $\hat{v}_j$  is the true value of  $v_j$ , which does not necessarily need to be known. The solution to the system of the  $2n$  first-order partial differential equations is a power law of the observables:

$$f_k = \prod_{i=1}^N (O_i - \hat{O}_i)^{b_i^k}, \quad (5)$$

which leads for each  $k$  to a system of  $j = 1, \dots, n$  linear algebraic equations for the  $i = 1, \dots, N$  unknowns  $b_i^k$ . Clearly there are  $k = 1, \dots, M = N - n$  non-trivial independent solutions:

$$\sum_{i=1}^m a_{ij} b_i^k = 0. \quad (6)$$

After removing the nuisance parameters, we are left with  $N - n$  observables. This means that we lose one observable per each nuisance parameter we eliminate and therefore that we can only eliminate  $N - 1$  nuisance parameters. Compared to the case in which a prior is imposed on the nuisance parameters and one marginalizes over them, in this approach the resulting statistical errors on the  $\mu_j$  are expected to increase. However, we now obtain a set of observables which is independent of the systematics (and thus we do not rely on any choice of prior for the  $v_j$ ). This is generally true and does not depend on our assumption that the relevant quantities can be modelled as power laws. Note also that any combination of solutions to the differential equations, equation (2), is also a solution, we simply need to find all independent solutions.

Exactly the same analysis can be repeated for the case in which the observables can be modelled by linear combinations of the nuisance parameters:

$$O_i - \hat{O}_i = g(\boldsymbol{\mu}) + \sum_{j=1}^n a_{ij} (v_j - \hat{v}_j), \quad (7)$$

such that the solution to the system of first-order differential equations is given by a linear combination of the observables,

$$f_k = \sum_{i=1}^N b_i^k (O_i - \hat{O}_i), \quad (8)$$

and we are once more led to a system of  $n$  linear algebraic equation for the  $N$  unknowns  $b_i^k$ , with  $N - n$  non-trivial solutions,

$$\sum_{i=1}^m a_{ij} b_i^k = 0. \quad (9)$$

In Section 3 we discuss applications of equation (6) valid for the linear and the power-law cases and present the non-trivial solutions in cases where we have more observables than nuisance parameters.

Now we concentrate on problems with fewer observables than nuisance parameters. In this case there is no exact solution for the system of equation (2). However, in some specific cases that are still of practical interest there may be approximate solutions. In particular, we aim at problems where observables have similar (but not identical) dependences on some of the nuisance parameters (i.e.  $a_{ij} \sim a_{il}$  for some  $i, j, l$ ). The main difference from the exact treatment above will be that we cannot impose the full condition equation (2), but we should rather minimize the impact of systematics on the new observables. If we shift a given nuisance parameter  $v_i$  by  $\Delta v_i$  the change in  $f_k$  is approximately given by  $df_k/dv_i \Delta v_i$ . The optimal  $f_k$  should minimize the square of this quantity with respect to the parameters. We avoid the trivial solution by using Lagrange multipliers, which in the particular cases of observables modelled by equation (4) or equation (7) leads to the function

$$\mathcal{L}_k = \sum_{j=1}^n \left( \frac{df_k}{dv_j} \right)^2 \Delta v_j^2 - \lambda_k \left( \sum_i (b_i^k)^2 - A_k^2 \right), \quad (10)$$

where  $\lambda_k$  is a Lagrange multiplier to be solved for,  $A_k$  is the norm of the vector  $b_i^k$  and  $\Delta v_j$  is the uncertainty on the  $j$  nuisance parameter.<sup>1</sup> Let us start by studying the power-law case. In order to minimize this function, we vary it with respect to  $b_i^k$  using the expression for  $f_k$  as a power law of observables given in equation (5), which leads to the eigenvalue equation

$$\sum_{i=1}^N \mathbf{M}_{ii} b_i^k = \sum_l \left[ f_k^2 \sum_{j=1}^n \left( \frac{\Delta v_j}{v_j} \right)^2 a_{ij} a_{lj} \right] b_l^k = \lambda_k b_i^k, \quad (11)$$

where the first equality defines the matrix  $\mathbf{M}$ . This is our central equation. The Lagrange multiplier  $\lambda_k$ , which will be the eigenvalues of the matrix  $\mathbf{M}_{ii}^k$ , measures how much the extremal (with respect to the nuisance parameters) solution  $f_k$  is affected by the nuisance parameters. This can be seen by multiplying equation (11) by  $b_i^k$ , summing over  $i$ , and noticing that the left-hand side of the resulting expression is precisely  $\sum_j (df_k/dv_j)^2 \Delta v_j^2$ , thus  $(\Delta f_k)^2 = \lambda_k \sum_i (b_i^k)^2$ . The eigenvectors correspond to independent combinations of observables, and, if we are interested in minimizing the impact of nuisance parameters, we should choose those eigenvectors corresponding to eigenvalues which are small with respect to typical entries of the matrix. In fact, the existence of a small eigenvalue is due to the similar dependence of the observables on a particular nuisance parameter.

We can easily verify that this description is more general and includes the previously discussed cases. If an eigenvalue is zero, there is an independent non-trivial solution which is unaffected by changes of the nuisance parameters. Indeed it can be easily seen from equation (11) that there will always be  $N - n$  null eigenvalues when  $n < N$ , and their eigenvectors span the space of solutions of equation (6). Other eigenvalues are expected to be of the order of typical entries of the matrix  $\lambda \sim f^2 a^2$ , representing a solution for which the change under variations of the nuisance parameters is large  $\Delta f^2 \sim f^2 a^2 \sum_i b_i^2$ . In the case where  $n > N$ , even if all the eigenvalues are different from zero, small eigenvalues of  $\mathbf{M}_{ii}^k$

<sup>1</sup> Note that even though specifying a value for this uncertainty  $\Delta v_j$  requires a detailed knowledge of the problem and an inaccurate choice may in general introduce a bias, those solutions which are completely independent of the nuisance parameters (i.e. those corresponding to zero eigenvalues, see below) do not depend on the value of  $\Delta v_j$  at all. In the same sense, those solutions which are approximately independent of the nuisance parameters (i.e. which correspond to small eigenvalues) are expected to depend only mildly on the precise value of  $\Delta v_j$ .

correspond to solutions which change only slightly under shifts of the nuisance parameters and one can take these to be the approximately unbiased combinations. There will be as many independent approximately unbiased quantities as there are small eigenvalues of the matrix. These correspond to the instances where the nuisance parameters appear in equation (4) with powers which are almost the same for each  $O_i$  ( $a_{ij} \sim a_{il}$ ).

We can repeat the analysis for the linear case. Minimizing the function defined in equation (10) and using the expression for  $f_k$  as a linear combination of observables we now obtain the following eigenvalue equation:

$$\sum_{l=1}^N \mathbf{M}_{li} b_l^k = \sum_i \left[ \sum_{j=1}^n (\Delta v_j)^2 a_{ij} a_{lj} \right] b_i^k = \lambda_k b_i^k, \quad (12)$$

where now the matrix  $\mathbf{M}$  has a different shape, but the same considerations as above apply, i.e. we look for small eigenvalues and their corresponding eigenvectors will be solutions that are mildly influenced by the nuisance parameters.

