

## Cosmological ensemble and directional averages of observables

Camille Bonvin<sup>1</sup>, Chris Clarkson<sup>2</sup>, Ruth Durrer<sup>3</sup>, Roy Maartens<sup>4,5</sup>, and Obinna Umeh<sup>4</sup>

<sup>1</sup>*Theory Division, CERN, 1211 Geneva, Switzerland*

<sup>2</sup>*Astrophysics, Cosmology & Gravity Centre and Department of Mathematics & Applied Mathematics, University of Cape Town, Cape Town 7701, South Africa*

<sup>3</sup>*Département de Physique Théorique & Center for Astroparticle Physics, Université de Genève, Quai E. Ansermet 24, CH-1211 Genève 4, Switzerland*

<sup>4</sup>*Physics Department, University of the Western Cape, Cape Town 7535, South Africa*

<sup>5</sup>*Institute of Cosmology & Gravitation, University of Portsmouth, Portsmouth PO1 3FX, United Kingdom*

E-mail: camille.bonvin@cern.ch, chris.clarkson@gmail.com,  
ruth.durrer@unige.ch, roy.maartens@gmail.com, umeobinna@gmail.com

(Dated: August 17, 2015)

We show that at second order, ensemble averages of observables and directional averages do not commute due to gravitational lensing – observing the same thing in many directions over the sky is not the same as taking an ensemble average. In principle this non-commutativity is significant for a variety of quantities that we often use as observables and can lead to a bias in parameter estimation. We derive the relation between the ensemble average and the directional average of an observable, at second order in perturbation theory. We discuss the relevance of these two types of averages for making predictions of cosmological observables, focusing on observables related to distances and magnitudes. In particular, we show that the ensemble average of the distance in a given observed direction is increased by gravitational lensing, whereas the directional average of the distance is decreased. For a generic observable, there exists a particular function of the observable that is not affected by second-order lensing perturbations. We also show that standard areas have an advantage over standard rulers, and we discuss the subtleties involved in averaging in the case of supernova observations.

### I. INTRODUCTION

Cosmological observations have become very precise. Especially for the analysis of the cosmic microwave background (CMB) data one has to take into account not only first-order perturbations but also second-order effects like lensing [1–4]. For other perturbed quantities like supernova distances and redshifts as a function of observed direction [5–10], cosmic shear [11, 12] and galaxy number counts [13–16], second-order perturbative expressions have recently been published and demonstrated to be possibly non-negligible. We need to include these second-order effects if we want to compare theory with very precise observations. Their measurement is also an opportunity to test general relativity since most of these effects are different in theories which modify gravity.

In this paper we show that special attention has to be given when comparing a second-order calculation to observations. In cosmology, since we have only one universe at our disposition, we often replace ensemble averages by averages over directions. Here we show that at second order, directional and ensemble average do not commute. This means that the ergodic assumption is broken by observation on the observer’s past light-cone: due to gravitational lensing, observing the same thing in many directions over the sky is not the same as taking an ensemble average.

The existence of different kinds of averages has already been discussed by Kibble and Lieu [17], for the particular case of the magnification  $\mu$ . They argued that the average over random directions in the sky is not the same as the average over a random distribution of sources. They showed that the ‘random-source average’ of the magnification  $\mu$  is exactly given by its background value – a result previously demonstrated by Weinberg [18] – but that the ‘random-direction average’ of the magnification is affected by perturbations. They found that the quantity that is invariant under random-direction average is the reciprocal magnification  $\mu^{-1}$ .

Here we extend this distinction to arbitrary observables. We argue that theoretically, we can only calculate ensemble averages, whereas observationally we usually average over directions. To compare theory with observation, we need therefore to first take the directional average of the observable and then the ensemble average. We give an explicit expression for the difference between this procedure and the ensemble average in a single line of sight. Our calculation

is valid up to second order in perturbation theory and in the regime of weak lensing<sup>1</sup>.

We apply our formalism to various observables. In particular we discuss the case of the distance, showing that the directional average of the distance (followed by an ensemble average) is *smaller* than its background value, whereas the ensemble average of the distance along a single line of sight is *larger* than its background value. We also discuss the case of isolated standard candles and standard rulers for which it is not so clear which averaging procedure to consider, since we do not usually have many sources at a given redshift. We illustrate the different biases in different variables and we discuss how to construct observables which minimise the bias from second-order perturbations. Finally we apply our formalism to the CMB angular power spectrum. We show that in multipole-space, the average over directions is automatically taken before the ensemble average.

The rest of the paper is organised as follows: in section II we show that ensemble average and directional average do not commute. We derive a general relation between the two types of averages, valid at second order in perturbation theory. In section III, we discuss various examples and in section IV we conclude.

## II. AVERAGING AT SECOND ORDER

We consider a cosmological experiment, i.e. an observation where we detect photons coming in at direction  $\mathbf{n}_o$  from a source situated at redshift  $z$ . Our observable can be the density of galaxies in a certain direction, the temperature of the CMB, the luminosity of a supernova, etc. We usually repeat the measurement over various directions in the sky, and we measure the mean of the observable and/or the correlation functions (averaged over all directions at a fixed angular separation).

To compare these measurements with theoretical predictions we have to perform ensemble averages. Cosmological perturbations are stochastic fields and only their ensemble average and their variance (or other higher-order correlation functions) can be calculated. The usual procedure is to assume that due to stochastic isotropy and the ergodic principle, the (measured) directional average of our observable is equal to its ensemble average. Here we show that this procedure is correct only at first order in perturbation theory. At second order, ensemble averages and averaging over directions are two distinct procedures, which do not commute.

