

Model breaking measure for cosmological surveysAdam Amara^{*} and Alexandre Refregier[†]*Department of Physics, Institute for Astronomy, ETH Zurich, Wolfgang-Pauli-Strasse 27,
CH-8093 Zurich, Switzerland*

(Received 22 September 2013; published 1 April 2014)

Recent observations have led to the establishment of the concordance Λ CDM model for cosmology. A number of experiments are being planned to shed light on dark energy, dark matter, inflation and gravity, which are the key components of the model. To optimize and compare the reach of these surveys, several figures of merit have been proposed. They are based on either the forecasted precision on the Λ CDM model and its expansion or on the expected ability to distinguish two models. We propose here another figure of merit that quantifies the capacity of future surveys to rule out the Λ CDM model. It is based on a measure of the difference in volume of observable space that the future surveys will constrain with and without imposing the model. This model breaking figure of merit is easy to compute and can lead to different survey optimizations than other metrics. We illustrate its impact using a simple combination of supernovae and baryon acoustic oscillation mock observations and compare the respective merit of these probes to challenge Λ CDM. We discuss how this approach would impact the design of future cosmological experiments.

DOI: 10.1103/PhysRevD.89.083501

PACS numbers: 98.80.Es

I. INTRODUCTION

Recent progress in cosmological observations have led to the establishment of the Λ CDM model as the standard model for cosmology. This simple model is able to fit a wide array of observations with about six parameters [1–3]. In spite of its success, several key ingredients of the model are not fully understood and have been introduced to fit the data rather than being derived from fundamental theory. These include dark matter [4], which (if attributed to particles) exists outside the standard model of particle physics and dark energy [5,6]. The other ingredients of the model are associated with inflation, which conditions the initial state of the Universe, and Einstein gravity, which has not been tested on cosmological scales.

Alternatives to the Λ CDM model are numerous and growing. However since the data is currently consistent with the Λ CDM model, progress in the field will likely be driven by the acquisition of new data that can be used to further challenge the model. In so doing, we hope to find evidence that will lead to a deeper understanding of physical processes and point us towards more fundamental alternative models. Significant amounts of current efforts in cosmology are thus focused on the design of future experiments that can optimally increase our cosmological knowledge. However, since there exists a wide array of equally compelling alternative models, finding a suitable metric with which to compare and optimize future experiments is challenging.

At present, the dominant metric for gauging the quality of planned experiments is the dark energy task force (DETF) figure of merit (FoM) [7]. This metric consists of expanding the simplest Λ CDM model so that the dark energy component is modeled as having an equation of state w , which is given by the ratio of pressure to density of dark energy. This equation is assumed to evolve linearly with scale factor a , $w(a) = w_0 + (1 - a)w_a$ [8,9]. The DETF FoM can then be derived from the determinant of the covariance matrix of the two dark energy parameters w_0 and w_a , which can be calculated using Fisher matrix methods [10]. Since the linear expansion of the equation of state is only one of many possible extensions beyond Λ CDM, relying solely on this optimization may lead to biases in experiment design.

An alternative approach, which was proposed by the follow-up committee known as the DETF FoM Working Group [11], is to consider a more general expression for the equation of state. This approach relies on principle component analysis (PCA) methods to find the fundamental modes that a given experiment can measure. In their report, the DETF FoM Working Group suggests a prescription where the equation of state is divided into 36 redshift bins out to $z \sim 10$. One difficulty, however, is that Fisher matrix calculations can be unstable. The final results, therefore, can depend on the user's choice of initial basis set, which once again may not be well motivated and can lead to unintended selection biases [12]. The DETF FoM Working Group also advocates to use alternative theoretical expansions that can be used to model possible deviations of gravity from Einstein's theory [13,14].

Numerous alternative metrics have been proposed in the literature. As well as further PCA based techniques

^{*}adam.amara@phys.ethz.ch[†]alexandre.refregier@phys.ethz.ch

[15–18], other methods have been proposed that include in the determinant calculations parameters of Λ CDM beyond those of the equation of state. These include the integrated parameter space optimization (IPSO) [19–21], and model selection methods based on forecasting the Bayes factor [22–26]. The latter approach relies on comparing two models and calculating the Bayes factor [$B_{01} \equiv p(d|M_0)/p(d|M_1)$], which quantifies the odds of which model (M_0 or M_1) is preferred by the data (d). This method still requires a choice of an alternative model to which the null model can be compared.

