Lawrence Berkeley National Laboratory

Title
Improved renormalization group computation of likelihood functions for cosmological data sets

Author
McDonald, Patrick

Improved renormalization group computation of likelihood functions for cosmological data sets

Patrick McDonald[1, *]

[1]*Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA*
(Dated: June 24, 2019)

Evaluation of likelihood functions for cosmological large scale structure data sets (including CMB, galaxy redshift surveys, etc.) naturally involves marginalization, i.e., integration, over an unknown underlying random signal field. Recently, I showed how a renormalization group method can be used to carry out this integration efficiently by first integrating out the smallest scale structure, i.e., localized structure on the scale of differences between nearby data cells, then combining adjacent cells in a coarse graining step, then repeating this process over and over until all scales have been integrated. Here I extend the formulation in several ways in order to reduce the prefactor on the method's linear scaling with data set size. The key improvement is showing how to integrate out the difference between specific adjacent cells before summing them in the coarse graining step, compared to the original formulation in which small-scale fluctuations were integrated more generally. I suggest some other improvements in details of the scheme, including showing how to perform the integration around a maximum likelihood estimate for the underlying random field. In the end, an accurate likelihood computation for a million-cell Gaussian test data set runs in two minutes on my laptop, with room for further optimization and straightforward parallelization.

I. INTRODUCTION

[1] presented a new method to evaluate large scale structure likelihood functions, inspired by renormalization group (RG) ideas from quantum field theory [e.g., 2, 3]. This paper is a followup to that one, so some of the pedagogical discussion and derivations there will not be repeated here. To recap the basics: the fact that structure in the Universe starts as an almost perfectly Gaussian random field and evolves in a computable way on the largest scales [e.g., 4–6] suggests a statistically rigorous first-principles likelihood analysis can be used to extract information on cosmological models from observational data sets [e.g., 7–11]. Generally, we have a data vector $\mathbf{o}$, some relatively small number of global cosmological parameters we want to measure, $\boldsymbol{\theta}$, and a random field we'd like to marginalize over, $\boldsymbol{\phi}$. ($\boldsymbol{\phi}$ could be a variety of different things, depending on the data set and theoretical setup, e.g., the underlying true temperature field for CMB, the linear regime density and/or potential fields for a galaxy redshift survey modeled by traditional perturbation theory, the evolving displacement field in the functional integral formulation of [6], etc.) Starting with Bayes' rule $L(\boldsymbol{\theta}, \boldsymbol{\phi}|\mathbf{o})L(\mathbf{o}) = L(\mathbf{o}|\boldsymbol{\theta}, \boldsymbol{\phi})L(\boldsymbol{\phi}, \boldsymbol{\theta}) = L(\mathbf{o}|\boldsymbol{\theta}, \boldsymbol{\phi})L(\boldsymbol{\phi}|\boldsymbol{\theta})L(\boldsymbol{\theta})$ we obtain

$$L(\boldsymbol{\theta}|\mathbf{o}) = \int d\boldsymbol{\phi} \ L(\boldsymbol{\theta}, \boldsymbol{\phi}|\mathbf{o}) = \int d\boldsymbol{\phi} \ L(\mathbf{o}|\boldsymbol{\theta}, \boldsymbol{\phi})L(\boldsymbol{\phi}|\boldsymbol{\theta}) \ , \tag{1}$$

where I have dropped $L(\mathbf{o})$ which has no parameter dependence and the prior $L(\boldsymbol{\theta})$ which plays no role in this discussion because it can be pulled out of the integral. I have highlighted the usual cosmological form where some of the cosmological parameters determine a prior on the signal field, $L(\boldsymbol{\phi}|\boldsymbol{\theta})$, and then there is some likelihood for the observable given $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$, $L(\mathbf{o}|\boldsymbol{\theta}, \boldsymbol{\phi})$. It is this $\boldsymbol{\phi}$ integral that we need to carry out. Generally, we can take at least part of $L(\boldsymbol{\phi}|\boldsymbol{\theta})$, $L_G(\boldsymbol{\phi}|\boldsymbol{\theta})$, to be Gaussian, defined by its covariance, $\mathbf{P}(\boldsymbol{\theta})$. In this case we have