We consider an arbitrary function of direction  $f(\mathbf{n}_o)$  in the sky. Here  $\mathbf{n}_o$  is the (lensed) observed direction. In an unperturbed universe,  $f$  does not depend on directions:  $f = f_0$ . In a perturbed universe, we expand  $f$  around  $f_0$ , in perturbations of order  $\epsilon$

$$f(\mathbf{n}_o) = f_0 \left[ 1 + \epsilon \delta_1(\mathbf{n}_o) + \frac{\epsilon^2}{2} \delta_2(\mathbf{n}_o) + \mathcal{O}(3) \right]. \quad (1)$$

Taking the expectation value of (1) and assuming Gaussianity (so that the expectation value of third-order perturbations vanishes) we get

$$\langle f(\mathbf{n}_o) \rangle = f_0 \left[ 1 + \epsilon \langle \delta_1(\mathbf{n}_o) \rangle + \frac{\epsilon^2}{2} \langle \delta_2(\mathbf{n}_o) \rangle + \mathcal{O}(4) \right]. \quad (2)$$

Naively we would set  $\langle \delta_1(\mathbf{n}_o) \rangle = 0$ . However, at second order in perturbation theory this is not correct. When we go to second order, we have to take into account that the observed photon direction  $\mathbf{n}_o$  is lensed from the unperturbed source direction (see Fig. 1),  $\mathbf{n} = \mathbf{n}_o + \boldsymbol{\alpha}$ , where  $\boldsymbol{\alpha} = \epsilon \boldsymbol{\alpha}_1 + \epsilon^2 \boldsymbol{\alpha}_2/2 + \mathcal{O}(3)$  is the deflection angle, which we assume to be small. As a consequence, the distribution of images is not statistically homogeneous and isotropic, meaning that  $\langle \delta_1(\mathbf{n}_o) \rangle \neq 0$ . More precisely we have

$$\epsilon \langle \delta_1(\mathbf{n}_o) \rangle = \epsilon \langle \delta_1(\mathbf{n} - \boldsymbol{\alpha}) \rangle = \epsilon \langle \delta_1(\mathbf{n}) \rangle - \epsilon^2 \langle \boldsymbol{\alpha}_1 \cdot \nabla \delta_1(\mathbf{n}) \rangle + \mathcal{O}(3) = -\epsilon^2 \langle \boldsymbol{\alpha}_1 \cdot \nabla \delta_1(\mathbf{n}_o) \rangle + \mathcal{O}(3). \quad (3)$$

In the last equality, we have used the fact that the distribution of the sources is statistically homogeneous and isotropic (a consequence of the statistical homogeneity and isotropy of the primordial fluctuations), so that  $\langle \delta_1(\mathbf{n}) \rangle = 0$ . Note that in expressions which are already second order, we do not need to distinguish between  $\mathbf{n}$  and  $\mathbf{n}_o$ .

The expectation value of the second-order expression for  $f$  is therefore of the form

$$\langle f(\mathbf{n}_o) \rangle = f_0 \left[ 1 - \epsilon^2 \langle \boldsymbol{\alpha}_1 \cdot \nabla \delta_1(\mathbf{n}_o) \rangle + \frac{\epsilon^2}{2} \langle \delta_2 \rangle + \mathcal{O}(4) \right]. \quad (4)$$

---

<sup>1</sup> We do not discuss here the more difficult problem of how caustics may affect observables [19].

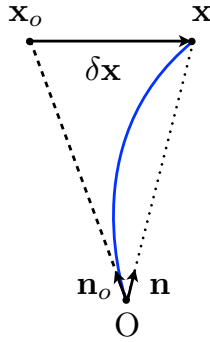


Figure 1: We measure the value of the observable  $f$  at the true comoving position of the source  $\mathbf{x}$ . This position is observed by photons coming in at the direction  $\mathbf{n}_o$ , corresponding to an unperturbed position in the sky  $\mathbf{x}_o \equiv \chi \mathbf{n}_o$ . The deflection vector  $\delta \mathbf{x}$  relates  $\mathbf{x}_o$  to  $\mathbf{x}$  and the deflection angle is defined through  $\alpha = \delta \mathbf{x} / \chi$ .

The second term in (4) directly follows from the non-random distribution of the images generated by gravitational lensing. This term can be simplified using

$$\langle \alpha_1 \cdot \nabla \delta_1(\mathbf{n}_o) \rangle = \langle \nabla \cdot (\alpha_1 \delta_1)(\mathbf{n}_o) \rangle - \langle \nabla \cdot \alpha_1(\mathbf{n}_o) \delta_1 \rangle = 2 \langle \kappa_1 \delta_1 \rangle, \quad (5)$$

where the second equality follows since a total divergence does not contribute to the average and since the first-order convergence is given by  $-2\kappa_1 = \nabla \cdot \alpha_1(\mathbf{n}_o)$  (see [4] for details). With this, the expectation value of  $f$  becomes

$$\langle f(\mathbf{n}_o) \rangle = f_0 \left[ 1 - 2\epsilon^2 \langle \kappa_1 \delta_1 \rangle + \frac{\epsilon^2}{2} \langle \delta_2 \rangle + \mathcal{O}(4) \right]. \quad (6)$$

Let us now see what happens if instead of calculating the expectation value of  $f$ , we first average over observed directions, and then take the ensemble average. From (1) we have<sup>2</sup>

$$\langle \bar{f} \rangle \equiv \frac{1}{4\pi} \left\langle \int d\Omega_{\mathbf{n}_o} f(\mathbf{n}_o) \right\rangle = f_0 \left[ 1 + \epsilon \langle \overline{\delta_1(\mathbf{n}_o)} \rangle + \frac{\epsilon^2}{2} \langle \overline{\delta_2(\mathbf{n}_o)} \rangle + \mathcal{O}(3) \right]. \quad (7)$$

The first-order term gives

$$\begin{aligned} \langle \overline{\delta_1(\mathbf{n}_o)} \rangle &= \frac{1}{4\pi} \left\langle \int d\Omega_{\mathbf{n}_o} \delta_1(\mathbf{n}_o) \right\rangle = \frac{1}{4\pi} \left\langle \int d\Omega_{\mathbf{n}} \left| \frac{\partial \mathbf{n}_o}{\partial \mathbf{n}} \right| \delta_1(\mathbf{n} - \alpha) \right\rangle \\ &\simeq \frac{1}{4\pi} \left\langle \int d\Omega_{\mathbf{n}} \left[ \delta_1(\mathbf{n}) - \epsilon \delta_1(\mathbf{n}) \nabla \cdot \alpha_1(\mathbf{n}) - \epsilon \alpha_1 \cdot \nabla \delta_1(\mathbf{n}) \right] \right\rangle. \end{aligned} \quad (8)$$

The second and third terms can be combined into a total derivative. Since  $\mathbf{n}$  is unperturbed, the directional average over  $\mathbf{n}$  commutes with the ensemble average, and we obtain

$$\langle \overline{\delta_1(\mathbf{n}_o)} \rangle \simeq \frac{1}{4\pi} \int d\Omega_{\mathbf{n}} \left[ \langle \delta_1(\mathbf{n}) \rangle - \epsilon \nabla \cdot \langle \delta_1(\mathbf{n}) \alpha_1 \rangle \right] = 0. \quad (9)$$

This result holds also at next order in perturbation theory. It relies only on the fact that the expectation value of total derivatives vanishes in a statistically homogeneous and isotropic universe<sup>3</sup>.