Let us now remark that, in some problems, the nuisance parameters or the observables are constrained to satisfy  $\ell$  relations, say  $g_i(v_1, \dots, v_n, O_1, \dots, O_N) = 0$ , for  $i = 1, \dots, \ell$ . This implies that they are not independent. The way to treat these cases is again through the use of Lagrange multipliers, such that when the nuisance parameters can be cancelled out exactly, the total derivative equation must be replaced by the following one

$$\frac{d}{dv_i} \left( f^k(O_1, \dots, O_N) + \sum_{i=1}^{\ell} \alpha_i g_i(v_1, \dots, v_n, O_1, \dots, O_N) \right) = 0, \quad (13)$$

where  $\alpha_i$  is the Lagrange multipliers. We will see an explicit example of this case in Section 3.6. The way to do it in the formalism of equation (10) is that instead of minimizing the function  $\mathcal{L}$  from equation (10), one has to include the constraints in the following way:

$$\begin{aligned} \mathcal{L}_k = & \sum_{j=1}^n \left( \frac{df_k}{dv_j} \right)^2 \Delta v_j^2 - \lambda_k \left( \sum_i (b_i^k)^2 - A_k^2 \right) \\ & + \sum_{i=1}^{\ell} \alpha_i g_i(v_1, \dots, v_n, O_1, \dots, O_N), \end{aligned} \quad (14)$$

where  $\alpha_i$  is the Lagrange multipliers used to impose the conditions  $g_i = 0$ . One then minimizes  $\mathcal{L}_k$  with respect to  $b_i^k$ ,  $\lambda_k$  and  $\alpha_i$ .

Before ending this section, let us make an important comment on equation (10). If one imposes an infinite uncertainty on a nuisance parameter, say  $\Delta v_1 \rightarrow \infty$  for example, typical eigenvalues of the matrix  $\mathcal{M}$  will be infinite. However, if there are combinations of observables that are completely independent of  $v_1$ , some eigenvalues will be finite since for those combinations  $df_k/dv_1 = 0$ , so that one should restrict to the space spanned by eigenvectors corresponding to those finite eigenvalues. In this sense looking for combinations of observables which are completely independent of some nuisance parameter  $v_1$  is equivalent to assigning an infinite uncertainty to  $v_1$ .

### 3 APPLICATIONS

In this section we illustrate several cases where our method is applicable. We start with the study of BAOs measurements where we assume that the sound horizon at radiation drag, which is common to all observations, is the only quantity affected by systematic effects. In this case the dependence on the systematic can be cancelled out

exactly. A similar problem is that of cancelling astrophysical inputs in dark matter observations through the use of observables with adjusted energy ranges so that a nuisance function can be cancelled, as done in Fox et al. (2011a,b). Next we consider a case where cancellation of the nuisance parameters cannot be exact: we will then find approximate unbiased combinations. This is the case for solar neutrino fluxes, where the neutrino flux predictions in detail depend on many nuisance parameters but the dependences are similar enough that approximate unbiased combinations can be found.

In the spirit of increasing complexity, in the next example we consider a case where there is one systematic common to all observations and on top of that a correction to this main trend that differs from observation to observation in a well known way. The method will cancel out the dependence on the common systematic and minimize that of the individual corrections. This is the case of the cosmic clocks. A similar case is that of SNeIa where the intrinsic magnitude of the SNe is unknown: its average value will be constant and in common to all SNe but its actual value depends on a hidden variable (probably metallicity of the host galaxy).

To conclude we consider a case that at first sight may go beyond the scope of our method, where the interesting parameters and the nuisance parameters are defined through correlation of the observables and not the observables themselves. This is the example of the instrumental noise in CMB angular power spectrum, where the observables are the temperature fluctuations (or their spherical harmonic transform  $a_{\ell m}$ ) but it is the angular power spectrum  $C_\ell$  that carries direct information on the cosmological parameters and is affected by noise bias.

#### 3.1 Measurement of the expansion history of the universe with baryon acoustic oscillations

BAO experiments will, eventually, measure the acoustic scale in the radial and perpendicular direction in the sky. Let us suppose that one has measurements of the BAOs at  $N_z$  different redshifts  $z_i$ . The observables produced by the astronomical surveys are

$$\widehat{D}_A(z_i) = D_A(z_i)/r_s = \delta\theta/(1+z_i), \quad (15)$$

$$\widehat{H}(z_i) = H(z_i)r_s = c\Delta z. \quad (16)$$

Here  $D_A(z)$  denotes the angular diameter distance to redshift  $z$ ,  $H(z)$  the Hubble parameter at redshift  $z$ ,  $\delta\theta$  denotes the angular measurement of the BAO scale,  $\Delta z$  denotes the radial measurement and  $r_s$  denotes the sound horizon at radiation drag, which is the BAO ‘standard ruler’. The sound horizon at radiation drag is determined by CMB observations, for standard cosmologies it is affected by a very small error but its estimate can be significantly biased for non-standard models (e.g. Eisenstein & White 2004; de Bernardis et al. 2009; Mangilli, Verde & Beltran 2010). This consideration motivates us to consider  $r_s$  a nuisance parameter. From equation (4) we see that we can write  $f_k$  as a power law of the observables:

$$f_k = \prod_{i=1}^{N_z} \widehat{D}_A(z_i)^{b_i^k} \prod_{j=N_z+1}^{2N_z} \widehat{H}(z_{j-N_z})^{b_j^k}, \quad (17)$$

so that the values of the exponents  $b_i^k$  will be given by solutions to the system of linear equations (6) (or equivalently finding the kernel space of the matrix in equation 11), which we write explicitly for the particular case of two redshifts  $z_1, z_2$ ,

$$(b_1^k + b_2^k - b_3^k - b_4^k) \frac{f_k}{r_s} = 0, \quad (18)$$

with solutions

$$f_1 = \widehat{D}_A(z_1)\widehat{H}(z_1), \quad (19)$$

$$f_2 = \widehat{H}(z_2)/\widehat{H}(z_1), \quad (20)$$

$$f_3 = \widehat{D}_A(z_2)/\widehat{D}_A(z_1). \quad (21)$$

Any function of these quantities will also be a solution of the differential equation but will not contain new information. For example, suppose that one adds  $f_4 = \widehat{D}_A(z_2)\widehat{H}(z_2)$  to the set of solutions, this will simply be a combination of the solutions we listed, indeed  $f_4 = f_1 f_2 f_3$ . For the case of two redshifts, we obtain only 4 – 1 independent quantities after eliminating our single nuisance parameter  $r_s$  from our four observables.

For measurements of BAOs at  $N_z$  different redshifts, the  $2N_z - 1$  solutions will be given by all the ratios  $g = \widehat{D}_A(z_i)/\widehat{D}_A(z_1)$  and  $h = \widehat{H}(z_i)/\widehat{H}(z_1)$  plus the combination  $f = \widehat{D}_A(z_1)\widehat{H}(z_1)$ . Note that although it may seem that we obtained that  $z_1$  is the pivot scale, this is purely arbitrary as one may point out that the same quantities with any  $z_i$  as the pivot scale is also solutions of the differential equation. This is simply a consequence of the fact that any combination of our solutions is also a solution. Any pivot is equally good since the errors on the parameters will be independent of the choice of pivot under the assumption that they are Gaussian.