The end result can vary, depending on the FoM used. This is ultimately due to the fact that the FoMs are being used to ask subtly different questions. In an era where the total amount of data is growing, it is conceivable and fully expected that different FoMs will lead to similar optimization. However, as experiments begin to fill the entire available cosmic volume, the trade-offs are likely to become more subtle. Hence, care should be given to focus precisely on the questions that we want to address.

In this paper, we explore the motivational question: *Which experiment is most likely to find data that will falsify Λ CDM?* Given the success of Λ CDM so far, the detection of any deviation from this model would be a major discovery. These deviations may not necessarily emerge as a deviation from $w = -1$. As a result, to answer the motivational question above we formulate a new figure of merit, building on earlier work [27], which can be readily calculated using Gaussian approximations. In its purest form this figure of merit can be calculated using only (i) current data, (ii) the predictions from the simple Λ CDM model that we wish to challenge and (iii) the expected covariance matrix of the data for a future experiment. As part of our work, we also show how robust theoretical priors, such as light propagation on a metric, can also be included in the calculation, if so desired. While the DETF FOM and the Bayes ratio approach are, respectively, related to model fitting and model selection, our approach is related to the problem of model testing.

This paper is organized as follows. In Sec. II, we derive our new FoM and show the Gaussian approximation version of the calculation. In Sec. III, we investigate a simple cosmological toy-model example to illustrate our method. In this section we also compare our calculations to an FoM derived from the determinant of the Fisher matrix of the standard Λ CDM parameters. Finally, in Sec. IV we offer a discussion to summarize our findings.

II. FORMALISM

The basic principle of our approach is to make comparisons between the likely outcomes of future experiments in data space. In its purest form, this is a comparison between $p(D_f|D_c)$ and $p(D_f|D_c, \Theta)$, where the former is the probability of future data D_f , given only current data D_c and the latter is the probability of future data given current

data and the constraint that the standard model (with parameters Θ) being studied (in our case standard Λ CDM) must hold. In this empirical case, we can calculate the probability of future data by integrating over all possible values of the data (see derivation in the Appendix) such that

$$p(D_f|D_c) = \int p(D_f|T)p(T|D_c)dT, \quad (1)$$

where we have introduced the concept of “true” value T that corresponds to the value we obtain as the errors tend to zero. For the case where we assume a standard model holds, we can calculate the probability of future data by integrating over all possible values of the model parameters, Θ ,

$$p(D_f|D_c, \Theta) = \int p(D_f|\Theta)p(\Theta|D_c)d\Theta. \quad (2)$$

In both cases, we can calculate the probabilities of the underlying variables given today’s data. For instance, in the case of the model parameters,

$$p(\Theta|D_c) = \frac{p(D_c|\Theta)p(\Theta)}{p(D_c)}. \quad (3)$$

Given two density distributions [for instance Eqs. (1) and (2)] we will need to be able to quantitatively compare them. For this, the concept of information entropy, which quantifies the level of uncertainty, is useful. A robust measure for this purpose is the relative entropy, also known as the Kullback-Leibler (KL) divergence [28], between the two distributions. In this case, this can be calculated as

$$KL[p, q] = \int \ln\left(\frac{p(x)}{q(x)}\right)p(x)dx, \quad (4)$$

where $p(x)$ and $q(x)$ are the two probability distributions to be compared. This measure quantifies the difference of information in the two cases and provides a measure of the difference between the two distributions.

Using this measure our proposed figure of merit measure for model breaking is simply

$$\Phi = KL[p(D_f|D_c, \Theta), p(D_f|D_c)]. \quad (5)$$

A. The Gaussian case

The analysis outlined above is general and can be used to study probability distribution functions of arbitrary shape. However, due to their simplicity, probability distribution functions (PDFs) that are multivariate Gaussians are very attractive cases to study. In this case, the probabilities would be given by

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{k/2}|\mathbf{C}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right], \quad (6)$$

where k is the number of dimensions, $\boldsymbol{\mu}$ is the mean (i.e. peak) of the PDF and \mathbf{C} is the covariance matrix.

