

Breaking the spell of Gaussianity: forecasting with higher order Fisher matrices

Elena Sellentin,¹ Miguel Quartin^{2★} and Luca Amendola¹

¹*Institut Für Theoretische Physik, Ruprecht-Karls-Universität Heidelberg, Philosophenweg 16, D-69120 Heidelberg, Germany*

²*Instituto de Física, Universidade Federal do Rio de Janeiro, CEP 21941-972 Rio de Janeiro, RJ, Brazil*

Accepted 2014 April 4. Received 2014 April 2; in original form 2014 February 5

ABSTRACT

We present the new method DALI (Derivative Approximation for Likelihoods) for reconstructing and forecasting posteriors. DALI extends the Fisher matrix formalism but allows for a much wider range of posterior shapes. While the Fisher matrix formalism is limited to yield ellipsoidal confidence contours, our method can reproduce the often observed flexed, deformed or curved shapes of known posteriors. This gain in shape fidelity is obtained by expanding the posterior to higher order in derivatives with respect to parameters such that non-Gaussianity in the parameter space is taken into account. The resulting expansion is positive definite and normalizable at every order. Here, we present the new technique, highlight its advantages and limitations and show a representative application to a posterior of dark energy parameters from supernovae measurements.

Key words: methods: data analysis – methods: numerical – methods: statistical – cosmological parameters.

1 INTRODUCTION

In the last few years, the Fisher matrix (FM) formalism has been widely applied to forecast constraints on cosmological parameters from future experiments (see e.g. Tegmark, Taylor & Heavens 1997; Wang et al. 2010; Bassett et al. 2011; Bueno Belloso, García-Bellido & Sapone 2011; Abramo 2012; Debono 2013; Amendola et al. 2014). With its recipe-like structure and its many elementary maths operations, this technique knows how to entice scientists away from more complex methods such as the Markov chain Monte Carlo (MCMC; Christensen et al. 2001; Lewis & Bridle 2002; Dunkley et al. 2005; Akeret et al. 2012), nested sampling (Feroz & Hobson 2008; Feroz, Hobson & Bridges 2009) or full-grid analysis (Tegmark & Zaldarriaga 2000), although these other methods are known to reproduce the shape of posteriors much more faithfully. The omnipresence of the FM is mainly caused by its speedy execution. A fast posterior evaluation is indeed sometimes more than a convenience: in Amendola, Marra & Quartin (2013b) and Heneka, Marra & Amendola (2014), a blind search for systematics on the Union2.1 supernova (SNeIa) data set (Amanullah et al. 2010) required roughly 10^6 such evaluations, and the FM had to be employed whenever valid.

While speed certainly is an important asset for a forecasting technique, often one desires the essential shape of the posterior to be captured, such that degeneracy directions and regions of the parameter space that are not preferred by the data are represented

adequately. In this respect, the FM has often been criticized since it assumes that the posterior is a Gaussian function of the parameters, and therefore is bound to produce ellipsoidal confidence-level contours. Of these ellipses, the principal axes represent the local direction of parameter degeneracies, and the area of the ellipses is taken as a measure of the constraining power of an experiment (figure of merit; Albrecht et al. 2006; Amendola et al. 2013a). However, a mismatch between both the orientation and the size of these ellipses with respect to MCMC-generated posteriors has often been observed (Wolz et al. 2012; Khedekar & Majumdar 2013; Rodriguez et al. 2013). To which extent FMs are a trustworthy forecasting technique is consequently a debatable question.

These drawbacks of the FM originate from its assumption of the posterior being Gaussian in the parameters, which is exact only when the data are Gaussian and the model is linear in the parameters. This assumption is approximately true when one has collected enough data such that the central limit theorem kicks in. However, it is often the case that the amount of data is insufficient to warrant such an approximation, except perhaps close to the maximum of the posterior. In fact, for many parameters of dark-energy-related research, targeted parameters such as w_a (see definition below) are weakly constrained non-linear model parameters, such that the posterior contains a non-negligible amount of non-Gaussianity. Therefore, an obvious method to improve the description of the posterior beyond the scope of the FM is to tackle the Gaussian assumption. One recent investigation used invertible transformation of parameters in order to make the posterior more Gaussian (Kosowsky, Milosavljevic & Jimenez 2002; Joachimi & Taylor 2011).

★E-mail: miguel.q.fisica@gmail.com

Here, instead we build on the FM, but expand the posterior to higher orders. If the posterior P really is Gaussian in the parameters, the higher order derivatives of $\log P$ will be zero, such that the extended method falls back on to the FM and nothing is lost. If they are non-zero, a gain in shape fidelity is to be expected. As many posteriors have a smooth shape and resemble often a ‘surrealistic’ version of an ellipse, i.e. the ellipses are slightly curved, flexed or otherwise distorted, already the inclusion of just a few higher order derivatives promises good improvements.

The main problem in expanding over a Gaussian distribution is that the expansion is in general not guaranteed to be a true distribution, i.e. positive definite and normalizable. Edgeworth or Gram–Charlier series suffer indeed from this serious problem. Here, we find that a simple rearrangement of the terms in the Taylor series can guarantee that the expansion remains a true distribution at every order. The expansion turns out to be a derivative expansion rather than a parameter expansion, as we clarify below.

One of the most severe caveats of the FM is that it does not carry any information that allows us to check whether its assumption of (approximate) Gaussianity is fulfilled. Therefore, one must rely on alternative techniques if one is worried about the breakdown of the FM estimates. Besides correcting the shapes of contours, higher order corrections in the Taylor expansion of the posterior also serve as a fast (and simplest) double-check on the FM analysis.

This paper is organized as follows. In an attempt to clearly separate lengthy calculations from our argumentation line, we promote the appendices to a vital part of the paper. In Section 2.1, we develop the extended formalism of posterior reconstruction and focus on a parameter-independent covariance matrix. The derivative expansion is discussed in Section 2.2. In Section 3.1, we specialize the method to SNe data and we apply it in Section 3.2 to the Union2.1 SN catalogue and to a mock catalogue with 1000 SNe up to $z = 2$. We then discuss in Section 3.3 the issue of marginalization of parameters. We conclude in Section 4. The appendix contains in-depth calculations, a comparison between frequentist and Bayesian FM and an extension of our method to parameter-dependent covariance matrices.

2 INCLUDING NON-GAUSSIANITY INTO A POSTERIOR

2.1 Taylor expansion of the posterior

We consider a posterior¹ that depends on n parameters p_α , where α can take values $1, \dots, n$. Denoting with P the posterior distribution, we expand the log-likelihood $\mathcal{L} = -\log(P)$ as a function of the parameters p_α in Taylor series around the likelihood best fit, indicated by the subscript 0 as

$$\begin{aligned} -\mathcal{L} \equiv \log P &\approx \log P_0 + \frac{1}{2} (\log P)_{,\alpha\beta}|_0 \Delta p_\alpha \Delta p_\beta \\ &+ \frac{1}{3!} (\log P)_{,\alpha\beta\gamma}|_0 \Delta p_\alpha \Delta p_\beta \Delta p_\gamma \\ &+ \frac{1}{4!} (\log P)_{,\alpha\beta\gamma\delta}|_0 \Delta p_\alpha \Delta p_\beta \Delta p_\gamma \Delta p_\delta, \end{aligned} \quad (1)$$

¹ Our results hold for a likelihood exactly in the way they hold for a posterior. This is especially true when we use uniform priors, since then the difference between likelihood and posterior in practice vanishes.

where summation over repeated indices is implied, $\Delta p_\alpha = p_\alpha - \hat{p}_\alpha$ is the deviation of a parameter from its best-fitting \hat{p}_α and $_{,\alpha} \equiv \partial_{p_\alpha}$. The first-order derivatives vanish because we are at the maximum of the posterior. Expanding to the second order yields the Fisher approximation. From the third order onwards, non-Gaussianities are taken into account, which correct for misestimates of the posterior by the FM, and thereby lead to a deformation of its shape. We can write the approximation as

$$P = N \exp \left[-\frac{1}{2} F_{\alpha\beta} \Delta p_\alpha \Delta p_\beta - \frac{1}{3!} S_{\alpha\beta\gamma} \Delta p_\alpha \Delta p_\beta \Delta p_\gamma - \frac{1}{4!} Q_{\alpha\beta\gamma\delta} \Delta p_\alpha \Delta p_\beta \Delta p_\gamma \Delta p_\delta - \mathcal{O}(5) \right], \quad (2)$$

where N is a normalization constant,

$$\begin{aligned} F_{\alpha\beta} &= \mathcal{L}_{,\alpha\beta}, \\ S_{\alpha\beta\gamma} &= \mathcal{L}_{,\alpha\beta\gamma}, \\ Q_{\alpha\beta\gamma\delta} &= \mathcal{L}_{,\alpha\beta\gamma\delta}, \end{aligned} \quad (3)$$

and we shall neglect the fifth and higher order terms of the Taylor series. From now on all derivatives in the Taylor series are taken at the best-fitting value.