### 3.1.1 Implications for cosmological parameter estimation

In this subsection we use the above findings to forecast errors for the dark energy equation of state parameter  $w_X(z)$  using the widely used parametrization  $w_X = w_0 + w_a(1 - a)$ . If the reader is not interested in the details of this analysis, she or he can skip to Section 3.2. In terms of this parametrization, the expansion rate and luminosity distance can be written as

$$H(z) = \sqrt{\frac{\Omega_m h^2}{1 - \Omega_X}} \times \left[ (1+z)^3 (1 - \Omega_X) + \Omega_X (1+z)^{3(1+w_0+w_a)} e^{-3w_a z/(1+z)} \right]^{1/2}, \quad (22)$$

$$D_A(z) = \frac{c}{(1+z)} \int_0^z dz' \frac{1}{H(z')}. \quad (23)$$

Note that although  $\Omega_X$  appears explicitly in these expressions, flatness has been assumed.<sup>2</sup> BAOs surveys measure combinations proportional to  $D_A(z)/r_s$  and  $H(z)r_s$ . Using the Fisher matrix formalism, we can forecast errors for these combinations (Seo & Eisenstein 2007) combining measurements at five redshifts bins of width  $\Delta z = 0.1$  for  $z < 0.5$  for a survey with specifics similar to those of Baryon Oscillation Spectroscopic Survey (BOSS)-Sloan Digital Sky Survey III (SDSS-III)<sup>3</sup> and 15 redshifts bins at  $z > 0.5$  for a survey with specifics similar to those of *Euclid*.<sup>4</sup> We will compare three cases: keeping the full set of observations  $D_A(z_i)/r_s$ ,  $H(z_i)r_s$  (‘full’); keeping only the ratios to some pivot scale  $\widehat{D}_A(z_i)/\widehat{D}_A(z_p)$ ,  $\widehat{H}(z_i)/\widehat{H}(z_p)$  as is sometimes advocated (‘ratios’) and our ‘unbiased’ combinations  $\widehat{D}_A(z_i)/\widehat{D}_A(z_p)$ ,  $\widehat{H}(z_i)/\widehat{H}(z_p)$ ,  $\widehat{D}_A(z_p)\widehat{H}(z_p)$ . For the sake of simplicity, we keep the cosmology fixed and take into account only

<sup>2</sup> Although  $\Omega_X + \Omega_m = 1$ , the present Hubble rate, parametrized by  $h$ , is still a free parameter. We follow Seo & Eisenstein (2007) in using  $\Omega_X$  and  $\Omega_m h^2$  to parametrize the cosmology dependence.

<sup>3</sup> <http://www.sdss3.org/surveys/boos.php>

<sup>4</sup> <http://arxiv.org/abs/0912.0914>

errors on  $H_0$  and  $r_s$ . Using equations (22) and (23) one can use the Fisher matrix for  $D_A(z_i)/r_s$  and  $H(z_i)r_s$  in order to compute the Fisher matrix (and errors) for  $w_0, w_a, r_s, H_0$ . The results are given in Table 1. In the ‘full’ case we must assume a prior error on  $r_s$  (for example a CMB prior) which we take to have a 30 per cent error since for some non-standard models the (systematic) error on  $r_s$  can be estimated to be as large as that e.g. Mangilli et al. (2010).

From the table we see that the statistical errors are always smaller when using the ‘full’ set as was to be expected (our method is throwing away some information in order to obtain quantities which are insensitive to systematic errors, this implies that the statistical errors must grow). Let us remark that when we consider measurements at 20 different redshifts as specified above, we obtain that the errors on  $w_0$  and  $w_a$  change only by 0.4 and 1 per cent, respectively, between the ‘full’ and ‘unbiased’ sets. Thus, one is insensitive to systematic errors on  $r_s$  at the price of a modest increase in errors. The precise estimate of how much information is lost depends on how big the assumed uncertainty on  $r_s$  is (or how big the systematic errors are), as can be seen in the table for the case in which we assume a 2 per cent error on  $r_s$ . Finally, note that when one has measurements at fewer redshifts the difference between the ‘ratios’ and ‘unbiased’ cases increases dramatically, this is due to the fact that one is throwing away one observable from a set of just a handful of them, thus losing comparatively more information.

## 3.2 Dark matter

We discuss in this section another example of exact cancellation of nuisance parameters, or in general of nuisance functions. In particular, we consider energy-dependent observables, which have been used to cancel the dependences on systematic errors in neutrino detectors (Villante et al. 1999) or on the theoretical distribution of dark matter particles (Fox et al. 2011a,b).

Let us consider observables  $O_i(E)$  that depend on the integral of some function  $f(x)$ , that we may not know, within some known range  $[x_i, y_i]$  as

$$O_i(E_{\min} < E < E_{\max}) = g(\mu_i) F(x_i(E_{\min}), y_i(E_{\max})). \quad (24)$$

The function  $f(x)$  in the neutrino case is called the response function and contains the convolution of the neutrino cross-section, detector resolution and neutrino flux (Villante et al. 1999). The function  $f(x)$  in the case of dark matter detection corresponds to  $f(v)/|v|$ , the ratio of the dark matter velocity distribution to the dark matter speed when observing the recoil energy spectrum, and to the integral of that function times the dark matter speed when observing total rates (Fox et al. 2011a,b).

For simplicity, we concentrate on the case of two observables  $O_1$  and  $O_2$ . Let us assume that the observables have an energy range overlap such that we can find energies  $E_a, E_b, E_c$  and  $E_d$  such that  $x_1(E_a) = x_2(E_c) = x$  and  $y_1(E_b) = y_2(E_d) = y$ .

Our observables can be rewritten as

$$O_1(E_a < E < E_b) = g(\mu_1) F(x, y), \quad (25)$$

$$O_2(E_c < E < E_d) = g(\mu_2) F(x, y). \quad (26)$$

Following the lines discussed in the previous section, the dependence on the nuisance function  $f(x, y)$  is removed by the combination  $O_1(E_a < E < E_b)/O_2(E_c < E < E_d)$ , as advocated in the above references.

