Cosmological discordances: A new measure, marginalization effects, and application to geometry versus growth current data sets

Weikang Lin and Mustapha Ishak

Cosmological discordances: a new measure, marginalization effects, and application to geometry vs growth current data sets

Weikang Lin[*] and Mustapha Ishak[†]

Department of Physics, The University of Texas at Dallas, Richardson, Texas 75080, USA

The continuous progress toward more precise cosmological surveys and experiments has galvanized recent interest into consistency tests on cosmological parameters and models. At the heart of this effort is quantifying the degree of inconsistency between two or more cosmological data sets. We introduce an intuitive moment-based measure we call the *Index of Inconsistency* (IOI) and show that it is sensitive to the separation of the means, the size of the constraint ellipsoids, and their orientations in the parameter space. We find that it tracks accurately the inconsistencies when present. Next, we show that parameter marginalization can cause a loss of information on the inconsistency between two experiments and we quantify such a loss using the drop in IOI. In order to zoom on a given parameter, we define the relative residual IOI and the relative drop in IOI. While these two quantities can provide insights on the parameters that are most responsible for inconsistencies, we find that the full IOI applied to the whole parameter spaces is what must be used to correctly reflects the degree of inconsistency between two experiments. We discuss various properties of IOI, provide its eigen-mode decomposition, and compare it to other measures of discordance. Finally, we apply IOI to current geometry data sets (i.e. an improved Supernovae Type Ia compilation, Baryon Acoustic Oscillations from 6dF, SDSS MGS and Lyman-α forest, and high-ℓ CMB temperature data from Planck-2015) versus growth data sets (i.e. Redshift Space Distortions from WiggleZ and SDSS, Weak Lensing from CFHTLenS, CMB Lensing, Sunyav-Zeldovich effect, and low-ℓ CMB temperature and polarization data from Planck-2015). We find that a persistent inconsistency is present between the two data sets. This could reflect the presence of systematics in the data or inconsistencies in the underlying model.

PACS numbers: 98.80.Es,95.36.+x,98.80.-k

I. INTRODUCTION

The ΛCDM standard model of cosmology enjoys a number of observational confirmations and successes. However, it does come with two intriguing conundrums. The first one is that it requires a dark matter component counting for about 26% of matter-energy content in the Universe. The second one is that the expansion of the Universe is accelerating and we don't know what is driving this acceleration. Associated with this cosmic acceleration is a dark energy component that could count for about 69% of the energy budget in the Universe. These and other questions have motivated studies to consider if there are any problems with the underlying model, theory or assumptions. See for example, the following articles [1–9] and references therein.

One route to test this that is attracting more attention lately is to compare results and parameters from different experiments and then to look for any inconsistencies. Over a decade ago, reference [10] used simulated data sets to demonstrate how inconsistencies between the dark energy parameter spaces as constrained by the expansion versus the growth can signal a failure of the underlying gravity theory. A related method called "parameter splitting" was applied to real data in Refs. [11–13] and looked for inconsistencies between dark energy parame-

ters as constrained by expansion versus the growth. References. [14, 15] introduced a technique called the crossing function, which enables us to see if the best fit from one experiment is consistent with the constraints from another experiment.

Most recently, a number of papers have focused on possible discrepancies between cosmological parameters as constrained by different experiments. While some have faded away, some others seem to persist; see for example Refs. [16–20]. We take the point of view that whether due to systematic effects in the data or due to underlying physics, these discrepancies need a careful study. Otherwise, a joint analysis will be of questionable outcome. In the next section, we review briefly some of the inconsistencies or tensions that have appeared in the literature.

An important question is to quantify the degree of inconsistency or tension. Very often, the inconsistency between two experiments is qualitatively shown in marginalized one-dimensional (1-D) or two-dimensional (2-D) likelihood contour plots. However, as we demonstrate in this paper, this method can not accurately represent the inconsistency. First of all, it is not a quantitative method, rendering it difficult to be interpreted in an accurate way. Secondly, if the underlying model has three or more parameters this method fails to account for the full inconsistency. Moreover, as shown later in this work, listing all the marginalized 1-D or 2-D plots does not properly represent the full inconsistency neither. Therefore some measures have been introduced in the literature in order to quantify inconsistencies between

two experiments and we review most of them in Sec. X. An important criterion is that a measure must properly describe experimental discordances.

In this paper, we define and apply a measure of inconsistency that is found to track accurately inconsistencies. It is a moment-based quantity that we call the *index of inconsistency* (IOI). This can be applied to two or more experiments or data sets. We find that in the gaussian and weak prior limit, most other measures in the literature reduce to or contain a term of IOI. This measure properly describes the factors causing the inconsistency: i.e mean difference, constraint volumes and their orientations.

Importantly, we also show here that parameter marginalization can hide inconsistency. We argue that a good measure of inconsistency should reflect this. When zooming on specific parameters, we define the relative drop in IOI and the relative residual IOI. This allows one to track down the inconsistencies and find parameters that are most subject to these inconsistencies.

Finally, we apply the new measures and study the inconsistency between the geometry data sets and the growth data sets. For the geometry, we use the type Ia supernovae compilation in Ref. [21], the Baryon Acoustic Oscillation (BAO) from the Six Degree Field Glactic Survey (6dF) [22], the main galaxy sample from the Sloan Digital Sky Survey (SDSS-DR7) and the SDSS quasar-Lyman-α forest [18], and the high-ℓ CMB temperature data from Planck 2015 [23]. For the growth, we use the low-ℓ CMB temperature and polarization data form Planck 2015 [24], CMB lensing [24], thermal Sunyaev–Zel'dovich effect [25], cosmic shear from the Canada France Hawaii Lensing Survey (CFHTlens) [19], and the redshift space distortion (RSD) from the WiggleZ Dark Energy Survey (WiggleZ MPK) [26, 27] and the SDSS DR12 CMASS and LOWZ catalogs [28]. This application is motivated by the fact that modified gravity theory can lead to inconsistency on the dark energy properties when fitting ΛCDM or wCDM model separately into the geometry and the growth sets of experiments, and also in order to quantify the degree of inconsistency between these two data sets as reported in some previous works.

We organize our paper as follows: In Sec. II we review some inconsistencies among different cosmological experiments. In Sec. III we define IOI, demonstrate that it can properly describe the inconsistency between two experiments, discuss about its properties, and show that it correctly reflects the effects of marginalization. In Sec. X we view other measures of experimental inconsistency in the literature, and pointing out to some limitations. Then in Sec. XI we apply IOI along with some other measures in Gaussian and weak prior limit to the geometry and the growth experiments based on the ΛCDM and the wCDM models. Finally we summarize our work in Sec. XII. In this work, 'inconsistency', 'tension' and 'discordance' share the same meaning, but the quantities *index of inconsistency* (IOI), the *tension* ($\mathcal{T}$) and the *discordance* are different measures of inconsistency.

II. BRIEF SURVEY OF REPORTED TENSIONS OR INCONSISTENCIES

In this sections we list some inconsistencies or tensions between cosmological experiments as reported in the literature. Again, regardless if these are caused by different systematic effects in various data sets or caused by shortcomings in the cosmological model, they require careful examination.

Hubble constant tension: A precise locally measured Hubble constant by ladder distance observation is given as $H_0 = 73.24 \pm 1.74$ km s^{-1}/Mpc^{-1} by Refs. [29, 30], which is higher than the one derived from the Planck 2015 data $H_0 = 66.93 \pm 0.62$ km s^{-1}/Mpc^{-1} assuming ΛCDM model [31]. The tension is reported as at a 3.4-σ confidence level in Ref. [30]. Reference [16] used their measure called the *tension* and reported a high inconsistency between the local measurement of H_0 and the Planck result. They found that the odds for the two experiments to be consistent one with another are only $1 : 116$.

In order to locally measure the Hubble constant, one need to hierarchically calibrate the distances to different celestial objects, so papers have pointed out to the systematic effects involved. For example, Ref. [32] adopted a different outlier rejection criteria in the Cepheid samples and find that the high value of locally measured Hubble constant in Ref. [30] could be due to systematic error in the distance calibration. But recently authors in Ref. [29], using more Cepheid variables calibrated type Ia supernova, confirmed their earlier high value found in Ref. [30]. Also Ref. [33] using Gaia Data Release 1 reported a similarly high value of Hubble constant ($H_0 = 73.0$ km s^{-1}/Mpc^{-1}), different from the Planck measurement at $2.5 - 3.5$ σ level. So the tension of H_0 between local measurements and the Planck 2015 result still remains. Ref [34] independently measured the Hubble constant based on time delay strong lensing, and reported $H_0 = 71.9^{+2.4}_{-3.0}$ km s^{-1}/Mpc^{-1} in ΛCDM model consistent with the local measurement.

On the other hand, this tension might indicate a possible problem with the underlying model. For example, if ΛCDM model is extended to the wCDM model, it was claimed in Ref. [35] that a piecewise function of w can reconcile the derived Hubble constant from the Planck data with the one locally measured. Ref. [36] showed that varying 12 cosmological parameters can solve the current tension on H_0 between Planck and local measurement. But they also pointed out that tension remains if BAO and distances to supernova are included in the joint analysis.

Gravitational Lensing and Planck: Another persisting tension is between cosmic shear experiments and Planck for the determination of the amplitude of matter density fluctuations as for example parameterized by σ_8. CFHTlenS [19] found a lower amplitude than Planck 2015 [23]. In the marginalized σ_8 vs Ω_m plane, the CFHTlenS confidence contours are shifted to the upper-

left compared to the Planck 2015 results. Most recently, the KIDS-450 survey find a similar discrepancy at a 2.5-σ level for $S_8 = \sigma_8\sqrt{\Omega_m/0.3}$ [37] and claim a substantial discordance in the full parameter space compared to Plancl 2015. Finally, Ref. [38] used galaxy-galaxy lensing measurements of the BOSS CMASS sample using 250 square degrees of weak lensing from CFHTLenS and CS82 and found also a lower value of S_8 than that of Planck 2015. They pointed out that this can be caused by systematics or new physics. Indeed, a number of systematic effects for lensing require more work such as for example intrinsic alignment of galaxies, baryonic effects, and photometric redshifts [39–44]. On the other hand, it was pointed out in Refs. [17, 36] that the tension between CFHTlenS and Planck can be solved by varying A_{lens} (lensing anomaly parameter).

BAO at z=2.34: The baryon acoustic oscillation in the Lyman-α forest measurement gives the ratio of hubble radius to the drag epoch sound horizon $D_H/r_d = 9.145 \pm 0.204$ at $z_{eff} = 2.34$ [18], which combines Lyα forest auto-correlation [45] and the quasar-Lyα cross-correlation [46] methods. If we take $r_d = 147.50 Mpc^{-1}$ form Planck 2015 [23], the inferred hubble parameter at $z = 2.34$ is $H(z = 2.34) \approx 222 \pm 5$ km s^{-1}/Mpc[1]. On the other hand, the Planck 2015 best fit gives $H_{Pl}(z = 2.34) = 236.6$ km s^{-1}/Mpc, which is higher than the one derived from BAO Lyman-α forest measurement. It is suggested that this discrepancy could be evidence for the interacting dark energy and dark matter model, see for example [47].

CMB Lensing anomaly: The ΛCDM model predicts a certain strength of the CMB lensing potential. In Ref. [48] introduced a lensing anomaly parameter defined as $A_{lens} \equiv C_\ell^{\psi,obs}/C_\ell^{\psi,\Lambda CDM}$, (where $C_\ell^{\psi,obs}$ and $C_\ell^{\psi,\Lambda CDM}$ are the observed and predicted lensing-potential power spectra) to test such an amplitude. If ΛCDM is a consistent model on the CMB temperature, polarization and lensing observation, this lensing anomaly parameter must be unity, $A_{lens} = 1$. Early the WMAP (along with ACBAR) data give $A_{lens} = 3.1^{+1.8}_{-1.5}$ at 2σ confidence level [48], meaning the observed lensing potential is significantly higher than expected. However this high value of A_{lens} did not persist in the Planck data, and dropped to 1.22 ± 0.1 at 1σ confidence level [23]. So although there is still a small gap between A_{lens} and unity, it became more consistent with the prediction of the ΛCDM model. Nevertheless, it was shown that modified-gravity equations for the two scalar potentials (in the Newtonian Gauge) can solve this small lensing anomaly [49]. They claimed that this lensing anomaly gives a preference for modified gravity at a 95% confidence level. Reference [50] found that adding compensated isocurvature perturbations to the ΛCDM model can solve such a CMB lensing anomaly.

---

[1] Since $D_H(z) = 1/H(z)$.

Redshift Space Distortion (RSD) and Cluster Abundance vs Planck: Compared to what is inferred from Planck 2015 best fit, RSD measurements generally give smaller growth of large structure parameterized as $f\sigma_8$, where $f = d\ln\delta/d\ln a$ is the growth rate; see for examples Refs. [13, 51, 52]. The two most precise galaxy cluster abundance measurements also seem to have significant tensions with the Planck 2015 results in the marginalized σ_8-Ω_m plane. One is the X-ray as mass tracer measurement from the Chandra cluster cosmology project [53], and the other is the thermal Sunyaev-Zel'dovich effect from Planck itself [25]; also see Ref. [13].

Geometry versus Growth: When combined together, geometry probes and growth probes provide powerful combinations to constrain cosmology. Moreover, when one is contrasted with another they constitute a useful consistency test of the underlying theory [10, 54, 55] or a mean to detect different systematics in the data sets. References [11–13] used a technique called parameter splitting, in which they separate parameters constraining the dark energy properties (e.g., Ω_Λ and w) into a geometry set (e.g., Ω_Λ^{geom} and w^{geom}) and a growth set (e.g., Ω_Λ^{grow} and w^{grow}). Reference [12] found that RSD data generally favor a higher w^{grow}, and Ref. [13] found that a subset of cluster abundance data (from Refs. [25, 53]) mostly cause the deviations of dark-energy meta-parameters.

III. INDEX OF INCONSISTENCY (IOI): MOTIVATION AND DEFINITION

Different works on measures of inconsistency in the literature adapt different notations. For a consistent discussion throughout the paper, we use the following notation so Bayes' theorem reads

$$\mathscr{P}(\boldsymbol{\lambda}; Q) = \frac{\mathcal{L}(Q; \boldsymbol{\lambda})\mathcal{P}(\boldsymbol{\lambda})}{E(Q)}, \qquad (1)$$

where $\mathscr{P}$ is the posterior probability distribution, $\mathcal{L}$ the likelihood, $\mathcal{P}$ the prior and E the evidence. Most of the experiments discussed here are assumed to give Gaussian likelihoods $\mathcal{L}^{(i)}(\boldsymbol{Q}; \boldsymbol{\lambda})$ on the parameters. For Gaussian likelihoods,

$$\mathcal{L}^{(i)} = \mathcal{L}_{max}^{(i)} \exp\left(-\tfrac{1}{2}(\boldsymbol{\lambda} - \boldsymbol{\mu^{(i)}})^T \boldsymbol{L^{(i)}}(\boldsymbol{\lambda} - \boldsymbol{\mu^{(i)}})\right), \quad (2)$$

where $\boldsymbol{\lambda}$ is the parameter vector, $\boldsymbol{\mu^{(i)}}$ the mean ($i = 1, 2$ for two experiments), and $\boldsymbol{L^{(i)}}$ the fisher matrix. For mildly non-Gaussian distributions, the above equation is treated as an approximation to the real distribution, with $\boldsymbol{\mu^{(i)}}$ and $\boldsymbol{L^{(i)}}$ given by the mean and inverse of covariance matrix of the real distribution. The mean $\boldsymbol{\mu^{(i)}}$ is a function of the data vector $\boldsymbol{Q}$. Experiments are assumed to have the same Gaussian prior distribution

$$\mathcal{P} = \frac{\sqrt{|\boldsymbol{P}|}}{(2\pi)^{n/2}} \exp\left(-\tfrac{1}{2}(\boldsymbol{\lambda} - \boldsymbol{\mu^{(p)}})^T \boldsymbol{P}(\boldsymbol{\lambda} - \boldsymbol{\mu^{(p)}})\right), \quad (3)$$

where $|\boldsymbol{P}| \equiv \det(\boldsymbol{P})$ and $\boldsymbol{P}$ is the inverse of covariance matrix of the prior. A weak prior limit is taken as $\boldsymbol{P} \to \boldsymbol{0}$. The Gaussian and weak prior limit refers to a situation where likelihoods are Gaussian on the parameters and the prior is weak. Posteriors are normalized, and are denoted as $\mathscr{P}^{(i)}$. The inverse of the covariance matrix of a posterior is denoted as $\boldsymbol{F}^{(i)}$, and the mean as $\bar{\boldsymbol{\mu}}^{(i)}$. Some common notations in this work are summarized in Tables I and II.

A. Motivation for IOI

In Bayesian parameter estimation, experiments give probability density distributions of the parameters in a given model. Different experiments usually give different distributions. The joint distribution is the one obtained by simultaneously analysing the data from two or more experiments. Figure 1 shows two toy 1-D Gaussian distributions (red and blue) given by two different experiments, and their joint distribution (black dotted) on parameter x.

We can see in Fig. 1 that the red distribution favors the value of x at 4, while the blue one favors at 10, so there is an inconsistency between the two experiments. One might want to use the difference of the two means to quantify the inconsistency, i.e., $\delta = x_{blue} - x_{red} = 10 - 4 = 6$. But this method cannot be right, because the mean difference is not invariant under a parameter scaling. If we let $x \to x' = 2x$, the new means become 20 and 8, and the difference becomes 12. The inconsistency between two experiments for a model should not depend on the scaling.

One then realizes that the mean difference need to be normalized by the uncertainties of the distributions, since the uncertainties will be scaled inversely as the mean difference. And this normalization makes perfect sense: if the uncertainties are very large compared to the mean difference, the inconsistency will actually be small. But which distribution's uncertainty should be used to normalize the mean difference? Intuition tells us that we should not use just one but somehow both of them. One might want to use the uncertainty of the joint distribution. For example the figure of bias (Fob) defined in Ref. [56] reduces to this type of form in one dimension. But this choice has problems for the reasons that follow. Suppose the uncertainties of the red and the blue distributions are $\sigma_{(1)}$ and $\sigma_{(2)}$, the uncertainty of the joint distribution is given by $\sigma^{-2} = \sigma_{(1)}^{-2} + \sigma_{(2)}^{-2}$. Normalizing δ by σ will give a quantity like $\frac{\delta^2}{\sigma^2} = \delta^2(\frac{1}{\sigma_{(1)}^2} + \frac{1}{\sigma_{(2)}^2})$. If one distribution has a much smaller uncertainty than the other, say $\sigma_{(2)} \ll \sigma_{(1)}$, that quantity can be approximated as $\frac{\delta^2}{\sigma^2} \simeq \frac{\delta^2}{\sigma_{(2)}^2}$. So in this situation the mean difference is normalized only by σ_2, and it diverges as $\sigma_{(2)} \to 0$. This is very counter-intuitive. If one uncertainty is very small but the other is very large, the inconsistency should be
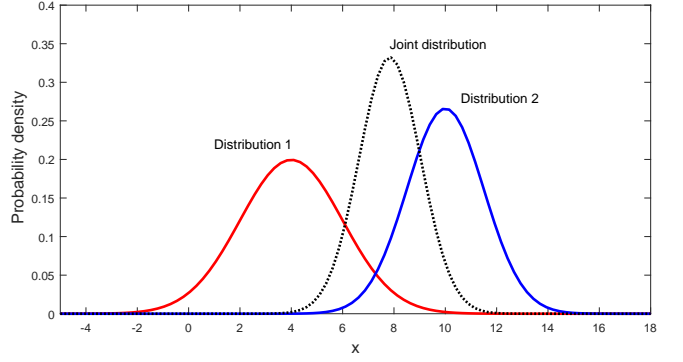

FIG. 1. Two toy one-dimensional Gaussian probability density distributions (red and blue) and their joint distribution (black dotted).

small. That is because the distribution with a large uncertainty can extend to the mean of the other distribution that has a small uncertainty. So if $\sigma_{(2)} \ll \sigma_{(1)}$, the mean difference should be normalized by $\sigma_{(1)}$ instead of $\sigma_{(2)}$. Therefore, instead of $\delta^2(\frac{1}{\sigma_{(1)}^2} + \frac{1}{\sigma_{(2)}^2})$, it is a quantity like $\frac{\delta^2}{\sigma_{(1)}^2 + \sigma_{(2)}^2}$ that should be used.

Here, we approach the inconsistency in a slightly different way from above. Indeed, the above discussion did not consider the mean of the joint distribution. Instead of using the difference between the means of the two distributions, we use two differences. One difference $\delta^{(1)}$ is between the first distribution mean and the joint distribution mean, the other difference $\delta^{(2)}$ is between the second distribution mean and the joint distribution mean. Then we normalize the first difference by the first uncertainty, and the second difference by the second uncertainty. Finally we take the average of the normalized mean differences (squared) as the measure of inconsistency. More precisely, we are use the following quantity to measure the inconsistency,

$$\text{Inconsistency} = \frac{1}{2}\left(\frac{(\delta^{(1)})^2}{\sigma_{(1)}^2} + \frac{(\delta^{(2)})^2}{\sigma_{(2)}^2}\right). \qquad (4)$$

The meaning of Eq. (4) is as follows: it is the average of two terms, each of which measures the 'difficulty' for the corresponding distribution to 'support' or 'favor' the mean of the joint distribution. The mean of the joint distribution will be closer to the mean of the distribution with a smaller uncertainty. For a metaphor, if two people are separated, one is at location A and the other is at location B. We can use the minimum time taken by them to meet to quantify the degree of how they are separated. Surely both of them need to move, and the one who travels faster or moves easier needs to make a longer distance.

We will show in Sec. III B that the above quantity ac-

TABLE I. Table of notations: Probability distributions, their means and elements of means, Fisher matrices and elements of the Fisher matrices for the likelihood of the ith experiment, prior, and the posterior of the ith experiment. Likelihoods are *not* normalized in the parameter space, while the Prior and Posteriors are.

Distributions	Notations	Inverse of covariance matrix	Elements of Fisher matrices	Means	Elements of means
ith Likelihood	$\mathcal{L}^{(i)}$	$\boldsymbol{L^{(i)}}$	$L^{(i)}_{jk}$	$\boldsymbol{\mu^{(i)}}$	$\mu^{(i)}_j$
Prior	$\mathcal{P}$	$\boldsymbol{P}$	P_{jk}	$\boldsymbol{\mu^{(p)}}$	$\mu^{(p)}_j$
ith Posterior	$\mathscr{P}^{(i)}$	$\boldsymbol{F^{(i)}}$	$F^{(i)}_{jk}$	$\boldsymbol{\bar{\mu}^{(i)}}$	$\bar{\mu}^{(i)}_j$

TABLE II. Other frequently used notations in this work.

Parameter vector	Observable vector	Mean difference	Covariance matrix	G matrix
$\boldsymbol{\lambda}$	$\boldsymbol{Q}$	$\boldsymbol{\delta}$	$\boldsymbol{C}$	$\boldsymbol{G} = (\boldsymbol{C^{(1)}} + \boldsymbol{C^{(2)}})^{-1}$

tually turns out to be

$$\text{Inconsistency} = \frac{1}{2}\frac{\delta^2}{\sigma^2_{(1)} + \sigma^2_{(2)}}, \tag{5}$$

which has the same form as the one we logically obtained in the previous paragraph.

So far we have only considered 1-D distributions, but let's "extrapolate" the result to multi-dimensional distributions. For multi-dimensional Gaussian distributions, the uncertainty (squared) is specified by the covariance matrix $\boldsymbol{C}$. So, it is reasonable to guess the general form of Eq. 5 to be $\frac{1}{2}\boldsymbol{\delta}^T(\boldsymbol{C^{(1)}} + \boldsymbol{C^{(2)}})^{-1}\boldsymbol{\delta}$, and it turns out to be an proper guess as we demonstrate in Sec. III B.

B. Definition of IOI

We first define the term $\frac{1}{2}\Delta\chi^2(\boldsymbol{\mu})$ as,

$$\tfrac{1}{2}\Delta\chi^2(\boldsymbol{\mu}) \equiv \tfrac{1}{2}\left(\Delta\chi^2_{(1)}(\boldsymbol{\mu}) + \Delta\chi^2_{(2)}(\boldsymbol{\mu})\right), \tag{6}$$

with $\Delta\chi^2_{(i)}(\boldsymbol{\mu}) = \chi^2_{(i)}(\boldsymbol{\mu}) - \chi^2_{(i)}(\boldsymbol{\mu^{(i)}})$. We can see that $\frac{1}{2}\Delta\chi^2(\boldsymbol{\mu})$ between two experiments is defined as an average of two terms, namely,

$\Delta\chi^2_{(1)}(\boldsymbol{\mu})$: The 'difficulty' for the 1st experiment to support the mean of the joint analysis.

$\Delta\chi^2_{(2)}(\boldsymbol{\mu})$: The 'difficulty' for the 2nd experiment to support the mean of the joint analysis.