Here, the $n \times n$ matrix $F_{\alpha\beta}$ is the usual FM. We dub the $n \times n \times n$ tensor $S_{\alpha\beta\gamma}$ the Flexion² tensor and the scalar

$$S \equiv S_{\alpha\beta\gamma} \Delta p_\alpha \Delta p_\beta \Delta p_\gamma \quad (4)$$

just ‘the Flexion’. Likewise, we call $Q_{\alpha\beta\gamma\delta}$ the Quarxion tensor and

$$Q \equiv Q_{\alpha\beta\gamma\delta} \Delta p_\alpha \Delta p_\beta \Delta p_\gamma \Delta p_\delta \quad (5)$$

just ‘the Quarxion’. Finally, for simplicity, we dub

$$F \equiv F_{\alpha\beta} \Delta p_\alpha \Delta p_\beta \quad (6)$$

just ‘the Fisher’. We therefore refer to the expansion equation (2) up to fourth order as the Fisher–Flexion–Quarxion approximation. Any non-zero Flexion or Quarxion tensor implies immediately that the posterior is not exactly Gaussian in the parameters, and the larger their components are, the larger is the non-Gaussianity.

In the frequentist approach, the FM is defined as the data average of $F_{\alpha\beta}$, i.e.

$$F_{\alpha\beta}^F \equiv \langle \mathcal{L}_{,\alpha\beta} \rangle. \quad (7)$$

In the Bayesian approach, the data are no longer random variables and no averaging takes place. We have instead the alternative definition

$$F_{\alpha\beta}^B \equiv \mathcal{L}_{,\alpha\beta}|_{\text{BF}}, \quad (8)$$

that is, the FM is evaluated at the parameter maximum-likelihood best fit. This point is sometimes neglected in the literature and in Appendix A we comment on the difference between these two definitions. Nevertheless, when making a forecast for a future experiment, the maximum-likelihood parameter set is chosen beforehand (it is the fiducial set) and the two definitions coincide. In this paper, we assume the frequentist definition because it allows for several simplifications and because the whole Fisher approach (and the extension here proposed) is most useful when doing forecasts.

² We borrow the term “flexion” from the weak-lensing literature (Goldberg & Bacon 2005; Bacon et al. 2006), where it also refers to third-order corrections to the shapes of images, which typically flex the shape of sources from ellipses towards a banana-shaped image.

Note that the exponential in the FM approximation contains only a quadratic form. The argument of the exponential function is consequently always negative, which ensures that the probability stays finite. This handy feature is not necessarily true for the Quarxions and never true for the Flexions: the Flexion is cubic in the Δp and will therefore always become negative at large enough Δp . Whenever negative Flexion and Quarxions terms become larger than the Fisher, the argument of the exponential becomes positive and the Fisher–Flexion–Quarxion approximation diverges at large Δp . This is a fundamental problem in many expansions around a Gaussian, such as the Edgeworth or the Gram–Charlier.

It is however possible to solve this problem by expanding in derivatives rather than in Δp , as we show next.

2.2 DALI: the derivative expansion

We consider now cases in which the parameters appear only in a theoretical model μ that is compared to a data set, and not in the covariance matrix of the parameter space. We label the theoretical prediction corresponding to the i th data point as μ_i ; notation can be simplified by introducing the model vector μ . In this paper, Latin indices generally run over the data and Greek indices over the parameters.

Averaging over possible data sets generated from a given fiducial, we find that the FM is given by (see Appendix B)

$$F_{\alpha\beta} = \langle \mathcal{L}_{,\alpha} \mathcal{L}_{,\beta} \rangle, \quad (9)$$

i.e. no second derivatives appear. With $M = C^{-1}$ being the inverse of the parameter-independent and positive definite covariance matrix in the data space, we find in Appendix B the Flexion tensor to be

$$S_{\alpha\beta\gamma} = \langle \mathcal{L}_{,\alpha\beta} \mathcal{L}_{,\gamma} \rangle + \langle \mathcal{L}_{,\gamma\alpha} \mathcal{L}_{,\beta} \rangle + \langle \mathcal{L}_{,\beta\gamma} \mathcal{L}_{,\alpha} \rangle \\ = \mu_{,\alpha\beta} M \mu_{,\gamma} + \text{cycl.} \quad (10)$$

The Quarxion tensor is

$$Q_{\alpha\beta\gamma\delta} = \mu_{,\alpha\gamma\delta} M \mu_{,\beta} + \mu_{,\delta\gamma} M \mu_{,\beta\alpha} \\ + \mu_{,\alpha\beta\delta} M \mu_{,\gamma} + \mu_{,\delta\beta} M \mu_{,\gamma\alpha} \\ + \mu_{,\alpha\gamma\beta} M \mu_{,\delta} + \mu_{,\beta\gamma} M \mu_{,\delta\alpha} \\ + \mu_{,\delta\gamma\beta} M \mu_{,\alpha}. \quad (11)$$

It is obvious from their definition in equation (3) that Flexions and Quarxions are symmetric under index permutation. Both Flexion and Quarxion tensors also transform under parameter-space transformations in the same way as the FM: to wit, with a series of simple Jacobian transformations. When taking the full Flexion or Quarxion term, all the distinct terms of the same type in equations (10) and (11) become indistinguishable. For instance, by renaming the indices,

$$\mu_{,\alpha\beta} M \mu_{,\gamma} \Delta p_{\alpha} \Delta p_{\beta} \Delta p_{\gamma} = \mu_{,\gamma\alpha} M \mu_{,\beta} \Delta p_{\gamma} \Delta p_{\alpha} \Delta p_{\beta} \quad (12)$$

$$= \mu_{,\beta\gamma} M \mu_{,\alpha} \Delta p_{\beta} \Delta p_{\gamma} \Delta p_{\alpha}. \quad (13)$$

Therefore, we can simplify

$$S = 3 \mu_{,\alpha\beta} M \mu_{,\gamma} \Delta p_{\alpha} \Delta p_{\beta} \Delta p_{\gamma},$$

$$Q = (4 \mu_{,\alpha\gamma\delta} M \mu_{,\beta} + 3 \mu_{,\delta\gamma} M \mu_{,\beta\alpha}) \Delta p_{\alpha} \Delta p_{\beta} \Delta p_{\gamma} \Delta p_{\delta}. \quad (14)$$

Although some terms in Q are positive definite (e.g. $\mu_{,\delta\gamma} M \mu_{,\beta\alpha}$), it appears that neither S nor Q are globally positive definite; this problems shows up at all orders. However, as anticipated, the expansion can be arranged also in a different way, namely in order of

derivatives. That is, to second order in the μ derivatives, we have

$$P = N \exp \left[-\frac{1}{2} \mu_{,\alpha} M \mu_{,\beta} \Delta p_{\alpha} \Delta p_{\beta} \right. \\ \left. - \left(\frac{1}{2} \mu_{,\alpha\beta} M \mu_{,\gamma} \Delta p_{\alpha} \Delta p_{\beta} \Delta p_{\gamma} \right. \right. \\ \left. \left. + \frac{1}{8} \mu_{,\delta\gamma} M \mu_{,\beta\alpha} \Delta p_{\alpha} \Delta p_{\beta} \Delta p_{\gamma} \Delta p_{\delta} \right) + \mathcal{O}(3) \right]. \quad (15)$$