For the second-order term in (7), we may neglect the perturbations of the direction (which contribute at third order only) so that  $\delta_2(\mathbf{n}_o) \simeq \delta_2(\mathbf{n})$ , and directional and ensemble average of  $\delta_2$  commute

$$\langle \overline{\delta_2(\mathbf{n}_o)} \rangle = \langle \delta_2 \rangle + \mathcal{O}(4). \quad (10)$$

<sup>2</sup> Here we assume for simplicity that we average over the whole sky, but the argument holds for any patch of the sky. The boundary terms which appear are vectors which disappear after ensemble averaging.

<sup>3</sup> Note that the result does not hold if we enter the strong lensing regime where caustics and multiple images appear, and where the map  $\mathbf{n}_o \mapsto \mathbf{n}$  is not one to one. However, the regions where this happens contribute a negligible area for most purposes.

With this we obtain

$$\langle \bar{f} \rangle = f_0 \left[ 1 + \frac{\epsilon^2}{2} \langle \delta_2 \rangle + \mathcal{O}(4) \right] = \langle f(\mathbf{n}) \rangle. \quad (11)$$

Comparing (11) with (6) we see that

$$\langle f(\mathbf{n}) \rangle = \langle \bar{f} \rangle = \langle f(\mathbf{n}_o) \rangle + 2f_0\epsilon^2 \langle \kappa_1 \delta_1 \rangle + \mathcal{O}(4). \quad (12)$$

This equation provides a general relation between the ensemble and the directional average of any observable, valid at second order in perturbation theory, in the weak lensing regime. In particular, it shows that ensemble and directional averages commute only if the variable under consideration has vanishing correlation with the convergence (or similarly with the deflection angle). An important consequence is that an observable whose ensemble average is invariant under second-order perturbations (like for example the magnification  $\mu$ , see [17, 18]) will automatically receive corrections when we take its directional average (and vice-versa).

In practice, if we measure a directional average,  $\bar{f}$ , the ensemble average which we have to compare our measurements with is  $\langle \bar{f} \rangle$ . However, if we measure just one realisation  $f(\mathbf{n}_o)$  in a fixed observed direction  $\mathbf{n}_o$ , we should compare our measurements with  $\langle f(\mathbf{n}_o) \rangle$ .

Before we discuss important examples, let us also note the following: if we observe a power of  $f$ , say  $f^p$ , a Taylor expansion up to second order yields

$$\langle f^p(\mathbf{n}) \rangle = \langle \bar{f}^p \rangle = f_0^p \left[ 1 + \frac{p\epsilon^2}{2} \langle \delta_2 \rangle + \frac{p(p-1)\epsilon^2}{2} \langle \delta_1^2 \rangle + \mathcal{O}(4) \right], \quad \text{whereas} \quad (13)$$

$$\langle f^p(\mathbf{n}_o) \rangle = f_0^p \left[ 1 - 2p\epsilon^2 \langle \kappa_1 \delta_1 \rangle + \frac{p\epsilon^2}{2} \langle \delta_2 \rangle + \frac{p(p-1)\epsilon^2}{2} \langle \delta_1^2 \rangle + \mathcal{O}(4) \right]. \quad (14)$$

It follows that, choosing  $p-1 = -\langle \delta_2 \rangle / \langle \delta_1^2 \rangle$ , we can avoid second-order corrections to  $\langle \bar{f}^p \rangle$ , while choosing  $p-1 = (4\langle \kappa_1 \delta_1 \rangle - \langle \delta_2 \rangle) / \langle \delta_1^2 \rangle$  we can avoid second-order corrections to  $\langle f^p(\mathbf{n}_o) \rangle$ . Hence there is an optimal power of a given variable which removes corrections to the mean at second order.

More generally, for an arbitrary function  $F(f)$  we obtain

$$\langle F(f(\mathbf{n})) \rangle = \langle \overline{F(f)} \rangle = F(f_0) + \frac{\epsilon^2}{2} \left[ F''(f_0) f_0^2 \langle \delta_1^2 \rangle + F'(f_0) f_0 \langle \delta_2 \rangle \right] + \mathcal{O}(4), \quad \text{whereas} \quad (15)$$

$$\langle F(f(\mathbf{n}_o)) \rangle = F(f_0) + \frac{\epsilon^2}{2} \left[ F''(f_0) f_0^2 \langle \delta_1^2 \rangle + F'(f_0) f_0 \{ \langle \delta_2 \rangle - 4\langle \kappa_1 \delta_1 \rangle \} \right] + \mathcal{O}(4). \quad (16)$$

Hence choosing  $F''(f_0) f_0 / F'(f_0) = -\langle \delta_2 \rangle / \langle \delta_1^2 \rangle$  avoids contributions from second-order perturbations to the mean over directions,  $\langle \overline{F(f)} \rangle$ , while choosing  $F''(f_0) f_0 / F'(f_0) = (4\langle \kappa_1 \delta_1 \rangle - \langle \delta_2 \rangle) / \langle \delta_1^2 \rangle$  avoids second-order contributions to  $\langle F(f(\mathbf{n}_o)) \rangle$ .

### III. APPLICATION TO SPECIFIC OBSERVABLES

We consider various examples and show explicitly how the two averaging procedures lead to different contributions from second-order perturbations.

#### A. Distances, Angular Sizes and Standard Areas

The angular diameter distance up to second-order has been fully calculated in [6–9]. Here we are interested in the dominant second-order terms: we want to know what is the maximum impact that second-order contributions can have on the mean distance. As discussed in [4], this means that we need to take into account only the terms with the maximal number of transverse derivatives. We can neglect contributions proportional to the gravitational potential and its time and radial derivatives, relative to transverse derivative of the gravitational potential:

$$\Psi, \partial_t \Psi, \partial_r \Psi \ll \partial_a \Psi, \quad (17)$$

where  $\partial_a \Psi \equiv \mathbf{e}_a^i \partial_i \Psi$  and  $\mathbf{e}_a = (\mathbf{e}_\theta, \mathbf{e}_\varphi)$  are orthogonal to the photon propagation  $\mathbf{n}_o$ .