So $\frac{1}{2}\Delta\chi^2(\boldsymbol{\mu})$ is the averaged 'difficulty' for the two experiments to support the joint mean. This definition of $\frac{1}{2}\Delta\chi^2$ is equivalent to $\frac{1}{2}\Delta\chi^2 \equiv \frac{1}{2}\chi^2_{min} - (\frac{1}{2}\chi^2_{(1),min} + \frac{1}{2}\chi^2_{(2),min})$, where $\chi^2 = \chi^2_{(1)} + \chi^2_{(2)}$ and 'min' means minimum of χ^2 (and maximum of likelihood).

As shown in appendices A 1 and A 2, in the Gaussian limit where $\Delta\chi^2_{(i)} = (\boldsymbol{\lambda} - \boldsymbol{\mu^{(i)}})^T\boldsymbol{L^{(i)}}(\boldsymbol{\lambda} - \boldsymbol{\mu^{(i)}})$, the mean of the joint likelihood is given by a Fisher-matrix-weighted average,

$$\boldsymbol{\mu} = \boldsymbol{L}^{-1}(\boldsymbol{L^{(1)}}\boldsymbol{\mu^{(1)}} + \boldsymbol{L^{(2)}}\boldsymbol{\mu^{(2)}}), \tag{7}$$

where $\boldsymbol{L} = \boldsymbol{L^{(1)}} + \boldsymbol{L^{(2)}}$. And then $\frac{1}{2}\Delta\chi^2$, in this limit, can be computed explicitly as

$$\begin{aligned} \tfrac{1}{2}\Delta\chi^2 &\xrightarrow{\text{Gaussian}} \\ &\tfrac{1}{2}(\boldsymbol{\mu^{(2)}} - \boldsymbol{\mu^{(1)}})^T(\boldsymbol{C^{(1)}} + \boldsymbol{C^{(2)}})^{-1}(\boldsymbol{\mu^{(2)}} - \boldsymbol{\mu^{(1)}}). \end{aligned} \tag{8}$$

Thus, we define IOI for two experiments to be the Gaussian limit of $\frac{1}{2}\Delta\chi^2$ [Eq. (6)] and given by

$$\text{IOI} \equiv \tfrac{1}{2}\boldsymbol{\delta}^T\boldsymbol{G}\boldsymbol{\delta}, \tag{9}$$

where $\boldsymbol{\delta} \equiv \boldsymbol{\mu^{(2)}} - \boldsymbol{\mu^{(1)}}$, and $\boldsymbol{G} \equiv [(\boldsymbol{L^{(1)}})^{-1} + (\boldsymbol{L^{(2)}})^{-1}]^{-1} = (\boldsymbol{C^{(1)}} + \boldsymbol{C^{(2)}})^{-1}$. This result is the same as our guided guess in Sec. III A. In the cases of Gaussian likelihoods, IOI is the same as $\frac{1}{2}\Delta\chi^2$. But in general cases, we define IOI as the moment-based quadratic quantity $\frac{1}{2}\boldsymbol{\delta}^T\boldsymbol{G}\boldsymbol{\delta}$, with $\boldsymbol{L^{(i)}}$ obtained by the inverse of the covariance matrices $(\boldsymbol{C^{(i)}})^{-1}$ and $\boldsymbol{\mu^{(i)}}$ given by the means. *So the definition of IOI is only motivated by $\frac{1}{2}\Delta\chi^2$, but not $\frac{1}{2}\Delta\chi^2$ itself.* IOI is a moment-based quadratic quantity.

If we have priors assigned to the two experiments, we can extend the definition of IOI to include priors in the means and covariance matrices as follows:

$$\begin{aligned} \boldsymbol{\mu^{(i)}} &\xrightarrow{\text{with prior}} \boldsymbol{\bar{\mu}^{(i)}} = \frac{1}{\boldsymbol{F^{(i)}}}(\boldsymbol{L^{(i)}}\boldsymbol{\mu^{(i)}} + \boldsymbol{P}\boldsymbol{\mu^{(p)}}), \\ \boldsymbol{\delta} &\xrightarrow{\text{with prior}} \boldsymbol{\bar{\delta}} = \boldsymbol{\bar{\mu}^{(2)}} - \boldsymbol{\bar{\mu}^{(1)}}, \\ \boldsymbol{G} &\xrightarrow{\text{with prior}} \boldsymbol{\bar{G}} = [(\boldsymbol{F^{(1)}})^{-1} + (\boldsymbol{F^{(2)}})^{-1}]^{-1}, \\ \text{IOI} &\xrightarrow{\text{with prior}} \tfrac{1}{2}\boldsymbol{\bar{\delta}}^T\boldsymbol{\bar{G}}\boldsymbol{\bar{\delta}}, \end{aligned} \tag{10}$$

where $\boldsymbol{F^{(i)}} = \boldsymbol{L^{(i)}} + \boldsymbol{P}$ for Gaussian distributions, and is an approximation in mildly non-Gaussian cases. Thus Eq. (9) is the weak prior limit ($\boldsymbol{P} \to 0$) of Eq. (10). However, we suggest to use Eq. (9) to compare two experiments for the following reasons. First, we usually have only weak priors on parameters. Second, the priors might be already biased or in tension with the experiments considered. Two experiments should be compared with very
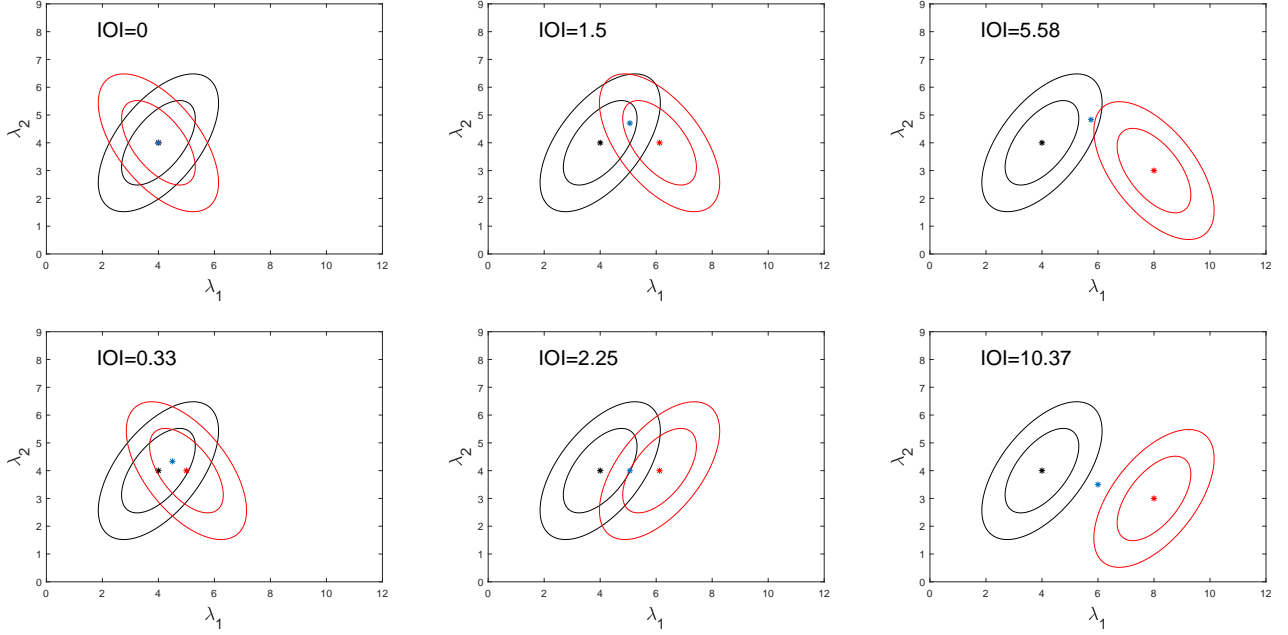
FIG. 2. Examples of IOIs. The centers of the ellipses are the likelihood maxima. The extra (blue) point is the likelihood maximum of the joint likelihood. The two in the middle have the same deviation of likelihood maxima, but they do not have the same IOI because of the different degenerate directions. Same is also true for the two on the right.

TABLE III. Jeffreys' scales as interpretation of the values of IOI. Jeffreys' scales are an empirical scales that originally classify the ranges of the Bayesian evidence ratio. But comparing the numerical values of IOI and the visual separations of likelihood contours in Fig. 2, we find that Jeffreys' scales are appropriate for the classification and interpretation of IOI. Higher IOI represents higher inconsistency. Since IOI is positive definite, the original interpretation of the negative values of Bayesian evidence ratio does not apply to IOI (see Table V for such a difference).

Ranges	IOI< 1	1 <IOI< 2.5	2.5 <IOI< 5	IOI> 5
Interpretation	no significant inconsistency	weak inconsistency	moderate inconsistency	strong inconsistency

weak priors. Including strong priors may affect the comparison of two experiments. But of course, we can use Eq. (9) to compare and see if the prior is consistent with the experiments to be compared, by simply replacing $L^{(2)}$ and $\mu^{(2)}$ with P and $\mu^{(p)}$ in Eq. (9).

For illustration, we show in Fig. 2 some examples of toy likelihood contours with their corresponding values of IOI. Those examples show that IOI can properly represent inconsistency between two experiments, i.e., it increases whenever the experimental inconsistency increases graphically. We use Jeffreys' scale (shown in Table III; see Ref. [57]) as an interpretation of the values of IOI. Note that Jeffreys' scales were originally empirical scales for Bayesian evidence ratio. Here IOI is not an evidence ratio but we use this empirical scale as it seems to give sensible meanings for the inconsistencies compared to what is shown in Fig. 2. Moreover, reference [58] proposed a quantity called *tension* to measure experimental inconsistency. They used Jeffreys' scale as the interpretation of the *tension*. We will show further in the paper

that the *tension* reduces to IOI in the Gaussian and weak prior limit. So we expect that IOI and the *tension* are similar quantities for nearly Gaussian distributions, and using the same scale is good for comparison. However, we note again that one should not translate IOI into a probability ratio, although we propose Jeffreys' scales as the interpretation of IOI.

The next logical step is to generalize Eq. (9) to define IOI for comparison of N experiments as

$$\frac{1}{N}\sum_i \Delta\chi^2_{(i)}(\mu) \xrightarrow{\text{Gaussian}} \frac{1}{N}\Big(\sum_i \mu^{(i)\,T}L^{(i)}\mu^{(i)} - \mu^T L\mu\Big),$$
(11)

$$\text{IOI} \equiv \frac{1}{N}\Big(\sum_{i=1}^{N} \mu^{(i)\,T}L^{(i)}\mu^{(i)} - \mu^T L\mu\Big),$$
(12)

where μ is the mean of the joint likelihood combining all experiments and it is also given by the Fisher-matrix-

weighted average,

$$\boldsymbol{\mu} = \boldsymbol{L}^{-1}\Big(\sum_i \boldsymbol{L}^{(i)}\boldsymbol{\mu}^{(i)} \Big), \qquad (13)$$

where $\boldsymbol{L} = \sum \boldsymbol{L}^{(i)}$. Each term in the sum in the left-hand side of Eq. (11) has the same interpretation as that in Eq. (6), which is measuring the 'difficulty' for the ith experiment to support the mean of the joint analysis. So $\frac{1}{N}\sum \Delta\chi^2_{(i)}(\mu)$ represents the averaged 'difficulty' for the N experiments to support their joint mean. We have again taken the Gaussian limit of Eq. (11) and defined IOI for N experimental as the moment-based quantity Eq. (12). But since Eq. (12) contains more than two means, it cannot be reduced into a quadratic form as in the case of two experiments. For the case of two experiments, the two different forms Eq. (9) and Eq. (12) are equivalent; see appendix A 2.

Before closing this subsection, we give some remarks. Our IOI and other measures of inconsistency in the literature are a model-dependent description of inconsistency between experiments. For example, in Sec. XI we will see that IOI between the geometry experiments growth experiments for the ΛCDM model is different from that for the wCDM model. Reference [58] suggested that a better model should provide more consistent explanations on different experiments. With IOI, it means a better model should give a smaller IOI on two or more experiments.

C. Inconsistency measures and Gaussianity

Current cosmological experiments usually provide approximately Gaussian probability distributions on parameters. In those cases, it is usual to approximate the actual probability distribution by a Gaussian one based on its moments. The mean and covariance matrix of the approximate Gaussian distribution are taken to be those of the actual distribution. If a measure does not correctly describe inconsistency in Gaussian cases, one would not expect it to correctly describe inconsistency in general. Therefore constructing a correct measure of inconsistency for Gaussian and nearly Gaussian cases is an important step to the measure for general cases, and this is what we aim to accomplish in this work. We argue that IOI works correctly in Gaussian cases, and we expect it to give meaningful information in mildly non-Gaussian cases. So IOI has an important role: one can compare a measure with IOI to see whether that measure properly describes experimental inconsistency in Gaussian cases.

However, we do not expect IOI to describe inconsistency properly in highly non-Gaussian cases. In fact, it is even difficult to define inconsistency when distributions are non-Gaussian. There might be other criteria to consider for general cases, which need to be further explored in future works. We note that even if a measure reduces to IOI in Gaussian limit, it does not necessarily mean such a measure can describe inconsistency properly. For

example, in Gaussian limit $\Delta\chi^2$ is the same as IOI, but it can give us misleading results in some non-Gaussian cases as we will explain. $\Delta\chi^2$ focuses on its minimum value but ignores the overall distribution. For example, in Fig. 3 the 1-D probability distribution $P_{(2)}(\lambda)$ given by experiment 2 is Gaussian, while $P_{(1)}(\lambda)$ given by experiment 1 is not. In fact, $P_{(1)}(\lambda)$ is only slightly different from a Gaussian distribution: there is a narrow local peak at $\lambda = -5$ which is also the global maximum of $P_{(1)}(\lambda)$; the rest of $P_{(1)}(\lambda)$ is overall Gaussian. When we compare the two experiments, intuition tells us we can ignore that narrow peak. That is because the integrated probability around that peak is small compared to the rest, which gives it a negligible probability weight. So even for such a mildly non-Gaussian distribution, $\Delta\chi^2$ does not work. One would expect the situation shown in Fig. 3 can be approximated by two Gaussian distributions. That is exactly what the moment-based IOI does. Compared to the case without the narrow peak, $P_{(1)}(\lambda)$ with that peak only slightly changes the parameters' means and covariance and the value of IOI.
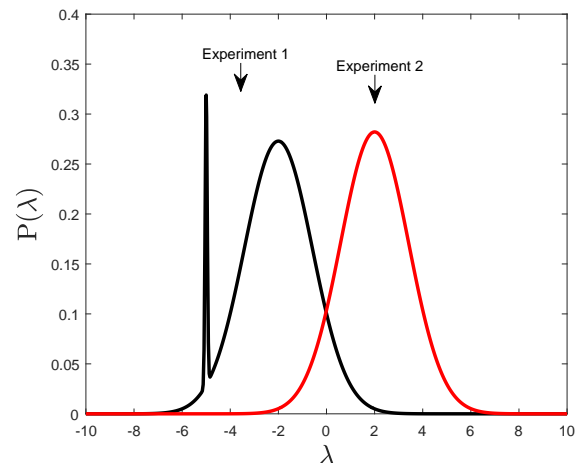

FIG. 3. An example of non-Gaussian case. $P_{(2)}(\lambda)$ given by experiment 2 is Gaussian, while $P_{(1)}(\lambda)$ given by experiment 1 is slightly different from Gaussian. The narrow peak of $P_{(1)}(\lambda)$ should be ignored when we compare the two experiments. Using $\Delta\chi^2$ as a measure of inconsistency only considers the maximum that may have insignificant probability weight, but ignores the overall distribution. The moment-based IOI cares about the distribution as a whole.

It is worth noting that IOI is invariant under a linear parameter transformation, $\boldsymbol{\lambda}_{new} = \boldsymbol{M}\boldsymbol{\lambda}_{old}$ with a transformation matrix $\boldsymbol{M}$. Under such a transformation, the mean difference becomes $\boldsymbol{\delta} \to \boldsymbol{M}\boldsymbol{\delta}$ and the fisher matrices become $\boldsymbol{L}^{(i)} \to (\boldsymbol{M}^{-1})^T \boldsymbol{L}^{(i)} \boldsymbol{M}^{-1}$. Then IOI trans-

forms as,

$$\text{IOI}_{\text{old}} \to \text{IOI}_{\text{new}}$$

$$
\begin{aligned}
\text{IOI}_{\text{new}} =&\tfrac{1}{2}\boldsymbol{\delta}^T \boldsymbol{M}^T \Big[\big((\boldsymbol{M}^{-1})^T \boldsymbol{L}^{(1)} \boldsymbol{M}^{-1}\big)^{-1} \\
& + \big((\boldsymbol{M}^{-1})^T \boldsymbol{L}^{(2)} \boldsymbol{M}^{-1}\big)^{-1} \Big]^{-1} \boldsymbol{M}\boldsymbol{\delta} \\
=&\tfrac{1}{2}\boldsymbol{\delta}^T \boldsymbol{M}^T (\boldsymbol{M}^T)^{-1} \big[(\boldsymbol{L}^{(1)})^{-1} \\
& + (\boldsymbol{L}^{(2)})^{-1}\big]^{-1} \boldsymbol{M}^{-1}\boldsymbol{M}\boldsymbol{\delta} \\
=&\tfrac{1}{2}\boldsymbol{\delta}^T \boldsymbol{G}\boldsymbol{\delta} = \text{IOI}_{\text{old}} \, .
\end{aligned}
\tag{14}
$$

However, IOI is not invariant under a general invertible parameter transformation which should be a desirable property for a measure of inconsistency. IOI is only nearly invariant under an invertible parameter transformation that can be approximated by a linear transformation in the parameter region of interest. The parameter region of interest here means the parameter space spanned by several confidence levels around the joint mean. Constructing a measure based on IOI, which can describe inconsistency for general cases and is general-parameter-invariant, is out of the scope of this paper and is left for future work.

IV. FACTORS AFFECTING IOI AND INCONSISTENCY MEASURES

Experimental inconsistencies are affected by at least three factors: the likelihood mean deviation, the volume of the covariance matrix (how big the iso-likelihood ellipses are), and the degeneracy directions (orientations of the ellipses). The first two factors are related in an obvious way to degree of inconsistency. Basically a larger mean deviation and a smaller covariance matrix volume give a larger experimental inconsistency. The third factor is a little bit more subtle and can be seen as follows. Let's consider the two right panels in Fig. 2 for example. We can see that those two panels have the same mean deviation and the same covariance matrix volumes, but the upper obviously has a smaller inconsistency due to different constraint orientations. So how does IOI relate to and describe different constraint orientations? We recall that IOI incorporates the joint mean to specify the inconsistency. For a given mean deviation and constraint volumes, the orientations of those ellipse determine the location of the joint mean. If the ellipse orientations make the joint mean to locate closer to the mean of each experiment mean, then IOI will be smaller, and vice-versa. A different set of ellipse orientations leads to a different matrix $\boldsymbol{G}$ and a different IOI. So for the case of two experiments, the third factor of inconsistency is reflected by the matrix $\boldsymbol{G}$. The values of IOI can then describe such a change of the graphical inconsistency due to the change of ellipse orientations.

We discussed how the lower right panel in Fig. 2 has a larger inconsistency than upper right panel and explained that this is because the ellipse orientations in
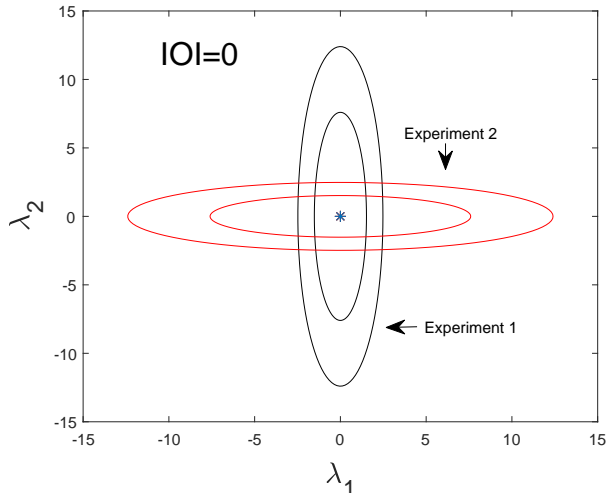

FIG. 4. Two toy experiments with different degenerate directions. As we explain in the text, inconsistency between experiments should be treated separately from small overlapping area caused by degeneracy breaking.

the lower panel make the joint mean further away from each individual mean. However, it is not correct to associate the larger inconsistency in the lower panel with the smaller "overlap" between the two ellipses compared to the upper panel. A smaller overlap does not necessarily mean a larger inconsistency, as we will explain. We recall that the purpose of (in)consistency check is to see whether we can jointly study two or more experiments. Two experiments can have a small overlap but still be consistent with each other. Take Fig. 4 for example. The two toy experiments have orthogonal degenerate directions, which give them a small overlap. The small overlap here does not prevent us from jointly study the two experiments, so we say the two experiments are consistent with each other. Thus, this kind of small overlap should not be counted as an increase of inconsistency. Some might worry that the small overlap here means the two distributions are different. But (in)consistency test is not to look for difference between experiments, but to see if they share anything in common. To see this again, let us assume for a moment that the small overlap does suggest an increasing of inconsistency. Imagine that we have a third experiment with a diagonal (45°) degenerate direction going through the overlap of the previous two experiments. Now the three experiments also have a small overlap and a lot of differences. But do they have a large inconsistency? Obviously not.

Finally, it is worth clarifying that different ellipse orientations affect simultaneously the inconsistency between two experiments and the constraining power of the joint analysis from these two experiments. The latter is related to the "overlap area" of the two experiments. In fact, it can be described by the well-known figure of merit

(FOM); see e.g., Ref. [59]

$$\text{FOM} = |\boldsymbol{C}|^{-1/2}\,, \qquad (15)$$

where $\boldsymbol{C}$ is the covariance matrix of a distribution. FOM quantifies how good a constraint is. For two distributions with certain constraints powers (i.e., certain FOM_1 and FOM_2) and a fixed prior, different ellipse orientations change how strong the joint constraint is, i.e., how large FOM_j is. For Gaussian cases (and certain FOM_1 and FOM_2), FOM_j only depends on ellipse orientations but not on how the ellipses are separated. In contrast, IOI incorporates ellipse orientations and mean separation to specify experimental inconsistency. So while FOM describes how powerful the joint analysis is, IOI tells us whether we can do such a joint analysis or not.

V. MARGINALIZATION CAN HIDE INCONSISTENCY

It is usual to investigate the experimental inconsistency in some marginalized 1-D or 2-D likelihood contour plots, and see if the constrains from different experiments coincide. But there are two problems for this common method. First, if the two experiments have more than three common parameters, we cannot show the inconsistency of them by a plot. Second, since we marginalize over other parameters, we would like to know if the process of marginalization bring up or down the inconsistency between the two experiments. We will show graphically that marginalization hides and lowers artificially the full inconsistency. So after marginalization two experiments will look more consistent one with another but they actually may not be so. Using a measure of discordance solves the first problem, since a measure can be obtained in arbitrary parameter dimension. But we are going to show that only IOI can correctly track this effect of marginalization.

We recall that marginalization over the ith parameter corresponds to deleting the ith component from each mean along with the ith column and ith row from each covariance matrix. Parameter marginalization ignores the detailed information of the probability distribution on the parameters being marginalized over [60]. For the inconsistency of two independent and Gaussian distributions, intuitively the loss of information on inconsistency is due to the following two factors. First, in the case where we have more parameters, an unmatched numerical value of any parameter constrained by two experiments introduces some experimental inconsistency. Marginalization over some unmatched parameters obviously hides the inconsistency raised by those parameters. Second, even the mean values of those parameters are the same (i.e., no unmatched numerical values of those parameters), marginalization can also hide some inconsistency due to the correlations among parameters. One example has been already given by Fig. 2 in our Sec. III B. We can see that the upper right panel in Fig. 2 has a

smaller inconsistency than the lower panel. This is because the orientations of the contour ellipses in the upper right panel allow them to overlap more, leading to smaller inconsistency. However, if we marginalize both panels into 1-D plots, they both have identical 1-D plots. This example shows that ignoring the orientations of contour ellipses is a reason that marginalization hides inconsistency.

To further demonstrate the second point above, we show in Fig. 5 two toy experiments. These two toy experiments have the same mean of λ_2 (here it is the second element of $\boldsymbol{\lambda}$), but they have different means in λ_1. If we marginalize over λ_2, we will make the two experiments looked more consistent with each other as shown in the right panel of Fig. 5. This is because in this case both experiments have broad marginalized uncertainties in λ_1, but the actual degeneracy direction of the experiments is not along the direction of λ_1. Marginalizing over λ_2 hides the fact that the two experiments actually have little overlap shown in higher dimension. This effect is caught by the drop in IOI. The total IOI is 20 meaning a significant inconsistency between the two toy experiments, but after marginalizing over λ_2, IOI becomes 0.24 meaning no significant inconsistency.