$$L(\boldsymbol{\theta}|\mathbf{o}) = \int d\boldsymbol{\phi} \ e^{-\frac{1}{2}\boldsymbol{\phi}^t \mathbf{P}^{-1}\boldsymbol{\phi} - \frac{1}{2}\mathrm{Tr}\ln(2\pi\mathbf{P}) + \ln L_{\mathrm{NG}}(\boldsymbol{\phi}|\boldsymbol{\theta}) + \ln L(\mathbf{o}|\boldsymbol{\theta}, \boldsymbol{\phi})} \ , \tag{2}$$

where I have used $\ln\det(\mathbf{P}) = \mathrm{Tr}\ln(\mathbf{P})$ and defined $\ln L_{\mathrm{NG}}(\boldsymbol{\phi}|\boldsymbol{\theta}) \equiv \ln L(\boldsymbol{\phi}|\boldsymbol{\theta}) - \ln L_G(\boldsymbol{\phi}|\boldsymbol{\theta})$. (Even for what we call non-Gaussian initial conditions [e.g., 12–17], the observable can often if not always be written as a function of an underlying Gaussian random field, i.e., no L_{NG} needed, and in other scenarios like [6] where the natural $\boldsymbol{\phi}$ is not Gaussian, there is still a natural Gaussian piece.) Less generally but still often usefully (e.g., for primary CMB and large scale galaxy clustering ignoring primordial non-Gaussianity) we can take $\ln L_{\mathrm{NG}} = 0$ and $L(\mathbf{o}|\boldsymbol{\theta}, \boldsymbol{\phi})$ to be

---

Gaussian by assuming $\mathbf{o}$ is linearly related to $\boldsymbol{\phi}$, i.e., $\mathbf{o} = \boldsymbol{\mu} + \mathbf{R}\boldsymbol{\phi} + \boldsymbol{\epsilon}$ where $\boldsymbol{\mu}$ is the mean vector, $\mathbf{R}$ is a linear response matrix, and $\boldsymbol{\epsilon}$ is Gaussian observational noise with covariance matrix $\mathbf{N}$. Then we have

$$L_{\text{Gaussian}}(\boldsymbol{\theta}|\mathbf{o}) = \int d\boldsymbol{\phi}\ e^{-\frac{1}{2}\boldsymbol{\phi}^t \mathbf{P}^{-1}\boldsymbol{\phi} - \frac{1}{2}\operatorname{Tr}\ln(2\pi\mathbf{P}) - \frac{1}{2}(\mathbf{o}-\boldsymbol{\mu}-\mathbf{R}\boldsymbol{\phi})^t \mathbf{N}^{-1}(\mathbf{o}-\boldsymbol{\mu}-\mathbf{R}\boldsymbol{\phi}) - \frac{1}{2}\operatorname{Tr}\ln(2\pi\mathbf{N})} \tag{3}$$

$$= e^{-\frac{1}{2}(\mathbf{o}-\boldsymbol{\mu})^t \mathbf{C}^{-1}(\mathbf{o}-\boldsymbol{\mu}) - \frac{1}{2}\operatorname{Tr}\ln(2\pi\mathbf{C})}$$

where in the last line the integration has been carried out analytically, with $\mathbf{C} \equiv \mathbf{N} + \mathbf{R}\mathbf{P}\mathbf{R}^t$. Even this analytic integration does not really solve the Gaussian problem, however, as the time to calculate $\mathbf{C}^{-1}$ and $\det(\mathbf{C})$ (or its derivatives) by brute force numerical linear algebra routines scales like N^3, where N is the size of the data set, which becomes prohibitively slow for large data sets. The RG approach of [1] addresses the Gaussian scenario by doing the $\boldsymbol{\phi}$ integral in a different way that produces the result directly as a number instead of these matrix expressions, and can also be applied to non-Gaussian scenarios. Note that, as discussed in [1], the approach can also be used to directly compute derivatives of $\ln L(\boldsymbol{\theta}|\mathbf{o})$ with respect to $\boldsymbol{\theta}$, not just the value at one choice of $\boldsymbol{\theta}$, by passing the derivative inside the $\boldsymbol{\phi}$ integral to produce a new integral. Traditional power spectrum estimation can be done by taking $\boldsymbol{\theta}$ to parameterize $\mathbf{P}(\boldsymbol{\theta})$ by amplitudes in k bands.