**Table 1.** Fisher matrix analysis for the BAOs. We repeat the analysis for the cases in which the Fisher is computed from each the ‘full’ set of observables  $\{D_A(z_i)/r_s, H(z_i)r_s\}$ , the ‘ratios’  $\{D_A(z_i)/D_A(z_p), H(z_i)/H(z_p)\}$  including external priors on  $r_s$ , and our optimally ‘unbiased’ set  $\{D_A(z_i)/D_A(z_p), H(z_i)/H(z_p), D_A(z_p)H(z_p)\}$ , as specified in the first column; we have not specified the pivot redshift  $z_p$  since any choice gives the same result. We have assumed a typical variance  $\sigma_{r_s}/r_s$  of 0.02 or 0.3, as specified in the second column. We show the case in which one has measurements at 20 redshifts up to redshift  $z \approx 2$ , and also the case in which one has measurements at 10 redshifts up to redshift  $z \approx 0.8$ , as stated in the third column. In the fourth column we give the full Fisher matrix with the order of the indexes given by  $1 = w_0, 2 = w_a, 3 = \Omega_X, 4 = \Omega_m h^2, 5 = r_s$ , thus for example the element in the first row and second column of a matrix gives an estimate for the correlation between  $w_0$  and  $w_a$ . The  $2 \times 2$  submatrix for  $w_0, w_a$  in the unbiased and ratios cases never depends on the assumed variance for  $r_s$  or  $\Omega_m h^2$  by construction. For most of the cases we assumed a prior on  $\Omega_X$  of  $\sigma_{\Omega_X}/\Omega_X = 0.01$ , except for the last two rows.

set	$\sigma_{r_s}/r_s$	redshifts	$F_{ij}^{-1}(w_0, w_a, \Omega_X, \Omega_m h^2, r_s)$
full	0.3	20	$\begin{pmatrix} 0.0579 & -0.215 & -0.00801 & 4.45 \times 10^{-7} & 0.981 \\ - & 0.982 & 0.0371 & -2.87 \times 10^{-6} & -6.34 \\ - & - & 0.00146 & -1.14 \times 10^{-7} & -0.252 \\ - & - & - & 1.5 \times 10^{-6} & -0.000893 \\ - & - & - & - & 57.6 \end{pmatrix}$
unbiased	0.3	20	$\begin{pmatrix} 0.0584 & -0.218 & -0.00813 & 0 & 0 \\ - & 1 & 0.0379 & 0 & 0 \\ - & - & 0.00149 & 0 & 0 \\ - & - & - & 1.5 \times 10^{-6} & 0 \\ - & - & - & - & 2025 \end{pmatrix}$
ratios	0.3	20	$\begin{pmatrix} 0.0587 & -0.216 & -0.00821 & 0 & 0 \\ - & 1.01 & 0.0375 & 0 & 0 \\ - & - & 0.00151 & 0 & 0 \\ - & - & - & 1.5 \times 10^{-6} & 0 \\ - & - & - & - & 2025 \end{pmatrix}$
full	0.02	20	$\begin{pmatrix} 0.0435 & -0.121 & -0.0043 & 0.0000136 & 0.133 \\ - & 0.38 & 0.0132 & -0.0000877 & -0.859 \\ - & - & 0.000505 & -3.49 \times 10^{-6} & -0.0341 \\ - & - & - & 1.49 \times 10^{-6} & -0.000121 \\ - & - & - & - & 7.8142 \end{pmatrix}$
unbiased	0.3	10	$\begin{pmatrix} 0.198 & -0.472 & -0.00305 & 0 & 0 \\ - & 1.63 & 0.036 & 0 & 0 \\ - & - & 0.00201 & 0 & 0 \\ - & - & - & 1.5 \times 10^{-6} & 0 \\ - & - & - & - & 2025 \end{pmatrix}$
ratios	0.3	10	$\begin{pmatrix} 0.21 & -0.454 & -0.00303 & 0 & 0 \\ - & 1.66 & 0.0361 & 0 & 0 \\ - & - & 0.00201 & 0 & 0 \\ - & - & - & 1.5 \times 10^{-6} & 0 \\ - & - & - & - & 2025 \end{pmatrix}$
full	0.3	20	$\begin{pmatrix} 0.171 & -0.738 & -0.0285 & 2.05 \times 10^{-6} & 4.53 \\ - & 3.41 & 0.132 & -0.0000103 & -22.8 \\ - & - & 0.00519 & -4.07 \times 10^{-7} & -0.897 \\ - & - & - & 1.5 \times 10^{-6} & -0.000842 \\ - & - & - & - & 169 \end{pmatrix}$
unbiased	0.3	20	$\begin{pmatrix} 0.182 & -0.793 & -0.0307 & 0 & 0 \\ - & 3.69 & 0.143 & 0 & 0 \\ - & - & 0.00563 & 0 & 0 \\ - & - & - & 1.5 \times 10^{-6} & 0 \\ - & - & - & - & 2025 \end{pmatrix}$

### 3.3 Solar neutrinos

The standard solar model (SSM) depends on  $\sim 20$  input parameters, including the solar age and luminosity, the opacity, the rate of diffusion, the zero-age abundances of key elements (He, C, N, O, Ne, Mg, Si, S, Ar, Fe) and the nuclear cross-sections (S-factors) for the pp chain and CN cycle reactions. The observable quantities in this case are the solar neutrino fluxes<sup>5</sup>  $\phi_i$ , in particular the  ${}^7\text{Be}$  or  ${}^8\text{B}$  neutrino fluxes, which have been independently measured by the solar neutrino detectors. The observed neutrino fluxes depend on elements abundances, solar structure parameters and on the S-factors, but the quantities of interest are actually the S-factors and the rest are poorly determined – nuisance – parameters.

To be more specific, the multidimensional set of variations in SSM input parameters  $\{\Delta\beta_j\}$  from the SSM best values  $\{\beta_j^{\text{SSM}}\}$  often collapses to a one-dimensional dependence of the neutrino fluxes on the solar core temperature ( $T_c$ ), where  $T_c$  is an implicit function of the variations  $\{\Delta\beta_j\}$  (Bahcall & Ulmer 1996). The correlations between  $\phi_i$  and  $T_c$  are strong but not exact.

The sensitivity to parameter variations can be expressed in terms of the logarithmic partial derivatives  $\alpha(i, j)$  evaluated for each neutrino flux  $\phi_i$  and each SSM input parameter  $\beta_j$ ,

$$\alpha(i, j) \equiv \frac{\partial \ln [\phi_i / \phi_i^{\text{SSM}}]}{\partial \ln [\beta_j / \beta_j^{\text{SSM}}]}, \quad (27)$$

where  $\phi_i^{\text{SSM}}$  and  $\beta_j^{\text{SSM}}$  denote the SSM best values. This information, in combination with the assigned uncertainties in the  $\beta_j$ , then provides an estimate of the uncertainty in the SSM prediction of  $\phi_i$ . Here we employ the logarithmic partial derivatives of Pena-Garay & Serenelli (2008) and Serenelli, Haxton & Pena-Garay (2011), corresponding to the higher metal composition (Grevesse & Sauval 1998) SSM,

$$\begin{aligned} f({}^7\text{Be}) &= \frac{\phi({}^7\text{Be})}{\phi({}^7\text{Be})^{\text{SSM}}} \\ &= \left[ L_{\odot}^{3.558} O^{1.267} A^{0.786} D^{0.136} \right] \\ &\quad \times \left[ S_{11}^{-1.07} S_{33}^{-0.441} S_{34}^{0.878} S_{17}^{0.0} S_{e7}^{0.0} S_{114}^{-0.001} \right] \\ &\quad \times \left[ x_{\text{C}}^{0.004} x_{\text{N}}^{0.002} x_{\text{O}}^{0.053} x_{\text{Ne}}^{0.044} x_{\text{Mg}}^{0.057} x_{\text{Si}}^{0.116} x_{\text{S}}^{0.083} x_{\text{Ar}}^{0.014} x_{\text{Fe}}^{0.217} \right], \end{aligned} \quad (28)$$