We calculate the angular diameter distance up to second order with these simplifications. Here we just present the result, the detailed calculation is given in [4]. The angular diameter distance is given by the determinant of the 2x2 magnification matrix  $\mathcal{D}_{ab}$ , defined as<sup>4</sup>

$$\mathcal{D}_{ab} = \nabla_a \alpha_b = \frac{\lambda}{1+z} \begin{pmatrix} 1 - \kappa - \gamma^{(1)} & -\gamma^{(2)} - \omega \\ -\gamma^{(2)} + \omega & 1 - \kappa + \gamma^{(1)} \end{pmatrix}, \quad (18)$$

where  $\omega$  is a curl component which vanishes at first order,  $-2\kappa = \nabla \cdot \alpha$  is the convergence and  $\gamma^{(1)}$  and  $\gamma^{(2)}$  are the shear components (from now on, we absorb the smallness parameter  $\epsilon$  in the perturbation variables). The magnification matrix obeys a second-order differential equation, which can be solved order by order in perturbation theory. The angular diameter distance is given by

$$d_A^2 \simeq \frac{\chi^2}{(1+z)^2} [1 - 2\kappa + \kappa^2 - |\gamma|^2], \quad (19)$$

where  $|\gamma|^2 \equiv (\gamma^{(1)})^2 + (\gamma^{(2)})^2$  and we have neglected  $\omega^2$  which is of order 4 in perturbation theory. We have also set  $\lambda = \chi = \eta_0 - \eta$ , i.e. we neglect the perturbations of the affine parameter which contain less transverse derivatives than the deflection angle;  $\eta_0$  and  $\eta$  denote respectively the conformal time today and at the source position. Expanding  $\kappa$  to second order,  $\kappa = \kappa_1 + \kappa_2/2$ , and taking the square root of (19),

$$d_A = d_0 \left[ 1 + \delta_1 + \frac{1}{2} \delta_2 + \mathcal{O}(3) \right] = d_0 \left[ 1 - \kappa_1 - \frac{1}{2} \kappa_1^2 - \frac{1}{2} \kappa_2 + \frac{1}{2} (\kappa_1^2 - |\gamma_1|^2) + \mathcal{O}(3) \right], \quad (20)$$

where  $d_0 = \chi/(1+z)$  is the background distance (recall that we neglect perturbations in the redshift, which contain less transverse derivatives and are therefore subdominant). As  $|\gamma|$  enters squared, we need it only to first order,  $|\gamma| = |\gamma_1| + \mathcal{O}(2)$ .

We are interested in the average of  $d_A$ . As demonstrated in [4], the second-order convergence  $\kappa_2$  can be written as a total divergence, so it does not contribute: both the ensemble average and the directional average of  $\kappa_2$  vanish:

$$\langle \kappa_2 \rangle = \langle \overline{\kappa_2} \rangle = 0. \quad (21)$$

Furthermore, as we show in [4], the combination  $\kappa_1^2 - |\gamma_1|^2$  is also a total divergence, so that

$$\langle \kappa_1^2 - |\gamma_1|^2 \rangle = \langle \overline{\kappa_1^2 - |\gamma_1|^2} \rangle = 0. \quad (22)$$

If we take the ensemble average of the directional average, we obtain therefore

$$\langle \overline{d_A} \rangle = \langle d_A(\mathbf{n}) \rangle = d_0 \left( 1 - \frac{1}{2} \langle \kappa_1^2 \rangle \right). \quad (23)$$

This shows that lensing *decreases the directional average* of the distance with respect to the background value  $d_0$ .

On the other hand, if we calculate the ensemble average of the distance, we have a remaining second-order term given by (5). Combining this second-order term with the first-order square, we obtain

$$\langle d_A(\mathbf{n}_o) \rangle = d_0 \left( 1 + \frac{3}{2} \langle \kappa_1^2 \rangle \right). \quad (24)$$

This is the result discussed in [3]. It shows that lensing *increases the ensemble average* of the distance with respect to the background value  $d_0$ .

The difference between (23) and (24) can be interpreted in the following way. If we consider many realisations of a random but fixed line of sight, the fact that there are structures between the source and the observer will, on average, increase the distance. This means that structures generate with higher probability a de-focusing on a random line of sight. However, if we average the distance over directions, under-densities which lead to de-focusing are competing with over-densities which lead to focusing. Since lensing also changes the distribution of lines of sight within a given

---

<sup>4</sup> Since photon propagation is conformally invariant,  $\mathcal{D}$  can be calculated in a non-expanding universe. For an expanding universe, we multiply  $\mathcal{D}$  by the scale factor  $a(\eta)$  (see appendix A of [11]), which we rewrite as  $1/(1+z)$  since perturbations in the redshift are negligible relative to the terms with four transverse derivatives.

solid angle, the contribution from under-densities does not exactly cancel the contribution from over-densities. On average, the latter dominates, leading to a decrease of the distance.

Taking an arbitrary power of  $d_A$  to second order and making use of (13) and (14), with  $\delta_1 = -\kappa_1$  and  $\delta_2 = -\kappa_1^2 + \text{divergence}$ , we obtain

$$\langle d_A^p(\mathbf{n}) \rangle = \langle \overline{d_A^p} \rangle = d_0^p \left[ 1 + \frac{p(p-2)}{2} \langle \kappa_1^2 \rangle + \mathcal{O}(4) \right], \quad \text{whereas} \quad (25)$$

$$\langle d_A^p(\mathbf{n}_o) \rangle = d_0^p \left[ 1 + \frac{p(p+2)}{2} \langle \kappa_1^2 \rangle + \mathcal{O}(4) \right]. \quad (26)$$