The relative entropy between two multivariate Gaussian distributions, e.g. $p(\mathbf{x})$ and $q(\mathbf{x})$, with the covariance matrices \mathbf{C}_p and \mathbf{C}_q is given by [29]

$$KL = \frac{1}{2} \ln |\mathbf{C}_p \mathbf{C}_q^{-1}| + \frac{1}{2} \text{tr} \mathbf{C}_p^{-1} ((\boldsymbol{\mu}_q - \boldsymbol{\mu}_p)(\boldsymbol{\mu}_q - \boldsymbol{\mu}_p)^T + \mathbf{C}_q - \mathbf{C}_p). \quad (7)$$

For the simplest case, where the two distributions have the same mean, this reduces to

$$KL = \frac{1}{2} \ln |\mathbf{C}_p \mathbf{C}_q^{-1}| + \frac{1}{2} \text{tr} \mathbf{C}_p^{-1} (\mathbf{C}_q - \mathbf{C}_p). \quad (8)$$

B. Calculating the covariance matrix

Given the covariance matrix for current data, \mathbf{C}_c , we can compute the covariance matrix, \mathbf{C}_m , of the parameters of the model that need to be adhered to. This can be done by calculating the Fisher matrix, \mathbf{C}_m^{-1} , through a matrix rotation, as

$$\mathbf{C}_m^{-1} = \mathbf{Y} \mathbf{C}_c^{-1} \mathbf{Y}^T, \quad (9)$$

where \mathbf{Y} is the Jacobian matrix of derivatives such that $\mathbf{Y}_{ij} = \partial \mathbf{D}_i / \partial \Theta_j$. In principle, it is also possible to have constraints on the Θ parameters for external data that will not change in the future. To make the predictions for future error bars, we can then project back to the covariance in the observables, \mathbf{C}_x , based on existing errors

$$\mathbf{C}_x = \mathbf{Y}^T \mathbf{C}_m \mathbf{Y}. \quad (10)$$

Finally, to calculate the full covariance matrix for future data, \mathbf{C}_1 , of $p(D_f | D_c, M)$, we need to account for the error bars associated with the future experiment, given by the matrix \mathbf{C}_f . The full matrix corresponding to the operation in Eq. (2) is then

$$\mathbf{C}_1 = \mathbf{C}_x + \mathbf{C}_f. \quad (11)$$

For the case of the purely empirical predictions, where there is no model and data vector entries are independent of each other, the Jacobian matrices \mathbf{Y} become the identity matrix \mathbf{I} , which greatly simplifies the equations above. For instance, in the case where no external data set is used, the covariance matrix of $p(D_f | D_c, M)$ becomes

$$\mathbf{C}_0 = \mathbf{C}_c + \mathbf{C}_f. \quad (12)$$

In this case the model breaking figure of merit defined in Eq. (5) and using (8) reduces to

$$\Phi = \frac{1}{2} \ln |\mathbf{C}_1 \mathbf{C}_0^{-1}| + \frac{1}{2} \text{tr} \mathbf{C}_1^{-1} (\mathbf{C}_0 - \mathbf{C}_1). \quad (13)$$

In the case where we also want to consider shifts in mean values one would use an analogous expression with extra terms coming from Eq. (7).

The two cases above are the extreme examples: (i) one where all the data points are correlated with all other data point when projected through a model and (ii) the case where all the data points are independent from each other. It is possible to construct an intermediate case that we call a minimal model that defines a weak correlation between subsets of the data. For example, in the cosmological setting we could introduce a correlation between data taken at the same redshift, while making the data from two different redshifts fully independent. In this case, the Jacobian would be constructed using the derivatives with respect to the data, i.e. $\mathbf{Y}_{ij} = \partial \mathbf{D}_i / \partial \mathbf{D}_j$ and \mathbf{C}_0 would be modified accordingly for the model breaking figure of merit in Eq. (13).

III. COSMOLOGICAL EXAMPLE

To demonstrate our approach, we briefly explore a simple cosmological example. For this we will focus on geometrical tests, namely supernovae flux decrements and tangential and radial measurements of the baryon acoustic oscillation (BAO) scales.