The decisive advantage of this expression is that now the expansion is a true distribution, i.e. normalizable and positive definite, since the highest order term in Δp , $\mu_{,\delta\gamma} M \mu_{,\beta\alpha} \Delta p_{\alpha} \Delta p_{\beta} \Delta p_{\gamma} \Delta p_{\delta} = (\mu_{,\delta\gamma} \Delta p_{\gamma} \Delta p_{\delta})^2 M$, is positive definite (if, as we assumed from the start, the data inverse covariance matrix M is itself positive definite). Remarkably, this is true at every order; for instance, at third order, we have

$$P = N \exp \left[-\frac{1}{2} \mu_{,\alpha} M \mu_{,\beta} \Delta p_{\alpha} \Delta p_{\beta} - \left(\frac{1}{2} \mu_{,\alpha\beta} M \mu_{,\gamma} \Delta p_{\alpha} \Delta p_{\beta} \Delta p_{\gamma} \right. \right. \\ \left. \left. + \frac{1}{8} \mu_{,\delta\gamma} M \mu_{,\beta\alpha} \Delta p_{\alpha} \Delta p_{\beta} \Delta p_{\gamma} \Delta p_{\delta} \right) \right. \\ \left. - \left(\frac{1}{6} \mu_{,\delta} M \mu_{,\beta\alpha\gamma} \Delta p_{\alpha} \Delta p_{\beta} \Delta p_{\gamma} \Delta p_{\delta} \right. \right. \\ \left. \left. + \frac{1}{3!2!} \mu_{,\alpha\beta\delta} M \mu_{,\gamma\tau} \Delta p_{\alpha} \Delta p_{\beta} \Delta p_{\gamma} \Delta p_{\delta} \Delta p_{\tau} \right. \right. \\ \left. \left. + \frac{1}{3!3!2!} \mu_{,\alpha\beta\gamma} M \mu_{,\delta\tau\sigma} \Delta p_{\alpha} \Delta p_{\beta} \Delta p_{\gamma} \Delta p_{\delta} \Delta p_{\tau} \Delta p_{\sigma} \right) + \mathcal{O}(4) \right], \quad (16)$$

where again one sees that the leading term, the last one in equation (16), is positive definite. Note that the derivative expansion requires only derivatives of the order of $N/2$ (for N even) or $(N+1)/2$ (for N odd) for an expansion of the order of N in Δp , rather than $N-1$ as in the expansion (2). The numerical coefficient for a term of the order of N in Δp formed with n_1 and $n_2 = N - n_1$ derivatives is $(n_1! n_2!)^{-1}$ for $n_1 \neq n_2$ and $[2(n_1!)^2]^{-1}$ for $n_1 = n_2$ (see Appendix C).

The approximated posteriors (15) and (16) are the main products of this paper: they represent true distributions and the second- and third-derivative correction, respectively, over the Fisher approximation. We baptize this new posterior reconstruction method DALI: Derivative Approximation for Likelihoods. For the sake of clear referencing, we further call the approximation (equation 15) in which the leading terms are second derivatives the ‘doublet-DALI’ and the approximation that has third derivatives as leading order (equation 16) the ‘triplet-DALI’.

The derivative expansion can actually be directly obtained in a very simple way. We label the i th data point of the data set by m_i , and combine them into a vector \mathbf{m} . Let us start from the standard Gaussian likelihood exponent

$$\frac{1}{2} [\mathbf{m} - \mu(p_1, \dots, p_n)]^T M [\mathbf{m} - \mu(p_1, \dots, p_n)]. \quad (17)$$

Now, we expand to second order around the best-fitting \hat{p}_{α} as

$$\mu \approx \hat{\mu} + \mu_{,\alpha} \Delta p_{\alpha} + \frac{1}{2} \mu_{,\alpha\beta} \Delta p_{\alpha} \Delta p_{\beta}, \quad (18)$$

where $\hat{\mu} \equiv \mu(\hat{p}_1, \dots, \hat{p}_n)$, so we obtain

$$\begin{aligned} & \frac{1}{2} [\mathbf{m} - \mu(p_1, \dots, p_n)] M [\mathbf{m} - \mu(p_1, \dots, p_n)] \approx \\ & \frac{1}{2} [\mathbf{m} - \hat{\mu}] M [\mathbf{m} - \hat{\mu}] \\ & - (\mathbf{m} - \hat{\mu}) M \left(\mu_{,\alpha} \Delta p_\alpha + \frac{1}{2} \mu_{,\alpha\beta} \Delta p_\alpha \Delta p_\beta \right) \\ & + \frac{1}{2} \mu_{,\alpha} M \mu_{,\beta} \Delta p_\alpha \Delta p_\beta + \frac{1}{2} \mu_{,\alpha} M \mu_{,\beta\gamma} \Delta p_\alpha \Delta p_\beta \Delta p_\gamma \\ & + \frac{1}{8} \mu_{,\alpha\beta} M \mu_{,\gamma\delta} \Delta p_\alpha \Delta p_\beta \Delta p_\gamma \Delta p_\delta. \end{aligned} \quad (19)$$

The first term on the rhs is an irrelevant constant that can be absorbed in the normalization; the second term averages out to zero, while the remaining terms are indeed as in equation (15). It is worth remarking again that the expansion equations (2) and (15) are mathematically equivalent; it is only when arranged in order of derivatives rather than in powers of Δp that they differ at each finite order.

In Appendix C, we extend this formalism to parameter-dependent correlations. We leave however tests of this case in realistic cosmological scenarios to future work.

2.3 Speed and complexity

The one incontrovertible advantage of the FM is its speed. A quick order of estimate of the complexity of the DALI approximation can be obtained by observing that the expensive computations needed for the matrices are the evaluations of the vectors of the derivatives. For n parameters, there are n possible first derivatives, so the complexity rises linearly with n . For the ‘doublet’ correction (equation 15), one needs also the second derivatives, of which there are $n(n+1)/2$ distinct ones for n parameters, and similarly $(n^3 + 3n^2 + 2n)/6$ for the ‘triplet’ correction. Since every numerical derivative of order p requires (at lowest accuracy) $p+1$ evaluations of the posterior, the complexity for large n goes like n^2 and $(2/3)n^3$ for the doublet and triplet, respectively. In comparison, grids or MCMC routines evaluate the full likelihood (which implies generating theoretical predictions of the data at every point in parameter space) typically thousands of times already for e.g. four parameters. Therefore, only for $\mathcal{O}(1000)$ [$\mathcal{O}(100)$] parameters does the doublet [triplet] require roughly the same $\mathcal{O}(10^6)$ evaluations of a typical Monte Carlo run in large parameter spaces. In practice, the evaluation of the posterior is thus significantly faster with DALI, as most forecasts in cosmology rely on less than dozen free parameters, and the posterior can be numerically costly to compute. Note, however, that only Gaussian posteriors are again Gaussians with less dimensions if they are marginalized. This analytical result makes marginalizations with FM extremely fast. For non-Gaussian posteriors, for which DALI is interesting, there exists no general analytical marginalization. Therefore, DALI will be slower in this respect than FM – a price that one has to pay, if the non-Gaussianity of a posterior shall be captured.

3 DALI METHOD AT WORK

3.1 Specialization to SNe

We consider now an application of our method to SNe data. The measurable quantity is the distance modulus, which is related to the

dimensionless luminosity distance by

$$\mu_i = 5 \log \hat{d}(z_i), \quad (20)$$

where the index i denotes the dependence on a given redshift. The likelihood function for the SNe after marginalization of the Hubble constant and the absolute luminosity is (Amendola & Tsujikawa 2010)

$$\mathcal{L} = -\log L = \frac{1}{2} \left(S_2 - \frac{S_1^2}{S_0} \right), \quad (21)$$

where the sums are

$$S_n = \sum_i \frac{(m_i - \mu_i)^n}{\sigma_i^2}, \quad (22)$$

where m_i is a measurement at redshift z_i and the corresponding theoretical mean μ_i . The log-likelihood can be written as

$$\mathcal{L} = \frac{1}{2} X_i M_{ij} X_j, \quad (23)$$

where $X_i = m_i - \mu_i$ and the inverse covariance matrix is

$$M_{ij} = s_i s_j \delta_{ij} - \frac{s_i^2 s_j^2}{S_0}, \quad (24)$$

(no sum) where $s_i = 1/\sigma_i$. If one assumes $s_i = 1/\sigma$ (constant), then the covariance matrix is

$$M_{ij} = \sigma^{-2} \left(\delta_{ij} - \frac{1}{N} \right). \quad (25)$$

So finally we have

$$\begin{aligned} F_{\alpha\beta}^{\text{SN}} &= \left\langle \left(\frac{\partial \mu_i}{\partial p_\alpha} M_{ij} X_j \right)^2 \right\rangle, \\ &= 25 \frac{\partial \log \hat{d}_i}{\partial p_\alpha} M_{ij} \frac{\partial \log \hat{d}_j}{\partial p_\beta}. \end{aligned} \quad (26)$$

Similarly, the Flexion and Quarxion tensors and the DALI expansion are then obtained by replacing μ_i with $5 \log \hat{d}_i$.

Note that a parameter that appears additively in μ_i , like the offset, will not enter the DALI terms; therefore, the analytic marginalization of the posterior affects only the Fisher term and remains analytic also in DALI.