$$\begin{aligned} f({}^8\text{B}) &= \frac{\phi({}^8\text{B})}{\phi({}^8\text{B})^{\text{SSM}}} \\ &= \left[ L_{\odot}^{7.130} O^{2.702} A^{1.380} D^{0.280} \right] \\ &\quad \times \left[ S_{11}^{-2.73} S_{33}^{-0.427} S_{34}^{0.846} S_{17}^{1.0} S_{e7}^{-1.0} S_{114}^{0.005} \right] \\ &\quad \times \left[ x_{\text{C}}^{0.025} x_{\text{N}}^{0.007} x_{\text{O}}^{0.111} x_{\text{Ne}}^{0.083} x_{\text{Mg}}^{0.106} x_{\text{Si}}^{0.211} x_{\text{S}}^{0.151} x_{\text{Ar}}^{0.027} x_{\text{Fe}}^{0.510} \right], \end{aligned} \quad (29)$$

where each parameter on the left-hand side represents a  $\beta_j / \beta_j^{\text{SSM}}$ . The luminosity, opacity, solar age and the diffusion parameters are denoted by  $L_{\odot}$ ,  $O$ ,  $A$  and  $D$ , while  $S$  and  $x$  denote S-factor and abundance ratios. The errors assigned to the solar model inputs are (0.4, 2.5, 0.44, 15.0, 0.9, 4.3, 5.1, 7.5, 2.0, 7.2, 29.7, 32.0, 38.7, 53.9, 11.5, 11.5, 9.2, 49.6, 11.5) per cent (Serenelli et al. 2011) for all the quantities  $L_{\odot}$ ,  $O$ ,  $A$ ,  $D$ ,  $S_j$   $x_q$  in the order as they appear in equations (28) and (29).

<sup>5</sup> The index  $i$  here labels the nuclear reaction that produces the neutrinos.

A new re-evaluation of the solar composition (Asplund, Grevesse & Sauval 2005), which leads to significantly smaller abundances than previously estimated, has led to the lack of matching of helioseismological data and SSM predictions. It is therefore relevant to test observables that are weakly dependent both on solar composition and more theoretical inputs like diffusion and opacity. In our analysis, we consider  $L_{\odot}$ ,  $O$ ,  $A$ ,  $D$  and  $x$  as nuisance parameters. The matrix  $\mathbf{M}_{ii}^k$  of equation (11) for the two combination of observables  $f({}^7\text{Be})$  and  $f({}^8\text{B})$  is

$$\begin{pmatrix} 35.69 & 74.42 \\ 74.42 & 156.34 \end{pmatrix},$$

whose smaller eigenvalue (0.22) has the corresponding eigenvector is proportional to (2.098, 1). Therefore the combination of observables,

$$\frac{f({}^7\text{Be})^{2.10}}{f({}^8\text{B})} = \left[ S_{11}^{0.485} S_{33}^{-0.498} S_{34}^{0.996} S_{17}^{-1.0} S_{e7}^{1.0} S_{114}^{-0.007} \right], \quad (30)$$

minimizes the impact of the errors due to composition, opacity and diffusion inputs (0.47 per cent error) and therefore optimizes the determination of nuclear cross-sections at solar core temperatures.

The formalism used here is very quick in testing the robustness of the method to changes in the assumed errors on the nuisance parameters. We have checked that the optimal observable does not significantly change by modifying the errors on some of the nuisance parameters. For example, if we double the errors on the abundance ratios, the power in  $f({}^7\text{Be})$  changes from 2.098 to 2.099 and the error from 0.47 to 0.92 per cent, while if we double the errors in theoretical inputs like opacity and diffusion, the power in  $f({}^7\text{Be})$  changes from 2.098 to 2.105 and the error from 0.47 to 0.48 per cent.

### 3.4 The expansion history of the universe from cosmic clocks

A direct method to determine the expansion history of the universe is to use cosmic clocks (Jimenez & Loeb 2002). The determination of the Hubble parameter is done by estimating the differential age of ‘clocks’, namely passively evolving galaxies. In order to estimate the differential age one approach is to use the spectral feature around the 4000 Å break (Moresco et al. 2011), although it is desirable to exploit the whole galaxy spectrum (Jimenez et al. 2003; Simon, Verde & Jimenez 2005; Stern et al. 2010). This feature, called  $D4000$ , depends both on the age and on metallicity of the stellar population. Although the dependence of  $D4000$  on the metallicity is weaker than the dependence on age, the metallicity is usually poorly known, and thus acts as a nuisance parameter in obtaining the differential age. To a good approximation the  $D4000$  feature can be written as

$$D4000 \propto \text{age}^{\alpha} Z^{\beta}, \quad (31)$$

where  $Z$  is the metallicity of the ‘clock’ and  $\alpha$ ,  $\beta$  are the exponents of the corresponding power laws. For spectra which have a poor signal-to-noise ratio there will be a large systematic error on  $Z$  due to the fact that spectral lines will be harder to identify. This systematic can be modelled to be a signal-to-noise ratio dependent shift  $Z^{\text{measured}} = \Delta(\gamma) Z^{\text{true}}$ , where  $\gamma$  is the signal-to-noise ratio. Of course there might be other systematics affecting what we have called  $Z^{\text{true}}$ , but this is beyond the scope of this paper. The biggest error is expected to be the systematic shift in the estimated metallicity  $\Delta(\gamma)$ , which we model as a power law  $\Delta(\gamma) = \Theta \gamma^b$ , where we have parametrized our ignorance with  $\Theta$  and  $b$  (Moresco et al. 2011). We can rewrite

**Table 2.** Measurements of the spectral feature around the 4000 Å break for different galaxies. These are examples of the types of data that one should obtain, and the results from actual measurements are expected to be similarly distributed. We show the measured value of the  $D4000$  feature in the first column, the error on each measurement on the second column as given by its dispersion, the redshift  $z$  of the galaxy in the third column and the signal-to-noise ratio of the spectrum in the fourth column.

$D4000$	$\sigma_d$	$z$	$\gamma$
2.00	0.03	0.15	100
1.95	0.03	0.2	100
1.90	0.03	0.25	100
1.81	0.05	0.4	72
1.82	0.05	0.5	73
1.76	0.05	0.6	75
1.83	0.03	0.7	85
1.70	0.025	0.9	93
1.60	0.04	1.05	75
1.50	0.05	1.28	70

this as

$$\Delta(\gamma) = \Theta B^{\log(\gamma_i)}, \quad (32)$$

where  $B = e^b$ . We then describe the ‘cosmic clocks’ in the following way:

$$d_i = \text{age}_i^\alpha \bar{Z}^\beta (\Theta B^{\log(\gamma_i)})^\beta \quad \text{for each } i, \quad (33)$$

where  $d_i$  is the label corresponding to a given absorption line (the  $D4000$  feature in this example), the index  $i$  runs over galaxies that have approximately the same metallicity,  $\bar{Z}$  is some central value of the metallicity,  $\gamma_i$  is the signal-to-noise ratio of each measurement and  $\beta$  is some known power – thus if the original sample shows a wide range of metallicity it should be split in metallicity bins before it can be described by equation (33). We now wish to find quantities which are independent of the nuisance parameters:  $\Theta$  and  $B$ . In this case the total derivative condition 2) becomes

$$\frac{df_k}{d\Theta} = \sum_{i=1}^N \frac{\partial f_k}{\partial d_i} \beta \frac{d_i}{\Theta} = 0, \quad (34)$$

$$\frac{df_k}{dB} = \sum_{i=1}^N \frac{\partial f_k}{\partial d_i} \beta \log(\gamma_i) \frac{d_i}{B} = 0, \quad (35)$$

which, if  $f_k$  is a power law  $f_k = \prod_i d_i^{b_i^k}$ , is simply

$$\sum_{i=1}^N b_i^k = 0, \quad (36)$$

$$\sum_{i=1}^N b_i^k \log(\gamma_i) = 0. \quad (37)$$

We have done this explicitly for the example data in Table 2 for 10 galaxies. This sample has characteristics not too dissimilar from those of the next  $H(z)$  release (Moresco et al., in preparation). The resulting combinations are the eight that satisfy

$$b_9^k = \frac{1}{5}(-30b_1 - 30b_2 - 30b_3 - 2b_4 - 3b_5 - 5b_6 - 15b_7 - 23b_8), \quad (38)$$