In spite of the fact that fairly fast methods to evaluate at least the Gaussian likelihood [Eq. (3)] have existed for a long time [e.g., 18–22], more often in practice data analysts compute summary statistics not explicitly based on likelihood functions [e.g., 23, 24], calibrating their parameter dependence and covariance by computing the same statistics on mock data sets. It is not entirely clear why existing likelihood-based methods are not used more often, and in [1] I was cautious about advocating immediate implementation of the RG approach. One question was if the prefactor on the linear scaling of computation time with data set size for this method might be so large as to make it significantly slower than others. This paper demonstrates that this is not a significant obstacle. At two minutes to accurately compute the likelihood function for a million-cell Gaussian test data set, the method is as fast as any that takes more than a few well-preconditioned conjugate gradient maximum likelihood solutions for the same data set (i.e., as fast as any method I know of, barring the possibility that my Julia implementation of conjugate gradient maximum likelihood is unfairly slow). The only reason not to implement this is if you believe the whole idea of likelihood-based analysis is a distraction. That would not necessarily be an entirely unreasonable position. E.g., if you believe that there is a lot of reliable cosmological constraining power to be gained from the deeply non-linear regime, heuristic summary statistics/"machine learning," combined with exhaustive mocks/simulations is probably the only way to extract it. To me, however, the likelihood+RG approach proposed here seems like an appealing path to large scale analysis, especially for incorporating weakly nonlinear information (e.g., without the need to explicitly estimate a bispectrum and its covariance).

This paper lays out a series of essentially technical improvements to the basic approach presented in [1]. See that paper for a derivation of the general RG equation and some more pedagogical discussion. Some of that basics are explained in less detail here when they can be read there.

II. REVISED FORMULATION

A. Master RG equation

Consider the general functional integral over some field $\boldsymbol{\phi}$,

$$I \equiv \int d\boldsymbol{\phi}\ e^{-S(\boldsymbol{\phi})} \equiv \int d\boldsymbol{\phi}\ e^{-\frac{1}{2}\boldsymbol{\phi}^t \mathbf{Q}^{-1}\boldsymbol{\phi} - \frac{1}{2}\operatorname{Tr}\ln(2\pi\mathbf{Q}) - S_I(\boldsymbol{\phi})}\ . \tag{4}$$

The connection to our cosmological likelihood functions, Eq. (2), is obvious, but not necessary for this subsection. Suppose that $\mathbf{Q} \to 0$, i.e., $\mathbf{Q}^{-1}$ goes to infinity (all its eigenvalues). In that limit the $\mathbf{Q}$ part of I becomes a representation of the delta function and it is clear that $I(\mathbf{Q} \to 0) \to \exp[-S_I(0)]$, i.e., the integral can be done trivially. Generally, however, $\mathbf{Q}$ is not sufficiently small so if we want to do the integral this way we need to change $\mathbf{Q}$ to take it to zero. But we can't simply change $\mathbf{Q}$ because that will change the value of I, the integral we are trying to perform. If we want to change $\mathbf{Q}$ while preserving I we need to simultaneously change S_I. The renormalization group equation tells us how to do this. Guided by, e.g., [3], [1] showed that we can preserve the value of I if the following differential equation is satisfied:

$$S_I' = \frac{1}{2}\frac{\partial S_I}{\partial \boldsymbol{\phi}^t}\mathbf{Q}'\frac{\partial S_I}{\partial \boldsymbol{\phi}} - \frac{1}{2}\operatorname{Tr}\left[\mathbf{Q}'\frac{\partial^2 S_I}{\partial \boldsymbol{\phi}\partial \boldsymbol{\phi}^t}\right]\ , \tag{5}$$

where we parameterize the evolution by λ, i.e., $\mathbf{Q} = \mathbf{Q}(\lambda)$, $S_I = S_I(\lambda)$, and the prime means derivative with respect to λ, where $\mathbf{Q}(\lambda = 0)$ and $S_I(\lambda = 0)$ represent the original elements of the integral. (Note that, relative to Eq. (7) of [1], I have moved the normalization constant $\mathcal{N}$ into S_I, after extracting $\mathrm{Tr} \ln \mathbf{Q}$ from it to keep the integral unit normalized when $S_I = 0$.) This formula is pure math, i.e., it assumes essentially nothing about $\mathbf{Q}$, $\mathbf{Q}'$, and $S_I(\boldsymbol{\phi})$. Typically λ will represent a length scale, where structure in $\mathbf{Q}$ has already been erased on smaller scales, and $\mathbf{Q}'$ is doing the job of erasing it on scale λ, but Eq. (5) applies to any infinitesimal change in $\mathbf{Q}$.