3.2 Applying DALI to the SN catalogues

In order to demonstrate the potential of DALI, we show how accurately it can recover the ‘banana-shaped’ posterior of the SN Union2.1 catalogue (Amanullah et al. 2010). This catalogue comprises the distance moduli of 580 SNeIa, which we use for the data points $m_i(z_i)$ of equation (22), together with their respective errors σ_i . We compare this data set with the distance moduli obtained from a flat w cold dark matter (CDM) cosmology with the Chevallier–Polarski–Linder (CPL) parametrization for the dark energy equation of state (Chevallier & Polarski 2001; Linder 2003) as

$$w(a) = w_0 + w_a(1 - a). \quad (27)$$

We choose the fiducial parameters to be the best-fitting parameters of the SN posterior found in Amanullah et al. (2010) for the w CDM model and evaluate the distance moduli at the redshifts of the Union2.1 catalogue.

In Fig. 1, we depict in grey solid contours the non-approximated posterior (obtained with a grid method), which we will frequently refer to as the ‘full’ posterior. Here, and in all other figures, the

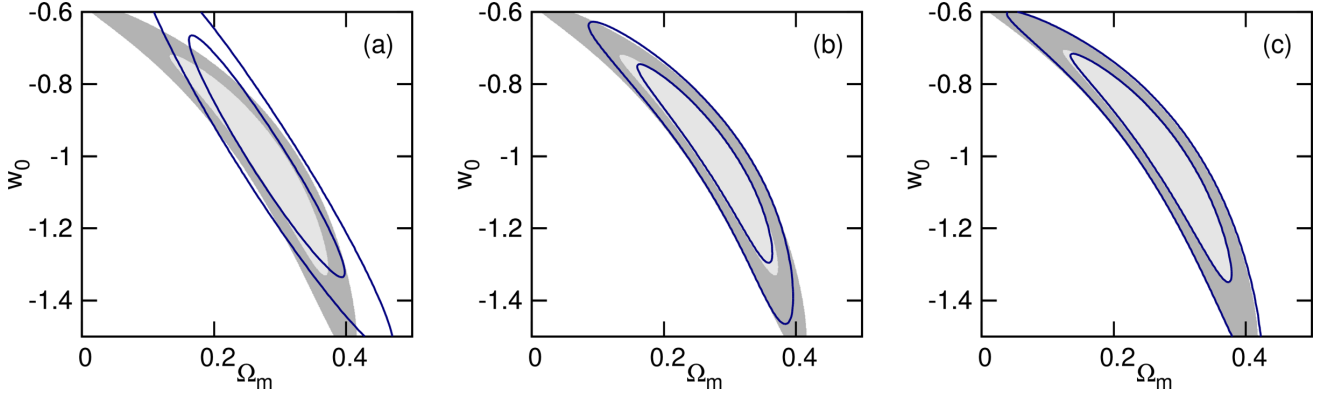


Figure 1. Comparison of the full, non-approximated posterior of the SNeIa Union2.1 catalogue (grey) with different approximations (dark blue). In this plot, only we fix $w_a = 0$ (i.e. assume what is often called ‘ w CDM’ model). The confidence contours are drawn at the 1σ and 2σ confidence levels. Panel (a): the FM approximation; panel (b): equation (15), the doublet-DALI approximation of the posterior includes well the non-Gaussianities; panel (c): equation (16), the triplet-DALI approximation captures the non-Gaussianities even better.

contours are drawn at 1σ and 2σ (we follow standard procedure and use 1σ and 2σ as shorthand notation for 68.3 and 95.4 per cent confidence levels). The improvement of the shape fidelity by successively adding higher order derivatives to the posterior can be seen as one inspects panel (a) [FM], (b) [doublet-DALI] and (c) [triplet-DALI] of Fig. 1.

As the observational campaigns for dark energy proceed and more data are collected, the posteriors are expected to become more and more Gaussian. To investigate the use of DALI in this respect, we mock a future SN catalogue with 1000 SNeIa, uniformly distributed in the redshift range $0 < z < 2$. We use a flat CPL-cosmology with $\Omega_m^{\text{fid}} = 0.285$, $w_0^{\text{fid}} = -1$ and $w_a^{\text{fid}} = 0$ as fiducial. In the $\{\Omega_m, w_0\}$ plane, such a catalogue yields a posterior of similar shape to Fig. 1, which the doublet- and triplet-DALI can recover nicely.

We further demonstrate the potential of our method on a posterior with higher non-Gaussianity: in Fig. 2, we marginalized the posterior of Fig. 1 over w_a in the range $(-\infty, +\infty)$, and our method can recover the shape of this heavily non-Gaussian posterior quite accurately. In Fig. 3, we marginalized instead over Ω_m . Note that the upper half of the Fisher ellipse extends far into the parameter space of positive w_a , which corresponds to a completely different expansion history of the universe, one that is ruled out at many σ by SN data. Changing from flat priors to Gaussian priors also affects the shapes of posteriors. We therefore show in Figs 4 and 5 that the DALI contours improve, as expected, with the marginalization over

two particular cases of Gaussian priors: the former Gaussian in w_a , the latter in Ω_m (in each case, we keep the priors on the remaining variables uniform).

3.3 Marginalization

The FM has four very useful properties: (i) it allows one to evaluate the $n\sigma$ confidence-level contours (which in that case are just ellipses) analytically; (ii) the Gaussian approximation allows one to trivially achieve marginalization over parameters by dropping lines and columns from its inverse matrix; (iii) fixing parameters at their best-fitting values is similarly achieved by dropping the corresponding column and line from the FM; (iv) the FM of the product of two posteriors is the sum of the two posteriors’ FMs. Only the last two properties are shared by the DALI method. The other two must be dealt with numerically.

In the vast majority of cases, we are interested in one- or two-dimensional contour plots of the posterior marginalized in all other parameters. Marginalizations in the DALI must be carried out numerically in an n -dimensional space. This is clearly a disadvantage of the DALI method when compared to standard FM, as without any further simplifications the numerical complexity will grow with the number n of parameters in the same way as in standard numerical integrations, which can be based on either grids or Monte Carlo

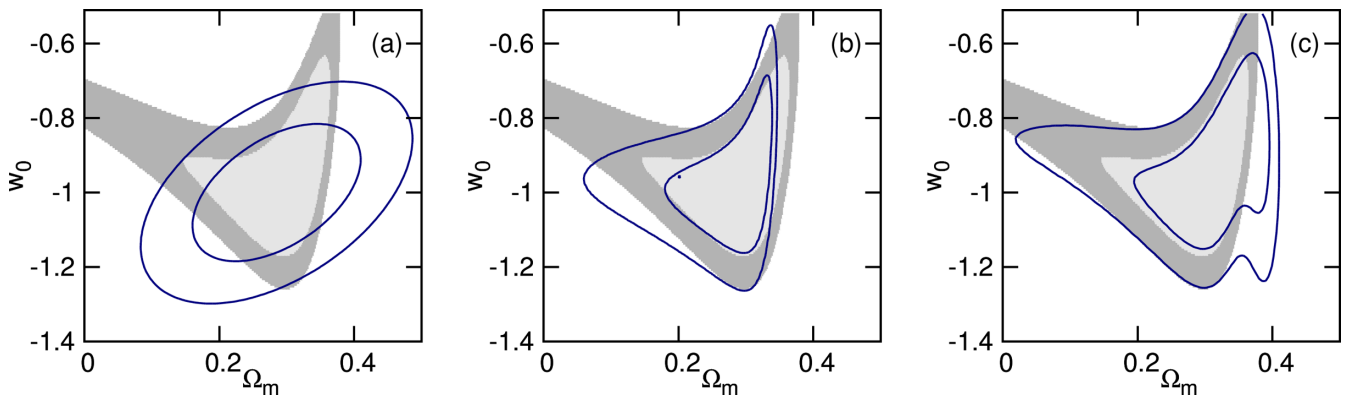


Figure 2. Same as Fig. 1 but for the mock catalogue of 1000 SNeIa (see text) and marginalizing over w_a , which results in a heavily non-Gaussian grey posterior. Again, the DALI methods capture the shape of the posterior much better than the FM. Note that the doublet-DALI is a very good compromise between speed and shape accuracy.