A. Background cosmology

Within the standard Λ CDM concordance model, the geometry measure can be derived from the line of sight comoving distance, χ ,

$$\chi(a) = c \int \frac{da}{a^2 H(a)}, \quad (14)$$

where c is the speed of light, a is the scale factor and $H(a)$ is the Hubble function. The Hubble function can be easily calculated in the Λ CDM and in the late time Universe by the Friedmann equation,

$$H^2(a) = H_0^2 \left(\frac{\Omega_m}{a^3} + \frac{\Omega_k}{a^2} + \Omega_\Lambda \right), \quad (15)$$

where Ω_m is the matter over density, Ω_Λ is the dark energy density and Ω_k is the curvature. The curvature term can be defined through the relation $\Omega_m + \Omega_\Lambda + \Omega_k = 1$. With this, it is clear that we can describe the geometry measures through three free parameters: h , Ω_m and Ω_Λ , where we use the standard approach of recasting the Hubble constant, H_0 , as the dimensionless quantity through $h = H_0 / 100 \text{ Km s}^{-1} \text{ Mpc}^{-1}$.

Observed distance measures are typically determined through the angular diameter distance (D_A) and the luminosity distance (D_L), which can be related to each other through the scale factor

$$D_A = a^2 D_L = ar(\chi), \quad (16)$$

where $r(\chi)$ is the comoving angular diameter distance. The supernovae technique measures the distance modulus (ΔD_M) as a function of redshift [5], where the distance modulus is determined from the flux ratio between the absolute and apparent fluxes of SNe. This can then be linked to the radial comoving distance through the luminosity distance D_L ,

$$\Delta D_M = 5 \log \left(\frac{D_L}{10 \text{ pc}} \right). \quad (17)$$

We also find it useful to define a new quantity,

$$R_{DM} = \left(\frac{D_L}{10 \text{ pc}} \right), \quad (18)$$

which contains the same information in a form closer to flux ratios rather than magnitude differences.

Baryon acoustic oscillation (BAO) studies rely on using galaxy surveys to measure the same acoustic peaks that are seen in the CMB, thereby using the scale set by these peaks as a standard ruler. The measurements can be made perpendicular to the line of sight (r_p) and along the line of sight (r_{pa}), which can be linked to the observed angular scale $\Delta\theta$ and redshift extent Δz [21],

$$\Delta\theta = \frac{ar_s}{D_A}, \quad (19)$$

and

$$\Delta z = \frac{Hr_s}{c}, \quad (20)$$

where r_s is the sound horizon [1] that, for simplicity, we set to 140 Mpc in this study. In this framework, the observable quantities are $\mathbf{O} = \{R_{DM}(a), \Delta\theta(a) \text{ and } \Delta z(a)\}$ and the model parameters are $\Theta = \{h, \Omega_m, \Omega_\Lambda\}$. With this in place, calculating $p(D_f|D_c, \Theta)$ is straightforward once the covariance matrices for current and future measurements, C_c and C_f , have been specified using Eqs. (9) and (12). The different levels of theoretical assumption can be viewed as having the hierarchy illustrated in Fig. 1. Most figures of merit build from extensions of the Λ CDM model (i.e. from the inside out), while our approach consists of the comparison of no or little theoretical assumptions with the Λ CDM model (i.e. outside in). It is possible to calculate a meaningful model breaking FoM using only (i) the simplest Λ CDM model being tested, (ii) current data and (iii) prediction of future error bars (corresponding to the outermost ring). This calculation would not include any priors on the classes of alternative theories. However, such priors can be added explicitly, as illustrated in Fig. 1.

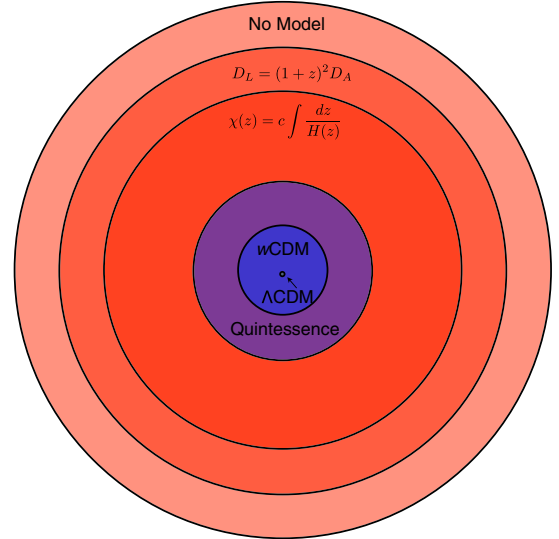


FIG. 1 (color online). Illustration of a hierarchy of model spaces. The center shows the most restricted point representing the Λ CDM model. Building out from the center, we show increasingly more flexible models, starting with the $w_0 - w_a$ expansion of the equation of state and then more generic quintessence models. The outer ring of the figure shows the constraints coming purely from the data, i.e. no model. Most figures of merit build out from the center. Our model challenging approach builds from the outside inwards by comparing the central Λ CDM point with the outer minimal theory layers.