From the above discussion, we can see that parameter marginalization hides experimental inconsistency when present. We need to explore the following two questions. Does IOI always drop after parameter marginalization when it should? When marginalizing over a parameter that has no inconsistency associated with it and it does not correlate with other parameter, then IOI remain the same after marginalizing over that parameter. We find that it is case if those two conditions are satisfied. We prove analytically in appendix B 1 that IOI never increases after parameter marginalization. The general condition for IOI to remain the same after marginalizing over λ_k is,

$$\sum_{i,j \neq k} \big(\widetilde{\boldsymbol{C_{kk}}}^{-1}\big)_{ij} C_{ki}\delta_j = \delta_k\,, \text{ (Equality Condition)} \quad (16)$$

where C_{ij}'s are the elements of $\boldsymbol{C} = \boldsymbol{C^{(1)}} + \boldsymbol{C^{(2)}}$, and $\widetilde{\boldsymbol{C_{kk}}}$ is a matrix obtained by taking out the kth row and kth column from $\boldsymbol{C}$. If two experiments give the same mean value of λ_k (i.e., $\delta_k = 0$), and if λ_k does not correlate with other parameters (i.e., $C_{ki} = 0$ for $i \neq k$), the equality condition [Eq. (16)] is satisfied since both sides are 0. Then IOI will remain the same after marginalizing over λ_k. So IOI does track exactly the effect of marginalization. This is a unique feature of IOI among all the measures of inconsistency[2].

However, the condition $\delta_k = 0$ & $C_{ki} = 0$ is only a subset of the equality condition [Eq. (16)], (although such subset is more common and practical). Equation (16)

---

[2] In Gaussian cases, it is also a feature of $\Delta\chi^2$ and the *tension*.
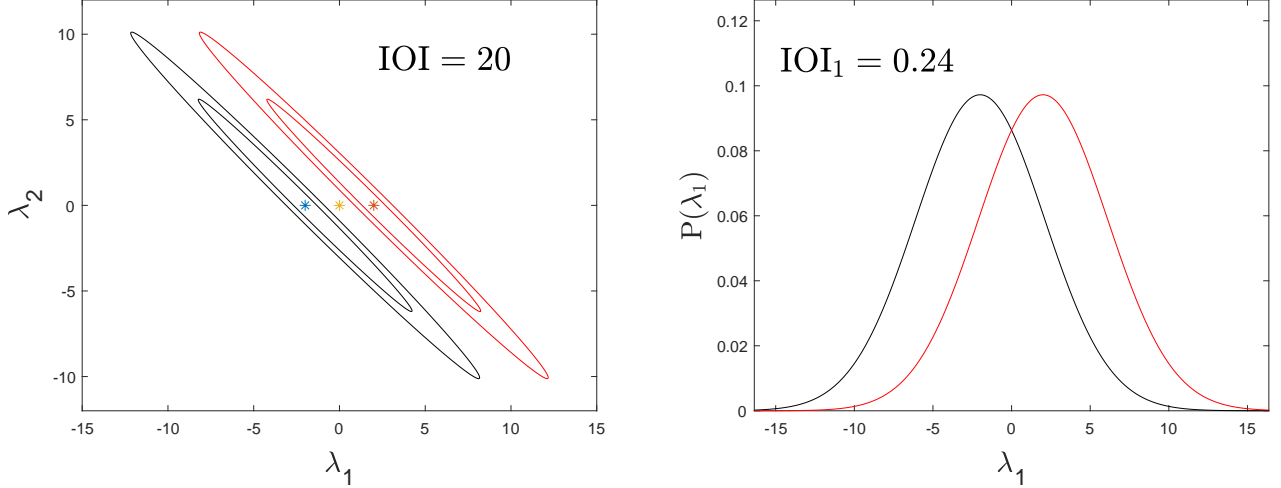
FIG. 5. Toy experiments to show marginalization artificially lowers inconsistency. The left panel shows the 2-D contour plot of the posteriors of two toy experiments. The left panel shows the 1-D probability distributions of the two experiments marginalized over the second parameter. Clearly the marginalized 1-D plot have more overlapped region than the 2-D plot, indicating an artificial decrease of inconsistency between the two experiments. The decrease of the 1-D IOI also reflects such a fact but on needs to consider rather the full IOI on the left.

suggests that there are some other conditions for the experimental inconsistency to be preserved after parameter marginalization. The equality condition (16) can be satisfied without $\delta_k = 0$ or $C_{ki} = 0$. An example of such situation and a full discussion are provided in appendix B 3. In other words, if the drop of IOI after marginalizing over one parameter is small, it does not necessarily mean we are safe to marginalize over that parameter. As we discuss in the next section, we should also check whether the residual IOI_i after marginalizing over all the other parameters is also small. If both are small, we can marginalize safely over that parameter to analyze the inconsistency for the remaining parameter space of the model.

VI. ZOOMING ON ONE PARAMETER: RELATIVE DROP IN IOI AND RELATIVE RESIDUAL IOI

The usual way to zoom the experimental inconsistency of a given parameter is to plot and compare the two marginalized 1-D likelihoods in the same graph. Examples are the marginalized 1-D plots in Fig. 10 and Fig. 11 in Sec. XI. As we show in Sec. V, this is not an accurate description of experimental inconsistency. An important point is that parameters are correlated with each other (in the language of IOI, the correlation is specified by the combined covariance matrix $C = C^{(1)} + C^{(2)}$ instead of $C^{(1)}$, $C^{(2)}$ or the joint covariance matrix $C_{joint} = \left[(C^{(1)})^{-1} + (C^{(2)})^{-1}\right]^{-1}$), but marginalized 1-D plots do not reflect this fact. So if the two experiments yield a consistent result on one parameter in this way,

it does not mean it is safe to marginalize such parameter and analyze the inconsistency of the other parameters. Even all marginalized 1-D plots show quite consistent constraints on all parameters, it does not mean the two experiments are actually consistent with each other. Adding a measure to an 1-D marginalized plot does not improve the situation at all, because a numerical value associated to an 1-D plot does not reflect the effect of marginalization either. All the above points suggest that the usual way of viewing the experimental inconsistency of parameters in marginalized plots is not accurate, and a measure of inconsistency for the full parameter space should be used.

On the other hand, we still want to see if two experiments agree on one particular parameter. For example, we want to see if the geometry and the growth sets of experiments give consistent properties of dark energy like its equation of state w. For IOI, an (obvious) one-parameter inconsistency IOI_i can be defined as the residual IOI of parameter λ_i after marginalizing over all the other parameters in the model. But as we show below, this IOI_i does not fully specify the inconsistency due to parameter λ_i. Indeed, to better specify the inconsistency due to one particular parameter in a model, we propose a supplemental measure based on IOI. We define the drop of IOI after marginalizing over λ_i itself, that is

$$\Delta\text{IOI}_i = \text{IOI} - \text{IOI}_{marg,i}. \qquad (17)$$

This definition is motivated by the fact that IOI never increases whenever we perform a parameter marginalization. Since other measures may increase in general, this supplemental measure of one-parameter inconsistency is unique among the other measures in the literature.

TABLE IV. Summary of the two measures of one-parameter inconsistency.

Notations	Meanings	Relative measures
IOI_i	The residual IOI after marginalizing over all the other parameters except λ_i	$\mathrm{IOI}_i^{\mathrm{Rel}} = \frac{\mathrm{IOI}_i}{\mathrm{IOI}}$
$\Delta\mathrm{IOI}_i$	The drop of IOI after marginalizing over λ_i	$\Delta\mathrm{IOI}_i^{\mathrm{Rel}} = \frac{\Delta\mathrm{IOI}_i}{\mathrm{IOI}}$

With the two measures of one-parameter inconsistency, we can see if it is safe to marginalize over a parameter to analyze the experimental inconsistency. The residual IOI_i can be directly associated with the corresponding marginalized 1-D plot, while the drop $\Delta\mathrm{IOI}_i$ can tells us the degree of inconsistency hidden by marginalizing over λ_i. If two experiments do not have inconsistency due to λ_i, both IOI_i and $\Delta\mathrm{IOI}_i$ should be small on Jeffreys' scales. On the other hand, if one of them is large, it means there exists inconsistency associated with that parameter, and it is not safe to marginalize over it to analyze the inconsistency of the other parameters. In Sec. B 3 we describe in detail different situations regarding different magnitudes of IOI_i and $\Delta\mathrm{IOI}_i$, and we summarize them in Table XII. In sum, marginalizing over a parameter that has a large $\Delta\mathrm{IOI}_i$ or IOI_i will either hide the inconsistency due to correlation or ignore the shape of the original distributions. The requirement is that both measures must be small.

The two measures of one-parameter inconsistency IOI_i and $\Delta\mathrm{IOI}_i$ can be compared to Jeffreys' scales (Table III). However, we often have situations where all IOI_i's and $\Delta\mathrm{IOI}_i$'s are not significant on Jeffreys' scales but they are still not insignificant compared to the full IOI. In this situation, it is not good to marginalize over any parameter to analyze the experimental inconsistency. Therefore, when we decide whether we can marginalize over a parameter, we can compare IOI_i and $\Delta\mathrm{IOI}_i$ to the full IOI. For that, we define the relative residual $\mathrm{IOI}_i^{\mathrm{Rel}}$ and relative drop $\Delta\mathrm{IOI}_i^{\mathrm{Rel}}$ for the parameter λ_i,

$$\mathrm{IOI}_i^{\mathrm{Rel}} = \frac{\mathrm{IOI}_i}{\mathrm{IOI}}, \qquad \Delta\mathrm{IOI}_i^{\mathrm{Rel}} = \frac{\Delta\mathrm{IOI}_i}{\mathrm{IOI}}. \qquad (18)$$

Both $\mathrm{IOI}_i^{\mathrm{Rel}}$ and $\Delta\mathrm{IOI}_i^{\mathrm{Rel}}$ need to be much smaller than 1 for us to safely marginalize over λ_i. The definitions of the two measures of one-parameter inconsistency are summarized in Table IV.

Note that the summation of all IOI_i's or $\Delta\mathrm{IOI}_i$'s is not the full IOI. That is

$$\sum_i \mathrm{IOI}_i \neq \mathrm{IOI}, \qquad \sum_i \Delta\mathrm{IOI}_i \neq \mathrm{IOI}. \qquad (19)$$

Consequently,

$$\sum_i \mathrm{IOI}_i^{\mathrm{Rel}} \neq 1, \qquad \sum_i \Delta\mathrm{IOI}_i^{\mathrm{Rel}} \neq 1. \qquad (20)$$

Thus we cannot interpret $\mathrm{IOI}_i^{\mathrm{Rel}}$ or $\Delta\mathrm{IOI}_i^{\mathrm{Rel}}$ as the fractional inconsistency for λ_i. Normalizing IOI_i and $\Delta\mathrm{IOI}_i$ by deriving the full IOI is only a way to see whether the inconsistency due to λ_i is significant to the full inconsistency of a model.

The analytic expression for $\Delta\mathrm{IOI}_i$ is given by (see appendix B 2)

$$\Delta\mathrm{IOI}_i = \frac{\left[\delta_i - \sum_{m,n\neq i} \left(\widetilde{\mathbf{C_{ii}}}^{-1}\right)_{mn} C_{im}\delta_n\right]^2}{C_{ii} - \sum_{m,n\neq i} \left(\widetilde{\mathbf{C_{ii}}}^{-1}\right)_{mn} C_{im}C_{in}}. \qquad (21)$$

However, despite its analytical neatness, the above expression is not as practically useful as the definition of $\Delta\mathrm{IOI}_i$ itself as given by [Eq. (17)].

VII. IOI EIGEN-MODE DECOMPOSITION

Since IOI is in a quadratic form, it possible to define eigen-modes of inconsistency as the eigen-modes of the matrix $\boldsymbol{G}$. By diagonalizing $\boldsymbol{G}$ we obtain N eigen-modes with eigen-values g_i. Note that the eigen-modes are characterized by $\boldsymbol{G}$, instead of $\boldsymbol{C^{(i)}}$ in the usual principal component analysis. Suppose $\boldsymbol{M}$ is the orthogonal transformation matrix from the original parameter space to the eigen-modes, i.e. $\boldsymbol{\xi} = \boldsymbol{M}\boldsymbol{\lambda}$ and $\boldsymbol{\delta^{\xi}} = \boldsymbol{M}\boldsymbol{\delta}$, then Eq. (9) in the eigen-modes of $\boldsymbol{G}$ reads,

$$\mathrm{IOI} = \sum_i \tfrac{1}{2} g_i (\delta^{\xi}_i)^2. \qquad (22)$$

From the above equation, we can see that the full IOI is a summation of N terms, each of which involves the mean difference of one and only one eigen-mode parameter. We can then identify each term as the inconsistency of each eigen-mode parameter, i.e, $\mathrm{IOI}_i^{\mathrm{eig}} = \tfrac{1}{2} g_i (\delta^{\xi}_i)^2$ (assumed in a descending order), and we have

$$\mathrm{IOI} = \sum_i \mathrm{IOI}_i^{\mathrm{eig}}. \qquad (23)$$

A nice feature of the inconsistency eigen-modes is that these $\mathrm{IOI}_i^{\mathrm{eig}}$'s equal to the drop $\Delta\mathrm{IOI}_i$ after marginalizing over the corresponding parameter, and also equal to the residual IOI_i of λ_i after marginalizing over all the other parameters, that is

$$\Delta\mathrm{IOI}_i = \mathrm{IOI}_i = \mathrm{IOI}_i^{\mathrm{eig}}. \qquad (24)$$

Recall Eqs. (23) and (24) are only valid for eigen-modes of inconsistency. In general cases, the summation of $\Delta\mathrm{IOI}_i$'s or IOI_i's does not equal to IOI of the full parameter space.
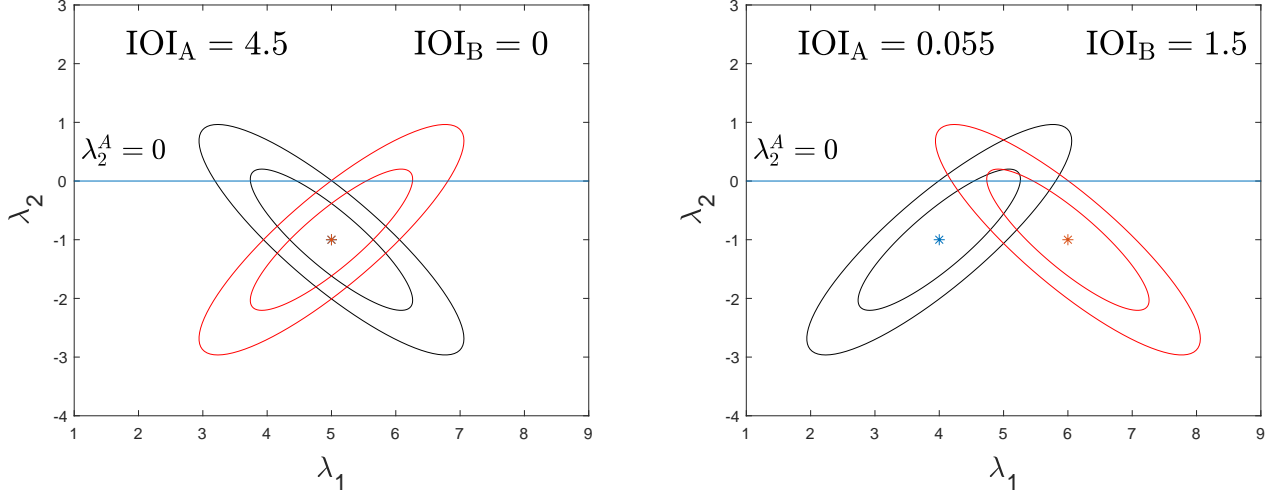
FIG. 6. Toy experiments to show IOI can be smaller or larger for model extension. Model A is nested in model B, and model A has a fixed value of $\lambda_2 = \lambda_2^A = 0$. In the left (right) panel model B has a larger (smaller) IOI than model A. See the text for detailed explanation.

Let us further clarify the meaning of inconsistency eigen-modes. First, this is not like the principal component analysis in which the uncertainties of eigen-modes are specified by the eigen-values of a Fisher matrix. Here the eigen-mode inconsistency is specified by $\mathrm{IOI}_i^{\mathrm{eig}} = \frac{1}{2}g_i(\delta^\xi_i)^2$, rather than g_i alone. A larger g_i does not necessarily mean a larger inconsistency. Secondly, if we rotate the parameter space and make one parameter along the deviation vector $\boldsymbol{\delta}$, we can express the full IOI as $\frac{1}{2}G_{\delta\delta}\delta^2$, where $G_{\delta\delta}$ (or $(\boldsymbol{G})_{\delta\delta}$) is the diagonal element of the rotated $\boldsymbol{G}$ corresponding to the deviation direction, and $\delta = |\boldsymbol{\delta}|$ is the magnitude of the deviation vector. So the largest inconsistency 'mode' seems to be just along the deviation vector $\boldsymbol{\delta}$. However, the numerical value of $\frac{1}{2}G_{\delta\delta}\delta^2$ is not preserved after marginalizing over other parameters. It is the element $(\boldsymbol{G}^{-1})_{\delta\delta}$ that is preserved after marginalization. And since $(\boldsymbol{G})_{\delta\delta}$ is not $\left[(\boldsymbol{G}^{-1})_{\delta\delta}\right]^{-1}$ in general, $\frac{1}{2}G_{\delta\delta}\delta^2$ changes its value after marginalizing over any other parameter. In contrast, the eigen-modes of inconsistency here have a feature that each $\mathrm{IOI}_i^{\mathrm{eig}}$ is preserved after marginalizing over any other parameter.

VIII. IOI IN NESTED MODELS (FIXING PARAMETERS)

We have shown that marginalization can hide experimental inconsistency and lowers IOI, but it does not mean that models with more parameters will necessarily have a higher IOI. Especially if model A is nested in model B (or alternatively said, model B is an extension to model A), it does not necessarily mean model B will have a higher IOI than model A. It was shown in Ref. [58] with their own measure called the *tension* that

extending a model does not necessarily increase or decrease experimental inconsistency. Here we use IOI to show this point in an explicit way. If model A is nested in model B, one (some) of the parameters in model B is fixed in model A. This actually has two effects on the likelihoods. First, fixing a parameter will take out the corresponding elements from the Fisher matrices of model B, the covariance matrices and consequently the matrix $\boldsymbol{G}$ will be changed. Second, the mean values are forced to be shifted onto a parameter sub-space of one (or some) constant parameter(s). These two effects can be shown more explicitly with an example in two dimensions. If the extension model B has means $\boldsymbol{\mu^{(1)}}$ and $\boldsymbol{\mu^{(2)}}$ from two experiments), and Fisher matrices $\boldsymbol{L^{(1)}}$ and $\boldsymbol{L^{(2)}}$. When we fix, e.g., λ_2 to λ_2^A, for each likelihood we can manipulate the term $(\boldsymbol{\lambda}-\boldsymbol{\mu^{(i)}})^T\boldsymbol{L^{(i)}}(\boldsymbol{\lambda}-\boldsymbol{\mu^{(i)}})$ as follows

$$
\begin{aligned}
&(\boldsymbol{\lambda}-\boldsymbol{\mu^{(i)}})^T\boldsymbol{L^{(i)}}(\boldsymbol{\lambda}-\boldsymbol{\mu^{(i)}}) \\
=&L_{11}^{(i)}(\lambda_1-\mu_1^{(i)})^2 + 2L_{12}^{(i)}(\lambda_1-\mu_1^{(i)})(\lambda_2^A-\mu_2^{(i)}) \\
&\qquad\qquad + L_{11}^{(i)}(\lambda_2^A-\mu_2^{(i)})^2 \\
=&L_{11}^{(i)}\Big[\lambda_1-\mu_1^{(i)} - \frac{L_{12}^{(i)}}{L_{11}^{(i)}}(\mu_2^{(i)}-\lambda_2^A)\Big]^2 \\
&\qquad\qquad\qquad + \text{terms independent of } \lambda_1,
\end{aligned}
\tag{25}
$$

where $L_{jk}^{(i)}$ is the (j,k) element of $\boldsymbol{L^{(i)}}$. After taking the exponential of the above, the terms independent of λ_1 become only a multiplying factor that can be taken care of by normalization. So fixing parameter λ_2 to λ_2^A changes the squared uncertainty of λ_1 from $C_{11}^{(i)}$ to $1/L_{11}^{(i)}$, and shifts the means of λ_1 from $\mu_1^{(i)}$ to $\mu_1^{(i),A} = \mu_1^{(i)} + \frac{L_{12}^{(i)}}{L_{11}^{(i)}}(\mu_2^{(i)}-\lambda_2^A)$.

The shifts of means change IOI the most: it may ei-

ther bring the means closer or push them further away. As a result, an extension may have a larger or smaller IOI than the model being extended. Figure 6 shows the two possibilities. In both panels in Fig. 6, model B is an extension to model A. Model B has two parameters, λ_1 and λ_2. Parameter λ_2 is fixed to 0 in model A. In the left panel, the two toy experiments give two consistent distributions for model B, and IOI_B is 0. But since we fix λ_2 to $\lambda_2^A = 0$, the distributions are now on the straight line. The two means are now pushed further apart in model A. The extension model B is more consistent than model A when comparing the two experiments, and accordingly IOI_B is smaller than IOI_A. On the other hand, in the right panel the means are brought closer when fixing λ_2 to $\lambda_2^A = 0$. The extension model B is then less consistent than model A, and IOI_B is accordingly larger than IOI_A. Therefore, an extension may have larger or smaller IOI.

We classify model B to be either a more consistent extension to model A (if IOI_B is smaller than IOI_A), or a less consistent extension (if IOI_B is larger than IOI_A). But since IOI depends on particular experiments, whether an extension is more or less consistent is relative and depends on the experiments under consideration.

IX. NUISANCE AND UNSHARED PARAMETERS

When we compare two experiments, we often come across situations where one experiment can constrain some parameters in a model but the other cannot; or each experiment has its own systematic/nuisance parameters. In those situations, the two experiments give two probability distributions with different parameter dimensions. It is then necessary to marginalize over the uncommon parameters to get two distributions with the same dimension. Does parameter marginalization in these cases hide some experimental inconsistencies? The answer is no. Marginalizing over any systematic parameters or unshared parameters does not affect IOI even in the case where the inconsistencies can be caused by those parameters.

For a real example, in Sec. XI we will compare the geometry experiments and the growth experiments in the ΛCDM model. Since the geometry experiments do not constraint the parameter τ (optical depth), we need to fix it when fitting the ΛCDM model to the geometry experiments. We then marginalize over τ for the growth experiments to compare the two sets of experiments. If the geometry experiments do not constrain or very poorly constrain τ, it is equivalent to the situation where: 1) it gives very wide (weak) uncertainty on τ, and $\left(\boldsymbol{L^{geom}}\right)_{\tau\tau} \to 0$; 2) τ has no correlation with other parameters in the geometry set, and $\left(\boldsymbol{L^{geom}}\right)_{\tau i} \to 0$ for $i \neq \tau$. No matter what constraint of τ is given by the growth experiments, the geometry experiments do not 'object' such a constraint, and there is no conflict between the two sets on the value of τ. This does not necessarily mean that the joint mean of τ would be the same as the mean given by the growth experiments. In fact they are different in general, and the joint mean is given by the Fisher-matrix-weighted average [Eq. (7)]. However, the form of $\boldsymbol{L^{geom}}$ now implies that the joint mean of τ does not depends on the fixed value of τ for in geometry set, or the mean given by the geometry set (if it gave one).

Mathematically, we can see more clearly from the analytical expression of ΔIOI_τ given by Eq. (B10). If $\left(\boldsymbol{L^{geom}}\right)_{\tau i} \to 0$ for $i \neq \tau$ and $\left(\boldsymbol{L^{geom}}\right)_{\tau\tau} \to 0$, we have $C_{\tau\tau}^{geom} \to \infty$. Every term in Eq. (B10) is finite except for $C_{\tau\tau} = C_{\tau\tau}^{geom} + C_{\tau\tau}^{grow} \to \infty$. So the denominator of Eq. (B10) goes to ∞ and $\Delta\text{IOI}_\tau \to 0$. In another word, the drop of IOI after marginalizing over τ is 0. Also $\text{IOI}_\tau \to 0$ as well since $C_{\tau\tau} \to \infty$ and $\text{IOI}_\tau = \frac{1}{2}\delta_\tau^2/C_{\tau\tau} \to 0$ for finite δ_τ. Since both ΔIOI_τ and IOI_τ are zero, there is no problem to marginalize over τ to analyze the inconsistency of the whole model. For the same reason, it is also safe to marginalize over all the nuisance parameters of each experiment.

X. COMPARISON WITH OTHER MEASURES OF INCONSISTENCY IN THE LITERATURE

In this section we compare IOI to some other measures of inconsistency/consistency in the literature. Their Gaussian and weak prior limits will be derived for the comparison. We also examine how such measures track or not inconsistencies in various particular cases.