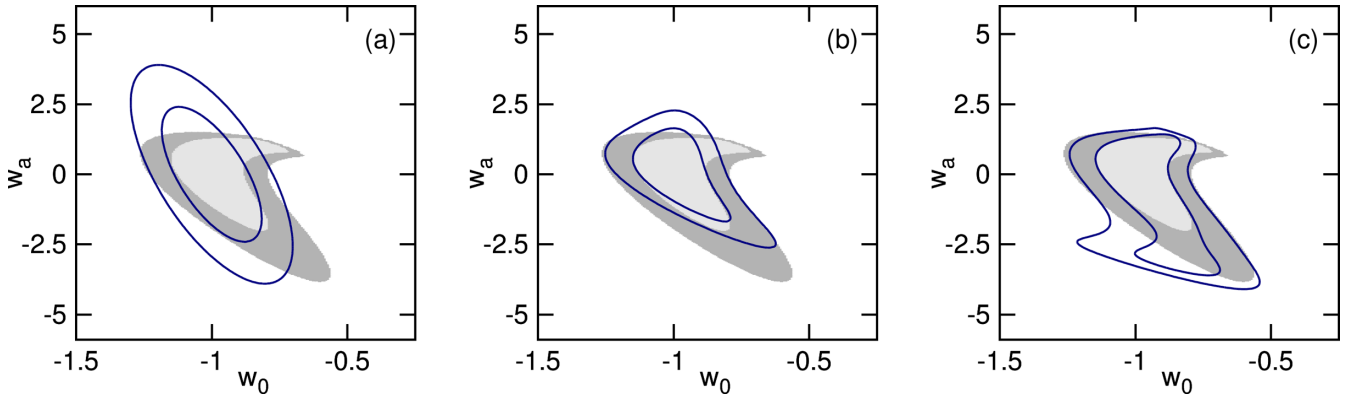


Figure 3. Same as Fig. 2 but this time marginalized over Ω_m in the interval $[0, 1]$. Note that the upper half of the Fisher ellipse covers parameter ranges with high w_a . This indicates that Fisher does not capture the underlying physics well. In both derivative expansions, the posterior does not cover these regions.

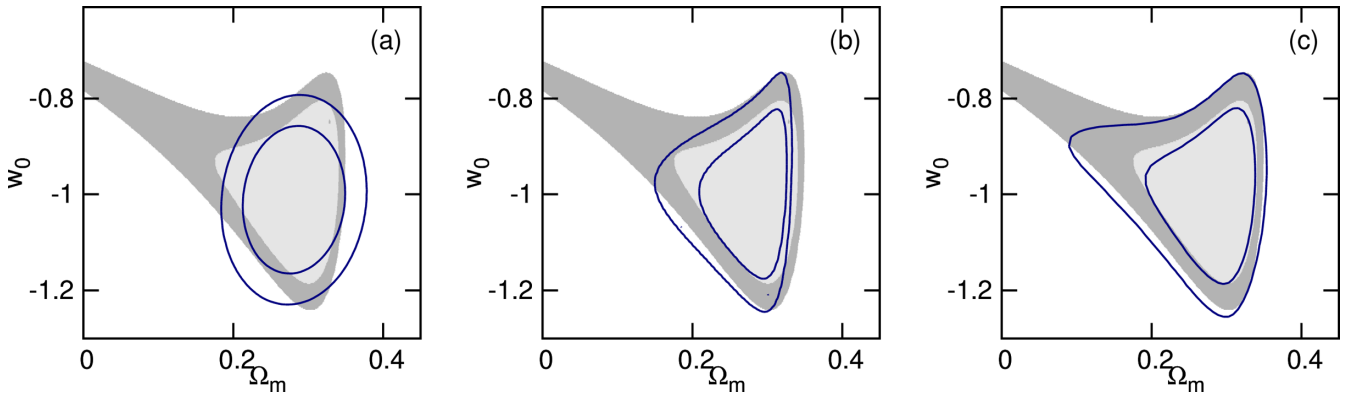


Figure 4. Same as Fig. 2 but marginalized with a Gaussian prior of $\sigma_{w_a} = 1.0$, instead of a flat prior. For this smoother prior, the triplet-Dali contours do not leak out of the grey underlying posterior shape.

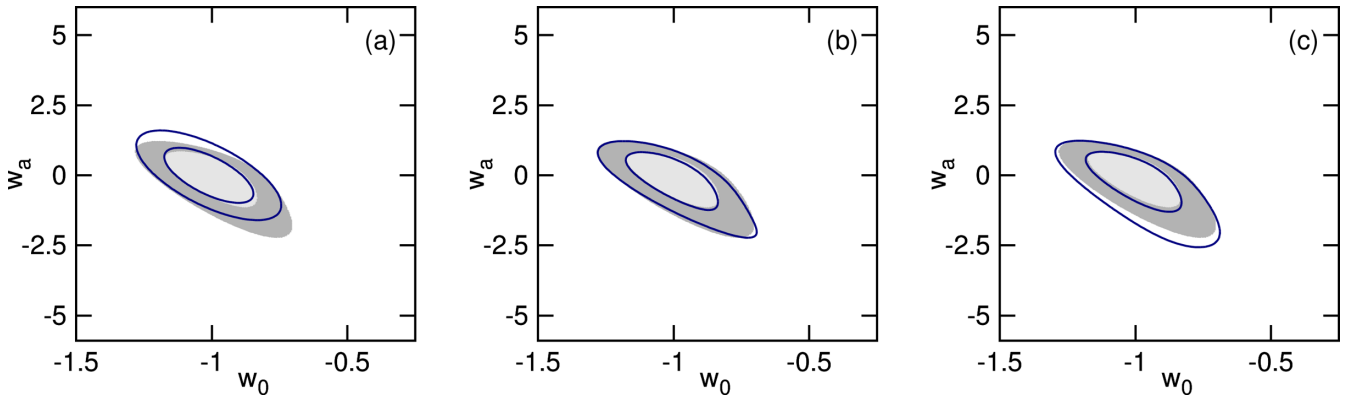


Figure 5. Same as Fig. 3 but marginalized with a Gaussian prior of $\sigma_{\Omega_m} = 0.03$, instead of a flat prior. Also for this rather tight prior, the marginalization leads to a noticeable amount of non-Gaussianity, which can be well captured by DALI.

methods. Nevertheless, since the needed derivatives are only evaluated at the best fit (the fiducial), marginalizations can be carried out without evaluating the posterior, i.e. without running over the data for each parameter set. As discussed above in Section 2.2, this makes the DALI method much faster than standard grids or (except for a *very large* number of parameters) MCMC's.

4 CONCLUSIONS

Our new DALI method of posterior reconstruction was developed to eliminate the drawback of the FM approach, while making only

small concessions in manners of speed. We achieved this goal by expanding the posterior up to second or third order in parameter derivatives such that the approximation comprises a significant amount of the non-Gaussianity in the parameter space. The new terms give a fast measure of how much non-Gaussianity the posterior contains and how accurately the FM reproduces the posterior. The gain in shape fidelity when using the DALI method results in a more faithful reconstruction of the posterior.

As an additional application, the DALI method could help MCMC routines to determine beforehand the high-probability regions to explore. The speed of MCMC methods in fact have been

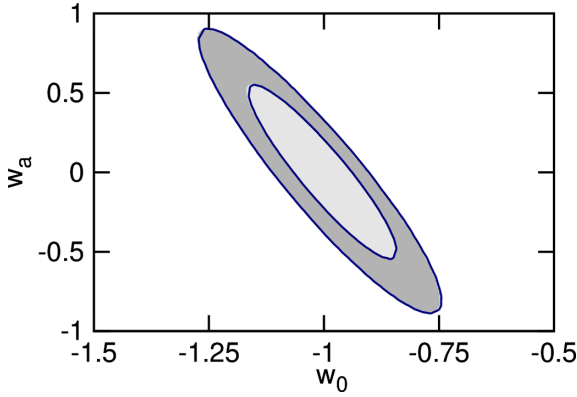


Figure 6. Same as Figs 2 and 3 but for a fixed value of $\Omega_m = \Omega_m^{\text{fid}}$. If the posterior is highly Gaussian, the DALI method falls back on to the FM and ellipsoidal confidence contours emerge.

known to be dependent on the shape of the so-called proposal distribution from which the random walk steps are selected. Usually, a simple multivariate Gaussian distribution is used, based on the FM expansion (Dunkley et al. 2005). Another option is to run a first crude MCMC run and use the rough posterior estimate as a proposal distribution (Lewis & Bridle 2002). The DALI method offers a third alternative, one which we expect to allow for faster convergence than a simple multivariate Gaussian.

The DALI method can also be employed to gauge quantitatively how good is the FM approximation of Gaussianity. For example, the posterior is very ellipsoidal in the parameters w_a and w_0 , as can be seen in Fig. 6. The DALI method then falls back on to the FM – with the important advantage of having checked that the assumption of a Gaussian posterior is justified. In fact, the FM in itself contains no information that allows one to carry out such a check, and authors sometimes run a full MCMC in order to compare the final contours (Wolz et al. 2012; Rodriguez et al. 2013). A full MCMC run is obviously a somewhat costly numerical procedure (and may involve some trial-and-error), which although completely justifiable for final forecasts in expensive surveys is often not the most convenient one when fast results are desired. We nevertheless leave a more detailed exploration of how to best use the DALI method as a measurement of non-Gaussianity for future work.