**Table 3.** Explicit solutions of equations (34) and (35) satisfying conditions (38) and (39) for  $f_k = \prod_i d_i^{b_i^k}$ , using the measurements given in Table 2.  $k$  is the row index, and the column index  $i$  is ordered as in Table 2, i.e.  $i = 1$  correspond to the first row in that table,  $i = 2$  to the second and so on.

$b_1$	$b_2$	$b_3$	$b_4$	$b_5$	$b_6$	$b_7$	$b_8$	$b_9$	$b_{10}$
1	0	0	0	0	0	0	0	-6	5
0	1	0	0	0	0	0	0	-6	5
0	0	1	0	0	0	0	0	-6	5
0	0	0	1	0	0	0	0	-2/5	-3/5
0	0	0	0	1	0	0	0	-3/5	-2/5
0	0	0	0	0	1	0	0	-1	0
0	0	0	0	0	0	1	0	-3	2
0	0	0	0	0	0	0	1	-23/5	18/5

$$b_{10}^k = 5b_1 + 5b_2 + 5b_3 - \frac{3b_4}{5} - \frac{2b_5}{5} + 2b_7 + \frac{18b_8}{5}. \quad (39)$$

We explicitly give a set of eight independent solutions in Table 3. One may then use these combinations to produce robust measurements of the redshift as a function of cosmic time and apply it for example to the case of dark energy. Of course this cancellation of the systematics depends on the way we model them, but there is no method to treat completely unknown systematic errors. This method can be generalized to the more general case where the full spectrum is used (Jimenez et al. 2003; Simon et al. 2005; Stern et al. 2010), although in this case the effect of the metallicity as a systematic will be smaller.

### 3.5 Hubble diagram from SN1a: removing second parameter dependence

The above example was motivated by the cosmic clocks problem, but to show that the solution found is general enough to be applied in different contexts, we consider a different problem in this section. SNe1a are considered to be standard candles, so that, for a sample of SNe spanning a wide redshift interval their apparent magnitude can be used to infer the luminosity distance as a function of redshift and thus constrain cosmological parameters.

In reality, SNe1a are standardizable candles: an empirical algorithm (Phillips 1993) relates peak luminosity and stretch of the light curve. Using this procedure one can exploit SNe1a as if they were standard candles; the scatter around the luminosity distance relation gets dramatically reduced, and thus obtain a luminosity distance  $d_L$  using the distance modulus  $\mu$ :

$$\mu = m - M = 25 + 5 \log d_L, \quad (40)$$

where  $d_L$  encloses the dependence on cosmological parameters,  $m$  denotes the SN apparent magnitude and  $M$  the absolute magnitude. The distance modulus is related to the peak SN magnitude in  $r$  band by

$$m_{\text{peak}} = M_{\text{stand}} - \alpha(s - 1) + \mu(z) + K(s, z) + A_r, \quad (41)$$

where  $M_{\text{stand}}$  is the standardized peak  $B$  absolute magnitude,  $\alpha$  is the Phillips (1993) parameter relating stretch and peak luminosity,  $s$  is a typical stretch,  $\mu(z)$  is the distance modulus,  $A_r$  is the  $r$ -band Milky Way extinction along the SN’s line of sight and  $K(s, z)$  is an approximate  $K$ -correction to the  $r$  band at  $z = 0$ . Of course, one could use any optical band besides  $r$ . All these corrections have a residual scatter and there is evidence that some of the residual scatter could be due to some ‘second parameter’ (e.g. Aubourg et al. 2008; Brandt et al. 2010; Sullivan et al. 2010) such as metallicity



of the host galaxy. There are thus two sources of errors: (a) the fact that, after all the corrections have been applied the average absolute magnitude  $\bar{M}$ , although common to all SNe in the sample, is unknown and (b) that there is a residual scatter around  $M$  which could, in principle, not be purely statistical but depends on extra parameters. For future data this may become the limiting factor in improving cosmological constraints from SNe observations.

Thus we can recast the above problem as

$$\hat{m}_i = \bar{M} + \Delta M_i + \mu_i, \quad (42)$$

where  $i = 1, N$  runs over the SN in the sample,  $\Delta M_i$  includes all residual systematic errors,  $\hat{m}_i$  the ( $k$ -corrected, extinction corrected, stretch corrected) observable quantities and  $\mu_i$  enclose the cosmological dependence and is the quantity of interest. As these are logarithmic quantities, the reader will immediately realize that this problem is similar to the cosmic clock one where  $\log \Delta = \Delta M_i$ ,  $Z = \log \bar{M}$  and  $\log \alpha = \mu_i$ . One can completely cancel the dependence on  $\bar{M}$  by taking linear combinations  $f_k = \sum_i b_i^k m_i$  that satisfy

$$\sum_i b_i^k = 0, \quad (43)$$

similarly to the BAO case. One thus obtains combinations that are robust under changes in the estimated value of  $\bar{M}$ .

We can do even better and attempt to cancel the remaining systematics  $\Delta M_i$ . There are indications that the absolute magnitude depends also on the metallicity of the SN (or host galaxy) and that this is the leading contribution to  $\Delta M_i$ , so we will assume that  $\Delta M_i$  depends on the host galaxy metallicity  $Z$  and model  $\Delta M(Z)$  as a polynomial<sup>6</sup> of some power  $\ell$ :

$$\Delta M(Z) = \Theta_1(Z - Z_0) + \Theta_2(Z - Z_0)^2 + \dots + \Theta_\ell(Z - Z_0)^\ell, \quad (44)$$

where  $\Theta_i$  for  $i = 1, \dots, \ell$  and  $Z_0$  is unknown parameter. This can also be thought of as a Taylor expansion of the dependence of  $\Delta M$  on  $Z$ . This polynomial can be explicitly expanded and rewritten in the following form:

$$\Delta M(Z) = \tilde{\Theta}_0 + \tilde{\Theta}_1 Z + \tilde{\Theta}_2 Z^2 + \dots + \tilde{\Theta}_\ell Z^\ell, \quad (45)$$

where  $\tilde{\Theta}_i$  is combinations of  $\Theta_j$  and  $Z_0$ . There is an ongoing effort to measure the metallicities, so that here we take them to be part of our space of observables. Therefore, our nuisance parameters will be  $\bar{M}$  and  $\tilde{\Theta}_i$  for  $i = 0, 1, \dots, \ell$ . We find that the total derivative conditions can then be written as

$$\frac{df_k}{d\tilde{\Theta}_0} \propto \frac{df_k}{d\bar{M}} \propto \sum_{i=1}^N \frac{\partial f_k}{\partial \hat{m}_i} = 0, \quad (46)$$

$$\frac{df_k}{d\tilde{\Theta}_1} = \sum_{i=1}^N \frac{\partial f_k}{\partial \hat{m}_i} Z_i = 0, \quad (47)$$

$$\frac{df_k}{d\tilde{\Theta}_2} = \sum_{i=1}^N \frac{\partial f_k}{\partial \hat{m}_i} Z_i^2 = 0, \quad (48)$$