A. The *Robustness*

The *robustness* was introduced in Refs. [61, 62], and has been applied, e.g., in Ref. [63] to investigate the internal inconsistency of type Ia supernovae dataset. It is defined as the Bayesian evidence ratio between the combined model and the split model in order to quantify the consistency between two experiments as follows. Combined model: the two experiments are jointly fit by a set of parameters; Split model: each experiment is individually fitted by one set of parameters. The robustness is then defined as

$$R = \frac{E_{comb}}{E_{split}}, \qquad (26)$$

where E is the corresponding Bayesian evidence, so

$$E_{comb} = \int d^N \lambda\, \mathcal{L}(\boldsymbol{Q}; \boldsymbol{\lambda})\mathcal{P}(\boldsymbol{\lambda})$$
$$= \int d^N \lambda\, \mathcal{L}^{(1)}(\boldsymbol{Q^{(1)}}; \boldsymbol{\lambda})\mathcal{L}^{(2)}(\boldsymbol{Q^{(2)}}; \boldsymbol{\lambda})\mathcal{P}(\boldsymbol{\lambda}) \quad (27)$$

$$E_{split} = E_1 \times E_2$$
$$= \int d^N \lambda\, \mathcal{L}^{(1)}(\boldsymbol{Q^{(1)}}; \boldsymbol{\lambda})\mathcal{P}(\boldsymbol{\lambda})$$
$$\times \int d^N \lambda'\, \mathcal{L}^{(2)}(\boldsymbol{Q^{(2)}}; \boldsymbol{\lambda'})\mathcal{P}(\boldsymbol{\lambda'}). \quad (28)$$

The values of R are interpreted by Jeffreys' scales listed in Table V.

In the Gaussian limit, we can calculate the *robustness* analytically. Differently from the treatment in Ref. [62], we do not set one of the means to be $\boldsymbol{0}$ to find

$$R = \sqrt{\frac{|\boldsymbol{F^{(1)}}\boldsymbol{F^{(2)}}|}{|\boldsymbol{F}\boldsymbol{P}|}} \exp\Big[-\frac{1}{2}\big(\bar{\boldsymbol{\mu}}^{(1)\,T}\boldsymbol{F^{(1)}}\bar{\boldsymbol{\mu}}^{(1)}$$
$$+ \bar{\boldsymbol{\mu}}^{(2)\,T}\boldsymbol{F^{(2)}}\bar{\boldsymbol{\mu}}^{(2)} - \boldsymbol{\mu}^{(p)\,T}\boldsymbol{P}\boldsymbol{\mu}^{(p)} - \bar{\boldsymbol{\mu}}^T\boldsymbol{F}\bar{\boldsymbol{\mu}}\big)\Big], \quad (29)$$

where,

$$\boldsymbol{F^{(i)}} = \boldsymbol{L^{(i)}} + \boldsymbol{P},$$
$$\boldsymbol{F} = \boldsymbol{L^{(1)}} + \boldsymbol{L^{(2)}} + \boldsymbol{P},$$
$$\bar{\boldsymbol{\mu}}^{(i)} = \frac{1}{\boldsymbol{F^{(i)}}}\big(\boldsymbol{L^{(i)}}\boldsymbol{\mu}^{(i)} + \boldsymbol{P}\boldsymbol{\mu}^{(p)}\big), \quad (30)$$
$$\bar{\boldsymbol{\mu}} = \frac{1}{\boldsymbol{F}}\big(\boldsymbol{L^{(1)}}\boldsymbol{\mu}^{(1)} + \boldsymbol{L^{(2)}}\boldsymbol{\mu}^{(2)} + \boldsymbol{P}\boldsymbol{\mu}^{(p)}\big).$$

The major steps of the above calculation can be found in our appendix A 3. In the weak prior limit Eq. (29) can be further reduced to,

$$\ln R = \tfrac{1}{2}\ln\big(\frac{|\boldsymbol{L^{(1)}}\boldsymbol{L^{(2)}}|}{|\boldsymbol{L}\boldsymbol{P}|}\big) - \text{IOI}. \quad (31)$$

So the *robustness* in the case of Gaussian and weak prior is related to the negative of IOI with an additional term $\tfrac{1}{2}\ln\big(\frac{|\boldsymbol{L^{(1)}}\boldsymbol{L^{(2)}}|}{|\boldsymbol{L^{(2)}}\boldsymbol{P}|}\big)$ that accounts for the Occam's razor factor (favoring the model with fewer assumptions). Similar relation [to Eq.,(31)] is pointed out in Ref. [61] where $\ln R$ is related to $\frac{1}{2}\Delta\hat{\chi}^2$ (same as $\frac{1}{2}\Delta\chi^2$ with our notation) which in the Gaussian case is defined as IOI in this work. Ref. [64] noted that the dependence of prior does not go away in the weak prior limit and $\ln R$ diverges as $|\boldsymbol{P}| \to 0$. This seems to imply that two experiments are always consistent with each other when a very weak prior is used, which can't be true.

The dependence of $|\boldsymbol{P}|$ can be eliminated by a normalization. So for example, the authors of Ref. [62] define $R_N = R/R_*$ where $R_* = \frac{|\boldsymbol{F^{(1)}}|}{(|2\boldsymbol{L^{(1)}}+\boldsymbol{P}||\boldsymbol{P}|)}$, so that $R_N = 1$ if the second experiment gives the same likelihood as the first one. Then from Eq. (31), R_N in Gaussian and weak

prior limit reads

$$\ln R_N = \tfrac{1}{2}\ln\big(\frac{|2\boldsymbol{L^{(2)}}|}{|\boldsymbol{L^{(1)}} + \boldsymbol{L^{(2)}}|}\big) - \text{IOI}, \quad (32)$$

and the dependence on $|\boldsymbol{P}|$ is eliminated. But other problems seem to persist. For example, consider two Gaussian likelihoods with the same mean, i.e., $\boldsymbol{\delta} = \boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)} = 0$. And suppose their Fisher matrices are given by

$$\boldsymbol{L^{(1)}} = \begin{pmatrix} 1 & 0 \\ 0 & \epsilon \end{pmatrix}, \quad \boldsymbol{L^{(2)}} = \begin{pmatrix} \epsilon & 0 \\ 0 & 1 \end{pmatrix}, \quad \text{with } \epsilon \ll 1. \quad (33)$$

Similar likelihood contours of these two toy experiments are shown in Fig. 4, but now the contours are much narrower. The first constraint ellipse is very elongated along the y direction, while the other is very elongated along the x direction. Although there is little overlap between the two experiments, they have the same parameter mean so there should be no tension between them[3]. But the normalized *robustness* turns out to be (to the leading order of ϵ)

$$R_N \approx 2\sqrt{\epsilon} \ll 1. \quad (34)$$

So R_N implies these two toy experiments are not 'robust' or not consistent with each other, which seems to be in contradiction with the fact that they are actually consistent with each other.

The term $\frac{|\boldsymbol{L^{(1)}}\boldsymbol{L^{(2)}}|}{|\boldsymbol{L}||\boldsymbol{P}|}$ in the original *robustness* also disfavors orthogonal constraints. This may be easier to see this if we relate the *robustness* to IOI and FOM. From the Gaussian and weak prior limit of the *robustness* [Eq. (31)] and the definition of FOM [Eq. (15)], we can see that

$$-\ln R = \ln\Big(\frac{\text{FOM}_j \times \text{FOM}_p}{\text{FOM}_1 \times \text{FOM}_2}\Big) + \text{IOI}, \quad (35)$$

where the subscripts "1", "2", "j" and "p" stand for "first", "second", "joint" and "prior". In Sec. IV we argued that different constraint orientations do lead to different experimental inconsistencies, which have already been taken care of by IOI. For a fixed FOM_p, the term $\frac{\text{FOM}_j \times \text{FOM}_p}{\text{FOM}_1 \times \text{FOM}_2}$ can describe how powerful the joint constraint is compared to the two individual constraints. It is true that this term does also depend on the constraint orientations, and it is larger for orthogonal constraints than parallel constraints. But it should not be taken into accounted as indication of consistency or inconsistency, because it is independent of how far away the two individual constraints are separated from each other.

[3] A real and similar example of this situation is the experiments of BAO and distances to supernova.

TABLE V. Jeffreys' scales - empirical scales that classify the ranges of the Bayesian evidence ratio. Positive $\ln R$ ($R > 1$) favors the combined model while negative $\ln R$ favors the split model. If $|\ln R| < 1$, empirically it means there is no enough evidence to tell which model is supported by the data. Higher value of $|\ln R|$ has higher preference.

ranges	$|\ln R| < 1$	$1 < |\ln R| < 2.5$	$2.5 < |\ln R| < 5$	$|\ln R| > 5$
meanings	no preference	weak evidence	moderate evidence	strong evidence

B. The *Tension*

Based on Bayesian evidence ratio, a quantity called the *tension* $\mathcal{T}$ was introduced in Ref. [58]. This measure was used in Ref. [16], e.g., to study the tension between the direct measurement of H_0 and the one derived from Planck CMB observation. The $\mathcal{T}$ is defined as follow. For two posteriors $\mathscr{P}^{(1)}$ and $\mathscr{P}^{(2)}$, we transfer them while keeping their shapes to get two new posteriors $\mathscr{P}_{sh}^{(1)}$ and $\mathscr{P}_{sh}^{(2)}$. Peaks of $\mathscr{P}_{sh}^{(1)}$ and $\mathscr{P}_{sh}^{(2)}$ coincide at the same location. Then the *tension* $\mathcal{T}$ is defined as

$$\mathcal{T} \equiv \frac{\int \mathscr{P}_{sh}^{(1)} \mathscr{P}_{sh}^{(2)} d^N\lambda}{\int \mathscr{P}^{(1)} \mathscr{P}^{(2)} d^N\lambda}. \qquad (36)$$

As Ref. [58] explained, the *tension* $\mathcal{T}$ is defined as the joint Bayesian evidence ratio between a virtual hypothesis and a actual hypothesis. It was also proposed in Ref. [58] to interpret $\ln \mathcal{T}$ using Jeffreys' scales.

In the case of Gaussian likelihoods and a Gaussian posterior, the integral of posterior product in Eq. (36) can be calculated analytically, (see our appendix A 3 for details)

$$\int \mathscr{P}^{(1)} \mathscr{P}^{(2)} d^N\lambda = |\bar{\boldsymbol{G}}|^{1/2} \exp\left(-\frac{1}{2}\bar{\boldsymbol{\delta}}\bar{\boldsymbol{G}}\bar{\boldsymbol{\delta}}\right), \qquad (37)$$

where (again) $\bar{\boldsymbol{G}} = \left[(\boldsymbol{F}^{(1)})^{-1} + (\boldsymbol{F}^{(2)})^{-1}\right]^{-1}$, $\boldsymbol{F}^{(i)} = \boldsymbol{L}^{(i)} + \boldsymbol{P}$, $\bar{\boldsymbol{\delta}} = \bar{\boldsymbol{\mu}}^{(1)} - \bar{\boldsymbol{\mu}}^{(2)}$ and $\bar{\boldsymbol{\mu}}^{(i)} = (\boldsymbol{F}^{(i)})^{-1}(\boldsymbol{L}^{(i)}\boldsymbol{\mu}^{(i)} + \boldsymbol{P}\boldsymbol{\mu}^{(p)})$. Note that $\bar{\boldsymbol{G}}$ and $\bar{\boldsymbol{\mu}}^{(i)}$ are different from $\boldsymbol{G}$ and $\boldsymbol{\mu}^{(i)}$, and the formers have the prior involved. In the weak prior limit, $\bar{\boldsymbol{G}} \xrightarrow{\text{weak prior}} \boldsymbol{G}$ and $\bar{\boldsymbol{\mu}}^{(i)} \xrightarrow{\text{weak prior}} \boldsymbol{\mu}^{(i)}$. The shifted-posterior-product integral is similar except that $\bar{\boldsymbol{\delta}} = \bar{\boldsymbol{\mu}}^{(1)} - \bar{\boldsymbol{\mu}}^{(2)} = 0$, so $\int \mathscr{P}_{sh}^{(1)} \mathscr{P}_{sh}^{(2)} d^N\lambda = |\bar{\boldsymbol{G}}|^{1/2}$. Therefore, the *tension* $\mathcal{T}$ becomes

$$\ln \mathcal{T} = \frac{1}{2}\bar{\boldsymbol{\delta}}^T \bar{\boldsymbol{G}}\bar{\boldsymbol{\delta}} \xrightarrow{\text{weak prior}} \frac{1}{2}\boldsymbol{\delta}^T \boldsymbol{G}\boldsymbol{\delta} = \text{IOI}. \qquad (38)$$

So $\ln \mathcal{T}$ reduces to IOI in the Gaussian and weak prior limit. Since the shifted posteriors always have a larger overlapping region for Gaussian cases, $\mathcal{T} > 1$, which can also be seen from Eq. (38).

It was also shown in Ref. [58] that the *tension* is another information on the problem of model selection. For example, Table 2 in Ref. [58] shows that while there is not enough Bayesian evidence favoring the ΛCDM model over the wCDM model, the latter significantly reduces the tension on H_0 between Planck and the local measurements.

Authors in Ref. [64] commented that the *tension* $\ln \mathcal{T}$ is of the order N (dimension of the parameter space)
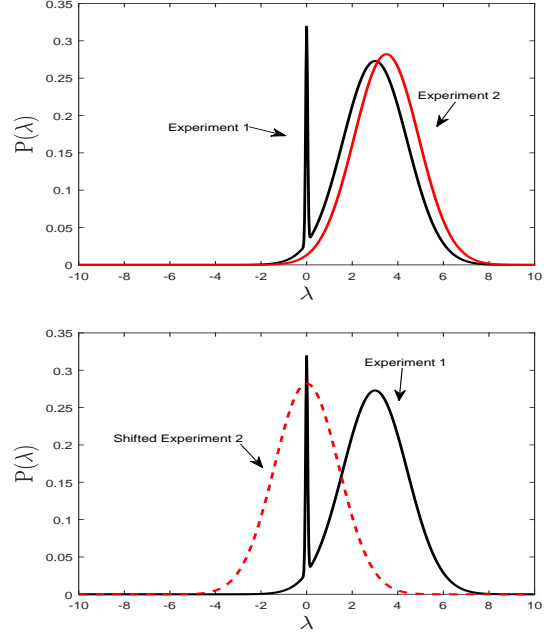

FIG. 7. Demonstration of the problem of the *tension* in a non-Gaussian case. Upper panel: two original one-dimensional probability distributions. The one ($P_{(1)}(\lambda)$) given by experiment 1 is non-Gaussian, while the other ($P_{(2)}(\lambda)$) is Gaussian. Lower panel: the distribution $P_{(2)}(\lambda)$ is shifted to coincide its maximum with that of $P_{(1)}(\lambda)$. Due to the small overlap of the shifted distributions in the lower panel, the numerator in Eq. (36) is smaller than the denominator, which makes $\ln \mathcal{T}$ negative.

and could overestimate the inconsistency between two experiments. Since $\ln \mathcal{T}$ and IOI are the same in Gaussian and weak prior limit, that comment would also apply to IOI. But we find here otherwise. First, Table 2 in Ref. [58] showed that some extensions of the ΛCDM model increase the *tension* while the others decrease it. So increasing the degree of freedoms does not necessarily increases or decrease $\ln \mathcal{T}$. Second, we can always rotate the parameter space and let the first axis be along the deviation vector $\boldsymbol{\delta}$. Then $\ln \mathcal{T}$ consists of only one term $\frac{1}{2}g_\delta\delta^2$ in the Gaussian and weak prior limit (see Sec. VII). So it should be of the order unity.

While the *tension* reduces to IOI in the Gaussian and weak prior limit, the difference between them lies in non-Gaussian cases and in the fact that they originate from distinct concepts (see Table VI). One advantage of the *tension* compared to IOI is that the *tension* is defined

for general distributions. The *tension* may give us a better sense of inconsistency in general cases (although we don't know yet what a good measure is in non-Gaussian cases), while IOI may not work in highly non-Gaussian cases. However, in some cases, it may be better to use IOI. Recall that the definition of the *tension* involves the shifted distributions such that they have coinciding maxima. This may give to the *tension* an ambiguity in choosing the shift in general cases. Consider the two 1-D distributions shown in the upper panel of Fig. 7. The two experiments give two similar distributions. But $P_{(1)}(\lambda)$ is not Gaussian, and it has a narrow peak at $\lambda = 0$ which is unfortunately also the maximum of $P_{(1)}(\lambda)$. If we shift the two distributions to make their maxima coincide, we will get the lower panel in Fig. 7. Since the shifted distributions have less overlap than the original distributions, the numerator in Eq. 36 can be smaller than the denominator. We can then get $\mathcal{T} < 1$ and $\ln \mathcal{T} < 0$. The negative value of $\ln \mathcal{T}$ here certainly can not be interpreted as 'no tension'. In fact, the 'majorities' of the two distribution may be very different and exhibiting a large tension, but $\ln \mathcal{T}$ may still be negative due to the possibly small overlap of the shifted distributions. The problem is that the numerator in Eq. (36) is not guaranteed to be larger than the denominator. Better shifted distributions can be defined as those who maximize the numerator in Eq. (36), though such definition may not be as practical as the original one. However, distributions in cosmology are usually unimodal, making two maxima to coincide usually also maximize the numerator in Eq. (36). Our extreme example here is only to show the possible ambiguity in the choice of distribution shift for general cases, but the measure *tension* should work just fine in the usual and common cases.

C. The *Discordance*

The *discordance* was constructed from overlapping iso-likelihood contours and was proposed in Ref. [65] to measure the inconsistency between two distributions in the following way. First, for a given specific percentage, we can draw two iso-likelihood contours of the two distributions. Depending on how far away the two distributions are separated from each other, the two iso-likelihood contours may or may not overlap. If the two distributions are very far away from each other, the two iso-likelihood contours need to be large (corresponding to large included percentage) in order to overlap with each other. So the percentage within each overlapping iso-likelihood contours traces the "distance" of two distributions. The *discordance* is defined as the minimum percentage σ_{eq} of the overlapping iso-likelihoods that include equal probability. A larger σ_{eq} corresponds to a greater "distance" between the two distributions, and a larger inconsistency. If the maxima of two distributions coincide at the same location, σ_{eq} will be 0; on the other hand if they are separated very far away and very inconsistent with each

other, $\sigma_{eq} \to 1$. The range of *discordance* is $[0, 1]$, and it is compared to 1 to measure the level of inconsistency.

Reference [65] also discussed the effect of marginalization on the value of *discordance*. But instead of decreasing after marginalization, the *discordance* tends to increase. In their Appendix A (in Ref. [65]) they showed a simple example of 2-D toy experiments. Both of their toy experiments provide uncorrelated parameter constraints. While the two experiments give different mean values of parameter x, the means of y are the same. After they marginalize over y, the *discordance* increases. This is different from the behavior of IOI and what we have shown graphically in Sec. V. We have argued that marginalizing over inconsistency-irrelevant parameters[4] does not affect the study of inconsistency between two experiments. In their example the two mean values of the y are the same, and there is no correlation between x and y, marginalizing over y should not change the inconsistency between the two toy experiments or the measure representing it. It was suggested in Ref. [65] that, for a model with N dimensions, it is the full N-dimensional σ_{eq} that should be used for a conservative measure of discordance. However, if two experiments have inconsistent results on only one parameter and if all parameters are uncorrelated, a model with a higher dimension will have a smaller *discordance*. In that case, a very high-dimensional model will have a very low *discordance*. The full-parameter *discordance* actually underestimates the inconsistency in that case.

Unlike for the other measures, we could not obtain an explicit expression of the *discordance*. However, there is a simple algebraic procedure to calculate it in Gaussian and weak prior limit that we provide below. When $|P| \to 0$ so that the prior can be treated as a constant in the parameter range of interest, the integrated probability included within the iso-likelihood contour of the same $\Delta\chi^2$ is given by

$$P_N(< \Delta\chi^2) = \frac{\gamma(\frac{N}{2}, \frac{\Delta\chi^2}{2})}{\Gamma(\frac{N}{2})} \,, \tag{39}$$

where Γ is the gamma function and γ is the lower incomplete gamma function. The overlapping iso-likelihoods with minimum σ_{eq} will be tangentially touching with each other at point $\boldsymbol{\lambda}_{eq}$. The gradients of two iso-likelihoods at $\boldsymbol{\lambda}_{eq}$ must be of opposite directions, that is

$$(1-x)\frac{\partial \Delta\chi^2_{(1)}}{\partial \boldsymbol{\lambda}_{eq}} = -x\frac{\partial \Delta\chi^2_{(2)}}{\partial \boldsymbol{\lambda}_{eq}} \,, \tag{40}$$

where $x \in [0, 1]$ is a number we need to solve for. In Gaussian cases with $\Delta\chi^2_{(i)} = (\boldsymbol{\lambda} - \boldsymbol{\mu}^{(i)})^T \boldsymbol{L}^{(i)} (\boldsymbol{\lambda} - \boldsymbol{\mu}^{(i)})$,

---

[4] Parameters that have the same mean values and do not correlate with other parameters, or unshared parameters

we can get $\boldsymbol{\lambda}_{eq}$ as a function of x from Eq. (40), that is

$$\boldsymbol{\lambda}_{eq} = \frac{1}{(1-x)\boldsymbol{L^{(1)}} + x\boldsymbol{L^{(2)}}}\left((1-x)\boldsymbol{L^{(1)}}\boldsymbol{\mu^{(1)}} + x\boldsymbol{L^{(2)}}\boldsymbol{\mu^{(2)}}\right).$$

(41)

Equation (41) can be viewed as a modified-Fisher-matrix-weighted average. There are three special cases: $\boldsymbol{\lambda}_{eq} = \boldsymbol{\mu^{(1)}}$ when $x = 0$; $\boldsymbol{\lambda}_{eq} = \boldsymbol{\mu}$ when $x = 0.5$; $\boldsymbol{\lambda}_{eq} = \boldsymbol{\mu^{(2)}}$ when $x = 1$. By definition we have

$$\Delta\chi^2_{(1)}(\boldsymbol{\lambda}_{eq}) = \Delta\chi^2_{(2)}(\boldsymbol{\lambda}_{eq}).$$

(42)

We can first solve Eq. (42) for x, substitute x in Eq. (41) to get $\boldsymbol{\lambda}_{eq}$, and substitute $\boldsymbol{\lambda}_{eq}$ in Eq. (39) to get

$$\sigma_{eq} = P_N(< \Delta\chi^2(\boldsymbol{\lambda}_{eq})).$$

(43)

We use this algorithm to make corresponding calculations further in the paper.

D. The *Surprise*

Based on relative entropy (or Kullback-Leibler divergence) in information theory, an measure, called *surprise*, was introduced in Refs. [64, 66, 67]. Reference [68] used this measure to study the tension between different distance indicators. For two posterior probability distributions, this quantity $S(\mathscr{P}^{(2)}||\mathscr{P}^{(1)})$ is defined as the difference between the relative entropy $D(\mathscr{P}^{(2)}||\mathscr{P}^{(1)})$ and its average $\langle D \rangle_{Q_2}$

$$D(\mathscr{P}^{(2)}||\mathscr{P}^{(1)}) = \int d^N\lambda \mathscr{P}^{(2)}(\boldsymbol{\lambda}) \ln\left(\frac{\mathscr{P}^{(2)}(\boldsymbol{\lambda})}{\mathscr{P}^{(1)}(\boldsymbol{\lambda})}\right),$$

(44)

$$S(\mathscr{P}^{(2)}||\mathscr{P}^{(1)}) = D(\mathscr{P}^{(2)}||\mathscr{P}^{(1)}) - \langle D \rangle_{Q_2}.$$

(45)

The average above is taken over a probability distribution $P(\boldsymbol{Q^{(2)}}|\boldsymbol{Q^{(1)}})$ of the second data $\boldsymbol{Q^{(2)}}$ given the first data $\boldsymbol{Q^{(1)}}$:

$$\langle D \rangle_{Q_2} = \int d\boldsymbol{Q^{(2)}} \ D \times P(\boldsymbol{Q^{(2)}}|\boldsymbol{Q^{(1)}}),$$

(46)

$$P(\boldsymbol{Q^{(2)}}|\boldsymbol{Q^{(1)}}) = \int d^N\lambda \mathcal{L}(\boldsymbol{Q^{(2)}}|\boldsymbol{\lambda})\mathscr{P}(\boldsymbol{\lambda}|\boldsymbol{Q^{(1)}}).$$

(47)

The uncertainty of D can be also calculated,

$$\sigma^2(D) = \langle D^2 \rangle_{Q_2} - \langle D \rangle^2_{Q_2}.$$

(48)

And the significance of $S(\mathscr{P}^{(2)}||\mathscr{P}^{(1)})$ is defined as

$$\text{significance of } S = \frac{S(\mathscr{P}^{(2)}||\mathscr{P}^{(1)})}{\sigma(D)}.$$

(49)

Two different cases are discussed in Ref. [66]: one is called "updating data" where the second experiment is used to updated the constraints from the first experiments as a prior; the other is called "replacing data"

where two experiments are used separately to set constraints on parameters. Since we are considering the inconsistency between two experiments with similar constraining powers, it is the second case that is more relevant to our work. The *surprise* and $\sigma^2(D)$ for second case in Gaussian limit is given in Ref. [64], which in our notation read

$$S(\mathscr{P}^{(2)}||\mathscr{P}^{(1)})$$

$$\xrightarrow{\text{Gaussian}} \frac{1}{2}\bar{\boldsymbol{\delta}}^T\boldsymbol{F^{(1)}}\bar{\boldsymbol{\delta}} - \frac{1}{2}\text{tr}(\boldsymbol{I_N} + (\boldsymbol{F^{(2)}})^{-1}\boldsymbol{F^{(1)}}) \quad (50)$$

$$\xrightarrow{\text{weak prior}} \frac{1}{2}\boldsymbol{\delta}^T\boldsymbol{L^{(1)}}\boldsymbol{\delta} - \frac{1}{2}\text{tr}(\boldsymbol{I_N} + (\boldsymbol{L^{(2)}})^{-1}\boldsymbol{L^{(1)}}),$$

$$\sigma^2(D) = \frac{1}{2}\text{tr}\left[(\boldsymbol{I_N} + (\boldsymbol{F^{(2)}})^{-1}\boldsymbol{F^{(1)}})^2\right], \quad (51)$$

where $\boldsymbol{I_N}$ is the N-dimensional identity matrix. The relative entropy (and the *surprise*) is divided by $\ln(2)$ when being measured in the unit of bits. For example, if the second experiment only gives a narrower constraint than the first one, $D(\mathscr{P}^{(2)}||\mathscr{P}^{(1)}) = 1$ bit roughly means that the second experiment improves and shrinks the constraint to half compared to the first experiment .