This shows that the ensemble average of  $d_A^p$  experiences no second-order corrections for  $p = -2$ , while the ensemble average of the mean over directions of  $d_A^p$  experiences no second-order corrections for  $p = +2$ . These results are completely consistent with [17], noting that  $\mu = d_0^2 / \det \mathcal{D}_{ab} = (d_0/d_A)^2$ . The background distance is then given by

$$d_0 = \frac{\chi}{1+z} = \left[ \frac{p+2}{4} \langle \overline{d_A^p} \rangle - \frac{p-2}{4} \langle d_A^p(\mathbf{n}_o) \rangle \right]^{1/p}. \quad (27)$$

From these results we can also calculate the observed angular sizes of standard objects on the sky determined via

$$\text{solid angle} \quad \langle \Delta\Omega(\mathbf{n}_o) \rangle = \langle d_A^{-2}(\mathbf{n}_o) \rangle \Delta S = \Delta\Omega_0, \quad (28)$$

$$\text{while} \quad \langle \overline{\Delta\Omega} \rangle = \langle d_A^{-2}(\mathbf{n}) \rangle \Delta S = \Delta\Omega_0 \left( 1 + 4\langle \kappa_1^2 \rangle \right). \quad (29)$$

$$\text{linear angle} \quad \langle \Delta\theta(\mathbf{n}_o) \rangle = \langle d_A^{-1}(\mathbf{n}_o) \rangle \Delta L = \Delta\theta_0 \left( 1 - \frac{1}{2} \langle \kappa_1^2 \rangle \right), \quad (30)$$

$$\text{while} \quad \langle \overline{\Delta\theta} \rangle = \langle d_A^{-1}(\mathbf{n}) \rangle \Delta L = \Delta\theta_0 \left( 1 + \frac{3}{2} \langle \kappa_1^2 \rangle \right). \quad (31)$$

Here  $\Delta S$  and  $\Delta L$  are the standard area and standard length, and we have used (11), i.e.,  $\langle \overline{f} \rangle = \langle f(\mathbf{n}) \rangle$ . Only one of these is invariant: the expectation value of the angular size of a ‘standard area’. Any standard *ruler*, whether averaged over directions or not, will receive corrections.

The observables discussed here are all affected by the square of the convergence  $\langle \kappa_1^2 \rangle$ . The main contribution to  $\langle \kappa_1^2 \rangle$  can be calculated using the Limber approximation:

$$\langle \kappa_1^2 \rangle = \left[ \int_0^\chi d\chi' \frac{\chi - \chi'}{\chi\chi'} \Delta_\Omega \Psi(\chi) \right]^2 \simeq 4\pi \sum_{\ell=0}^{\infty} \left[ \frac{\ell(\ell+1)}{2\ell+1} \right]^2 \int_0^\chi d\chi' \frac{(\chi - \chi')^2}{\chi'\chi'^2} g^2(\chi') (P_0 T^2) \left( k = \frac{\ell+1/2}{\chi'} \right), \quad (32)$$

where  $\Delta_\Omega$  is the angular Laplacian,  $T$  is the transfer function,  $P_0$  is the power spectrum of the primordial gravitational potential and  $g$  is the growth factor (see [3] for more detail). Figure 2 shows  $\langle \kappa_1^2 \rangle$  as a function of redshift for different values of cosmological parameters. At the last scattering surface,  $z \sim 1100$ ,  $\langle \kappa_1^2 \rangle$  reaches 0.6 percent.

In the discussion above, we have considered only the perturbations with four transverse derivatives of the gravitational potential. A crucial consequence is that in this case, the second-order convergence  $\kappa_2$  can be written as a total divergence, which vanishes on average. The only remaining contribution is therefore the square of the first-order convergence  $\langle \kappa_1^2 \rangle$ . However, the full relativistic expression for the distance contains also second-order terms with two transverse derivatives of the gravitational potential and second-order terms with no transverse derivatives [6–9].

The terms with no transverse derivatives are for example due to the integrated Sachs-Wolfe or the Shapiro time delay. These terms change the physical length of the photon geodesic between the source and the observer. They do not vanish on average and they affect therefore the mean distance to the source. Their amplitude is however of the order of the square of the gravitational potential,  $\Psi^2 \sim 10^{-10}$ , i.e. much smaller than the first-order convergence square in (23) and (24). In addition to the ISW and the Shapiro time delay, the distance receives also corrections from the Doppler terms. The square of the velocity is of the order  $10^{-6}$  and the Doppler terms are therefore relevant only at very low redshift  $z \lesssim 0.5$ .

The terms with two transverse derivatives describe a coupling between the longitudinal and transverse deflections. For example, some of these terms are due to the fact that we average the distance over directions  $\mathbf{n}_o$  at a fixed value of the redshift. Since the redshift is itself perturbed,  $z = z_0 + \delta z$ , we obtain contributions proportional to

$$d_A(z_0 + \delta z, \mathbf{n}_o) \sim d'_A(z_0, \mathbf{n}_o) \cdot \delta z \sim d_0 \kappa_1 \frac{\delta z_1}{1+z}, \quad (33)$$