Needless to say, the range of applicability of the DALI method is not restricted to cosmology and can be applied to any data set.

ACKNOWLEDGEMENTS

We would like to thank Michael Hobson, Marcos Lima, Claudia Quercellini and Jochen Weller for useful discussions. We thank Salvador Dali for inspiration. LA and ES acknowledge financial support from DFG through the TRR33 project ‘The Dark Universe’. MQ is grateful to Brazilian research agencies CNPq and FAPERJ for support.

REFERENCES

- Abramo L. R., 2012, MNRAS, 420, 2042
 Akeret J., Seehars S., Amara A., Refregier A., Csillaghy A., 2012, Astrophysics Source Code Library, record ascl:1303.003
 Albrecht A. et al., 2006, preprint (arXiv:astro-ph/0609591)
 Amanullah et al., 2010, ApJ, 716, 712
 Amendola L., Tsujikawa S., 2010, Dark Energy: Theory and Observations. Cambridge Univ. Press, Cambridge

- Amendola L. et al., 2013a, Living Rev. Relativ., 16, 6
 Amendola L., Marra V., Quartin M., 2013b, MNRAS, 430, 1867
 Amendola L., Fogli S., Guarnizo A., Kunz M., Vollmer A., 2014, Phys. Rev. D, 89, 063538
 Bacon D. J., Goldberg D., Rowe B., Taylor A., 2006, MNRAS, 365, 414
 Bassett B. A., Fantaye Y., Hlozek R., Kotze J., 2011, Int. J. Mod. Phys. D, 20, 2559
 Bueno Bellosio A., García-Bellido J., Sapone D., 2011, J. Cosmol. Astropart. Phys., 10, 10
 Chevallier M., Polarski D., 2001, Int. J. Mod. Phys. D, 10, 213
 Christensen N., Meyer R., Knox L., Luey B., 2001, Class. Quantum Gravity, 18, 2677
 Debono I., 2013, MNRAS, 437, 887
 Dunkley J., Bucher M., Ferreira P. G., Moodley K., Skordis C., 2005, MNRAS, 356, 925
 Feroz F., Hobson M., 2008, MNRAS, 384, 449
 Feroz F., Hobson M., Bridges M., 2009, MNRAS, 398, 1601
 Goldberg D. M., Bacon D. J., 2005, ApJ, 619, 741
 Heneka C., Marra V., Amendola L., 2014, MNRAS, 439, 1855
 Joachimi B., Taylor A., 2011, MNRAS, 416, 1010
 Khedekar S., Majumdar S., 2013, J. Cosmol. Astropart. Phys., 2, 30
 Kosowsky A., Milosavljevic M., Jimenez R., 2002, Phys. Rev. D, 66, 063007
 Lewis A., Bridle S., 2002, Phys. Rev. D, 66, 103511
 Linder E. V., 2003, Phys. Rev. Lett., 90, 091301
 Rodriguez C. L., Farr B., Farr W. M., Mandel I., 2013, Phys. Rev. D, 88, 084013
 Tegmark M., Zaldarriaga M., 2000, ApJ, 544, 30
 Tegmark M., Taylor A., Heavens A., 1997, ApJ, 480, 22
 Wang Y. et al., 2010, MNRAS, 409, 737
 Wolz L., Kilbinger M., Weller J., Giannantonio T., 2012, J. Cosmol. Astropart. Phys., 9, 9

APPENDIX A: FREQUENTIST AND BAYESIAN FM

Suppose there exists an observable $\mathbf{m} = [m_1, \dots, m_n]$ to which a theoretical prediction by a model μ corresponds that is a function of a parameter set: $\mu = \mu(p_1, \dots, p_m)$. In the FM formalism, the observed outcome is the mean value of the observables assumed as the null hypothesis. This method allows a quick way to estimate errors on cosmological parameters, given errors in observable quantities. The FM is defined as the Hessian of the log-likelihood function $\mathcal{L} = -\log(P)$ as

$$F_{\alpha\beta} = \langle -(\log P(\mathbf{m}, \mu))_{,\alpha\beta} \rangle. \quad (\text{A1})$$

This can be simplified as follows.

$$\begin{aligned} F_{\alpha\beta} &= \langle -(\log P(\mathbf{m}, \mu))_{,\alpha\beta} \rangle \\ &= \left\langle -\frac{P_{,\alpha\beta}}{P} + (\log P)_{,\alpha} (\log P)_{,\beta} \right\rangle \\ &= \langle (\log P)_{,\alpha} (\log P)_{,\beta} \rangle \end{aligned} \quad (\text{A2})$$

since

$$\left\langle \frac{P_{,\alpha\beta}}{P} \right\rangle = \int \frac{P_{,\alpha\beta}}{P} P d^n x = \partial_{\alpha\beta} \int P d^n x = 0. \quad (\text{A3})$$

In the case of Gaussian data, the likelihood for n data is

$$P = \frac{1}{(2\pi)^{n/2} \sqrt{|C|}} e^{-\frac{1}{2} (m_i - \mu_i) C_{ij}^{-1} (m_j - \mu_j)} \quad (\text{A4})$$

(note that in this appendix we use the covariance matrix C rather than its inverse M as in the main text). The FM is then (suppressing the

data indices i, j and implicitly summing over them unless otherwise specified; Tegmark et al. 1997)

$$F_{\alpha\beta} = \frac{1}{2} \text{Tr} [C_{,\alpha} C^{-1} C_{,\beta} C^{-1}] + \mu_{,\alpha} C^{-1} \mu_{,\beta}. \quad (\text{A5})$$

By the Cramer–Rao inequality, a model parameter p_α cannot have a variance smaller than $1/(F_{\alpha\alpha})^{1/2}$ (evaluated for unbiased estimators) when all other parameters are fixed, or a precision $(F^{-1})_{\alpha\alpha}^{1/2}$ when all other parameters are marginalized over. Note however that the Cramer–Rao inequality concerns variances and does not say anything about the relative size of the confidence regions.

The purely Bayesian definition of the FM is instead

$$F_{\alpha\beta}^B = -\log(P(\mathbf{m}, \boldsymbol{\mu}))_{,\alpha\beta} \Big|_{\text{BF}}, \quad (\text{A6})$$

where the derivatives have to be evaluated at the best-fitting values of the parameters, i.e. for parameters such that

$$P_{,\alpha} = 0. \quad (\text{A7})$$

This definition makes no reference to the average over the data, which in the Bayesian context are fixed once and for all by the current experiment. Expressions (A1) and (A6) are however in general different and the Cramer–Rao inequality does not hold in general for F^B . We can also write

$$\begin{aligned} F_{\alpha\beta}^B &= -\log(P(\mathbf{m}, \boldsymbol{\mu}))_{,\alpha\beta} \Big|_{\text{BF}} \\ &= -\frac{P_{,\alpha\beta}}{P} \Big|_{\text{BF}} + (\log P)_{,\alpha} (\log P)_{,\beta} \Big|_{\text{BF}} \\ &= -\frac{P_{,\alpha\beta}}{P} \Big|_{\text{BF}} \end{aligned} \quad (\text{A8})$$

due to equation (A7).

We show now that the only cases in which equations (A1) (evaluated on the best-fitting parameters) and (A6) coincide are (a) when the data are Gaussian and the parameters enter in a linear way in the mean and in the variance and (b) in the case of forecasting.