Despite its interesting concept, the outcome of the *surprise* S seems to depend on the choice of its argument order leading to some ambiguity. For example, let's consider two sets of toy experiments described by

$$\text{Left panel in Fig. 8}: \bar{\boldsymbol{\mu}}^{(2)} - \bar{\boldsymbol{\mu}}^{(1)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad (52)$$

$$\boldsymbol{F^{(1)}} = \begin{pmatrix} 4 & 0 \\ 0 & 128 \end{pmatrix}, \ \boldsymbol{F^{(2)}} = \begin{pmatrix} 4 & 0 \\ 0 & 128 \end{pmatrix}, \quad (53)$$

$$\text{Right panel in Fig. 8}: \bar{\boldsymbol{\mu}}^{(2)} - \bar{\boldsymbol{\mu}}^{(1)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad (54)$$

$$\boldsymbol{F^{(1)}} = \begin{pmatrix} 4 & 0 \\ 0 & 128 \end{pmatrix}, \ \boldsymbol{F^{(2)}} = \begin{pmatrix} 128 & 0 \\ 0 & 4 \end{pmatrix}. \quad (55)$$

These two sets of experiments are shown in Fig. 8.

The contour plots indicate that the two experiments shown on the left panel seems to have a smaller inconsistency than those on the right panel. This is for example, reflected quantitatively on the respective values of IOI with the right value (i.e. 1.9) being larger than the left one (i.e. 1.0). However if $S(\mathscr{P}^{(2)}||\mathscr{P}^{(1)})$ is used, it decreases from 0 on the left to -15 (or -21.6 bits) on the right implying a decrease of inconsistency. The significance of $S(\mathscr{P}^{(2)}||\mathscr{P}^{(1)})$ also has the same implication: 0 for the left panel and -0.64 for the right one. Now, since the *surprise* is not symmetric about the two experiments, one may choose a different argument order in S. So, if we instead use $S(\mathscr{P}^{(1)}||\mathscr{P}^{(2)})$ (i.e., switching the argument order), the *surprise* and its significance changes both from 0 on the left panel to ~ 47 and ~ 2 respectively on the right panel. It is unclear what should be a good rational to choose the order of arguments in
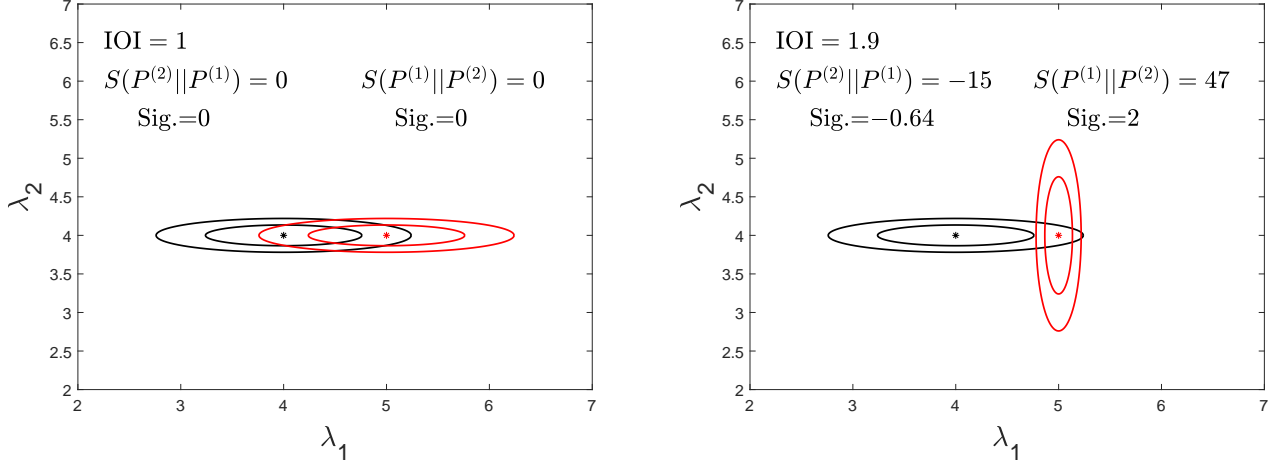
FIG. 8. Two sets of toy experiments described by Eqs. (52)-(53) (left panel) and Eqs. (54)-(55) (right panel). Experiment 1 is described by contours in black, and experiment two in red. The two experiments in the right panel seems to have a larger inconsistency, which is correctly described by the increase of IOI. Using different argument orders in the *surprise* leads to opposite conclusions about the inconsistency in the right panel.
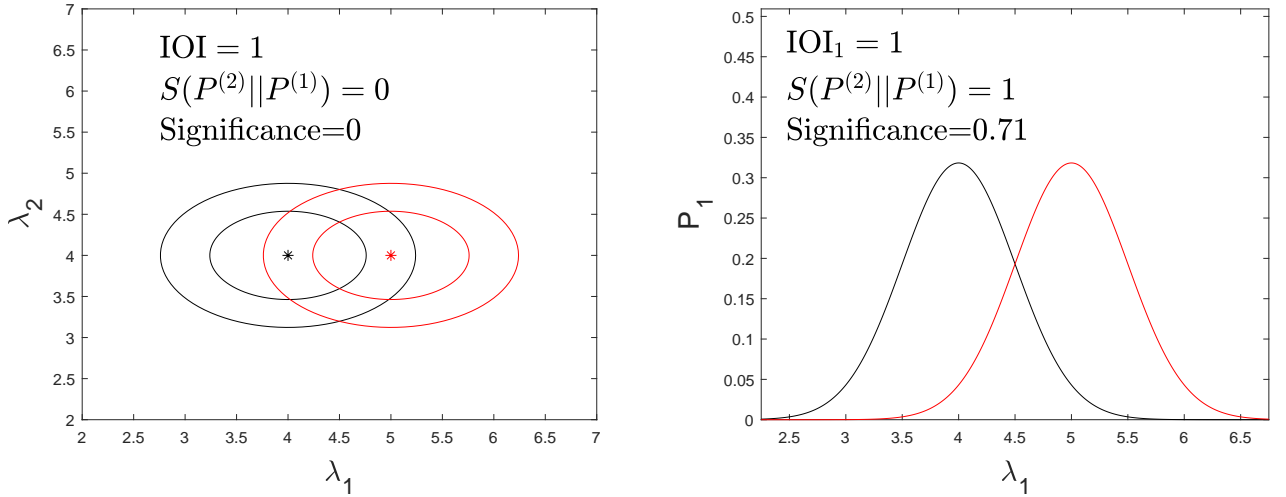

FIG. 9. The posterior probability contours of the two experiments described by (56) (left panel) and their marginalized 1D distribution over λ_2 (right panel). The *surprise* before marginalization is $S(\mathscr{P}^{(2)}||\mathscr{P}^{(1)}) = 0$, but after marginalization the *surprise* is $S_1 = 1$ with a significance 0.7. However, we argued in Sec. V that λ_2 should not be relevant to the inconsistency, a measure of the inconsistency should be preserved after marginalization over λ_2.

S. It could be argued that the choice should be physically motivated favoring the more trusted data set with less expected systematic uncertainties, thus with a prior preference of a given data set which could be viewed as consistent with a Bayesian statistics approach [69]. It remains unclear though what is the right choice if both data sets have their own significant systematic errors or when the inconsistency is rooted into a problem with the model or the theory itself.

Finally, the *surprise* and its significance are found to increase after parameter marginalization in, for example,

the following toy experiments:

$$\bar{\boldsymbol{\mu}}^{(2)} - \bar{\boldsymbol{\mu}}^{(1)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \boldsymbol{F}^{(1)} = \boldsymbol{F}^{(2)} = \begin{pmatrix} 4 & 0 \\ 0 & 8 \end{pmatrix}, \quad (56)$$

which are shown also in Fig. 9. We argued in Sec. V that λ_2 should not be relevant to the inconsistency, because the two experiments have the same mean of λ_2 and there is no correlation between λ_1 and λ_2. So after marginalizing over λ_2 a measure of inconsistency should be preserved, as is the case for IOI. As shown in Fig. 9, before marginalizing over λ_2, the *surprise* $S(\mathscr{P}^{(2)}||\mathscr{P}^{(1)})$ (and

its significance) is 0. After marginalizing over λ_2, it becomes 1 with a significance of 0.7. It can be argued this is a normal behaviour for the *surprise* because the results of λ_2 are more consistent than expected [69]. So marginalizing over λ_2 emphasises the inconsistency in λ_1 and that is reflected by the increase of the marginalized *surprise* [69]. But then when going from the right to the left panel, the inconsistencies as indicated by the values of the *surprise* and its significance on the left seem to go away. It is unclear if this is really the case since the inconsistency in λ_1 is still present.

E.   The Calibrated Evidence Ratio

Reference [70] introduced the calibrated evidence ratio (CER hereafter) by calibrating the *robustness* using its average. The level of CER can also be described by its significance. But differently from the case of the *surprise*, the average here is taken over a probability distribution of both the first and the second data $Q^{(1)}$ and $Q^{(2)}$, which is given by the joint evidence $E(Q^{(1)}, Q^{(2)})$,

$$\mathrm{CER} = \ln R - \langle \ln R \rangle_{Q_1, Q_2}, \qquad (57)$$

$$\sigma^2(R) = \left\langle \left(\ln R - \langle \ln R \rangle \right)^2 \right\rangle_{Q_1, Q_2}. \qquad (58)$$

In Gaussian cases, CER in N dimensions is given by

$$-\mathrm{CER} = \tfrac{1}{2}\bar{\delta}^T \bar{G}\bar{\delta} - \tfrac{N}{2}$$
$$\xrightarrow{\text{weak prior}} \tfrac{1}{2}\delta^T G \delta - \tfrac{N}{2} \qquad (59)$$
$$= \mathrm{IOI} - \tfrac{N}{2},$$

$$\sigma(R) = \sqrt{N/2}. \qquad (60)$$

So the opposite of CER is related to IOI in the Gaussian and weak prior limit by an additive constant, $-N/2$.

The CER measure did address some of the issues with its predecessors. It worth noting though that since IOI is positive definite, $-N/2$ is the minimum value of $-\mathrm{CER}$. Due to the term $-N/2$, $-\mathrm{CER}$ can be be more negative with higher dimensional paramter spaces. The significance defined as $-\mathrm{CER}/\sigma(R)$ goes as $-\sqrt{N/2}$ when IOI $= 0$. The term $-N/2$ (and $-\sqrt{N/2}$ for its significance) seems to suggest that a model with a higher dimension tends to have a smaller inconsistency. It can be argued that this is normal because a higher dimensional model allows larger scatter of the data, which leads to a larger expected $-\ln R$ [69]. Consider though the following situation: if two experiments have inconsistency on only one parameter (e.g., λ_1) and if all parameters are uncorrelated, a model with a higher dimension will have a smaller $-\mathrm{CER}$. This seems to underestimate the inconsistency that is due to one particular parameter.

It is also worth mentioning that from Eq. (59) we can see, in the Gaussian and weak prior limit, $-\mathrm{CER}$ has a close relation with the Akaike information criterion [72] for the following reason. The Akaike information criterion is to select a model with a smaller

AIC$= -2\ln\mathcal{L} + 2N$. For a model with N parameters, the corresponding splitting model has $2N$ parameters. The AIC difference between the combined and the split models is

$$\tfrac{1}{2}\Delta\mathrm{AIC} = \tfrac{1}{2}\mathrm{AIC}_{\mathrm{comb}} - \tfrac{1}{2}\mathrm{AIC}_{\mathrm{spl}}$$
$$= \mathrm{IOI} - \tfrac{1}{2}N \qquad (61)$$
$$= -\mathrm{CER}.$$

F.   Level of concordance

Analogous to the *robustness*, Ref. [71] most recently introduced the *level of concordance*, defined as $\exp\left[-\tfrac{1}{2}\mathcal{G}(Q^{(1)}, Q^{(2)})\right]$, to measure experimental (in)consistency. Instead of using Bayesian evidence ratio between the combined and the split models, $\mathcal{G}(Q^{(1)}, Q^{(2)})$ uses the deviance information criterion (DIC) (see Refs. [73, 74]), and it is defined as

$$\mathcal{G}(Q^{(1)}, Q^{(2)}) \equiv \mathrm{DIC}_{\mathrm{comb}} - \mathrm{DIC}_{\mathrm{spl}}, \qquad (62)$$

where DIC$= \chi^2_{min} + 2p_D$, and p_D is the effective number of parameters given by $p_D = \langle\chi^2\rangle - \chi^2_{min}$. Jeffreys' scale was used to to interpret this measure.

Since DIC approaches AIC in Gaussian limit (when parameters are well constrained) [74], $\mathcal{G}$ is equivalent to $-\mathrm{CER}$ (see Sec. X E for the connection between $-\mathrm{CER}$ and AIC) in Gaussian limit, i.e.,

$$\tfrac{1}{2}\mathcal{G}(Q^{(1)}, Q^{(2)}) = -\mathrm{CER} = \mathrm{IOI} - \tfrac{N}{2}. \quad \text{(Gaussian)} \quad (63)$$

Therefore, *level of concordance* measure seams to share the points mentioned above for the $-\mathrm{CER}$.

G.   Is there a connection between IOI, *Surprise* and $-\mathbf{CER}$

The definitions of the *robustness* $\ln(R)$, *tension* $\mathcal{T}$, *surprise* S, $-\mathrm{CER}$ and IOI are quite different. But we have seen in the previous sections that $\ln(R)$, $\mathcal{T}$ and $-\mathrm{CRE}$ are related to IOI by equations (31) & (32), (38) and (59). Is there a relation between the *surprise* S and IOI? We find that only in one dimension the significance of the *surprise* is simply related to IOI, and to the significance of $-\mathrm{CER}$.

In one dimension with Gaussian and weak prior, the *surprise* given by Eq. (50) reduces to

$$S(\mathscr{P}^{(2)}||\mathscr{P}^{(1)}) = \tfrac{1}{2}\delta^2/\sigma_{(1)}^2 - \tfrac{1}{2}(1 + \sigma_{(2)}^2/\sigma_{(1)}^2), \qquad (64)$$

the deviation of reduces to

$$\sigma(D(\mathscr{P}^{(2)}||\mathscr{P}^{(1)})) = \sqrt{\tfrac{1}{2}(1 + \sigma_{(2)}^2/\sigma_{(1)}^2)}, \qquad (65)$$

and the significance is given by

$$\text{significance of } S = \frac{S(\mathscr{P}^{(2)}||\mathscr{P}^{(1)})}{\sigma(D(\mathscr{P}^{(2)}||\mathscr{P}^{(1)}))}$$
$$= \sqrt{2}\left[\frac{1}{2}\frac{\delta^2}{(\sigma_{(1)}^2 + \sigma_{(2)}^2)} - \frac{1}{2}\right]. \qquad (66)$$

TABLE VI. Table of measures of discordance in the Gaussian and weak prior limit considered in this work. $\boldsymbol{\delta} = \boldsymbol{\mu^{(1)}} - \boldsymbol{\mu^{(2)}}$ and $\boldsymbol{G} = \left((\boldsymbol{L^{(1)}})^{-1} + (\boldsymbol{L^{(2)}})^{-1}\right)^{-1}$. We take the opposite of the *robustness*, the normalized *robustness* and CER to be the measures of inconsistency, since they are originally defined to be measures of consistency.

Quantities	Symbols	Relevant concept	Gaussian and weak prior limit
Index of inconsistency	IOI	$\Delta\chi^2_{(1)} + \Delta\chi^2_{(2)}$	$\frac{1}{2}\boldsymbol{\delta}^T\boldsymbol{G}\boldsymbol{\delta}$ (definition)
Robustness [61, 62]	$-\ln R$	Bayesian evidence ratio	$\text{IOI} + \frac{1}{2}\ln\left(\frac{|\mathbf{LP}|}{|\mathbf{L^{(1)}L^{(2)}}|}\right)$
Normalized *robustness* [62]	$-\ln R_N$	Bayesian evidence ratio	$\text{IOI} + \frac{1}{2}\ln\left(\frac{|\mathbf{L^{(1)}}+\mathbf{L^{(2)}}|}{|2\mathbf{L^{(2)}}|}\right)$
Tension [58]	$\ln\mathcal{T}$	Bayesian evidence ratio	IOI
Discordance [65]	σ_{eq}	Iso-likelihood contour	Algorithm provided by Eqs. (41)-(43)
Surprise [a] [64, 66, 67]	$S(\mathscr{P}^{(2)}\|\mathscr{P}^{(1)})$	Relative entropy	$\frac{1}{2}\boldsymbol{\delta}^T\boldsymbol{L^{(1)}}\boldsymbol{\delta} - \frac{1}{2}\text{tr}\left(\boldsymbol{I_N} + (\boldsymbol{L^{(2)}})^{-1}\boldsymbol{L^{(1)}}\right)$
and its deviation	$\sigma(D)$		$\sqrt{\frac{1}{2}\text{tr}\left[\left(\boldsymbol{I_N} + (\boldsymbol{L^{(2)}})^{-1}\boldsymbol{L^{(1)}}\right)^2\right]}$
Calibrated evid. ratio [70]	$-\text{CER}$	Bayesian evidence ratio	$\text{IOI} - N/2$
and its deviation	$\sigma(R)$		$\sqrt{N/2}$
Level of concordance [71]	$\mathcal{G}(\boldsymbol{Q^{(1)}},\boldsymbol{Q^{(2)}})$	Deviance information criterion	$\text{IOI} - N/2$

[a] The case of replacing data described in Ref. [66] is used in this work.

TABLE VII. Comparison of properties of the measures of discordance considered in this work in Gaussian and weak prior limit. In Gaussian and weak prior limit, $\ln\mathcal{T} = \text{IOI}$ and $\frac{1}{2}\mathcal{G}(\boldsymbol{Q^{(1)}},\boldsymbol{Q^{(2)}}) = -\text{CER}$. By saying whether or not a measure reflects the effects of marginalization, we mean if the measure can describe the hidden inconsistency after parameter marginalization. For measures like *robustness* and normalized *robustness* $-$CER, this column does not apply (N/A).

Quantities	Symmetric?	Follow inconsistency?	Give 0 if $\boldsymbol{\delta} = \mathbf{0}$?	Reflect effects of marginalization?	Proposed interpretation
IOI	✓	✓	✓	✓	Jeffreys' scales
Robustness	✓	✗	✗	N/A	Jeffreys' scales
Normalized $\ln R_N$	✗	✗	✗	✗	not specified
Discordance	✓	✓	✓	✗	compared to 100%
Surprise	✗	?[b]	✗	✗	bits	significance
$-$CER	✓	✓	✗	N/A	significance

[b] Due to the asymmetric behavior, if the *surprise* has the wrong choice of argument order, the trend will be incorrect.

In one dimension, the significance of the *surprise* becomes symmetric about two experiments. IOI in this case is given by $\frac{1}{2}\frac{\delta^2}{(\sigma^2_{(1)}+\sigma^2_{(2)})}$, and from (59) and (58) we can obtain the significance of $-$CER in this case as $\sqrt{2}\left(\text{IOI}-\frac{1}{2}\right)$, for $N = 1$. Therefore the three quantities are related by

$$\text{signif. of } S = \text{signif. of } (-\text{CER}) = \sqrt{2}\left(\text{IOI} - \tfrac{1}{2}\right). \quad (67)$$

In N dimension, due to the connection of $-$CER to AIC, the relation between IOI and $-$CER (and its significance) remains simple in the Gaussian and weak prior limit, and it is given by Eq. (59), and

$$\text{significance of } (-\text{CER}) = \sqrt{\tfrac{2}{N}}\left(\text{IOI} - \tfrac{N}{2}\right). \quad (68)$$

But it is not obvious to see the relation between IOI and *surprise*. The *surprise* and its significance become asymmetric for two- or multiple-dimensional parameter spaces, while IOI is always symmetrical.

Finally, we summarize the measures of inconsistency surveyed in this section in Table VI and Table VII.

XI. APPLICATION OF IOI TO COSMOLOGICAL EXPERIMENTS: GEOMETRY VS GROWTH

To test the internal consistency of the ΛCDM and wCDM models, we fit each of them separately to a set of geometry experiments and a set of growth experiments. This is partly motivated by the fact that different theories (e.g. modified gravity) have relations between the cosmic expansion background and the large scale structure growth, e.g. [10]. We use the publicly available Markov Chain Monte Carlo code COSMOMC [60] for parameter constraints. Note that, in this work, we will not use IOI for multiple experiments [Ref. (12)] to study the full experimental inconsistency among them. Rather, we will treat the geometry experiments as one set of experiments and the growth experiments as the other set, and then apply IOI to measure the (in)consistency between the two set of experiments [Eq. (9)].

TABLE VIII. Summary of geometry and growth experiments/data used in testing the model consistency.

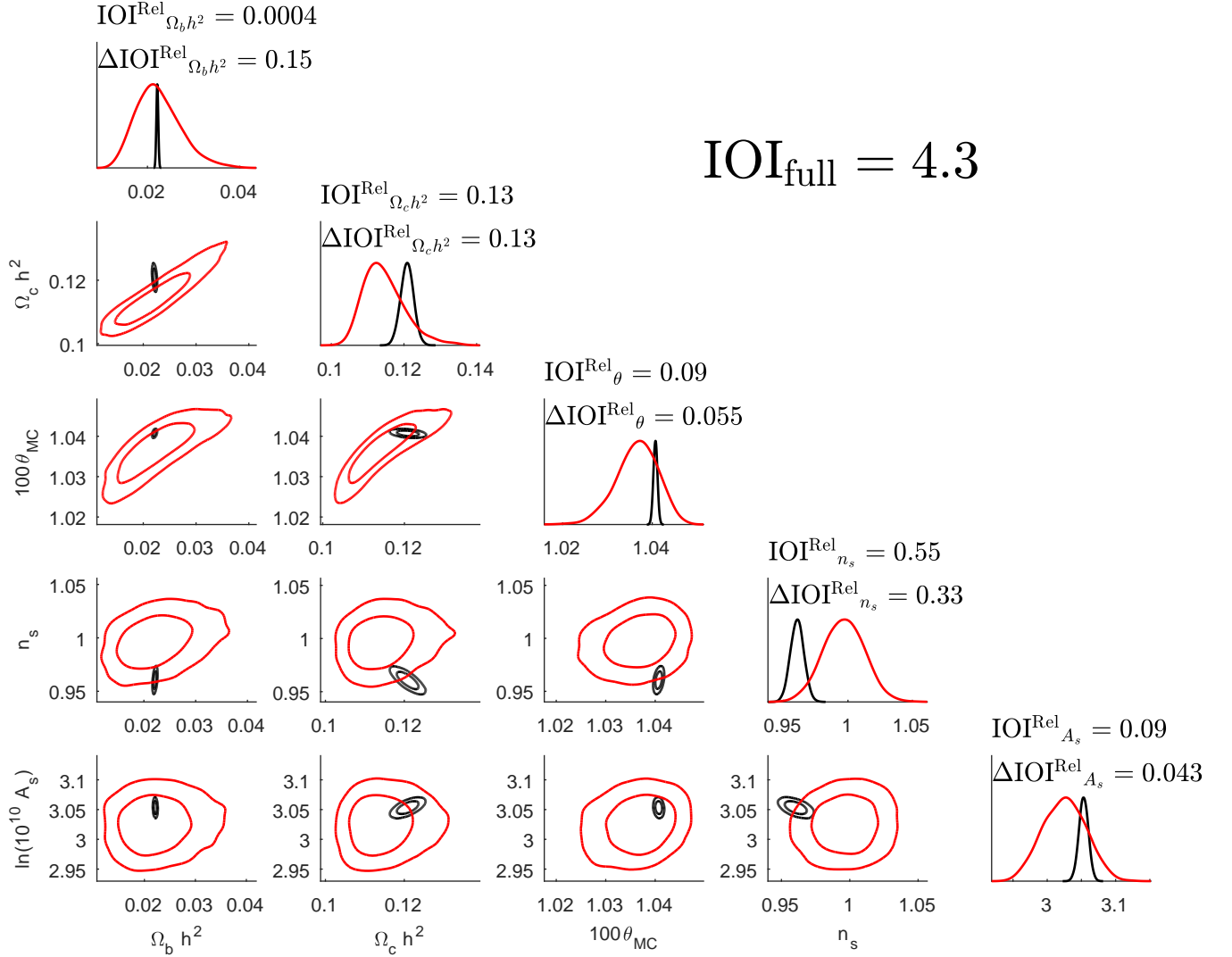| Geometry		Growth	
Supernovae Type Ia [21]		Planck 2015 low-ℓ CMB temperature and polarization [23]	
BAO	Six Degree Field Glactic Survey (6dF) ($z_{eff} = 0.106$) [22]	Planck 2015 CMB lensing [24]	
	SDSS main galaxy sample (MGS) ($z_{eff} = 0.15$) [75]	Sunyaev–Zel'dovich effect [25]	
	SDSS quasar-Lyman-α forest ($z_{eff} = 2.34$) [18]	CFHTlens galaxy weak lensing [19]	
Planck 2015 High-ℓ CMB temperature [23]		RSD	WiggleZ Dark Energy Survey [26, 27]
			SDSS DR12 CMASS and LOWZ catalogs [28]

FIG. 10. The marginalized 1-D probability distributions and 2-D likelihood contour plots for ΛCDM model constrained by the geometry versus the growth data sets. Black for the geometry set and red for the growth set. The two relative measures of one-parameter inconsistency ($\mathrm{IOI}_i^{\mathrm{Rel}}$ and $\Delta\mathrm{IOI}_i^{\mathrm{Rel}}$, see Eq. (18) for their definitions) are shown on top of each 1-D marginalized plot. *While these relative IOI measures can indicate the parameter(s) most associated with the inconsistency, **the important number here is the value of the full IOI** shown at the upper right corner of the figure.* Note that summation of all individual $\mathrm{IOI}_i^{\mathrm{Rel}}$'s or $\Delta\mathrm{IOI}_i^{\mathrm{Rel}}$'s is generally not 1. The full IOI between the two data sets is 4.34 and indicates a moderate inconsistency on Jeffreys' scales (see Table III).