B. Application

To demonstrate the approach outlined here, we construct a simple illustrative example. For this example we assume that three observables, R_{DM} , $\Delta\theta$ and Δz , have each been measured at ten points in the redshift range $z = [0.1, 2.1]$. These, therefore, would be a simplified example of what we would measure from a combination of SNe and BAO experiments. We set the current relative errors on the measurements coming for SNe (R_{DM}) to be 5% and the errors on the BAO measurements ($\Delta\theta$ and Δz) to be 10%. These errors are assumed to be independent and not coming from systematics. Figure 2 shows this configuration. The dotted curves in the figure show the predictions from a Λ CDM model with $h = 0.7$, $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$.

For our first example, we calculate figures of merit for future experiments where the errors are reduced by a given factor. Specifically, we consider four cases. The first is where all the measurements are improved by this factor and three other cases where only one of the probes has been improved. Next, we have decided to calculate the results for our figure of merit calculation, where we assume that there is an integral relation between the Hubble function and the distances [given by Eq. (14)] and that there is a relation between angular diameter and luminosity distances [as given by Eq. (16)]. However, we place no constraints on the functional form of H . This is illustrated by the third layer of Fig. 1. For simplicity in this illustrative example, we have

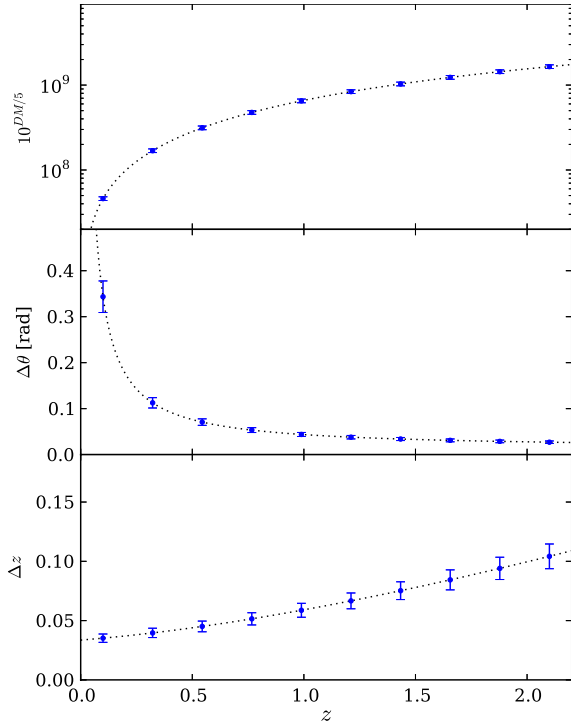


FIG. 2 (color online). Simple example with three observables, each measured in four redshift bins. The blue curve shows the baseline Λ CDM model used in this paper; the points show our toy-model example with 10% errors on all the observables.

assumed that the comoving angular diameter distance is equal to χ regardless of curvature. This calculation, therefore, effectively compares the allowed freedom of future data depending on whether or not the relation shown in Eq. (15) is imposed.

Since the comoving distance and the Hubble function are linked through an integral relationship, it is convenient to remap the data points onto a finer redshift grid so that the mapping from H to χ can be approximated by a matrix product involving a left triangular matrix. This mapping onto a finer grid can be done once the relationship between the fine and coarse grid are defined. For example, the coarse grid are averages over the finer grid, since this can be used to define the appropriate Jacobian for the mapping. For simplicity, we have assumed here that errors scale by the \sqrt{N} , where N is the number of fine points to one coarse data point.