B.   Application to Gaussian cosmological data

As in [1], I will demonstrate the calculation for a purely Gaussian example, i.e., $S_I(\boldsymbol{\phi})$ at most quadratic in $\boldsymbol{\phi}$. This is a special case only—Eq. (5) applies for any $S_I(\boldsymbol{\phi})$. The likelihood function will be Eq. (3), except for simplicity I will set $\boldsymbol{\mu} = 0$, i.e., I take

$$L(\boldsymbol{\theta}|\mathbf{o}) = \int d\boldsymbol{\phi} \; \frac{e^{-\frac{1}{2}\boldsymbol{\phi}^t \mathbf{P}^{-1}\boldsymbol{\phi} - \frac{1}{2}(\mathbf{o}-\mathbf{R}\boldsymbol{\phi})^t \mathbf{N}^{-1}(\mathbf{o}-\mathbf{R}\boldsymbol{\phi})}}{\sqrt{\det(2\pi\mathbf{P})\det(2\pi\mathbf{N})}} \; . \tag{6}$$

For the RG method to be efficient, the linear response matrix $\mathbf{R}$ and the observational noise $\mathbf{N}$ cannot be completely arbitrary. Ideally $\mathbf{R}$ should be fairly short range, e.g., a CMB telescope beam convolution or redshift space cells in which we have counted galaxies. Similarly, $\mathbf{N}$ should be short-range, e.g., diagonal for uncorrelated noise. The general approach can be adapted for special kinds of deviations from short range $\mathbf{R}$ or $\mathbf{N}$, but I will assume they are short range here. I generally assume the problem can be formulated to make $\mathbf{P}$ translation invariant (i.e., diagonal in Fourier space), although slow evolution in statistics can easily be accommodated. It is potentially useful to change integration variables to $\boldsymbol{\delta} \equiv \boldsymbol{\phi} - \boldsymbol{\phi}_0$, where $\boldsymbol{\phi}_0$ is some constant field specified by hand. We plan to make $\boldsymbol{\phi}_0$ the maximum likelihood field, but do not need to assume that. Substituting this into Eq. (6) and comparing to Eq. (4), understanding that $\boldsymbol{\phi}$ in Eq. (4) is a dummy variable so we can just as well replace it with $\boldsymbol{\delta}$, we see that the general integral I in Eq. (4) is equivalent to the the Gaussian cosmological $L(\boldsymbol{\theta}|\mathbf{o})$ if we define

$$\mathbf{Q}^{-1}(0) \equiv \mathbf{P}^{-1} + \mathbf{A}_\star \; . \tag{7}$$

and

$$\begin{aligned}
S_I(0) &\equiv \boldsymbol{\phi}_0^t \mathbf{P}^{-1}\boldsymbol{\delta} + \frac{1}{2}\boldsymbol{\phi}_0^t \mathbf{P}^{-1}\boldsymbol{\phi}_0 + \frac{1}{2}(\mathbf{o} - \mathbf{R}\boldsymbol{\phi}_0 - \mathbf{R}\boldsymbol{\delta})^t \mathbf{N}^{-1}(\mathbf{o} - \mathbf{R}\boldsymbol{\phi}_0 - \mathbf{R}\boldsymbol{\delta}) + \frac{1}{2}\mathrm{Tr}\ln(2\pi\mathbf{N}) \\
&\quad - \frac{1}{2}\boldsymbol{\delta}^t \mathbf{A}_\star \boldsymbol{\delta} + \frac{1}{2}\mathrm{Tr}\ln(\mathbf{I} + \mathbf{A}_\star \mathbf{P}) \\
&= \frac{1}{2}\boldsymbol{\delta}^t \left(\mathbf{R}^t \mathbf{N}^{-1}\mathbf{R} - \mathbf{A}_\star\right)\boldsymbol{\delta} - \left[(\mathbf{o} - \mathbf{R}\boldsymbol{\phi}_0)^t \mathbf{N}^{-1}\mathbf{R} - \boldsymbol{\phi}_0^t \mathbf{P}^{-1}\right]\boldsymbol{\delta} \\
&\quad + \frac{1}{2}\boldsymbol{\phi}_0^t \mathbf{P}^{-1}\boldsymbol{\phi}_0 + \frac{1}{2}(\mathbf{o} - \mathbf{R}\boldsymbol{\phi}_0)^t \mathbf{N}^{-1}(\mathbf{o} - \mathbf{R}\boldsymbol{\phi}_0) + \frac{1}{2}\mathrm{Tr}\ln(2\pi\mathbf{N}) + \frac{1}{2}\mathrm{Tr}\ln(\mathbf{I} + \mathbf{A}_\star \mathbf{P}) \; .
\end{aligned} \tag{8}$$