TABLE IX. Measures of inconsistency between geometry and growth experiments for ΛCDM and wCDM models. Measures of inconsistency are obtained by using the Gaussian and weak prior limits listed in Table VI. The weak prior limit of the *robustness* diverges, so we do not calculate it in this table. Quantities $\ln \mathcal{T}$ and $\mathcal{G}(\boldsymbol{Q^{(1)}}, \boldsymbol{Q^{(2)}})$ are not listed here, since they equal IOI and $-$CER respectively in the Gaussian and weak prior limit. The IOI between the two data sets is 4.34 and 5.41 for the ΛCDM and wCDM, respectively. These indicate moderate and strong inconsistency, respectively, on Jeffreys' scales. The other measures also indicate an overall moderate to strong (except $S(\mathrm{gm}||\mathrm{gr})$).

Measures	IOI	$-\ln R_N$	σ_{eq}	$S(\mathrm{gr}||\mathrm{gm})^{\mathrm{a}}$	Significance	$S(\mathrm{gm}||\mathrm{gr})$	Sig.	$-$CER	Sig.
ΛCDM	4.34	11.52	56.9%	-340.20	-0.60	3.06	1.46	1.84	1.17
wCDM	5.41	12.96	59.3%	-432.20	-0.63	4.05	1.51	2.41	1.39

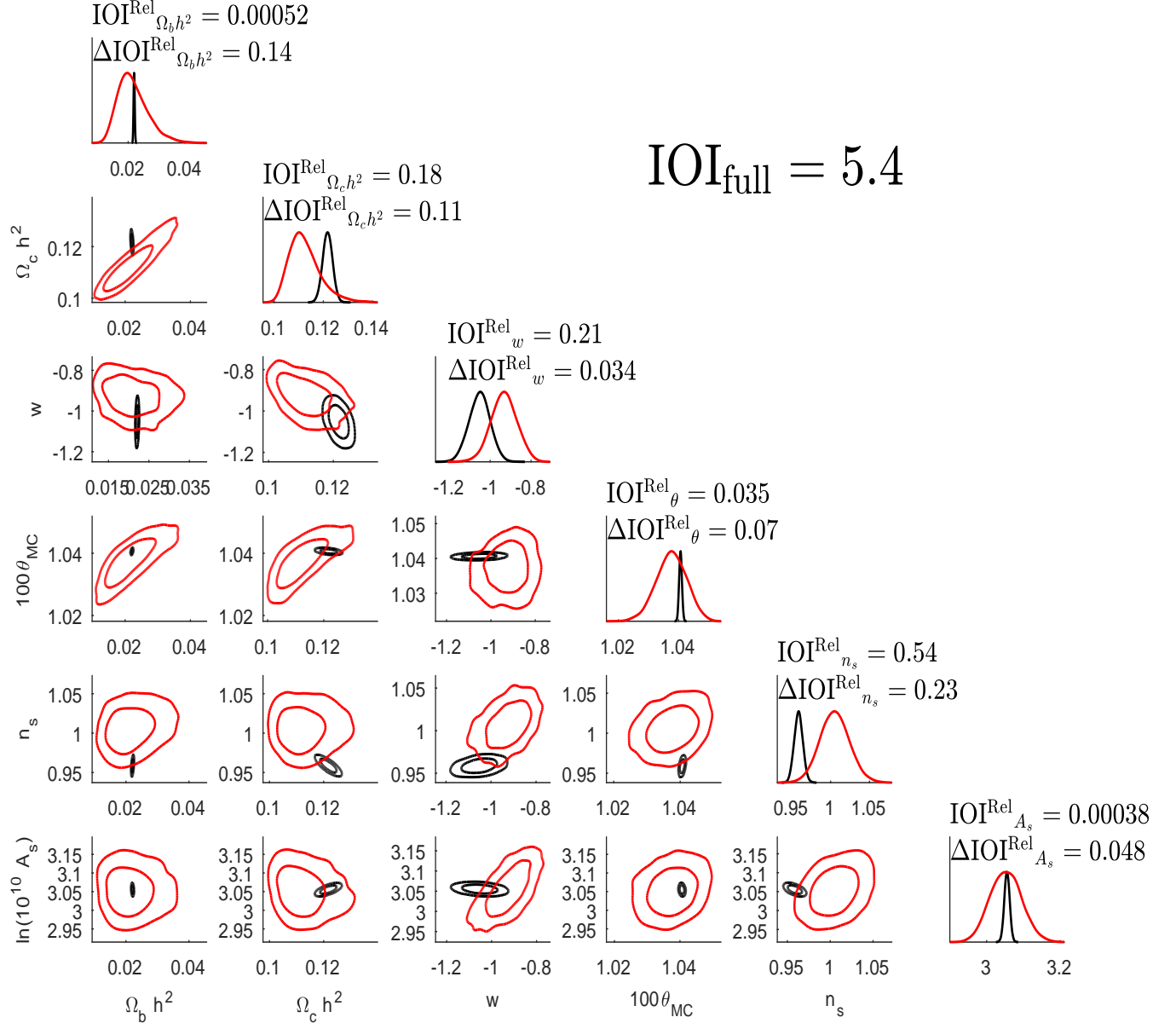$^{\mathrm{a}}$ 'gm' = 'geometry', 'gr' = 'growth'.

FIG. 11. Marginalized 1-D probability distributions and 2-D likelihood contour plots in a triangular displacing diagram for wCDM model constrained by the geometry and growth sets of experiments. The structure of this figure is the same as Fig. 10. **Again, the important number here is the value of the full IOI** shown at the upper right corner of the figure. Note that summation of all individual $\mathrm{IOI}_i^{\mathrm{Rel}}$'s or $\Delta\mathrm{IOI}_i^{\mathrm{Rel}}$'s is generally not 1. The full IOI between the two data sets is 5.34 for wCDM and indicates a strong inconsistency on Jeffreys' scales (see Table III).

A. Data sets

For the geometry set of experiments, we use the 740 type Ia supernovae catalog compiled in Ref. [21], the Baryon Acoustic Oscillation (BAO) standard ruler from the Six Degree Field Glactic Survey (6dF) [22], the main galaxy sample (MGS) of data release seven (DR7) from the Sloan Digital Sky Survey (SDSS) and the SDSS quasar-Lyman-α forest [18], and the high-ℓ CMB temperature data from Planck 2015 [23]. For the growth set of experiments, we use the low-ℓ CMB temperature and polarization data form Planck 2015 [24], CMB lensing [24], thermal Sunyaev–Zel'dovich effect [25], galaxy weak lensing cosmic shear from the Canada France Hawaii Lensing Survey (CFHTlens) [19], and the redshift space distortion from the WiggleZ Dark Energy Survey (WiggleZ MPK) [26, 27] and the SDSS DR12 CMASS and LOWZ catalogs [28]. Those two sets of experiments are summarized in Table VIII. Note that we put high-ℓ CMB temperature data in the geometry set as in Ref. [13], since its acoustic peaks strongly constrain the expansion history of the Universe.

From Table VIII we can see that there are some overlaps between the data sets used in this work and those used in Refs. [12, 13]. But there are some differences. First, for the RSD data from SDSS, while Ref. [13] separately used the measurements of Hr_s and $f\sigma_8$ as geometry and growth sets, we use them all in the growth set because these two measurements (as well as D_A/rs) are correlated. Second, we use the full data catalog of RSD from WiggleZ, while Ref. [13] used its compressed information on BAO and growth. Third, we use the combined Lyman-α BAO data at $z_{eff} = 2.34$ from Lyα forest autocorrelation and the quasar-Lyα cross-correlation [18], while Ref. [13] only used the data from cross-correlation.

B. Results from application to data: a moderate to strong inconsistency

For the ΛCDM model, we use the following parameters: $\Omega_b h^2$, $\Omega_c h^2$, θ, $\ln(10^{10} A_s)$, n_s, and τ (for the (wCDM) we addd w). We fit each model separately into the geometry set and growth set of experiments listed in Table VIII. Since the geometry experiments poorly constrain τ, we set it to the mean value obtained in Ref. [76] $\tau = 0.058$, while we allow it to vary in the fitting of the growth experiments. The constraints from the two sets of experiments are shown in Fig. 10 and 11. To quantify the inconsistency between the geometry and the growth experiments, we calculate IOI for the full parameter space (marginalized over τ). Additionally, to compare IOI to the other measures of inconsistency, we also calculate other measures in the Gaussian and weak prior limit, like the normalized *robustness* $\ln R_N$, the *discordance*, the *surprise* $S(\text{grow}||\text{geom})$ and $S(\text{geom}||\text{grow})$ (and their significance) and the $-$CER (and its significance). All results of inconsistency between the geometry

and the growth experiments for the ΛCDM and wCDM models are listed in Table IX.

We can see from Table IX that all the measures (except for $S(\text{grow}||\text{geom})$) overall suggest there are inconsistencies between geometry and growth data sets for ΛCDM and wCDM models. However, different measures provide different levels of inconsistency. Take for example the ΛCDM model. IOI suggests a moderate experimental inconsistency (in fact not too far from a significant one on Jeffreys' scales). We note that the normalized *robustness* gives a inconsistency much higher than other measures and ranked as very strong (decisive) if we insist to interpret it in Jeffreys' scales. However, this high inconsistency using the normalized *robustness* is artificial as we explain. The geometry experiments give much stronger constraints than the growth ones, so that $|\boldsymbol{L^{geom}}| \gg |\boldsymbol{L^{grow}}|$, $|\boldsymbol{L^{geom}} + \boldsymbol{L^{grow}}| \simeq |\boldsymbol{L^{geom}}|$ and the term $-\ln(\frac{2|\boldsymbol{L^{geom}}\boldsymbol{L^{grow}}|}{\boldsymbol{L^{geom}}(|\boldsymbol{L^{geom}} + \boldsymbol{L^{grow}}|)}) \gg 1$. Therefore the apparent inconsistency suggested by $\ln R_N$ is only due to the unmatched constraining powers of the two data sets. The quantity $-$CER is simply related to IOI by Eq. (59). It is suggested in Ref. [70] that $-$CER should be interpreted in terms of its significance. The significance of $-$CER gives a confidence level of inconsistency of 1.17σ. For the *surprise*, different orders of experiments in its argument lead to different results. While $S(\text{geom}||\text{grow})$ is positive and finite, $S(\text{grow}||\text{geom})$ is negative and large. The significance of the *surprise* also has such asymmetric problem. While the significance of $S(\text{grow}||\text{geom})$ gives no inconsistency, the significance of $S(\text{grow}||\text{geom})$ suggests 1.46σ inconsistency, which seems more consistent with the one suggested by IOI and Fig. 10. The value of *discordance* is 56.9%. Compared with 100%, the *discordance* also seems to suggest some inconsistencies.

Comparing the results between ΛCDM and wCDM models, we can see that all measures (except for $S(\text{grow}||\text{geom})$) suggest that wCDM model has a larger inconsistency regarding the geometry and the growth experiments. In the language of IOI in nested models (see Sec. VIII), we say that wCDM model is a less consistent extension to ΛCDM model when comparing growth versus geometry. The result that wCDM model has a larger experimental inconsistency is different from what was found in Ref. [58]. This may not be surprising, because Ref [58] were comparing Planck with local measurement on the Hubble constant and the age of the Universe, which are different experiments from what we are considering. The values of inconsistency measures are experiment dependent. Reference [77] also found that wCDM model is more favored than ΛCDM model when Planck, local measurement of H_0 and BAO datasets are analyzed jointly. Indeed, when IOI is applied to investigate the inconsistency between Planck and local measurement on H_0, we find IOI= 4.0 for ΛCDM and 0.59 for wCDM. These numbers are consistent with those provided in Table 2 in Ref. [58] (recall that IOI and the *tension* are the same in Gaussian limit). So wCDM does have a smaller IOI in this case. However, the reason might be simply

that Planck data has a poor constraint on H_0 if w is allowed to vary; $\sigma_{H_0} \sim 10$ km/s/Mpc. Recall that one of the three factors affecting IOI is the constraint volume: a larger constraint volume leads to a smaller IOI.

Finally in this subsection, we briefly investigate the roles of CFHTlens data and SDSS RSD data on the value of IOI for ΛCDM model. If we remove CFHTlens data alone, IOI drops from 4.3 to 3.2. If we remove SDSS RSD data alone, IOI drops from 4.3 to 3.8. Although the CFHTlenS data seems to cause more inconsistency, neither removing the CFHTlenS data nor the SDSS RSD data dramatically drops IOI. Therefore, the inconsistency between the geometry and the growth experiments does not seem to be caused by one particular data set. Furthermore, if we remove both CFHTlens and SDSS RSD data, IOI drops from 4.3 to 2.5, but a weak inconsistency still persists. In a follow up work, we will systematically explore the internal and cross inconsistencies between a number of data sets and more applications of IOI.

C. Relative drop of IOI and relative residual IOI in the geometry vs growth application

We have shown that parameter marginalization can hide experimental inconsistency in Sec. V. This is reflected by the drop of IOI after marginalizing over each parameter, which is shown in Tables XI and XI. The fact that IOI never increases offers us two ways to see the experimental inconsistency due to a particular parameter, which we apply here to the geometry and growth experiments.

The two measures of one-parameter inconsistency IOI_i and $\Delta\mathrm{IOI}_i$ described in Sec. VI can help us zoom into the parameter space, to see experimental inconsistency due to each parameter. For a given model, if two experiments do not have inconsistency due to λ_i, both IOI_i and $\Delta\mathrm{IOI}_i$ are small in Jeffreys' scales. But in some cases, even thought IOI_i and $\Delta\mathrm{IOI}_i$ are small in Jeffreys' scales, they are not negligible compared to the full IOI. So more conveniently, we can use the relative quantities $\mathrm{IOI}_i^{\mathrm{Rel}}$ and $\Delta\mathrm{IOI}_i^{\mathrm{Rel}}$, and compare them to unity to see the role of each parameter on experimental inconsistency. Only when both $\mathrm{IOI}_i^{\mathrm{Rel}}$ and $\Delta\mathrm{IOI}_i^{\mathrm{Rel}}$ are much smaller than 1, can we safely marginalize over λ_i to analyze the full inconsistency. We listed the two measures and their relative measures in Tables X and XI for ΛCDM model and wCDM model respectively.

Take the wCDM model for example, the geometry and the growth experiments have a significant (full) inconsistency (IOI=5.4). For individual parameters, the 1-D marginalized plots in Fig. 11 seems to show little inconsistency in $\Omega_b h^2$, θ and A_s, and the corresponding IOI_i's (and $\mathrm{IOI}_i^{\mathrm{Rel}}$) are small. But it does not mean it is safe to marginalize over these three parameters. For example, $\Delta\mathrm{IOI}_{bh^2}^{\mathrm{Rel}}$ is not much smaller than 1, so marginalizing over $\Omega_b h^2$ will hide some experimental inconsistency due to its correlation with the other parameters (specified by

$C = C^{(1)} + C^{(2)}$). On the other hand, since both $\mathrm{IOI}_{A_s}^{\mathrm{Rel}}$ and $\Delta\mathrm{IOI}_{A_s}^{\mathrm{Rel}}$ are small, the geometry and the growth experiments give consistent result on A_s, and it is safe to marginalize over it to study the full inconsistency. Similarly, it is also safe to marginalize over θ. For the ΛCDM model, from Table X it seems inappropriate to marginalize any parameter to analyze inconsistency, although θ and A_s seem to be least responsible for the inconsistency.

Which parameter have the most inconsistent results given by the geometry and the growth experiments? From Tables X and XI, we can see that for both ΛCDM model and wCDM model, the results of n_s are the most inconsistent. Both $\mathrm{IOI}_{n_s}^{\mathrm{Rel}}$ and $\Delta\mathrm{IOI}_{n_s}^{\mathrm{Rel}}$ are large. The next inconsistent parameter should be $\Omega_c h^2$, but the inconsistency caused by it is quite small compared to that due to n_s.

The definition of $\Delta\mathrm{IOI}_i$ is based on the fact that IOI never increases after a parameter marginalization. This drop of IOI represents the hidden inconsistency due to a parameter after we marginalize over it. This is a unique feature for IOI. To demonstrate it, we show as examples the drops of σ_{eq}, $\Delta S(\mathrm{geom}||\mathrm{grow})$ and its significance after marginalizing over each parameter in the last three rows of Tables X and XI. We can see that their drops are not always positive, suggesting that they are not able to represent the hidden inconsistency after a parameter marginalization.

It is however important to emphasize that although combining $\Delta\mathrm{IOI}_i$'s and IOI_i's can give us more insights into the inconsistency due to each parameter, they cannot fully describe the experimental inconsistency. It is the full IOI (recall this is not the summation of $\Delta\mathrm{IOI}_i$'s or IOI_i's) for the whole parameter space that should be calculated and used.

XII. SUMMARY

The ΛCDM model is considered to be the standard and concordance model in cosmology, but different experiments have been yielding constraints of parameters that do no perfectly agree with each other. Those inconsistencies could either be caused by systematics or the break down of the underlying model and theory. It is timely and important to develop tools to quantify the degree of inconsistency between different experiments.

We defined a moment-based measure that we call index of inconsistency (IOI). We suggested to interpret IOI using Jeffreys' scales listed in Table III. We showed that IOI tracks well the separation of the means, the volume of covariance ellipsoids, and the orientations of those ellipsoids. We discussed criteria that must be fulfilled by measures of inconsistency. For example, the numerical value of the measure must follow the graphical inconsistency of likelihood contour plots, represent experiments with definite consistency, and correctly reflect the effect of parameter marginalization. We find that IOI satisfies all those criteria. We provided an eigen-mode decompo-

TABLE X. Two measures of one-parameter inconsistency for the ΛCDM model. The first one is the residual IOI_i (and the *relative* residual between parentheses) for λ_i after marginalizing over all the other parameters. The second one is the inconsistency drop ΔIOI_i (and the *relative* drop between parentheses) after marginalizing over λ_i. Both $\text{IOI}_i^{\text{Rel}}$ and $\Delta\text{IOI}_i^{\text{Rel}}$ need to be much smaller than unity for us to safely marginalize over λ_i to analyze inconsistency. Four different situations are listed in Table XII regarding different values of IOI_i and ΔIOI_i. It seems that none of the parameters can be safely marginalized over, although θ and A_s seem to be least responsible for the inconsistency. The geometry and the growth experiments give quite inconsistent result on n_s since both $\text{IOI}_{n_s}^{\text{Ref}}$ and $\Delta\text{IOI}_{n_s}^{\text{Ref}}$ are large. The bottom three rows: drops of σ_{eq}, $S(\text{geom}||\text{grow})$ and its significance evaluated after marginalizing over one parameter. Note that their drops are not always positive after parameter marginalization as one would expect.

ΛCDM parameters	$\Omega_b h^2$	$\Omega_c h^2$	θ	A_s	n_s
IOI_i ($\text{IOI}_i^{\text{Rel}}$)	0.0017 (0.0004)	0.55 (0.13)	0.39 (0.09)	0.39 (0.09)	2.4 (0.55)
ΔIOI_i ($\Delta\text{IOI}_i^{\text{Rel}}$)	0.64 (0.15)	0.55 (0.13)	0.24 (0.06)	0.19 (0.04)	1.5 (0.33)
$\Delta\sigma_{eq}$	-10.7%	-14.5%	-6.6%	-13.5%	10.5%
$\Delta S(\text{geom}||\text{grow})$	0.97	1.03	-0.7	-0.13	1.05
Δ(significance of S)	0.62	0.47	-0.32	0.19	0.81

TABLE XI. Two measures of one-parameter inconsistency for the wCDM model. This is similar to Table X. For the wCDM model, it seems that θ and A_s are least responsible for the inconsistency, and we can marginalize over them. The geometry and the growth experiments give quite inconsistent result on n_s, as in the case of ΛCDM model. The bottom three rows are similar to those in Table X.

wCDM parameters	$\Omega_b h^2$	$\Omega_c h^2$	θ	w	A_s	n_s
IOI_i ($\text{IOI}_i^{\text{Rel}}$)	0.0028 (0.0005)	1.00 (0.18)	0.19 (0.035)	1.12 (0.21)	0.0021 (0.0004)	2.93 (0.54)
ΔIOI_i ($\Delta\text{IOI}_i^{\text{Rel}}$)	0.77 (0.14)	0.58 (0.11)	0.38 (0.07)	0.18 (0.034)	0.26 (0.048)	1.27 (0.23)
$\Delta\sigma_{eq}$	-4.2%	-7.1%	-2.0%	-8.1%	-8.7%	11.4%
$\Delta S(\text{geom}||\text{grow})$	1.39	0.66	-0.57	-0.63	0.65	0.2
Δ(significance of S)	0.41	0.03	-0.4	-0.81	0.07	-0.01

sition of IOI and also discussed the application of IOI for nested models.

We showed how marginalizing over a parameter can hide inconsistency between experiments when present. This is because every parameter can induce its own experimental inconsistency and also has correlations with other parameters. In fact, even if all marginalized distribution plots look consistent for two experiments, they might still be inconsistent. These facts show that descriptions of experimental inconsistency involving parameter marginalization is not accurate. We provided an analytical proof that IOI never increases whenever a parameter marginalization is performed but rather decreases in most cases. This work seems to be the first to quantify the loss of information on cosmological experimental inconsistency due to parameter marginalization.

In order to zoom on the inconsistency residing in particular parameter, we defined the relative drop in IOI and the relative residual IOI for a given parameter, $\text{IOI}_i^{\text{Rel}} = \frac{\text{IOI}_i}{\text{IOI}}$ and $\Delta\text{IOI}_i^{\text{Rel}} = \frac{\Delta\text{IOI}_i}{\text{IOI}}$. We explained that only when both relative measures are small one can safely marginalize over such a parameter when considering inconsistency measures. We also explained that in order to correctly quantify the inconsistency we should use the full IOI of all parameters in a model that are commonly constrained by two experiments.

We provided a comparative survey of other measures of inconsistency and their relationships, if any, with IOI.

We then applied IOI and other measures in the Gaussian and weak prior limit to quantify the inconsistency of the ΛCDM and wCDM models between geometry and growth experiments. As discussed in the previous section, IOI indicates a moderate to strong inconsistency between the two data sets when the ΛCDM and wCDM models are used with values of 4.3 an 5.4 on Jeffreys' scales. Overall, although different measures use different scales to quantity inconsistency, most of the measures (except for $S(\text{geom}||\text{grow})$) also suggest that there are moderate to strong experimental inconsistencies in ΛCDM and wCDM models with the wCDM model having a larger inconsistency between geometry and growth experiments. As we discussed in the previous section, this comparison and its results are different from two other works that compared Planck with local measurement of the Hubble constant and the age of the Universe, finding that wCDM brings some reconciliation. As we explained, the values of inconsistency measures are experiment dependent, as they should.

Finally, we used the IOI measure and found that CFTlenS data, SDSS RSD data and other growth data all contribute to the inconsistency. So this seems to indicate indicate that the inconsistency can not be associated with a single data set. In future work, we will explore in more detail the role of each data set in the experimental

inconsistencies.

Appendix A: Derivations and proofs

This appendix provides the calculations and derivations used in this work, including those used in the definition of IOI and its relationship to other measures of inconsistency.