The upper panel of Fig. 3 shows Φ as a function of the power of future surveys. As a comparison, the lower panel of the figure shows a calculation using the standard Fisher matrix methods, with the y axis showing the determinant of the 3×3 Fisher matrix of the future experiment relative to the Fisher matrix from the current data. This is close to IPSO optimization. Both optimization methods show the broad expected trend that higher precision measurements are better, but the details of the optimization are distinctly different. In the Fisher matrix optimization, we see that

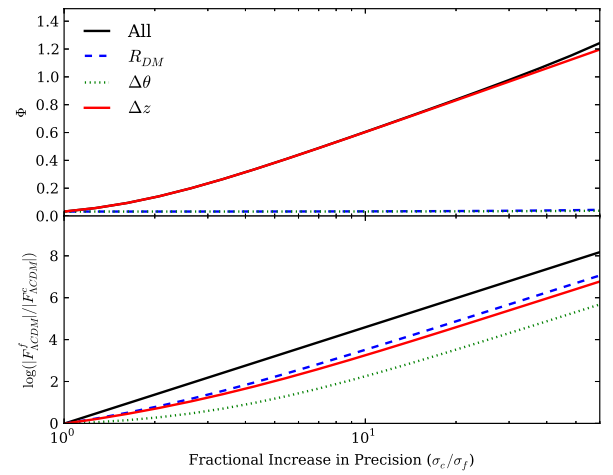


FIG. 3 (color online). The upper panel shows Φ , normalized by number of data points, between predictions using Λ CDM and those where the form of $H(z)$ is not specified. The lower panel shows the determinant of the 3×3 Fisher matrix of the Λ CDM ($h, \Omega_m, \Omega_\Lambda$). The results are shown as a function of increased precision of future experiments. The black curves show results when the measurements of all three probes are improved; the red curves show the results when only the Δz experiment is improved; the blue-dashed curves correspond to only improving the luminosity distance experiment (R_{DM}); and the green-dotted curves are for improvements in the angular diameter distance measurement ($\Delta\theta$).

constraints from R_{DM} and Δz are comparable (with a small preference for R_{DM}) and the constraints from $\Delta\theta$ are weaker. For the case where all the probes are improved, we see significant gains over the individual probe improvements. The optimization using our model breaking FoM with Φ (upper panel) strongly favors the Δz measurements. In fact, even in the case where the precision of all the probes is increased, this causes a negligible improvement in the figure of merit.

In practice, the data (both current and future) have finite resolutions in redshift. Our model breaking framework is thus sensitive to the space of functions that is (i) consistent with today's data and (ii) will cause a notable change in future data. As a result the framework will not be sensitive to variations on scales smaller than these two scales. However, additional smoothing constraints can be imposed—for instance, coming from a suite of well-motivated theories. We would put such constraints in the same category as imposing physics in the model classes figure (Fig. 1). In this case, these extra constraints should be added explicitly and justified clearly.

In our next analysis, we investigate the redshift sensitivity of the probes. Figure 4 shows the results when only the errors at one of the specific redshift (z_b) shown in Fig. 3 are improved by a factor of 10, i.e. $\sigma_c(z_b)/\sigma_f(z_b) = 10$. The Fisher matrix based optimization shows complex behavior, with the distance measure probes favoring improvements at lower redshifts while the measure based

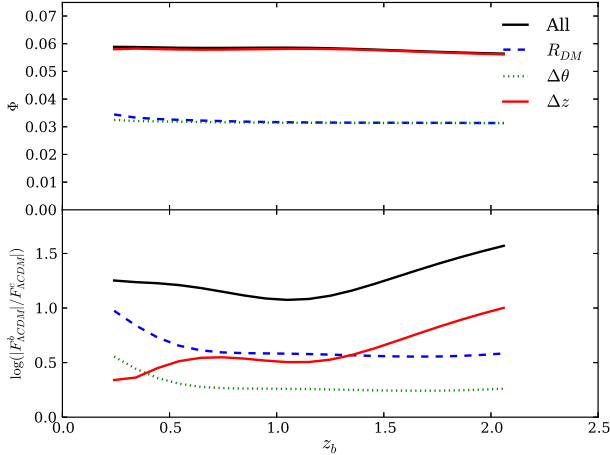


FIG. 4 (color online). Results for the Fisher matrix and our KL divergence based model breaking FoM when the errors at only one of the ten redshift points shown in Fig. 2 are reduced. The y axis and color scheme match those of Fig. 3. $\sigma_c(z_b)/\sigma_f(z_b) = 10$.

on $H(z)$ tends to favor higher redshifts. We also see that the relative importance of the different probes also depends strongly on the redshift range that is being targeted. On the other hand, the optimization based on KL divergence shows relatively simple trends. The ranking of the probes and their strengths is the same as that seen in Fig. 3, and there is almost no redshift preference. This implies that, for our simple model where the current relative errors are fixed as a function of redshift, improved measurements at all redshifts are equally favored. This can be important, since the cost of improving the errors at a given epoch is typically not independent of the redshift being targeted.