In fact, we have

$$(\log P)_{,\alpha} = -\frac{1}{2} \text{Tr} [C_{,\alpha} C^{-1} + C D_{,\alpha} - C^{-1} C_{,\alpha} C^{-1} D], \quad (\text{A9})$$

where we defined the data matrix of components as

$$D_{ij} = X_i X_j \quad (\text{A10})$$

and the data vector as

$$\mathbf{X} \equiv \mathbf{m} - \boldsymbol{\mu}. \quad (\text{A11})$$

Note that

$$D_{ij,\alpha} = -2\mu_{i,\alpha} X_j. \quad (\text{A12})$$

The best-fitting condition $(\log L)_{,\alpha} = 0$ gives

$$\text{Tr}[C^{-1} C_{,\alpha} C^{-1} D] = \text{Tr}[C_{,\alpha} C^{-1} + C^{-1} D_{,\alpha}]. \quad (\text{A13})$$

If C does not depend on the parameters, the best-fitting equation becomes

$$D_{,\alpha} = 0. \quad (\text{A14})$$

Finally, we have

$$\begin{aligned} F_{\alpha\beta}^B &= -(\log P)_{,\alpha\beta} \Big|_{\text{BF}} = \\ &= \frac{1}{2} \text{Tr} [C^{-1} C_{,\alpha\beta} (I - C^{-1} D) - C^{-1} C_{,\beta} (C^{-1} C_{,\alpha} \\ &\quad + C^{-1} D_{,\alpha}) + C^{-1} C_{,\alpha} C^{-1} C_{,\beta} C^{-1} D \\ &\quad + C^{-1} C_{,\beta} C^{-1} C_{,\alpha} C^{-1} D - 2\mu_{,\alpha\beta} C^{-1} \mathbf{X} \\ &\quad + 2\mu_{,\alpha} C^{-1} \mu_{,\beta} - C^{-1} C_{,\alpha} C^{-1} D_{,\beta}]_{\text{BF}}. \end{aligned} \quad (\text{A15})$$

Inserting the best-fitting condition (A13), we obtain

$$\begin{aligned} F_{\alpha\beta}^B &= -(\log P)_{,\alpha\beta} \Big|_{\text{BF}} \\ &= \frac{1}{2} \text{Tr} [C^{-1} C_{,\alpha\beta} (I - C^{-1} D) + C^{-1} C_{,\beta} C^{-1} C_{,\alpha} \\ &\quad - 2\mu_{,\alpha\beta} C^{-1} \mathbf{X} + 2\mu_{,\alpha} C^{-1} \mu_{,\beta}]_{\text{BF}} \\ &= F_{\alpha\beta(\text{BF})} + \Sigma_{\alpha\beta(\text{BF})} \end{aligned} \quad (\text{A16})$$

where

$$\Sigma_{\alpha\beta} = \frac{1}{2} \text{Tr} [C^{-1} C_{,\alpha\beta} (I - C^{-1} D)] - \mu_{,\alpha\beta} C^{-1} \mathbf{X}. \quad (\text{A17})$$

This shows that if the parameters enter linearly in C and in μ , the two FM coincide (always assuming Gaussian data). The matrix Σ expresses the difference between frequentist and Bayesian FMs. The first one is the one that ensures the Cramer–Rao inequality. The second one is the matrix that approximates the posterior.

Now, when we do forecasts, we generate mock data with variance given by C and mean given by μ . If we evaluate the average FM for many mock data, then we obtain

$$\langle D \rangle = C \quad (\text{A18})$$

$$\langle X \rangle = 0 \quad (\text{A19})$$

so that

$$\langle \Sigma_{\alpha\beta} \rangle = 0. \quad (\text{A20})$$

Then in doing a forecast we in general identify the two FMs, or rather we can say that the generation of mock data implements the frequentist approach. Analysing real data, however, one should use the Bayesian FM, because this is the approximation to the posterior.

APPENDIX B: PARAMETER-INDEPENDENT COVARIANCE MATRIX

We assume in this appendix that the parameters appear only in the theoretical model μ that is compared to a data set. The data covariance matrix shall be independent of parameters. In Appendix C, we extend our formalism to parameter-dependent correlations. Latin indices run over the data, Greek index over the parameters.

Averaging over possible data sets generated from a given fiducial (subscript 0), we have

$$\begin{aligned} F_{\alpha\beta} &\equiv \langle \mathcal{L}_{,\alpha\beta} \rangle_0, \\ S_{\alpha\beta\gamma} &\equiv \langle \mathcal{L}_{,\alpha\beta\gamma} \rangle_0, \\ Q_{\alpha\beta\gamma\delta} &\equiv \langle \mathcal{L}_{,\alpha\beta\gamma\delta} \rangle_0. \end{aligned} \quad (\text{B1})$$

Using the identities

$$\begin{aligned}\langle \mathcal{L}_{,\alpha} \rangle &= 0, \\ \left\langle \frac{P_{,\alpha\beta}}{P} \right\rangle &= 0, \\ \left\langle \frac{P_{,\alpha\beta\gamma}}{P} \right\rangle &= 0,\end{aligned}\quad (\text{B2})$$

we can show (see equation A2) that

$$F_{\alpha\beta} = \langle \mathcal{L}_{,\alpha} \mathcal{L}_{,\beta} \rangle, \quad (\text{B3})$$

so that no second derivatives appear. Note that

$$\frac{P_{,\alpha\beta}}{P} = -\mathcal{L}_{,\alpha\beta} + \mathcal{L}_{,\alpha} \mathcal{L}_{,\beta}. \quad (\text{B4})$$

The Flexion tensor is then

$$\begin{aligned}S_{\alpha\beta\gamma} &= -\left\langle \frac{P_{,\alpha\beta\gamma}}{P} \right\rangle + \left(\left\langle \frac{P_{,\alpha\beta} P_{,\gamma}}{P^2} \right\rangle + \text{cycl} \right) - 2 \left\langle \frac{P_{,\alpha} P_{,\beta} P_{,\gamma}}{P^3} \right\rangle \\ &= (\langle \mathcal{L}_{,\alpha\beta} \mathcal{L}_{,\gamma} - \mathcal{L}_{,\alpha} \mathcal{L}_{,\beta} \mathcal{L}_{,\gamma} \rangle + \text{cycl}) + 2 \langle \mathcal{L}_{,\alpha} \mathcal{L}_{,\beta} \mathcal{L}_{,\gamma} \rangle \\ &= (\langle \mathcal{L}_{,\alpha\beta} \mathcal{L}_{,\gamma} \rangle + \text{cycl}) - \langle \mathcal{L}_{,\alpha} \mathcal{L}_{,\beta} \mathcal{L}_{,\gamma} \rangle.\end{aligned}\quad (\text{B5})$$

We can make further progress by invoking the functional shape of the log-likelihood

$$\mathcal{L} = \text{const} + \frac{1}{2}(-\log \det M + X_i M_{ij} X_j), \quad (\text{B6})$$

where $M = C^{-1}$ is the inverse of the covariance matrix in the parameter space.

If the parameters are only in μ , we have

$$\mathcal{L}_{,\alpha} = -\mu_{,\alpha} M X, \quad (\text{B7})$$

$$\mathcal{L}_{,\alpha\beta} = -\mu_{,\alpha\beta} M X + \mu_{,\alpha} M \mu_{,\beta}. \quad (\text{B8})$$

When taking the data averages (denoted by $\langle \rangle$), all odd powers of X_i give zero and, since the data are Gaussian,

$$\langle X_j X_m \rangle = M_{jm}^{-1}, \quad (\text{B9})$$

$$\langle X_i X_j X_\ell X_m \rangle = M_{ij}^{-1} M_{\ell m}^{-1} + M_{i\ell}^{-1} M_{jm}^{-1} + M_{im}^{-1} M_{j\ell}^{-1}, \quad (\text{B10})$$

$$\begin{aligned}\langle X_i X_j X_\ell X_m X_k X_n \rangle &= M_{ij}^{-1} M_{\ell m}^{-1} M_{kn}^{-1} + \text{dist. perm.} \\ &= 15 \text{ terms,}\end{aligned}\quad (\text{B11})$$

(where only the distinguishable permutations have to be counted, i.e. permutations that produce identical terms, e.g. M_{ij} and M_{ji} , must be discarded). This means that

$$\langle \mathcal{L}_{,\alpha} \mathcal{L}_{,\beta} \mathcal{L}_{,\gamma} \rangle = 0, \quad (\text{B12})$$

and the Flexions matrix follows to be

$$\begin{aligned}S_{\alpha\beta\gamma} &= \langle \mathcal{L}_{,\alpha\beta} \mathcal{L}_{,\gamma} \rangle + \text{cycl} \\ &= \mu_{i,\alpha\beta} M_{ij} \mu_{k,\gamma} M_{km} \langle X_j X_m \rangle + \text{cycl} \\ &= \mu_{,\alpha\beta} M \mu_{,\gamma} + \text{cycl}.\end{aligned}\quad (\text{B13})$$

The Quarxions can be easily calculated from equation (B6) and turn out to be

$$\begin{aligned}Q_{\alpha\beta\gamma\delta} &= \langle \mathcal{L}_{,\alpha\beta\gamma\delta} \rangle \\ &= \mu_{,\alpha\gamma\delta} M \mu_{,\beta} + \mu_{,\delta\gamma} M \mu_{,\beta\alpha} \\ &\quad + \mu_{,\alpha\beta\delta} M \mu_{,\gamma} + \mu_{,\delta\beta} M \mu_{,\gamma\alpha} \\ &\quad + \mu_{,\alpha\gamma\beta} M \mu_{,\delta} + \mu_{,\beta\gamma} M \mu_{,\delta\alpha} \\ &\quad + \mu_{,\delta\gamma\beta} M \mu_{,\alpha} - \langle \mu_{\alpha\beta\gamma\delta} M X \rangle.\end{aligned}\quad (\text{B14})$$

The last term averages out to zero due to the Gaussian data, such that no fourth-order derivatives survive and what we are left with for the Quarxions is equation (11).