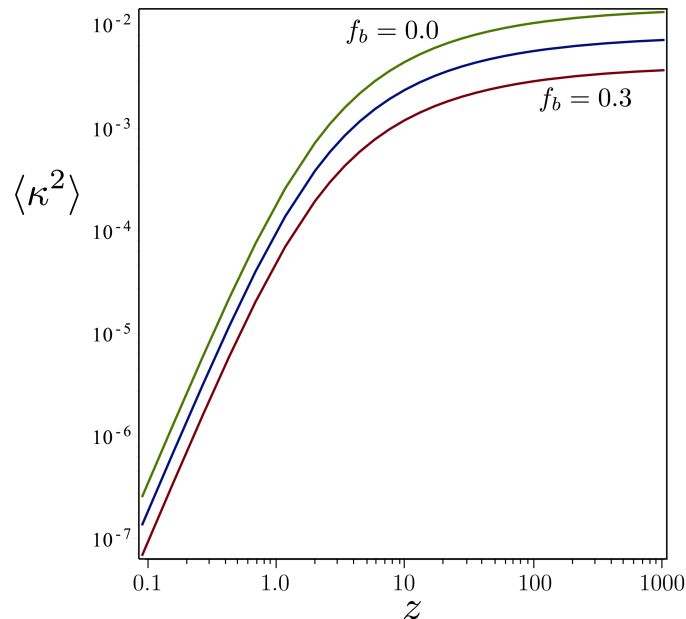


Figure 2: The main contribution to  $\langle \kappa_1^2 \rangle$  comes from terms with 4 transverse derivatives, shown here for  $\Omega_m = 0.3$ ,  $h = 0.7$  and  $n_s = 0.96$ . It is extremely sensitive to small-scale structure, illustrated here by considering different baryon fractions  $f_b \equiv \Omega_b/\Omega_m = 0, 0.15, 0.3$ . Higher  $f_b$  damps small-scale power below 100 Mpc through Silk damping, resulting in a smaller  $\langle \kappa_1^2 \rangle$ . At  $z \sim 1100$  an approximate dependence on model parameters is  $\langle \kappa_1^2 \rangle \approx 0.42 \Omega_m^{1.1 f_b + 1.8} h^{2.5 f_b + 3} n_s^5 e^{-12.6 \Omega_b h^2}$ . Note that here we have used the linear power spectrum to calculate  $\langle \kappa_1^2 \rangle$ . At high redshift, this is a good approximation, but at low redshift the linear expression underestimates the effect.

i.e. contributions due to the fact that the distance is integrated on a perturbed surface, at a fixed redshift from the observer. As seen from Fig. 2,  $\langle \kappa_1^2 \rangle$  reaches  $6 \times 10^{-3}$  at very large redshift. The terms with two transverse derivatives are therefore roughly of the order  $\kappa_1 \times \int \Psi \sim \sqrt{0.006} \times 10^{-5} \sim 8 \times 10^{-7}$ . It is therefore also justified to neglect these types of terms relative to the square of the convergence.

Finally, let us mention that the distance contains also first-order contributions proportional to the gravitational potential at the observer,  $\Psi_0$ , and to the peculiar velocity at the observer,  $\mathbf{v}_o \cdot \mathbf{n}_o$ . Whereas the velocity term vanishes under directional average, the potential term remains and contributes at the order of  $10^{-5}$ . This contribution is almost three orders of magnitude smaller than  $\langle \kappa_1^2 \rangle$  at high redshift. Below redshift 1 however, where  $\langle \kappa_1^2 \rangle$  is much smaller, this contribution may be relevant (see [20] for a detailed analysis of the impact of this term on supernovae measurements).

## B. Fluxes, Magnitudes and Standard Candles

For a standard(-izable) candle such as a type Ia supernova, the relevant quantities are not the geometrical variables angle and area distance, but rather the flux of photons at the observer and the luminosity distance. Observers typically plot the distance modulus  $\mu_m = m - M$  to a supernova at redshift  $z$  of observed magnitude  $m$  and true magnitude  $M$ , inferred via the observed flux  $\mathcal{F}$  and defined as follows<sup>5</sup>,

$$\mathcal{F}(z) = \frac{L}{4\pi d_L^2(z)}, \quad (34)$$

$$\mu_m(z) - 25 = 5 \log \left[ \frac{d_L(z)}{25 \text{ Mpc}} \right]. \quad (35)$$

Here  $z$  is the redshift of the supernova (we neglect its perturbation which is justified only for  $z \gtrsim 0.5$ , see [7, 9, 10]), and  $L$  is its intrinsic luminosity. Observers usually do not have many supernovae with the same redshift in different

<sup>5</sup> The distance modulus, denoted here by  $\mu_m$ , should not be confused with the lensing magnification  $\mu$ .

directions. Therefore, they directly fit the observed curve  $\mu_m(z)$  to the corresponding curve for some background cosmology, without taking into account perturbations. First-order perturbations to the distance have been discussed and taken into account as a systematic error [21–23]. Here we discuss a shift of the mean value due to second-order fluctuations. We concentrate on redshifts  $z > 0.5$ . The effect on the Hubble constant from close-by supernovae with  $z < 0.1$  is discussed in [10].

The luminosity distance is related to the angular diameter distance through  $d_L(z) = (1+z)^2 d_A(z)$ . Neglecting the perturbations in the redshift we have

$$\frac{\delta d_L}{d_L} = \frac{\delta d_A}{d_A} = -\kappa_1 - \frac{1}{2}\kappa_2 - \frac{1}{2}|\gamma_1|^2. \quad (36)$$

Inserting this into (34) and (35) and expanding up to second order we obtain

$$\mathcal{F}(z) = \mathcal{F}_0 \left( 1 + 2\kappa_1 + \kappa_2 + 3\kappa_1^2 + |\gamma_1|^2 \right), \quad (37)$$

$$\mu_m(z) = \mu_{m0}(z) - 5\kappa_1 - \frac{5}{2} \left( \kappa_2 + \kappa_1^2 + |\gamma_1|^2 \right). \quad (38)$$

The impact of second-order lensing on supernovae measurements then really depends on how observations are performed. If the observed flux of each supernova is directly compared with the intrinsic luminosity to extract the luminosity distance, then the relevant quantity is the expectation value of the flux. This is unaffected by second-order lensing, since  $\langle \kappa_2 \rangle = 0$  and  $\langle \kappa_1^2 - |\gamma_1|^2 \rangle = 0$ :

$$\langle \mathcal{F}(\mathbf{n}_o) \rangle = \mathcal{F}_0. \quad (39)$$

On the other hand, if the distance modulus  $\mu_m$  is used directly to extract the luminosity distance, then second-order lensing will systematically increase the expectation value

$$\langle \mu_m(z, \mathbf{n}_o) \rangle = \mu_{m0}(z) + 5 \langle \kappa_1^2 \rangle, \quad (40)$$

leading to an overestimate of the distance  $d_L$ . Considering Fig. 2, we find that for  $z < 2$  the shift in  $\mu_m$  is less than 0.003 and therefore will produce a shift in the dark energy equation of state around the percent level. However, since it is a shift with a definite sign leading to a slight overestimation of  $\mu_m$  and hence of supernovae distances, this can bias parameter estimation if not taken into account.