IV. DISCUSSION

We have developed a new formalism for calculating the discovery potential of future experiments. This new figure of merit offers a simple and robust alternative to metrics such as the DETF FoM, which focuses on the determinant of the covariance matrix on the dark energy equation of state parameters w_0 and w_a as calculated using Fisher matrix methods. One of the difficulties with the DETF FoM is that a decomposition into w_0 and w_a is not derived from fundamental theory and in fact there has been considerable effort to expand this figure of merit to include more generic $w(a)$ and to rely on principal component analysis methods to capture the most significant modes. Other approaches have been to consider expansions of other *ad hoc* parameters. However, the problem is that since these expansions of the model are not driven by fundamental theory, it becomes difficult to make informed choices about experiment design if different metrics point to different optimal configurations. In addition, the discovery of deviation from Λ CDM in any of the sector of the model, and not only in $w(a)$, would be of profound importance.

The formalism that we present here allows us to calculate a figure of merit for future experiment configurations based on three ingredients: (1) existing data, (2) the standard model to be tested (without extra parameters) and (3) the predictions of the errors for the future experiment. This method then effectively sets out to compare the model, which in cosmology is Λ CDM, against a no model case, which shows all of the allowed data space even in the absence of the model. We have shown that this is a well posed statistical problem and the KL divergence (relative entropy) between the two allowed PDFs in data space allows us to maximize the possibility that a future experiment will measure data that cannot be fit by the standard model. Furthermore, we have shown how physical constraints can be incorporated by including the relationships between the data points that these physical effects introduce.

One of the advantages of the Fisher approach to experiment optimization is that calculations are relatively fast and can be done through matrix manipulations of the experiment covariance matrices. We have shown in this work that we are able to make similar simplifications for the calculation of the KL divergence, which makes them also straightforward to calculate. This is a significant improvement over our earlier work [27], which relied on costly integrals using Monte Carlo methods.

Using our new method, we investigate a simple illustrative example of optimizing measurements of the SNe flux decrement and the radial and tangential BAO scale. These would be typical measurements for SNe and galaxy survey experiments. We demonstrate that the optimization of these experiments can depend on the choice of metric. In particular, the choice of metrics becomes important when comparing experiments with comparable information content. In 2006, the DEFT divided cosmology experiments into a number of stages. Stage II corresponded to on-going surveys at the time. Stage III were the next generation experiments (which are now being exited), and stage IV represented longer-term projects that are still in the planning and preparatory stages. We believe that in the design of stage III surveys, the choice of metric was not a critical step. This is because widely different designs were being considered with large ranges in information content. At this point, it is possible for all reasonable metrics to lead to the same optimization. For instance, fixing all other properties, such as depth, increasing the area of survey is always better regardless of the FoM. However, as we transition from stage III to stage IV, we are reaching fundamental limits, since, for instance, we begin to map out large fractions of the available cosmic volume. In this phase, the optimizations will become more subtle as the choice of optimization metric becomes increasingly important.

APPENDIX: DERIVATION OF $p(D_F|D_C)$

Let us consider the independent measurement D_C and D_F of a set of observables with a current and future data set,

respectively. Let Θ be a set of parameters of a model which makes predictions about the observables. The probability distribution of these variables is fully described by their joint probability distribution function $p(\Theta, D_C, D_F)$. Our aim is to derive Eq. (2), i.e. the conditional probability $p(D_F|D_C)$ in terms of $p(\Theta|D_C)$ and $p(D_F|\Theta)$ which are assumed to be given.

From the definition of conditional and joint probabilities, we get

$$p(D_F|D_C) = \frac{p(D_F, D_C)}{p(D_C)} = \int dy \frac{p(\Theta, D_C, D_F)}{p(D_C)} \quad (\text{A1})$$

and

$$p(\Theta, D_C, D_F) = p(D_C, D_F|\Theta)p(\Theta). \quad (\text{A2})$$

Using the latter in the former equation gives

$$p(D_F|D_C) = \int d\Theta \frac{p(D_C, D_F|\Theta)p(\Theta)}{p(D_C)}. \quad (\text{A3})$$

Since the current and future measurements are assumed to be independent, given a model Θ , $p(D_C, D_F|\Theta) = p(D_C|\Theta)p(D_F|\Theta)$. Thus,

$$p(D_F|D_C) = \int d\Theta \frac{p(D_C|\Theta)p(D_F|\Theta)p(\Theta)}{p(D_C)}. \quad (\text{A4})$$

Since $p(D_C|\Theta)p(\Theta) = p(\Theta|D_C)p(D_C)$, this becomes

$$p(D_F|D_C) = \int d\Theta p(\Theta|D_C)p(D_F|\Theta) \quad (\text{A5})$$

in accordance with Eq. (2). Following a similar argument, Eq. (1) can be derived by considering the true value, T , that would be measured as the errors tend to zero instead of model parameters.