The reason for subtracting $\frac{1}{2}\boldsymbol{\delta}^t \mathbf{A}_\star \boldsymbol{\delta}$ from $S_I(0)$ and adding it to the $\mathbf{Q}^{-1}$ term (adding zero overall, with $\mathbf{A}_\star$ an as yet unspecified matrix) will become clear below.

As in [1], the evolving Gaussian $S_I(\lambda)$ is represented numerically by the evolving coefficients $\mathbf{A}(\lambda)$, $\mathbf{b}(\lambda)$, and $\mathcal{N}(\lambda)$ of the general form

$$S_I(\lambda) \equiv \frac{1}{2}\boldsymbol{\delta}^t \; \mathbf{A}(\lambda) \; \boldsymbol{\delta} - \mathbf{b}^t(\lambda) \; \boldsymbol{\delta} + \mathcal{N}(\lambda) \; . \tag{9}$$

Comparison to Eq. (8) for $S_I(0)$ sets the initial conditions for $\mathbf{A}$, $\mathbf{b}$, and $\mathcal{N}$:

$$\mathbf{A}(0) \equiv \mathbf{R}^t \mathbf{N}^{-1}\mathbf{R} - \mathbf{A}_\star \tag{10}$$

$$\mathbf{b}(0) \equiv \mathbf{R}^t \mathbf{N}^{-1}(\mathbf{o} - \mathbf{R}\boldsymbol{\phi}_0) - \mathbf{P}^{-1}\boldsymbol{\phi}_0 \tag{11}$$

and

$$\mathcal{N}(0) \equiv \frac{1}{2}\boldsymbol{\phi}_0^t \mathbf{P}^{-1}\boldsymbol{\phi}_0 + \frac{1}{2}(\mathbf{o} - \mathbf{R}\boldsymbol{\phi}_0)^t \mathbf{N}^{-1}(\mathbf{o} - \mathbf{R}\boldsymbol{\phi}_0) + \frac{1}{2}\mathrm{Tr}\ln(2\pi\mathbf{N}) + \frac{1}{2}\mathrm{Tr}\ln(\mathbf{I} + \mathbf{A}_\star \mathbf{P}) \; . \tag{12}$$

Plugging Eq. (9) into Eq. (5) we find the flow equations for $\mathbf{A}$, $\mathbf{b}$, and $\mathcal{N}$:

$$\mathbf{A}' = \mathbf{A}\mathbf{Q}'\mathbf{A} \tag{13}$$

$$\mathbf{b}' = \mathbf{A}\mathbf{Q}'\mathbf{b} \tag{14}$$

$$\mathcal{N}' = \frac{1}{2}\mathbf{b}^t\mathbf{Q}'\mathbf{b} - \frac{1}{2}\mathrm{Tr}\left[\mathbf{A}\mathbf{Q}'\right] . \tag{15}$$

Note that if ϕ_0 is the maximum likelihood field (for given values of $\mathbf{P}$, $\mathbf{R}$, etc.), $\mathbf{b} = \mathbf{b}(0) = 0$. If the problem happened to be statistically homogeneous (translation invariant), we could set $\mathbf{A}_\star = \mathbf{R}^t\mathbf{N}^{-1}\mathbf{R}$ to make $\mathbf{A} = \mathbf{A}(0) = 0$. In that case there would be no evolution—$\mathcal{N}(0)$ would simply be the answer. This is the point of $\mathbf{A}_\star$, i.e., if we choose it to be as close as possible to $\mathbf{R}^t\mathbf{N}^{-1}\mathbf{R}$, we can reduce the RG evolution to be a minimal correction due to statistical inhomogeneities. The limitation, i.e., why $\mathbf{A}_\star$ generally can only approximate $\mathbf{R}^t\mathbf{N}^{-1}\mathbf{R}$, is that $\mathbf{A}_\star$ must maintain the symmetries necessary to allow us to efficiently evaluate $\mathrm{Tr}\ln\left(\mathbf{I} + \mathbf{A}_\star\mathbf{P}\right)$ in Eq. (12), e.g., in Fourier space, to set the initial value of $\mathcal{N}$.