1. IOI for arbitrary numbers of experiments

We consider combination of likelihoods that do not correlate with each other. A Gaussian likelihood for the ith experiment reads

$$
\begin{aligned}
\mathcal{L}^{(i)} &= \exp\left[-\tfrac{1}{2}\chi^2_{(i)}(\boldsymbol{\lambda})\right] \\
&= \mathcal{L}^{(i)}_{max}\exp\left[-\tfrac{1}{2}\Delta\chi^2_{(i)}(\boldsymbol{\lambda})\right] \\
&= \mathcal{L}^{(i)}_{max}\exp\left[-\tfrac{1}{2}(\boldsymbol{\lambda}-\boldsymbol{\mu^{(i)}})^T\boldsymbol{L^{(i)}}(\boldsymbol{\lambda}-\boldsymbol{\mu^{(i)}})\right].
\end{aligned}
\tag{A1}
$$

For arbitrary number of experiments, the likelihood of a joint analysis is defined as

$$
\begin{aligned}
\mathcal{L} &= \prod_i \mathcal{L}^{(i)} \\
&= \left(\prod_i \mathcal{L}^{(i)}_{max}\right)\exp\left[\sum_i -\frac{1}{2}(\boldsymbol{\lambda}-\boldsymbol{\mu^{(i)}})^T\boldsymbol{L^{(i)}}(\boldsymbol{\lambda}-\boldsymbol{\mu^{(i)}})\right].
\end{aligned}
\tag{A2}
$$

Note that $\boldsymbol{\mu^{(i)}}$ is the mean of the ith likelihood, not the ith component of $\boldsymbol{\mu}$. The sum in Eq. (A2) can be reduced to

$$
\begin{aligned}
&\sum_i (\boldsymbol{\lambda}-\boldsymbol{\mu^{(i)}})^T\boldsymbol{L^{(i)}}(\boldsymbol{\lambda}-\boldsymbol{\mu^{(i)}}) \\
&= \boldsymbol{\lambda}^T\boldsymbol{L}\boldsymbol{\lambda} - 2\boldsymbol{\mu}^T\boldsymbol{L}\boldsymbol{\lambda} + \sum_i \boldsymbol{\mu^{(i)}}^{\,T}\boldsymbol{L^{(i)}}\boldsymbol{\mu^{(i)}} \\
&= (\boldsymbol{\lambda}-\boldsymbol{\mu})^T\boldsymbol{L}(\boldsymbol{\lambda}-\boldsymbol{\mu}) + \sum_i \boldsymbol{\mu^{(i)}}^{\,T}\boldsymbol{L^{(i)}}\boldsymbol{\mu^{(i)}} - \boldsymbol{\mu}^T\boldsymbol{L}\boldsymbol{\mu},
\end{aligned}
\tag{A3}
$$

where,

$$
\boldsymbol{L} \equiv \sum \boldsymbol{L^{(i)}},
\tag{A4}
$$

$$
\boldsymbol{\mu} \equiv \boldsymbol{L}^{-1}\sum \boldsymbol{L^{(i)}}\boldsymbol{\mu^{(i)}},
\tag{A5}
$$

and we have used the fact that $\left(\boldsymbol{L^{(i)}}\right)^T = \boldsymbol{L^{(i)}}$. Since the last two term do not depend on $\boldsymbol{\lambda}$, we can combine them with $\prod \mathcal{L}^{(i)}_{max}$, and Eq. (A2) becomes

$$
\mathcal{L} = \mathcal{L}_{max}\exp\left[-\frac{1}{2}(\boldsymbol{\lambda}-\boldsymbol{\mu})^T\boldsymbol{L}(\boldsymbol{\lambda}-\boldsymbol{\mu})\right],
\tag{A6}
$$

where,

$$
\mathcal{L}_{max} = \left(\prod_i \mathcal{L}^{(i)}_{max}\right)\exp\left[-\frac{1}{2}\left(\sum_i \boldsymbol{\mu^{(i)}}^{\,T}\boldsymbol{L^{(i)}}\boldsymbol{\mu^{(i)}} - \boldsymbol{\mu}^T\boldsymbol{L}\boldsymbol{\mu}\right)\right].
\tag{A7}
$$

So the likelihood of a joint analysis is also Gaussian [Eq. (A6)], with its mean, Fisher matrix and maximum value given by Eqs. (A4) - (A7). The above result is also applicable to the case where one of the likelihoods is replaced by a Gaussian prior, where one of the Likelihoods is replaced by a prior with $\mathcal{L}^{(i)}_{max} = \mathcal{P}_{max} = \frac{|\boldsymbol{P}|^{1/2}}{(2\pi)^{N/2}}$ and set $\boldsymbol{L^{(i)}} = P$.

Note that we have defined IOI as

$$
\frac{1}{N}\sum_i \Delta\chi^2_{(i)}(\boldsymbol{\mu}) \xrightarrow{\text{Gaussian}} \text{IOI}.
\tag{A8}
$$

The maximum value of the joint likelihood $\mathcal{L}$ is

$$
\begin{aligned}
\mathcal{L}_{max} &= \mathcal{L}(\boldsymbol{\mu}) \\
&= \left(\prod \mathcal{L}^{(i)}_{max}\right)\exp\left(-\frac{1}{2}\sum_i \Delta\chi^2_{(i)}(\boldsymbol{\mu})\right) \\
&\xrightarrow{\text{Gaussian}} \left(\prod \mathcal{L}^{(i)}_{max}\right)\exp\left(-\frac{N}{2}\text{IOI}\right),
\end{aligned}
\tag{A9}
$$

By comparing Eq. (A7) and Eq. (A9) we have

$$
\text{IOI} \equiv \frac{1}{N}\left(\sum_i \boldsymbol{\mu^{(i)}}^{\,T}\boldsymbol{L^{(i)}}\boldsymbol{\mu^{(i)}} - \boldsymbol{\mu}^T\boldsymbol{L}\boldsymbol{\mu}\right),
\tag{A10}
$$

for multiple experiments.

It is important to emphasize that, $\frac{1}{N}\sum\Delta\chi^2_{(i)}$ is only a motivation for our definition of IOI. It is true that IOI and $\frac{1}{N}\sum\Delta\chi^2_{(i)}$ are the same for Gaussian likelihoods, but for non-Gaussian cases, we define IOI as a moment-based quantity [Eq. (A10)], with $\boldsymbol{L^{(i)}}$ given by the inverse of the covariance matrix and $\boldsymbol{\mu^{(i)}}$ given by the mean of the ith likelihood.

2. Special cases: two likelihoods

More often we will compare two experiments to see their inconsistency. For the case of two experiments ($i = 1, 2$), and the right-hand side of Eq. (A10) becomes

$$\sum_i \boldsymbol{\mu}^{(i)\,T} \boldsymbol{L}^{(i)} \boldsymbol{\mu}^{(i)} - \boldsymbol{\mu}^T \boldsymbol{L} \boldsymbol{\mu}$$

$$= \boldsymbol{\mu}^{(1)\,T} \boldsymbol{L}^{(1)} \boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)\,T} \boldsymbol{L}^{(2)} \boldsymbol{\mu}^{(2)} - (\boldsymbol{\mu}^{(1)\,T} \boldsymbol{L}^{(1)} + \boldsymbol{\mu}^{(2)\,T} \boldsymbol{L}^{(2)}) \boldsymbol{L}^{-1} \boldsymbol{L} \, \boldsymbol{L}^{-1} (\boldsymbol{L}^{(1)} \boldsymbol{\mu}^{(1)} + \boldsymbol{L}^{(2)} \boldsymbol{\mu}^{(2)})$$

$$= \boldsymbol{\mu}^{(1)\,T} \boldsymbol{L}^{(1)} \boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)\,T} \boldsymbol{L}^{(2)} \boldsymbol{\mu}^{(2)} - (\boldsymbol{\mu}^{(1)\,T} \boldsymbol{L} - \boldsymbol{\delta}^T \boldsymbol{L}^{(2)}) \boldsymbol{L}^{-1} \boldsymbol{L} \, \boldsymbol{L}^{-1} (\boldsymbol{L}^{(1)} \boldsymbol{\delta} + \boldsymbol{L} \boldsymbol{\mu}^{(2)}) \,,$$

$$= \boldsymbol{\mu}^{(1)\,T} \boldsymbol{L}^{(1)} \boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(1)\,T} \boldsymbol{L}^{(1)} \boldsymbol{\delta} + \boldsymbol{\mu}^{(2)\,T} \boldsymbol{L}^{(2)} \boldsymbol{\mu}^{(2)} + \boldsymbol{\delta}^T \boldsymbol{L}^{(2)} \boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)\,T} \boldsymbol{L} \boldsymbol{\mu}^{(2)} + \boldsymbol{\delta}^T \boldsymbol{L}^{(2)} \boldsymbol{L}^{-1} \boldsymbol{L}^{(1)} \boldsymbol{\delta}$$

$$= \boldsymbol{\mu}^{(1)\,T} \boldsymbol{L}^{(1)} \boldsymbol{\mu}^{(2)} + \boldsymbol{\mu}^{(1)\,T} \boldsymbol{L}^{(2)} \boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)\,T} \boldsymbol{L} \boldsymbol{\mu}^{(2)} + \boldsymbol{\delta}^T \boldsymbol{G} \boldsymbol{\delta} \,,$$

$$= \boldsymbol{\delta}^T \boldsymbol{G} \boldsymbol{\delta} \,, \tag{A11}$$

where $\boldsymbol{\delta} \equiv \boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}$ and in the second last row we have used $\boldsymbol{L}^{(2)} \boldsymbol{L}^{-1} \boldsymbol{L}^{(1)} = \boldsymbol{L}^{(1)} \boldsymbol{L}^{-1} \boldsymbol{L}^{(2)} = \left[(\boldsymbol{L}^{(1)})^{-1} + (\boldsymbol{L}^{(2)})^{-1} \right]^{-1} \equiv \boldsymbol{G}$. So the above derivation [Eq. (A11)] shows that Eq. (12) for IOI of multiple experiments reduces to Eq. (9) for two experiments. We will use

$$\text{IOI} \equiv \tfrac{1}{2} \boldsymbol{\delta}^T \boldsymbol{G} \boldsymbol{\delta} \tag{A12}$$

as the definition of IOI for two experiments.

3. Integrals of joint Gaussian distributions

The integral of a single Gaussian likelihood gives

$$\int d^N \lambda \mathcal{L}_{max}^{(i)} \exp\left(-\tfrac{1}{2} (\boldsymbol{\lambda} - \boldsymbol{\mu}) \boldsymbol{L}^{(i)} (\boldsymbol{\lambda} - \boldsymbol{\mu}) \right)$$
$$= \frac{(2\pi)^{\frac{N}{2}} \mathcal{L}_{max}^{(i)}}{\sqrt{|\boldsymbol{L}^{(i)}|}} \,. \tag{A13}$$

With the combined likelihood given by Eq. (A6), we immediately obtain,

$$\int d^N \lambda \mathcal{L} = \frac{(2\pi)^{\frac{N}{2}} \mathcal{L}_{max}}{\sqrt{|\boldsymbol{L}|}}$$
$$= \frac{(2\pi)^{\frac{N}{2}} \prod \mathcal{L}_{max}^{(i)}}{\sqrt{|\sum \boldsymbol{L}^{(i)}|}} \exp\left(-\tfrac{1}{2} \text{IOI} \right) \,. \tag{A14}$$

An useful case is to integrate the product of two normalized posteriors (as is used in Sec. X B), with $\mathscr{P}^{(i)} = \frac{|\boldsymbol{F}^{(i)}|^{1/2}}{(2\pi)^{N/2}} \exp\left[-\tfrac{1}{2} (\boldsymbol{\lambda} - \bar{\boldsymbol{\mu}}^{(i)})^T \boldsymbol{F}^{(i)} (\boldsymbol{\lambda} - \bar{\boldsymbol{\mu}}^{(i)}) \right]$, where $\boldsymbol{F}^{(i)} = \boldsymbol{L}^{(i)} + \boldsymbol{P}$ and $\bar{\boldsymbol{\mu}}^{(i)} = (\boldsymbol{F}^{(i)})^{-1} (\boldsymbol{L}^{(i)} \boldsymbol{\mu}^{(i)} + \boldsymbol{P} \boldsymbol{\mu}^{(p)})$. By setting $\boldsymbol{L}^{(i)} = \boldsymbol{F}^{(i)}$ and $\boldsymbol{\mu}^i = \bar{\boldsymbol{\mu}}^{(i)}$, we obtain from Eq. (A14) that

$$\int \mathscr{P}^{(1)} \mathscr{P}^{(2)} d^N \lambda$$
$$= \sqrt{\frac{|\boldsymbol{F}^{(1)} \boldsymbol{F}^{(2)}|}{|\boldsymbol{F}^{(1)} + \boldsymbol{F}^{(2)}|}} \exp\left(-\tfrac{1}{2} \bar{\boldsymbol{\delta}}^T \bar{\boldsymbol{G}} \bar{\boldsymbol{\delta}} \right) \tag{A15}$$
$$= |\bar{\boldsymbol{G}}|^{1/2} \exp\left(-\tfrac{1}{2} \bar{\boldsymbol{\delta}}^T \bar{\boldsymbol{G}} \bar{\boldsymbol{\delta}} \right) \,,$$

where, $\bar{\boldsymbol{G}} = ((\boldsymbol{F}^{(1)})^{-1} + (\boldsymbol{F}^{(2)})^{-1})^{-1}$ and $\bar{\boldsymbol{\delta}} = \bar{\boldsymbol{\mu}}^{(1)} - \bar{\boldsymbol{\mu}}^{(2)}$. Then in the weak prior limit, we just have to take those bars off.

Appendix B: Marginalization never increases IOI

1. Proof

We will adopt the following notations. An element of the matrix $\boldsymbol{C}$ is denoted as lower case C_{ij}, while $\widetilde{\boldsymbol{C_{ij}}}$ stands for a matrix obtained by taking out the ith row and the jth column of $\boldsymbol{C}$ and multiplied by $(-1)^{i+j}$. (Then $\widetilde{\boldsymbol{C_{ij,mn}}} = \widetilde{\boldsymbol{C_{mn,ij}}}$ is a matrix obtained by taking out the ith & mth rows and jth & nth columns of $\boldsymbol{C}$ and multiplied by $(-1)^{i+j+m+n}$.) $|\boldsymbol{C}|$ is the determinant of $\boldsymbol{C}$. Then $|\widetilde{\boldsymbol{C_{ij}}}|$ is the (i,j) cofactor $\boldsymbol{C}$. Let $\widetilde{\boldsymbol{\delta_i}}$ be the vector results from deleting the ith component from $\boldsymbol{\delta}$.

Suppose we are to marginalize λ_1. The full IOI is $\tfrac{1}{2} \boldsymbol{\delta}^T \boldsymbol{G} \boldsymbol{\delta}$, and the IOI after marginalizing over λ_1 is $\text{IOI}_{mar,1} = \tfrac{1}{2} (\boldsymbol{\delta}^T \boldsymbol{G} \boldsymbol{\delta})_{mar,1}$. The quantity $(\boldsymbol{\delta}^T \boldsymbol{G} \boldsymbol{\delta})_{mar,1}$ is calculated as follows: (1) delete the first element of $\boldsymbol{\delta}$ and get $\widetilde{\boldsymbol{\delta_1}}$; (2) delete the first row and the first column in $\boldsymbol{C} = \boldsymbol{C}^{(1)} + \boldsymbol{C}^{(2)}$ and let it be $\widetilde{\boldsymbol{C_{11}}}$; (3) then $(\boldsymbol{\delta}^T \boldsymbol{G} \boldsymbol{\delta})_{mar,1} = \widetilde{\boldsymbol{\delta_1}} (\widetilde{\boldsymbol{C_{11}}})^{-1} \widetilde{\boldsymbol{\delta_1}}$. What we are going to show is $\text{IOI} \geq \text{IOI}_{mar,1}$, that is, $\boldsymbol{\delta}^T \boldsymbol{G} \boldsymbol{\delta} \geq (\boldsymbol{\delta}^T \boldsymbol{G} \boldsymbol{\delta})_{mar,1}$.

We first make a linear parameter transformation that keep λ_1 unchanged. More explicitly, we let $\boldsymbol{\lambda} \to \boldsymbol{\lambda}' = \boldsymbol{M} \boldsymbol{\lambda}$ with a transformation matrix of the form

$$\boldsymbol{M} = \begin{pmatrix} 1 & 0 & 0 & \cdots \\ 0 & m_{22} & m_{23} & \cdots \\ 0 & m_{32} & m_{33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \xrightarrow{\text{symbolic}} \boldsymbol{M} = \begin{pmatrix} 1 & \boldsymbol{0} \\ \boldsymbol{0} & \widetilde{\boldsymbol{M}} \end{pmatrix} , \tag{B1}$$

where $\widetilde{\boldsymbol{M}} \equiv \widetilde{\boldsymbol{M_{11}}}$. Then the new deviation vector $\boldsymbol{\delta}'$ reads

$$\boldsymbol{\delta} = \begin{pmatrix} \delta_1 \\ \widetilde{\boldsymbol{\delta_1}} \end{pmatrix} \to \boldsymbol{\delta}' = \begin{pmatrix} \delta_1 \\ \widetilde{\boldsymbol{\delta_1}}' \end{pmatrix} = \begin{pmatrix} \delta_1 \\ \widetilde{\boldsymbol{M}} \widetilde{\boldsymbol{\delta_1}} \end{pmatrix} . \tag{B2}$$

Similarly, the new matrix $\boldsymbol{C}'$ reads,

$$\boldsymbol{C} = \begin{pmatrix} C_{11} & \boldsymbol{V_1}^T \\ \boldsymbol{V_1} & \widetilde{\boldsymbol{C_{11}}} \end{pmatrix} \to \boldsymbol{C}' = \begin{pmatrix} C_{11} & \boldsymbol{V_1}^T \widetilde{\boldsymbol{M}}^T \\ \widetilde{\boldsymbol{M}} \boldsymbol{V_1} & \widetilde{\boldsymbol{M}}^T \widetilde{\boldsymbol{C_{11}}} \widetilde{\boldsymbol{M}} \end{pmatrix} , \tag{B3}$$

where $(\boldsymbol{V_1})_i = C_{1i}$ for $i \neq 1$. So under $\boldsymbol{M}$, δ_1 and C_{11} stay the same, $\widetilde{\boldsymbol{\delta_1}}$ and $\boldsymbol{V_1}$ transform as vectors, and $\widetilde{\boldsymbol{C_{11}}}$

transforms as a rank-2 tensor. To simply the proof, we further let $\boldsymbol{C}'$ to take the following form,

$$
\boldsymbol{C} = \begin{pmatrix} C_{11} & C_{12} & C_{13} & \cdots \\ C_{21} & C_{22} & C_{23} & \cdots \\ C_{31} & C_{32} & C_{33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}
$$

$$
\rightarrow \ \boldsymbol{C}' = \begin{pmatrix} C_{11} & C'_{12} & C'_{13} & \cdots \\ C'_{12} & C'_{22} & 0 & \mathbf{0} \\ C'_{13} & 0 & C'_{33} & \mathbf{0} \\ \vdots & \mathbf{0} & \mathbf{0} & \ddots \end{pmatrix} . \tag{B4}
$$

We have used the fact that $C_{ij} = C_{ji}$. In other words, we transform the parameter space alone the direction of λ_1, so that the other (new) parameters are the eigen-modes of $\widetilde{C_{11}}$ in the projected $(N\text{-}1)$-dimensional parameter subspace orthogonal to λ_1. For the proof, we will work in the new parameter space and drop the primes. In this new parameter space, we have

$$
\boldsymbol{\delta}^T \boldsymbol{G} \boldsymbol{\delta} = \frac{1}{|\boldsymbol{C}|} \Big(\sum_{i,j} |\widetilde{\boldsymbol{C_{ij}}}| \delta_i \delta_j \Big)
$$

$$
= \frac{1}{|\boldsymbol{C}|} \Big(\sum_{i \neq 1} |\widetilde{\boldsymbol{C_{ii}}}| \delta_i^2 + \sum_{i \neq 1} \sum_{j \neq 1, i} |\widetilde{\boldsymbol{C_{ij}}}| \delta_i \delta_j \tag{B5}
$$

$$
+ 2 \sum_{i \neq 1} |\widetilde{\boldsymbol{C_{1i}}}| \delta_i \delta_1 + |\widetilde{\boldsymbol{C_{11}}}| \delta_1^2 \Big) .
$$

With our new matrix C of the form (B4), we can calculate those determinants explicitly, (let $\prod\limits_{i \neq 1} C_{ii} = A$, and $i, j \neq 1$)

$$
|\widetilde{\boldsymbol{C_{11}}}| = A ,
$$

$$
|\widetilde{\boldsymbol{C_{1i}}}| = -\frac{C_{1i}}{C_{ii}} A ,
$$

$$
|\boldsymbol{C}| = \big(C_{11} - \sum_{i \neq 1} \frac{C_{1i}}{C_{ii}} \big) A ,
$$

$$
|\widetilde{\boldsymbol{C_{ii}}}| = \big(C_{11} - \sum_{j \neq 1, i} \frac{C_{1j}}{C_{jj}} \big) \frac{A}{C_{ii}} ,
$$

$$
|\widetilde{\boldsymbol{C_{ij}}}|_{j \neq i} = -\frac{C_{1i} C_{1j}}{C_{ii} C_{jj}} A .
$$

After we substitute the above determinants into Eq. (B5) and cancel out A, $\boldsymbol{\delta}^T \boldsymbol{G} \boldsymbol{\delta}$ reads

$$
\boldsymbol{\delta}^T \boldsymbol{G} \boldsymbol{\delta} = \frac{1}{C_{11} - \sum\limits_{i \neq 1} \frac{C_{1i}^2}{C_{ii}}} \Big(\sum_{i \neq 1} \frac{C_{11}}{C_{ii}} \delta_i^2 - \sum_{i \neq 1} \sum_{j \neq 1, i} \frac{C_{1j}^2}{C_{ii} C_{jj}} \delta_i^2
$$

$$
- \sum_{i \neq} \sum_{j \neq 1, i} \frac{C_{1i} C_{1j}}{C_{ii} C_{jj}} \delta_i \delta_j - 2 \sum_{i \neq 1} \frac{C_{1i}}{C_{ii}} \delta_i \delta_1 + \delta_1^2 \Big) . \tag{B6}
$$

In our new parameter space, the IOI after marginalizing over λ_1 is rather easy to calculate and it is $2\mathrm{IOI}_{mar,1} = (\boldsymbol{\delta}^T \boldsymbol{G} \boldsymbol{\delta})_{mar,1} = \sum\limits_{i \neq 1} \frac{\delta_i^2}{C_{ii}}$. Taking the difference between $\boldsymbol{\delta}^T \boldsymbol{G} \boldsymbol{\delta}$ and $(\boldsymbol{\delta}^T \boldsymbol{G} \boldsymbol{\delta})_{mar,1}$, considering only the numerator of it, and rearranging terms, we have

The numerator of $\Big(\boldsymbol{\delta}^{\mathrm{T}} \mathrm{G} \boldsymbol{\delta} - (\boldsymbol{\delta}^{\mathrm{T}} \mathrm{G} \boldsymbol{\delta})_{\mathrm{mar},1} \Big)$

$$
= \sum_{i \neq 1} \frac{C_{11}}{C_{ii}} \delta_i^2 - \sum_{i \neq 1} \frac{C_{11}}{C_{ii}} \delta_i^2
$$

$$
+ \sum_{i \neq 1} \sum_{j \neq 1} \frac{C_{1j}^2}{C_{ii} C_{jj}} \delta_i^2 - \sum_{i \neq 1} \sum_{j \neq 1, i} \frac{C_{1j}^2}{C_{ii} C_{jj}} \delta_i^2
$$

$$
- \sum_{i \neq} \sum_{j \neq 1, i} \frac{C_{1i} C_{1j}}{C_{ii} C_{jj}} \delta_i \delta_j - 2 \sum_{i \neq 1} \frac{C_{1i}}{C_{ii}} \delta_i \delta_1 + \delta_1^2
$$

$$
= \sum_{i \neq 1} \frac{C_{1i}^2}{C_{ii}} \delta_i^2 - \sum_{i \neq} \sum_{j \neq 1, i} \frac{C_{1i} C_{1j}}{C_{ii} C_{jj}} \delta_i \delta_j - 2 \sum_{i \neq 1} \frac{C_{1i}}{C_{ii}} \delta_i \delta_1 + \delta_1^2
$$

$$
= \sum_{i, j \neq 1} \frac{C_{1i} C_{1j}}{C_{ii} C_{jj}} \delta_i \delta_j - 2 \sum_{i \neq 1} \frac{C_{1i}}{C_{ii}} \delta_i \delta_1 + \delta_1^2
$$

$$
= \Big(\sum_{i \neq 1} \frac{C_{1i}}{C_{ii}} \delta_i - \delta_1 \Big)^2 \geq 0 ,
$$

$$
\tag{B7}
$$

So $\boldsymbol{\delta}^T \boldsymbol{G} \boldsymbol{\delta} \geq (\boldsymbol{\delta}^T \boldsymbol{G} \boldsymbol{\delta})_{mar,1}$ and marginalization never increases IOI.