APPENDIX C: PARAMETER-DEPENDENT COVARIANCE MATRIX

If the parameters enter also the data covariance matrix M , we have, instead of equation (B7),

$$\begin{aligned}\mathcal{L}_{,\alpha} &= -\frac{1}{2} T_\alpha - \mu_{i,\alpha} M_{ij} X_j \\ &\quad + \frac{1}{2} X_i M_{i,\alpha} X_j,\end{aligned}\quad (\text{C1})$$

where we define

$$T_\alpha \equiv \text{Tr}(M^{-1} M_{,\alpha})$$

and, instead of equation (B8) for the second derivatives,

$$\begin{aligned}\mathcal{L}_{,\alpha\beta} &= -\frac{1}{2} \text{Tr}(M^{-1} M_{,\alpha\beta} - M^{-1} M_{,\alpha} M^{-1} M_{,\beta}) \\ &\quad - \mu_{i,\alpha\beta} M_{ij} X_j + \mu_{i,\alpha} M_{ij} \mu_{j,\beta} \\ &\quad + \frac{1}{2} X_i M_{ij,\alpha\beta} X_j - (\mu_{i,\alpha} M_{ij,\beta} + \mu_{i,\beta} M_{ij,\alpha}) X_j.\end{aligned}\quad (\text{C2})$$

With a further derivative, we obtain, in explicit notation,

$$\begin{aligned}S_{\alpha\beta\gamma} &= \left[\mu_{i,\alpha\beta} M_{ij} \mu_{j,\gamma} + \mu_{i,\alpha} M_{ij,\beta} \mu_{j,\gamma} \right. \\ &\quad \left. + \frac{1}{2} M_{ij,\alpha\beta} M_{\ell m,\gamma} M_{i\ell}^{-1} M_{jm}^{-1} + \text{cycl} \right] \\ &\quad - M_{ij,\alpha} M_{k\ell,\beta} M_{mn,\gamma} M_{ik}^{-1} M_{jm}^{-1} M_{\ell n}^{-1}.\end{aligned}\quad (\text{C3})$$

If $\mu_i = 0$, for instance when applying the formalism to density contrasts, then the Flexion tensor reduces to

$$\begin{aligned}S_{\alpha\beta\gamma} &= \left[\frac{1}{2} M_{ij,\alpha\beta} M_{\ell m,\gamma} M_{i\ell}^{-1} M_{jm}^{-1} + \text{cycl} \right] \\ &\quad - M_{ij,\alpha} M_{k\ell,\beta} M_{mn,\gamma} M_{ik}^{-1} M_{jm}^{-1} M_{\ell n}^{-1}.\end{aligned}\quad (\text{C4})$$

For the Quarxions, the result in tensor notation is

$$\begin{aligned}Q_{\alpha\beta\gamma\delta} &= \left[\mu_{,\alpha\beta\gamma} M \mu_{,\delta} + \mu_{,\alpha\beta} M \mu_{,\gamma\delta} + \mu_{,\alpha\beta} M_{,\gamma} \mu_{,\delta} + \mu_{,\alpha} M_{,\beta\gamma} \mu_{,\delta} \right. \\ &\quad + \frac{1}{2} M_{,\alpha\beta\gamma} M^{-1} M_{,\delta} M^{-1} - M_{,\alpha\beta} M^{-1} M_{,\gamma} M^{-1} M_{,\delta} M^{-1} \\ &\quad \left. + \frac{1}{2} M_{,\alpha\beta} M^{-1} M_{,\gamma\delta} M^{-1} + \text{dist. perm.} \right] \\ &\quad + 3 M_{,\alpha} M^{-1} M_{,\beta} M^{-1} M_{,\gamma} M^{-1} M_{,\delta} M^{-1}.\end{aligned}\quad (\text{C5})$$

Here, dist.perm. means all the distinguishable permutations (in equations C3 and C4, they coincide with cyclic permutations). For instance, among all the possible permutations of the term $\mu_{,\alpha\beta\gamma} M \mu_{,\delta}$, those that exchange $\alpha\beta\gamma$ give back the same term and are to be neglected: in this case, the possible $4! = 24$ permutations of $\alpha\beta\gamma\delta$ reduce by a factor of $3! = 6$ (the permutations of $\alpha\beta\gamma$), leaving only 4 terms, as in equation (B14). Similarly, for the term $\mu_{,\alpha} M_{,\beta\gamma} \mu_{,\delta}$, the two permutations of $\alpha\delta$ and the two of $\beta\gamma$ are to be neglected, leaving $24/4 = 6$ distinct permutations. For the terms without derivatives of M and n_1 derivatives in the first μ and n_2 in the second μ , the number of distinct permutations is therefore $N!/n_1!n_2!$ (where $N = n_1 + n_2$ is the total number of derivatives) if $n_1 \neq n_2$ and $N!/2(n_1!)^2$ otherwise. When inserted in the expansions (15) and (16), the $N!$ factor simplifies away.

To obtain the derivative expansion, we proceed as in Section 2.2. The standard Gaussian exponent including the M -dependent factor is

$$\frac{1}{2} \text{Tr}(\log M) - \frac{1}{2} [\mathbf{m} - \mu(p_\alpha)] M [\mathbf{m} - \mu(p_\alpha)]. \quad (\text{C6})$$

Now, we expand to second order in the derivatives around the best-fitting \hat{p}_α not only the μ term,

$$\mu \approx \hat{\mu} + \mu_{,\alpha} \Delta p_\alpha + \frac{1}{2} \mu_{,\alpha\beta} \Delta p_\alpha \Delta p_\beta \quad (\text{C7})$$

but also $G \equiv \log M$

$$G \approx \hat{G} + G_{,\alpha} \Delta p_\alpha + \frac{1}{2} G_{,\alpha\beta} \Delta p_\alpha \Delta p_\beta \quad (\text{C8})$$

$$M \approx \hat{M} \left(1 + G_{,\alpha} \Delta p_\alpha + \frac{1}{2} G_{,\alpha\beta} \Delta p_\alpha \Delta p_\beta \right), \quad (\text{C9})$$

where $\hat{M} = M(\hat{p}_\alpha)$. Then on averaging we obtain

$$\begin{aligned} & \left\langle \frac{1}{2} \text{Tr}(\log M) - \frac{1}{2} [\mathbf{m} - \mu(p_\alpha)] M [\mathbf{m} - \mu(p_\alpha)] \right\rangle \approx \\ & -\frac{1}{2} F_{\alpha\beta} \Delta p_\alpha \Delta p_\beta - \frac{1}{2} (\mu_{,\alpha\beta} M \mu_{,\gamma} + \mu_{,\alpha} M_{,\beta} \mu_{,\gamma}) \Delta p_\alpha \Delta p_\beta \Delta p_\gamma \\ & -\frac{1}{8} (\mu_{,\alpha\beta} M \mu_{,\gamma\delta} + 4 \mu_{,\alpha\beta} M_{,\gamma} \mu_{,\delta} \\ & + 2 \mu_{,\alpha} M_{,\beta\gamma} \mu_{,\delta}) \Delta p_\alpha \Delta p_\beta \Delta p_\gamma \Delta p_\delta \\ & -\frac{1}{4} \left(\frac{1}{2} \mu_{\alpha\beta} M_{,\gamma} \mu_{\delta\sigma} + \mu_{,\alpha} M_{,\beta\gamma} \mu_{,\delta\sigma} \right) \Delta p_\alpha \Delta p_\beta \Delta p_\gamma \Delta p_\delta \Delta p_\sigma \\ & -\frac{1}{16} \mu_{,\alpha\beta} M_{,\gamma\tau} \mu_{,\delta\sigma} \Delta p_\alpha \Delta p_\beta \Delta p_\gamma \Delta p_\delta \Delta p_\sigma \Delta p_\tau. \end{aligned} \quad (\text{C10})$$

The term asymptotically dominant is the last one. It is negative definite (as required for the normalizability condition) only if $M_{,\alpha\beta}$ is positive definite, i.e. when the covariance matrix is a convex function of the parameters.

This paper has been typeset from a \LaTeX file prepared by the author.