The discussion above assumes that the luminosity distance is extracted individually from each supernova. If the sample is large enough, one can instead split the supernovae into  $N$  bins of redshift and average the observed flux over all supernovae in the same bin. This average can be considered as an approximate angular average, or equivalently expectation over true source positions (per redshift bin) for fixed size angular patches:

$$\frac{1}{N} \sum_{i=1}^N \mathcal{F}_i(z, \mathbf{n}_o^i) = \frac{1}{4\pi} \sum_{i=1}^N \mathcal{F}_i(z, \mathbf{n}_o^i) \Delta\Omega_{\mathbf{n}_o} \simeq \langle \overline{\mathcal{F}(z)} \rangle = \langle \mathcal{F}(z, \mathbf{n}) \rangle = \mathcal{F}_0(z) \left( 1 + 4 \langle \kappa_1^2 \rangle \right), \quad (41)$$

where we used (12). We see that in this case the observed flux is increased by second-order lensing, leading to an underestimation of the luminosity distance. Hence even though averaging over supernovae in the same redshift bin has the advantage of decreasing the statistical uncertainty in the measurement of the flux, it has the disadvantage of introducing a systematic bias in the luminosity distance. This bias is also present if we average the distance modulus

$$\langle \overline{\mu_m(z)} \rangle = \mu_{m0}(z) - 5 \langle \kappa_1^2 \rangle, \quad (42)$$

and the distance is also underestimated in this case. From (25) we see that to fully exploit the potential of averaging over bins of redshift, without introducing any additional bias, we need to first extract the square of the luminosity distance for each supernova in the bin, and then take an average

$$\langle \overline{d_L^2} \rangle = d_{L0}^2. \quad (43)$$

Finally, let us note that in practice, since supernovae are not perfect standard candles, their intrinsic luminosity is calibrated on a training subset, for which both the flux and the distance are known. From this subset a relation between the intrinsic (peak) luminosity and the shape of the light-curve (for example its width) is derived [24]. This relation is then used to determine the intrinsic luminosity of the other supernovae. One can then wonder whether this relation is affected by second-order lensing, leading to a systematic bias in the determination of the intrinsic



luminosity and consequently to an error in the distance measurements. This is fortunately not the case. Indeed, in the training subset, both the distance and the flux are affected in a consistent way by lensing<sup>6</sup> (following (36) and (37)), leaving the intrinsic luminosity unchanged

$$L = 4\pi d_L^2 \mathcal{F} = 4\pi d_{L0}^2 \mathcal{F}_0 = L_0. \quad (44)$$

The shape of the light-curve is also unaffected, since lensing is not expected to vary during the time-scale of the supernova explosion. As a consequence, lensing can induce a constant shift in the amplitude of the light-curve, but it cannot generate a change in the shape. Therefore the relation between the shape of the light-curve and the intrinsic luminosity inferred from the training subset is not biased by second-order lensing.

### C. CMB angular power spectrum

As a last example, we discuss what happens with the CMB. In [4], we calculated the contribution of the shear and the convergence to the angular power spectrum of the CMB, up to second order in perturbation theory. We found that the lensed power spectrum  $\tilde{D}(\ell) \equiv \ell^2 \tilde{C}(\ell)$ , calculated for a constant magnification matrix, is related to the unlensed power spectrum  $D(\ell)$  through

$$\tilde{D}(\ell) = D(\ell) + \left[ \kappa + \kappa^2 + \frac{1}{4}|\gamma|^2 \right] \ell D'(\ell) + \frac{1}{2} \left[ \kappa^2 + \frac{1}{2}|\gamma|^2 \right] \ell^2 D''(\ell). \quad (45)$$

If we take a directional average of (45), as we do in [4], i.e. we average the shear and convergence over different parts of the sky, we obtain

$$\langle \overline{\tilde{D}(\ell)} \rangle = D(\ell) + \frac{5}{4} \langle \kappa_1^2 \rangle \ell D'(\ell) + \frac{3}{4} \langle \kappa_1^2 \rangle \ell^2 D''(\ell). \quad (46)$$

If instead we take an ensemble average, i.e. we average over all possible realisations of the shear and convergence fields, we obtain

$$\langle \tilde{D}(\ell) \rangle = D(\ell) - \frac{3}{4} \langle \kappa_1^2 \rangle \ell D'(\ell) + \frac{3}{4} \langle \kappa_1^2 \rangle \ell^2 D''(\ell). \quad (47)$$

The smoothing term  $\ell^2 D''(\ell)$  is the same in the two cases: it shows that lensing decreases the amplitude of the peaks. The pure displacement term  $\ell D'(\ell)$  on the other hand is different: it is positive if we take a directional average but negative if we take an ensemble average. This is completely equivalent to the effect on the distance described in section III A: the ensemble average of the distance is increased whereas the directional average is decreased. Since the CMB is averaged over directions, the correct averaging procedure is given by (46), leading to a shift of the peaks to lower multipoles<sup>7</sup>. As discussed in detail in [4], this shift is consistently included in standard Boltzmann codes (such as CAMB [25] or CLASS [26]), since it is due to the square of the first-order convergence. The only term in (45) neglected by current CMB analyses is the contribution from the second-order convergence  $\kappa_2$ . However, as already mentioned before, the dominant terms in  $\kappa_2$  (those with 4 transverse derivatives) vanish on average, and the subdominant terms are negligible, roughly  $10^4$  times smaller than the first-order convergence square (see discussion at the end of section III A).

Finally, let us mention that the standard way of calculating the CMB angular power spectrum is not through (46) but rather through a calculation of the lensed multipoles  $a_{\ell m}$ . This calculation automatically selects the correct averaging procedure since the  $a_{\ell m}$  are defined through a (weighted) average over directions

$$a_{\ell m} = \int d\Omega_{\mathbf{n}_o} Y_{\ell m}(\mathbf{n}_o) T(\mathbf{n}_o). \quad (48)$$

The angular power spectrum is given by

$$\langle a_{\ell m} a_{\ell' m'} \rangle = C_\ell \delta_{\ell\ell'} \delta_{mm'}. \quad (49)$$

The ensemble average is therefore automatically taken *after* the integral over directions.

---

<sup>6</sup> The distance to supernovae in the training subset is indeed usually measured through the host galaxy using its surface brightness fluctuation or the Tully-Fisher relation to determine the galaxy intrinsic luminosity. Photons emitted by the galaxy follow the same path as the supernova's photons and experience therefore the same lensing corrections.

<sup>7</sup> Note that the shift in the observed position of the peaks is governed both by the displacement term and by the smoothing term. Since the smoothing term dominates over the displacement term, the shift is also to lower multipoles in (47), but its amplitude is smaller.