In terms of these definitions, the result of formal analytic integration is

$$L(\boldsymbol{\theta}|\mathbf{o}) = e^{\frac{1}{2}\mathbf{b}^t\left(\mathbf{Q}^{-1}+\mathbf{A}\right)^{-1}\mathbf{b} - \mathcal{N} - \frac{1}{2}\mathrm{Tr}\ln(\mathbf{I}+\mathbf{A}\mathbf{Q})} . \tag{16}$$

We can use this formula once the components have been coarse-grained sufficiently to allow brute force linear algebra. To be clear: if we plug $\mathbf{A}(0)$, $\mathbf{b}(0)$, $\mathcal{N}(0)$, and $\mathbf{Q}(0)$ into this equation, it becomes precisely the analytic integration result in Eq. (3) (with $\boldsymbol{\mu} = 0$). The difference is that as these quantities evolve and are coarse grained their dimensions become smaller, with the result of the small-scale integration that has been performed stored in the simple number $\mathcal{N}$. See [1] for more discussion.

C.   Integrating out the difference between adjacent cells

In [1] I used

$$\mathbf{Q}^{-1}(\lambda) = \mathbf{Q}^{-1}(0) + \mathbf{K}(\lambda) \tag{17}$$

where $\mathbf{K}(\lambda \to \infty) \to \infty$ to suppress fluctuations. I mentioned the potentially cleaner possibility

$$\mathbf{Q}(\lambda) = \mathbf{Q}(0)\mathbf{W}(\lambda) \tag{18}$$

where $\mathbf{W}(\lambda \to \infty) \to 0$, e.g., $W(k, \lambda) \equiv e^{-k^2\lambda^2}$. Either of these was envisioned to suppress fluctuations in a smooth, homogeneous way (i.e., with no explicit connection to the data cell structure), starting from small scales to large. Once fluctuations were sufficiently suppressed on the scale of data cells, adjacent cells were combined, i.e., adjacent elements in $\mathbf{b}$ and the corresponding 2×2 block in $\mathbf{A}$ were summed. This worked well enough, but the number of elements that I needed to store in $\mathbf{A}$, which determines the speed of computation, seemed surprisingly large.

Here I introduce a new possibility, to more explicitly integrate out the fluctuations between pairs of cells that we are going to combine (see Appendix A for an alternative version of this idea). Given covariance matrix $\mathbf{Q}^1$ for some vector, we know that the covariance for a new vector where each adjacent pair of elements is replaced by one element with its average, $\mathbf{Q}^{2c}$, is simply given by the average of the appropriate 2×2 blocks of $\mathbf{Q}^1$, e.g., $Q_{11}^{2c} = \frac{1}{4}(Q_{11}^1 + Q_{12}^1 + Q_{21}^1 + Q_{22}^1)$, $Q_{12}^{2c} = \frac{1}{4}(Q_{13}^1 + Q_{14}^1 + Q_{23}^1 + Q_{24}^1)$, etc. This makes clear that if we define $\mathbf{Q}' \propto \mathbf{Q}^2 - \mathbf{Q}^1$, where $\mathbf{Q}^2$ is the matrix of equivalent dimension to $\mathbf{Q}^1$ but with the 2×2 blocks that will be compressed to $\mathbf{Q}^{2c}$ replaced by their average (e.g., $Q_{11}^2 = Q_{12}^2 = Q_{21}^2 = Q_{22}^2 = Q_{11}^{2c}$), we can straightforwardly evolve Eq. (5) from a starting $\mathbf{Q}^1$ to ending $\mathbf{Q}^2$, followed by a coarse graining combination of cells, and repeat. Formally, for each iteration what we are doing is defining $\mathbf{Q}(\lambda) = \mathbf{Q}^1 + \lambda(\mathbf{Q}^2 - \mathbf{Q}^1)$ so that $\mathbf{Q}' \equiv d\mathbf{Q}/d\lambda = \mathbf{Q}^2 - \mathbf{Q}^1$, and solving the differential equation (5) for λ running from 0 [where $\mathbf{Q}(\lambda = 0) = \mathbf{Q}^1$] to 1 [where $\mathbf{Q}(\lambda = 1) = \mathbf{Q}^2$].