-
- [1] G. Hinshaw *et al.*, *Astrophys. J. Suppl. Ser.* **208**, 19 (2013).
- [2] P. A. R. Ade *et al.* (Planck Collaboration), arXiv:1303.5076.
- [3] M. Bartelmann, *Rev. Mod. Phys.* **82**, 331 (2010).
- [4] J. L. Feng, *Annu. Rev. Astron. Astrophys.* **48**, 495 (2010).
- [5] J. A. Frieman, M. S. Turner, and D. Huterer, *Annu. Rev. Astron. Astrophys.* **46**, 385 (2008).
- [6] T. Padmanabhan, *Gen. Relativ. Gravit.* **40**, 529 (2008).
- [7] A. Albrecht *et al.*, arXiv:astro-ph/0609591.
- [8] M. Chevallier and D. Polarski, *Int. J. Mod. Phys. D* **10**, 213 (2001).
- [9] E. V. Linder, *Phys. Rev. Lett.* **90**, 091301 (2003).
- [10] R. A. Fisher, *Proc. Cambridge Philos. Soc.* **22**, 700 (1925).
- [11] A. Albrecht *et al.*, arXiv:0901.0721.
- [12] T. D. Kitching and A. Amara, *Mon. Not. R. Astron. Soc.* **398**, 2134 (2009).
- [13] J. A. Peacock *et al.* (unpublished).
- [14] Y.-S. Song, L. Hollenstein, G. Caldera-Cabral, and K. Koyama, *J. Cosmol. Astropart. Phys.* **04** (2010) 018.
- [15] A. Albrecht and G. Bernstein, *Phys. Rev. D* **75**, 103003 (2007).
- [16] R. G. Crittenden, L. Pogosian, and G.-B. Zhao, *J. Cosmol. Astropart. Phys.* **12** (2009) 025.
- [17] M. J. Mortonson, D. Huterer, and W. Hu, *Phys. Rev. D* **82**, 063004 (2010).
- [18] A. Hojjati, G.-B. Zhao, L. Pogosian, A. Silvestri, R. Crittenden, and K. Koyama, *Phys. Rev. D* **85**, 043508 (2012).
- [19] B. A. Bassett, *Phys. Rev. D* **71**, 083517 (2005).
- [20] B. A. Bassett, D. Parkinson, and R. C. Nichol, *The Astrophysical Journal* **626**, L1 (2005).
- [21] D. Parkinson, C. Blake, M. Kunz, B. A. Bassett, R. C. Nichol, and K. Glazebrook, *Mon. Not. R. Astron. Soc.* **377**, 185 (2007).
- [22] R. Trotta, *Mon. Not. R. Astron. Soc.* **378**, 819 (2007).
- [23] P. Mukherjee, D. Parkinson, P. S. Corasaniti, A. R. Liddle, and M. Kunz, *Mon. Not. R. Astron. Soc.* **369**, 1725 (2006).
- [24] C. Watkinson, A. R. Liddle, P. Mukherjee, and D. Parkinson, *Mon. Not. R. Astron. Soc.* **424**, 313 (2012).
- [25] R. Trotta, M. Kunz, and A. R. Liddle, *Mon. Not. R. Astron. Soc.* **414**, 2337 (2011).
- [26] P. Paykari and A. H. Jaffe, *Mon. Not. R. Astron. Soc.* **433**, 3523 (2013).
- [27] A. Amara and T. D. Kitching, *Mon. Not. R. Astron. Soc.* **413**, 1505 (2011).
- [28] S. Kullback and R. A. Leibler, *Ann. Math. Stat.* **22**, 79 (1951).
- [29] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*, Adaptive Computation and Machine Learning Series (MIT Press, Cambridge Massachusetts, 2006).