The condition for $\mathrm{IOI} = \mathrm{IOI}_{mar,1}$ (called equality condition from now) is $\sum\limits_{i \neq 1} \frac{C_{1i}}{C_{ii}} \delta_i = \delta_1$. Since we have transformed the parameter space along the λ_1 axis, this equality condition is not generic. To find the generic equality condition, we first note that the right-hand side of the equality condition δ_1 is invariant under the transformation. Therefore we can generalize the left-hand side to a quantity that is: (1) invariant under the transformation specified by Eq. (B1) and, (2) reduce to $\sum\limits_{i \neq 1} \frac{C_{1i}}{C_{ii}} \delta_i$ in the case when $\boldsymbol{C}$ takes the form of $\boldsymbol{C}'$ in Eq. (B3). An easy guess of that quantity is $\sum (\widetilde{\boldsymbol{C_{11}}}^{-1})_{ij} C_{1i} \delta_j$. And if we are marginalizing over the kth parameter, the generic equality condition becomes,

$$
\sum_{i, j \neq k} (\widetilde{\boldsymbol{C_{kk}}}^{-1})_{ij} C_{ki} \delta_j = \delta_k . \quad \text{(Equality Condition)} \tag{B8}
$$

We have tested and verified that if the above equality condition is satisfied, we indeed have $\mathrm{IOI} = \mathrm{IOI}_{mar,k}$. An intuitive subset of the equality condition (B8) is that $C_{ki} = \delta_j = 0$, which represents the two conditions mentioned in Sec. V: if the two experiments do not have conflict on a particular parameter, and if that parameter does not have correlation between the other parameters, experimental inconsistency will b preserved after marginalizing that parameter.
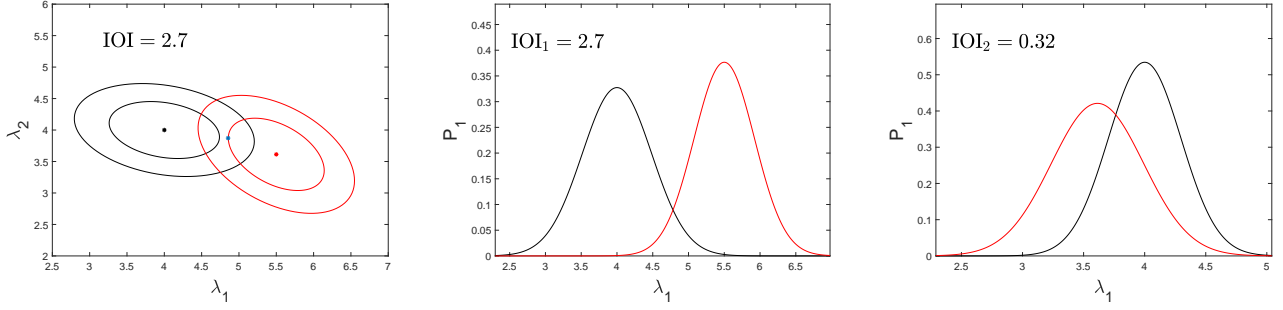
FIG. 12. An example of situation which satisfies the equality condition (B8). Left: A two-parameter model constrained by two experiments shown in black and red. Middle: The constraint of λ_1 marginalized over λ_2. Right: The constraint of λ_2 marginalized over λ_1. In this particular example, the condition $\frac{C_{12}}{C_{11}}\delta_1 = \delta_2$. If we marginalize over λ_2, IOI remains the same, i.e., IOI=IOI$_1$ even thought IOI$_2 \neq 0$.

2. The analytic expression for the drop of IOI

In Sec. VI, we have defined the drop of IOI after marginalizing over λ_k (ΔIOI$_k$) as a supplemental measure of the one-parameter inconsistency in a multi-parameter model. From the proof in Sec. B 1, especially from Eq. (B7) and Eq. (B6), we can obtain ΔIOI$_k$ analytically and it is,

$$\Delta\text{IOI}_k = \frac{\left(\delta_k - \sum_{i \neq k} \frac{C_{ki}}{C_{ii}}\delta_i\right)^2}{C_{11} - \sum_{i \neq k} C_{ki}^2/C_{ii}}. \qquad (B9)$$

The above expression is not generically right, since it is not in an invariant form under a linear transformation along the λ_k axis. To generalize the result [Eq. (B9)], we only need to find an invariant quantity that reduces to Eq. (B9) in the case when $\boldsymbol{C}$ takes the form of $\boldsymbol{C}'$ in Eq (B3). We find that the general expression of ΔIOI$_k$ is

$$\Delta\text{IOI}_k = \frac{\left[\delta_k - \sum_{i,j \neq k}\left(\widetilde{\boldsymbol{C_{kk}}}^{-1}\right)_{ij}C_{ki}\delta_j\right]^2}{C_{kk} - \sum_{i,j \neq k}\left(\widetilde{\boldsymbol{C_{kk}}}^{-1}\right)_{ij}C_{ki}C_{kj}}$$
$$\xrightarrow{\text{Symbolically}} \frac{\left(\delta_k - \mathbf{V_k}^T\left(\widetilde{\boldsymbol{C_{kk}}}^{-1}\right)\widetilde{\delta_\mathbf{k}}\right)^2}{C_{kk} - \mathbf{V_k}^T\left(\widetilde{\boldsymbol{C_{kk}}}^{-1}\right)\mathbf{V_k}}, \qquad (B10)$$

where $(\boldsymbol{V_k})_i = C_{ki}$ for $i \neq k$. We have tested and verified that Eq. (B10) is the same as the definition ΔIOI$_k \equiv$ IOI $-$ IOI$_{marg,k}$.

3. A discussion on different situations in parameter marginalization

In general, the equality equation can be satisfied without $C_{ki} = 0$ or $\delta_j = 0$. An example is shown by the three panels in Fig. 12. It is an coincidence there for IOI to be the same after marginalizing over λ_2. The reason is as follows. If $\delta_2 = 0$, we will have IOI>IOI$_1$ because marginalizing over λ_2 hides some inconsistency due to the correlation between λ_1 and λ_2 specified by $\boldsymbol{G}$. In Fig. 12, $\delta_2 \neq 0$, and the red ellipse is shifted a little downward compared to the case $\delta_2 = 0$. As a result, the joint mean is located at a point which is more easily "supported" by the two experiments, and consequently IOI becomes smaller. Meanwhile, IOI$_1$ does not depend on the vertical shift of both distributions. For a particular amount of vertical shift, IOI becomes the same as IOI$_1$.

If the drop of IOI is small after parameter marginalization, we may be able to marginalize over the corresponding parameter to analyse the inconsistency for the whole model. If so, we can analyse the model in a low parameter dimension. And hopefully it can be reduced to two or three dimension so that we can graphically show the full inconsistency. That is one of the motivations for us we propose a supplemental measure of the inconsistency for one parameter (ΔIOI$_i$) in Sec. VI. However, even when the drop is small, we still need to pay attention to whether it is a special case like the one shown in Fig. 12. In such a special case, marginalization over λ_2 ignores the original probability distributions, although we get the same IOI. Fortunately, it is easy to check whether it is a special case. For example, if the drop of IOI after marginalizing over λ_2 (i.e., ΔIOI$_2$) is small, we can check whether the IOI after marginalizing over all other parameters but λ_2 (i.e., IOI$_2$) is also small. If both ΔIOI$_2$ and IOI$_2$ are small, we can safely marginalize over λ_2 and analyze the inconsistency for the whole model. If not, marginalizing over λ_2 will either hide some inconsistencies or ignore the original shape of the probability distribution. Table XII summarizes the four situations regarding different combinations of ΔIOI$_i$ and IOI$_i$ and the corresponding effects of marginalizing over the λ_i. In the right panel of Fig. 12, IOI$_2$ is not 0 so the example shown in Fig. 12 is a special case. Combining IOI$_i$ and ΔIOI$_i$, we can get a lot of information about the full distribution.

TABLE XII. The four situations and effects of marginalizing over a parameter λ_i. Only when both ΔIOI_i and IOI_i are small can we safely marginalize over λ_i without losing any information about the inconsistency of the model.

Cases	effect
Both ΔIOI_i and IOI_i are small	safe to marginalize over λ_i to analyze the inconsistency for the whole model
IOI_i is small, but ΔIOI_i is large	marginalization will hide the inconsistency due the correlation (specified by $\boldsymbol{C} = \boldsymbol{C}^{(1)} + \boldsymbol{C}^{(2)}$) between λ_i and some other parameters
ΔIOI_i is small, but IOI_i is large	marginalization over λ_i will get a similar IOI as the whole model, but ignore the shape of the original probability distribution
Both IOI_i and ΔIOI_i are large	marginalization over λ_i will not be good for the analysis of the inconsistency

[1] T. Clifton, P. G. Ferreira, A. Padilla, and C. Skordis, "Modified gravity and cosmology," Phys. Rep. **513**, 1–189 (2012), arXiv:1106.2476.

[2] M. López-Corredoira, "Tests and Problems of the Standard Model in Cosmology," Foundations of Physics (2017), 10.1007/s10701-017-0073-8, arXiv:1701.08720.

[3] C. G. Tsagas, A. Challinor, and R. Maartens, "Relativistic cosmology and large-scale structure," Phys. Rep. **465**, 61–147 (2008), arXiv:0705.4397.

[4] K. Koyama, "Cosmological tests of modified gravity," Reports on Progress in Physics **79**, 046902 (2016), arXiv:1504.04623.

[5] A. Joyce, B. Jain, J. Khoury, and M. Trodden, "Beyond the cosmological standard model," Phys. Rep. **568**, 1–98 (2015), arXiv:1407.0059.

[6] M. Ishak, "Remarks on the Formulation of the Cosmological Constant/Dark Energy Problems," Foundations of Physics **37**, 1470–1498 (2007), astro-ph/0504416.

[7] A. Joyce, L. Lombriser, and F. Schmidt, "Dark Energy Versus Modified Gravity," Annual Review of Nuclear and Particle Science **66**, 95–122 (2016), arXiv:1601.06133.

[8] P. Bull et al., "Beyond Λ CDM: Problems, solutions, and the road ahead," Physics of the Dark Universe **12**, 56–99 (2016), arXiv:1512.05356.

[9] I. Debono and G. F. Smoot, "General Relativity and Cosmology: Unsolved Questions and Future Directions," Universe **2**, 23 (2016), arXiv:1609.09781 [gr-qc].

[10] M. Ishak, A. Upadhye, and D. N. Spergel, "Probing cosmic acceleration beyond the equation of state: Distinguishing between dark energy and modified gravity models," Phys. Rev. D **74**, 043513 (2006), astro-ph/0507184.

[11] S. Wang, L. Hui, M. May, and Z. Haiman, "Is modified gravity required by observations? An empirical consistency test of dark energy models," Phys. Rev. D **76**, 063503 (2007), arXiv:0705.0165.

[12] E. J. Ruiz and D. Huterer, "Testing the dark energy consistency with geometry and growth," Phys. Rev. D **91**, 063009 (2015), arXiv:1410.5832.

[13] J. L. Bernal, L. Verde, and A. J. Cuesta, "Parameter splitting in dark energy: is dark energy the same in the background and in the cosmic structures?" J. Cosmol. Astropart. Phys. **2**, 059 (2016), arXiv:1511.03049.

[14] A. Shafieloo, "Crossing statistic: Bayesian interpretation, model selection and resolving dark energy parametrization problem," J. Cosmol. Astropart. Phys. **5**, 024 (2012), arXiv:1202.4808 [astro-ph.CO].

[15] A. Shafieloo and D. K. Hazra, "Consistency of the Planck CMB data and ΛCDM cosmology," (2016), arXiv:1610.07402.

[16] J. L. Bernal, L. Verde, and A. G. Riess, "The trouble with H$_0$," J. Cosmol. Astropart. Phys. **10**, 019 (2016), arXiv:1607.05617.

[17] E. Di Valentino, A. Melchiorri, and J. Silk, "Beyond six parameters: Extending Λ CDM," Phys. Rev. D **92**, 121302 (2015), arXiv:1507.06646.

[18] É. Aubourg et al., "Cosmological implications of baryon acoustic oscillation measurements," Phys. Rev. D **92**, 123516 (2015), arXiv:1411.1074.

[19] C. Heymans et al., "CFHTLenS tomographic weak lensing cosmological parameter constraints: Mitigating the impact of intrinsic galaxy alignments," Mon. Not. R. Astron. Soc. **432**, 2433–2453 (2013), arXiv:1303.1808.

[20] T. Charnock, R. A. Battye, and A. Moss, "Planck confronts large scale structure: methods to quantify discordance," ArXiv e-prints (2017), arXiv:1703.05959.

[21] M. Betoule et al., "Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples," Astron. Astrophys. **568**, A22 (2014), arXiv:1401.4064.

[22] F. Beutler, C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, L. Campbell, Q. Parker, W. Saunders, and F. Watson, "The 6dF Galaxy Survey: baryon acoustic oscillations and the local Hubble constant," Mon. Not. R. Astron. Soc. **416**, 3017–3032 (2011), arXiv:1106.3366.

[23] P. A. R. Ade et al. (Planck Collaboration), "Planck 2015 results. XIII. Cosmological parameters," Astron. Astrophys. **594**, A13 (2016), arXiv:1502.01589.

[24] P. A. R. Ade et al. (Planck Collaboration), Astron. Astrophys. **594**, A15 (2016), arXiv:1502.01591.

[25] P. A. R. Ade and others (Planck Collaboration), "Planck 2015 results. XXIV. Cosmology from Sunyaev-Zeldovich cluster counts," Astron. Astrophys. **594**, A24 (2016), arXiv:1502.01597.

[26] Chris Blake et al., "The WiggleZ Dark Energy Survey: the selection function and z = 0.6 galaxy power spectrum," Mon. Not. R. Astron. Soc. **406**, 803–821 (2010), arXiv:1003.5721.

[27] David Parkinson, Signe Riemer-Sorensen, Chris Blake, Gregory B. Poole, Tamara M. Davis, et al., "The WiggleZ Dark Energy Survey: Final data release and cosmological results," Phys. Rev. D **86**, 103518 (2012), arXiv:1210.2130.

[28] S. Alam et al., "The Eleventh and Twelfth Data Re-

leases of the Sloan Digital Sky Survey: Final Data from SDSS-III," Astrophys. J. Supp. **219**, 12 (2015), arXiv:1501.00963 [astro-ph.IM].

[29] A. G. Riess, L. M. Macri, S. L. Hoffmann, D. Scolnic, S. Casertano, A. V. Filippenko, B. E. Tucker, M. J. Reid, D. O. Jones, J. M. Silverman, R. Chornock, P. Challis, W. Yuan, P. J. Brown, and R. J. Foley, "A 2.4% Determination of the Local Value of the Hubble Constant," Astrophys. J. **826**, 56 (2016), arXiv:1604.01424.

[30] A. G. Riess, L. Macri, S. Casertano, H. Lampeitl, H. C. Ferguson, A. V. Filippenko, S. W. Jha, W. Li, and R. Chornock, "A 3% Solution: Determination of the Hubble Constant with the Hubble Space Telescope and Wide Field Camera 3," Astrophys. J. **730**, 119 (2011), arXiv:1103.2976.

[31] R. Adam and others (Planck Collaboration), "Planck intermediate results. XLVI. Reduction of large-scale systematic effects in HFI polarization maps and estimation of the reionization optical depth," Astron. Astrophys. **596**, A107 (2016), arXiv:1605.02985.

[32] G. Efstathiou, "H_0 revisited," Mon. Not. R. Astron. Soc. **440**, 1138–1152 (2014), arXiv:1311.3461.

[33] S. Casertano, A. G. Riess, B. Bucciarelli, and M. G. Lattanzi, "A test of Gaia Data Release 1 parallaxes: implications for the local distance scale," Astron. Astrophys. **599**, A67 (2017), arXiv:1609.05175 [astro-ph.IM].

[34] V. Bonvin et al., "H0LiCOW - V. New COSMOGRAIL time delays of HE 0435-1223: H_0 to 3.8 per cent precision from strong lensing in a flat ΛCDM model," Mon. Not. R. Astron. Soc. **465**, 4914–4930 (2017), arXiv:1607.01790.

[35] Q.-G. Huang and K. Wang, "How the dark energy can reconcile Planck with local determination of the Hubble constant," European Physical Journal C **76**, 506 (2016), arXiv:1606.05965.

[36] E. Di Valentino, A. Melchiorri, and J. Silk, "Reconciling Planck with the local value of H_0 in extended parameter space," Physics Letters B **761**, 242–246 (2016), arXiv:1606.00634.

[37] H. Hildebrandt et al., "KiDS-450: cosmological parameter constraints from tomographic weak gravitational lensing," Mon. Not. R. Astron. Soc. **465**, 1454–1498 (2017), arXiv:1606.05338.

[38] A. Leauthaud et al., "Lensing is Low: Cosmology, Galaxy Formation, or New Physics?" ArXiv e-prints (2016), arXiv:1611.08606.

[39] M. A. Troxel and M. Ishak, "The intrinsic alignment of galaxies and its impact on weak gravitational lensing in an era of precision cosmology," Phys. Rep. **558**, 1–59 (2015), arXiv:1407.6990.

[40] D. Kirk et al., "Galaxy Alignments: Observations and Impact on Cosmology," Space Science Reviews **193**, 139–211 (2015), arXiv:1504.05465.

[41] M. A. Troxel and M. Ishak, "Cross-correlation between cosmic microwave background lensing and galaxy intrinsic alignment as a contaminant to gravitational lensing cross-correlated probes of the Universe," Phys. Rev. D **89**, 063528 (2014), arXiv:1401.7051.

[42] T. Eifler, E. Krause, S. Dodelson, A. R. Zentner, A. P. Hearin, and N. Y. Gnedin, "Accounting for baryonic effects in cosmic shear tomography: determining a minimal set of nuisance parameters using PCA," Mon. Not. R. Astron. Soc. **454**, 2451–2471 (2015), arXiv:1405.7423.

[43] J. N. Dossett, M. Ishak, D. Parkinson, and T. M. Davis, "Constraints and tensions in testing general relativity from Planck and CFHTLenS data including intrinsic alignment systematics," Phys. Rev. D **92**, 023003 (2015), arXiv:1501.03119.

[44] E. Krause, T. Eifler, and J. Blazek, "The impact of intrinsic alignment on current and future cosmic shear surveys," Mon. Not. R. Astron. Soc. **456**, 207–222 (2016), arXiv:1506.08730.

[45] T. Delubac et al., "Baryon acoustic oscillations in the Lyα forest of BOSS DR11 quasars," Astron. Astrophys. **574**, A59 (2015), arXiv:1404.1801.

[46] A. Font-Ribera et al., "Quasar-Lyman α forest cross-correlation from BOSS DR11: Baryon Acoustic Oscillations," J. Cosmol. Astropart. Phys. **5**, 027 (2014), arXiv:1311.1767.

[47] E. Abdalla, E. G. M. Ferreira, J. Quintin, and B. Wang, "Evidence for interacting dark energy from BOSS," ArXiv e-prints (2014), arXiv:1412.2777.

[48] E. Calabrese, A. Slosar, A. Melchiorri, G. F. Smoot, and O. Zahn, "Cosmic microwave weak lensing data as a test for the dark universe," Phys. Rev. D **77**, 123531 (2008), arXiv:0803.2309.

[49] E. Di Valentino, A. Melchiorri, and J. Silk, "Cosmological hints of modified gravity?" Phys. Rev. D **93**, 023513 (2016), arXiv:1509.07501.

[50] J. B. Muñoz, D. Grin, L. Dai, M. Kamionkowski, and E. D. Kovetz, "Search for compensated isocurvature perturbations with Planck power spectra," Phys. Rev. D **93**, 043008 (2016), arXiv:1511.04441.

[51] C. Blake et al., "The WiggleZ Dark Energy Survey: the growth rate of cosmic structure since redshift z=0.9," Mon. Not. R. Astron. Soc. **415**, 2876–2891 (2011), arXiv:1104.2948.

[52] B. A. Reid, H.-J. Seo, A. Leauthaud, J. L. Tinker, and M. White, "A 2.5 per cent measurement of the growth rate from small-scale redshift space clustering of SDSS-III CMASS galaxies," Mon. Not. R. Astron. Soc. **444**, 476–502 (2014), arXiv:1404.3742.

[53] A. Vikhlinin et al., "Chandra Cluster Cosmology Project III: Cosmological Parameter Constraints," Astrophys. J. **692**, 1060–1074 (2009), arXiv:0812.2720.

[54] K. Koyama, "Structure formation in modified gravity models," J. Cosmol. Astropart. Phys. **3**, 017 (2006), astro-ph/0601220.

[55] E. Bertschinger, "On the Growth of Perturbations as a Test of Dark Energy and Gravity," Astrophys. J. **648**, 797–806 (2006), astro-ph/0604485.

[56] C. Shapiro, "Biased Dark Energy Constraints from Neglecting Reduced Shear in Weak-Lensing Surveys," Astrophys. J. **696**, 775–784 (2009), arXiv:0812.0769.

[57] H. Jeffreys, *The Theory of Probability*, Oxford Classic Texts in the Physical Sciences (OUP Oxford, 1998).

[58] L. Verde, P. Protopapas, and R. Jimenez, "Planck and the local Universe: Quantifying the tension," Physics of the Dark Universe **2**, 166–175 (2013), arXiv:1306.6766 [astro-ph.CO].

[59] M. J. Mortonson, D. Huterer, and W. Hu, "Figures of merit for present and future dark energy probes," Phys. Rev. D **82**, 063004 (2010), arXiv:1004.0236.

[60] A. Lewis and S. Bridle, "Cosmological parameters from CMB and other data: A Monte Carlo approach," Phys. Rev. D **66**, 103511 (2002), astro-ph/0205436.

[61] P. Marshall, N. Rajguru, and A. Slosar, "Bayesian evidence as a tool for comparing datasets," Phys. Rev. D

73, 067302 (2006), astro-ph/0412535.

[62] M. C. March, R. Trotta, L. Amendola, and D. Huterer, "Robustness to systematics for future dark energy probes," Mon. Not. R. Astron. Soc. **415**, 143–152 (2011), arXiv:1101.1521.

[63] L. Amendola, V. Marra, and M. Quartin, "Internal robustness: systematic search for systematic bias in SN Ia data," Mon. Not. R. Astron. Soc. **430**, 1867–1879 (2013), arXiv:1209.1897 [astro-ph.CO].

[64] S. Seehars, S. Grandis, A. Amara, and A. Refregier, "Quantifying concordance in cosmology," Phys. Rev. D **93**, 103507 (2016), arXiv:1510.08483.

[65] N. MacCrann, J. Zuntz, S. Bridle, B. Jain, and M. R. Becker, "Cosmic discordance: are Planck CMB and CFHTLenS weak lensing measurements out of tune?" Mon. Not. R. Astron. Soc. **451**, 2877–2888 (2015), arXiv:1408.4742.

[66] S. Seehars, A. Amara, A. Refregier, A. Paranjape, and J. Akeret, "Information gains from cosmic microwave background experiments," Phys. Rev. D **90**, 023533 (2014), arXiv:1402.3593.

[67] S. Grandis, S. Seehars, A. Refregier, A. Amara, and A. Nicola, "Information gains from cosmological probes," J. Cosmol. Astropart. Phys. **5**, 034 (2016), arXiv:1510.06422.

[68] G.-B. Zhao *et al.*, "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Examining the observational evidence for dynamical dark energy," ArXiv e-prints (2017), arXiv:1701.08165.

[69] S. Grandis and D. Rapetti (private communication).

[70] S. Grandis, D. Rapetti, A. Saro, J. J. Mohr, and J. P. Dietrich, "Quantifying tensions between CMB and distance data sets in models with free curvature or lensing amplitude," Mon. Not. R. Astron. Soc. **463**, 1416–1430 (2016), arXiv:1604.06463.

[71] S. Joudaki *et al.*, "CFHTLenS revisited: assessing concordance with Planck including astrophysical systematics," Mon. Not. R. Astron. Soc. **465**, 2033–2052 (2017), arXiv:1601.05786.

[72] H. Akaike, "A new look at the statistical model identification," IEEE Transactions on Automatic Control **19**, 716–723 (1974).

[73] David J. Spiegelhalter, Nicola G. Best, Bradley P. Carlin, and Angelika Van Der Linde, "Bayesian measures of model complexity and fit," Journal of the Royal Statistical Society: Series B (Statistical Methodology) **64**, 583–639 (2002).

[74] A. R. Liddle, "Information criteria for astrophysical model selection," Mon. Not. R. Astron. Soc. **377**, L74–L78 (2007), astro-ph/0701113.

[75] A. J. Ross, L. Samushia, C. Howlett, W. J. Percival, A. Burden, and M. Manera, "The clustering of the SDSS DR7 main Galaxy sample - I. A 4 per cent distance measure at z = 0.15," Mon. Not. R. Astron. Soc. **449**, 835–847 (2015), arXiv:1409.3242.

[76] R. Adam and others (Planck Collaboration), "Planck intermediate results. XLVII. Planck constraints on reionization history," Astron. Astrophys. **596**, A108 (2016), arXiv:1605.03507.

[77] E. Di Valentino, A. Melchiorri, E. V. Linder, and J. Silk, "Constraining Dark Energy Dynamics in Extended Parameter Space," ArXiv e-prints (2017), arXiv:1704.00762.