#### IV. CONCLUSIONS

In this paper we have shown that at second order in perturbation theory, averaging over the observed directions and taking an ensemble average are two distinct procedures, which do not commute. This comes from the fact that the observed direction is lensed – hence it is itself perturbed – and its perturbation may well be correlated with the observable under consideration.

This is especially relevant for distance measurements, the perturbations of which are intimately related to lensing. In particular, we have shown that the distance to the last scattering surface is decreased by lensing if one takes an average over directions, whereas it is increased if one takes an ensemble average. In a companion paper [4] we argue that the directional average is relevant for CMB observations and that consequently second-order lensing shifts the position of the peaks to lower multipoles. However, as we showed there, the change in the distance captures the essence of the change to the peaks but does not accurately capture the shift or damping of the peaks. Consequently, the CMB sound horizon does not act as a standard ruler when lensing is present.

For the analysis of diffuse observables like the CMB (actually also the BAO's), correlation functions are calculated and fitted to the model. Moreover, in multipole space, the average over directions is always taken before the ensemble average, and no ambiguity arises concerning the averaging procedure. For standard rulers and candles which are point sources, it is much more subtle to remove the bias arising from lensing in going from the observable to the model, and the distance measure chosen must be carefully understood. In particular, the notion of standard ruler should be extended to a standard area as the expectation value of an observed solid angle is preserved under lensing, which is not the case for an observed linear angle.

We have also shown that for specific observables there exist functions which do not acquire corrections from second-order perturbations. It is therefore a good observational strategy to consider these functions of the observables in order to avoid systematic errors from second-order perturbations.

*Note added:* While this paper was being finalised, an independent paper on the same topic appeared [27].

**Acknowledgements:** We thank Giovanni Marozzi for interesting and helpful discussions. RD is supported by the Swiss National Science Foundation. CC, RM and OU are supported by the South African National Research Foundation. RM and OU are supported by the South African Square Kilometre Array Project and RM acknowledges support from the UK Science & Technology Facilities Council (grant ST/K0090X/1).

- 
- [1] A. Lewis and A. Challinor, Phys. Rept. **429**, 1 (2006).
  - [2] Planck Collaboration: P. A. R. Ade, et al. (2015) [arXiv:1502.01591].
  - [3] C. Clarkson, O. Umeh, R. Maartens and R. Durrer, JCAP **1411**, 036 (2014) [arXiv:1405.7860].
  - [4] C. Bonvin, C. Clarkson, R. Durrer, R. Maartens and O. Umeh, JCAP, in press (2015) [arXiv:1503.07831].
  - [5] I. Ben-Dayan, M. Gasperini, G. Marozzi, F. Nugier, and G. Veneziano, Phys. Rev. Lett. **110**, 021301 (2013) [arXiv:1207.1286].
  - [6] O. Umeh, C. Clarkson and R. Maartens, Class. Quantum Grav. **31**, 202001 (2014) [arXiv:1207.2109].
  - [7] I. Ben-Dayan, G. Marozzi, F. Nugier and G. Veneziano, JCAP **1211**, 045 (2012), [arXiv:1209.4326].
  - [8] O. Umeh, C. Clarkson and R. Maartens, Class. Quantum Grav. **31**, 205001 (2014) [arXiv:1402.1933].
  - [9] G. Marozzi, Class. Quantum Grav. **32**, 045004 (2015) [arXiv:1406.1135].
  - [10] I. Ben-Dayan, R. Durrer, G. Marozzi and D. J. Schwarz, Phys. Rev. Lett. **112**, 221301 (2014) [arXiv:1401.7973].
  - [11] F. Bernardeau, C. Bonvin and F. Vernizzi, Phys. Rev. D **81**, 083002 (2010) [arXiv:0911.2244].
  - [12] F. Bernardeau, C. Bonvin, N. Van de Rijdt and F. Vernizzi, Phys. Rev. D **86**, 023001 (2012) [arXiv:1112.4430].
  - [13] D. Bertacca, R. Maartens, and C. Clarkson, JCAP **1409**, 037 (2014) [arXiv:1405.4403].
  - [14] D. Bertacca, R. Maartens, and C. Clarkson, JCAP **1411**, 013 (2014) [arXiv:1406.0319].
  - [15] J. Yoo and M. Zaldarriaga, Phys. Rev. D **90**, 023513 (2014) [arXiv:1406.4140].
  - [16] E. Di Dio, R. Durrer, G. Marozzi and F. Montanari, JCAP **1412**, 017 (2014) [arXiv:1407.0376].
  - [17] T. W. B. Kibble and R. Lieu, Astrophys. J **632**, 718 (2005) [arXiv:astro-ph/0412275].
  - [18] S. Weinberg, Astrophys. J **208**, L1 (1976).
  - [19] G. F. R. Ellis, B. A. Basset and P. K. S. Dunsby, Class. Quantum Grav. **15**, 2345 (1998) [arXiv:gr-qc/9801092].
  - [20] R. Wojtak, T. M. Davis and J. Wiis (2015) [arXiv:1504.00718].
  - [21] C. Bonvin, R. Durrer and M. A. Gasparini, Phys. Rev. D **73**, 023523 (2006) [Erratum: ibid. D **85**, 029901 (2012)] [arXiv:astro-ph/0511183].
  - [22] L. Hui and P. B. Greene, Phys. Rev. D **73**, 123526 (2006) [arXiv:astro-ph/0512159].

- [23] SNLS Collaboration: T. Kronborg et al., *Astron. Astrophys.* **514**, 44 (2010), [arXiv:1002.1249];  
SNLS Collaboration: M. Sullivan et al., *Astrophys. J.* **737**, 102 (2011) [arXiv:1104.1444].
- [24] M. M. Phillips, *Astrophys. J.* **413**, L105 (1993);  
A. G. Riess, W. H. Press and R. P. Kirshner, *Astrophys. J.* **438** (1995).
- [25] A. Lewis, A. Challinor and A. Lasenby, *Astrophys. J.* **538**, 473 (2000) [arXiv:astro-ph/9911177].
- [26] D. Blas, J. Lesgourgues and T. Tram, *JCAP* **1107**, 034 (2011) [arXiv:1104.2933].
- [27] N. Kaiser and J. A. Peacock (2015) [arXiv:1503.08506].