<sup>6</sup> Although with present data there is only an indication of a trend of a dependence of the magnitude on the metallicity, this dependence in detail may be better quantified when more data become available. We assume here it can be modelled by a polynomial, of course the degree of the polynomial will depend on the nature and amount of the data and on the underlying – yet unknown –  $M$ - $Z$  relation.

$\vdots$

$$\frac{df_k}{d\tilde{\Theta}_\ell} = \sum_{i=1}^N \frac{\partial f_k}{\partial \hat{m}_i} Z_i^\ell = 0. \quad (49)$$

The solution to the set of differential equations is given by a linear combination  $f_k = \sum_i b_i^k \hat{m}_i$ , where the coefficient  $b_i^k$  is the  $N - \ell - 1$  that satisfy the following conditions:

$$\sum_{i=1}^N b_i^k = 0, \quad (50)$$

$$\sum_{i=1}^N b_i^k Z_i = 0, \quad (51)$$

$$\sum_{i=1}^N b_i^k Z_i^2 = 0, \quad (52)$$

$\vdots$

$$\sum_{i=1}^N b_i^k Z_i^\ell = 0. \quad (53)$$

We expect the results obtained here to be robust also under changes in the  $\Theta_\ell$ ,  $\bar{M}$  and  $Z_0$ .

Now let us compare the solution provided by this method with the standard approach. When analysing a SN sample it is customary to marginalize over  $\bar{M}$  as follows: for each choice of cosmological parameters ( $\alpha$ )  $\mu(z_i)$  is computed and used to extract  $\bar{M}(\alpha)$ ;  $\Delta M_i$  then encloses the statistical error. One can see that this procedure is equivalent to marginalize over  $H_0$ . In fact one can rewrite the apparent magnitude as

$$m_i = M - 5 \log H_0 + 25 + 5 \log(H_0 d_L)_i. \quad (54)$$

This procedure cannot be unbiased if the set of cosmological models/parameters scanned does not include the true underlying model. The solution provided by the present method instead cancels out the dependence on  $\bar{M}(\alpha)$  providing an unbiased answer, and can also mitigate the impact of remaining systematics (such as the influence of the metallicity on the absolute magnitude). Of course this cancellation of the remaining systematics depends on the way we model them, but there is no way of treating completely unknown systematic uncertainties.

### 3.6 CMB angular power spectrum

The angular power spectrum computed from the autocorrelation of a map produced by a given detector is given by  $C_\ell^{\text{meas}} = C_\ell^{\text{true}} + C_\ell^{\text{noise}}$ , where  $C_\ell^{\text{noise}}$  denotes the power spectrum of the detector noise and is called ‘noise bias’: a poor knowledge of the noise bias will bias the estimate of the CMB power spectrum. On the other hand the cross power spectrum for maps produced by different, uncorrelated, detectors does not have noise bias. In this subsection we wish to reproduce this well-known fact in the light of our approach. This is a simple example of a problem which is neither linear nor a power law, and can thus show how to apply our approach in more general cases.

Let us model this case as having two detectors that give two different measurements of the temperature perturbations of the CMB

with uncorrelated noise:

$$\begin{aligned} a_{\ell m}^{(1)} &= a_{\ell m} + N_{\ell m}^{(1)}, \\ a_{\ell m}^{(2)} &= a_{\ell m} + N_{\ell m}^{(2)}, \end{aligned} \quad (55)$$

where  $a_{\ell m}^{(1,2)}$  is the temperature anisotropies decomposed in spherical harmonics measured by each detector,  $a_{\ell m}$  is the ‘true’ temperature anisotropies and  $N_{\ell m}^{(1,2)}$  is the noise of each detector. We wish to apply the conditions given in equation (2) to this case. Since there is a different contribution of the noise per each  $m$ , it would appear that the problem has no solution as there are as many nuisance parameters as observables. However not all of them are independent since they are uncorrelated among themselves and with the ‘true’ temperature anisotropies, and must thus satisfy the following constraints:

$$\sum_m N_{\ell m}^{(1)} N_{\ell m}^{(2)} = 0, \quad (56)$$

$$\sum_m N_{\ell m}^{(1)} a_{\ell m} = 0, \quad (57)$$

$$\sum_m N_{\ell m}^{(2)} a_{\ell m} = 0. \quad (58)$$

One way to approach the problem is to use all these constraints in order to remain only with independent quantities. Though this gives the desired result, it is rather involved algebraically. We will instead use Lagrange multipliers and impose that the total derivative of the following function with respect to each  $N_{\ell m}^{(1,2)}$  be zero while keeping all of them independent:

$$f_k + \lambda_1^k \sum_m N_{\ell m}^{(1)} N_{\ell m}^{(2)} + \lambda_2^k \sum_m N_{\ell m}^{(1)} a_{\ell m} + \lambda_3^k \sum_m N_{\ell m}^{(2)} a_{\ell m}, \quad (59)$$

which gives the following two equations:

$$\frac{\partial f_k}{\partial a_{\ell m}^{(1)}} + \lambda_1^k a_{\ell m} + \lambda_2^k N_{\ell m}^{(2)} = 0, \quad (60)$$

$$\frac{\partial f_k}{\partial a_{\ell m}^{(2)}} + \lambda_1^k a_{\ell m} + \lambda_3^k N_{\ell m}^{(1)} = 0. \quad (61)$$

Imposing that the  $f_k$  depends explicitly only on  $a_{\ell m}^{(1,2)}$  implies  $\lambda_1^k = \lambda_2^k = \lambda_3^k$ , and the differential equations can be rewritten as

$$\frac{\partial f_k}{\partial a_{\ell m}^{(1)}} + \lambda_1^k a_{\ell m}^{(2)} = 0, \quad (62)$$

$$\frac{\partial f_k}{\partial a_{\ell m}^{(2)}} + \lambda_1^k a_{\ell m}^{(1)} = 0, \quad (63)$$

with the well-known solution

$$f_k = -\lambda_1^k \sum_m a_{\ell m}^{(1)} a_{\ell m}^{(2)}, \quad (64)$$

where any value of  $\lambda_1^k$  is equally good since it is simply a global factor. We have thus recovered the fact that the cross-correlations are independent of the noise.

## 4 CONCLUSIONS

We have presented an algorithm to minimize, or even completely cancel out, the effect of systematic uncertainties (nuisance parameters) that can somehow be modelled, even if the modelling is very rough or approximate. It is of use in interesting cases for which systematic errors are larger than statistical ones. The method was

inspired by renormalization group techniques, and this interpretation provides an elegant description of what nuisance parameters are: we identify the physics that is invariant under arbitrary rescalings of the nuisance quantities. This nuisance-independent physics is completely characterized by the *nuisance scaling dimensions* of the observables. In the limit case where nuisance can be totally washed out this recipe will unravel the underlying responsible *scale invariance*.

Our general approach is given in equations (1) and (2) and we report explicit recipes for cases where the observables have power law and linear dependences on nuisance parameters in equations (11) and (12), respectively, which apply also if only an approximate cancellation of the nuisance parameter is possible. Additional constraints on nuisance parameters or observables can be included (equation 14). The algorithm is general and can be applied to any experiment or observation. However, because of our own field of expertise, we have chosen to illustrate the method with some examples drawn from astrophysics and cosmology. In doing so we have discovered some interesting new results that can be used to analyse observations from large-scale galaxy and SN surveys. For the case of BAO experiments, we provide an optimal way to combine the observed quantities to reduce systematic uncertainties in the sound horizon scale  $r_s$ . For SN surveys we have provided a method to remove a dependence of the Hubble diagram of systematic uncertainties due to metallicity variability of the SN (or, as a proxy, that of the host galaxy). Finally, we show how for the cosmic clock method the feared age–metallicity degeneracy can be completely removed by choosing adequate combinations of the observed quantities. We chose to focus on these techniques because they are the most promising to unveil the nature of dark energy; removing or minimizing the dependence of this methods on poorly constrained systematics is crucial to be able to gain full advantage of the significant observational effort that is being invested in observational cosmology. Of course if a full knowledge of the central values of the nuisance parameters and their errors is known, our approach cannot be better than the Bayesian technique of marginalizing over the distribution of the nuisance parameters. On the other hand, if the central values and distributions of the nuisance parameters are not known, the Bayesian approach with an inaccurate prior for the nuisance parameters will introduce biases, while our technique aims at cancelling or minimize these possible biases. As we have emphasized before our method is general and we hope it will be used in other areas of the experimental sciences.

## ACKNOWLEDGMENTS

It is a pleasure to thank Roland de Putter and Ben Hoyle for useful discussions. JN is supported by FP7- IDEAS Phys.LSS 240117. LV acknowledges support from FP7-PEOPLE-2007-4-3-IRG n. 202182 and FP7- IDEAS Phys.LSS 240117. LV and RJ are supported by MICINN grant AYA2008-03531. CP-G is supported by the Spanish MICINN grant FPA-2007-60323 and the Generalitat Valenciana grant PROMETEO/2009/116. The work of CG was supported in part by grants: FPA 2009-07908, CPAN (CSD2007-00042) and HEPHACOS-S2009/ESP1473.

## REFERENCES

Asplund M., Grevesse N., Sauval A. J., 2005, in Barnes T. G., III, Bash F. N., eds, ASP Conf. Ser. Vol. 336, Cosmic Abundances as Records of Stellar Evolution and Nucleosynthesis. Astron. Soc. Pac., San Francisco, p. 25

- Aubourg É., Tojeiro R., Jimenez R., Heavens A., Strauss M. A., Spergel D. N., 2008, *A&A*, 492, 631
- Bahcall J. N., Ulmer A., 1996, *Phys. Rev. D*, 53, 4202
- Brandt T. D., Tojeiro R., Aubourg É., Heavens A., Jimenez R., Strauss M. A., 2010, *AJ*, 140, 804
- de Bernardis F., Bean R., Galli S., Melchiorri A., Silk J. I., Verde L., 2009, *Phys. Rev. D*, 79, 043503
- Eisenstein D. J., White M., 2004, *ApJ*, 664, 660
- Fox P. J. J., Kribs J. D., Tait T. M. P., 2011a, *Phys. Rev. D*, 83, 034007
- Fox P. J., Liu J., Weiner N., 2011b, *Phys. Rev. D*, 83, 103514
- Grevesse N., Sauval A. J., 1998, *Space Sci. Rev.*, 85, 161
- Jimenez R., Loeb A., 2002, *ApJ*, 573, 37
- Jimenez R., Verde L., Treu T., Stern D., 2003, *ApJ*, 593, 622
- Mangilli A., Verde L., Beltran M., 2010, *J. Cosmol. Astropart. Phys.*, 10, 9
- March M. C., Trotta R., Amendola L., Huterer D., 2011, *MNRAS*, 415, 143
- Moresco M., Jimenez R., Cimatti A., Pozzetti L., 2011, *J. Cosmol. Astropart. Phys.*, 3, 45
- Pena-Garay C., Serenelli A., 2008, preprint (arXiv:0811.2424)
- Phillips M. M., 1993, *ApJ*, 413, L105
- Seo H.-J., Eisenstein D. J., 2007, *ApJ*, 665, 14
- Serenelli A. M., Haxton W. C., Pena-Garay C., 2011, preprint (arXiv:1104.1639)
- Simon J., Verde L., Jimenez R., 2005, *Phys. Rev. D*, 71, 123001
- Song Y. S., Percival W. J., 2009, *J. Cosmol. Astropart. Phys.*, 10, 4
- Stern D., Jimenez R., Verde L., Kamionkowski M., Stanford S. A., 2010, *J. Cosmol. Astropart. Phys.*, 2, 8
- Sullivan M. et al., 2010, *MNRAS*, 406, 782
- Villante F. L., Fiorentini G., Lisi E., 1999, *Phys. Rev. D*, 59, 013006
- Wilson K. G., Kogut J., 1974, *Phys. Rev.*, 12, 75

## APPENDIX A: THE GENERAL PROBLEM

In this appendix we wish to suggest an interpretation of equation (2) that might prove useful when solving more general problems. As in the main text, we consider  $N$  observables  $O_i(\mu_1, \dots, \mu_m, \nu_1, \dots, \nu_n)$  that depend on  $m$  accurate or interesting quantities  $\mu_i$  and  $n$  ‘nuisance’ or ‘biased’ parameters  $\nu_i$ . Let us keep the quantities  $\mu_i$  fixed at some value, and assume that there are more observables than nuisance parameters  $N > n$ . The functions  $O_i(\nu_1, \dots, \nu_n)$  then define a mapping between  $\mathbb{R}^n$  and  $\mathbb{R}^N$ , which under certain smoothness assumptions defines a manifold (an  $n$ -dimensional hypersurface embedded in  $\mathbb{R}^N$ ). If this manifold is orientable, there will be at least one vector field  $v_i^1$  orthogonal to it, and in general there can be up to  $N - n$  such vector fields. The integrals of these vector fields (which exist at least locally) will be constant on the hypersurface:

$$\sum_{i=1}^N v_i^k \frac{\partial O_i}{\partial \nu_j} = \sum_{i=1}^N \frac{\partial f^k}{\partial O_i} \frac{\partial O_i}{\partial \nu_j} = 0, \quad (\text{A1})$$

where  $f_k$  is the integral of  $v_i^k$ . This is the same as equation (2). The problem is then to find such vector fields and integrate them.

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.