The obvious problem here is that generally $\mathbf{Q}^2 - \mathbf{Q}^1$ is a dense matrix, which we can't have if the method is to be fast. The key to the RG approach working is that elements of $\mathbf{Q}^2 - \mathbf{Q}^1$ will generally be small very far off-diagonal, i.e., physically we do not expect the correlation at wide separations to change much when the separation is changed by a small fractional amount. To put it another way, we do not expect to need to use small cells when measuring correlations at wide separations. This allows us to drop most elements of $\mathbf{Q}^2 - \mathbf{Q}^1$, keeping it, and $\mathbf{A}$ as influenced by it, sparse. The closest thing to an exception to this "no fine structure at large separations" rule that comes to mind

is the BAO feature—a relatively narrow bump at wide separation. Considering such a thing, we observe that it is only necessary for $\mathbf{Q}'$ to remain sparse, not strictly near-diagonal, i.e., we can if necessary include a strip of elements somewhere off-diagonal in $\mathbf{Q}'$, propagate this into $\mathbf{A}$, etc., as long as there are not too many of these elements.

Operationally, this program is surprisingly straightforward. I start by computing one full row of $\mathbf{Q}(0) = \left(\mathbf{P}^{-1} + \mathbf{A}_\star\right)^{-1}$. This is basically just a standard computation of a correlation function given a power spectrum, i.e., this matrix obeys translation invariance by construction, so its elements are a function only of separation, inverses can be done in Fourier space, and one row is all that is necessary to capture the full matrix. This $\mathbf{Q}(0)$ becomes $\mathbf{Q}^1$ described above and I compute the first two rows of $\mathbf{Q}^2$ (the 2×2 block-averaged matrix) directly from it. From this I compute the full sparse $\mathbf{Q}'$ including only elements above some threshold. I define the threshold to be some fraction of the maximum absolute value of $\mathbf{Q}'$, called $\epsilon_{\mathbf{Q}'}$, i.e., I keep elements with $|Q'_{ij}| > \epsilon_{\mathbf{Q}'}\max|\mathbf{Q}'|$. Note that this makes no assumption about the structure of $\mathbf{Q}'$, e.g., an off-diagonal stripe due to something like BAO will be propagated if it passes the threshold.

After evolving $\mathbf{A}$, $\mathbf{b}$, and $\mathcal{N}$ through Eqs. (13)-(15), they, along with $\mathbf{Q}$ as represented by a single row, are coarse-grained by factors of two (i.e., elements summed in the case of $\mathbf{b}$ and $\mathbf{A}$ and averaged in the case of $\mathbf{Q}$) and the next iteration proceeds exactly as before. All of the problem-specific details go into the construction of $\mathbf{Q}(0)$, $\mathbf{A}(0)$, $\mathbf{b}(0)$, and $\mathcal{N}(0)$—after that the algorithm proceeds essentially identically for any problem. After enough iterations the effective data set becomes small enough to finish the calculation by brute force using the analytic integral formula, Eq. (16).

Note that, while my test problems will be one dimensional, where factors of two coarse graining by combining adjacent pixels is the obvious thing to do, there is no obvious reason not to do this as well in higher dimensions. On a cartesian grid we can combine adjacent cells in one direction at a time. On a sphere, a hierarchical block of four HEALPixels [25] can be combined in two steps of pair combinations. However, it should also be possible to generalize the method to combine more than two cells at a time. $\mathbf{Q}^2$ as discussed above just needs to represent the appropriately averaged covariance.

D. Sparsification

While the $\mathbf{Q}'$ cut discussed above limits the range in $\mathbf{A}$ somewhat, in practice I find that the evolution of $\mathbf{A}$ produces many small elements that do not need to be fully propagated for accuracy and slow down the calculation significantly. In [1] I maintained the sparsity of $\mathbf{A}$ by computing elements only out to some maximum separation, taken to be a multiple of the RG distance scale λ. Here I suggest a potentially more generally adaptive method, along the lines of the element size cut discussed above involving $\epsilon_{\mathbf{Q}'}$. The key equation numerically is Eq. (13), because the matrix products there dominate the computation time. To control this, I introduce two more numerical parameters. When evaluating $\mathbf{AQ'A}$, I first trim $\mathbf{A}$ using another threshold parameter, $\epsilon_{\mathbf{A}}$, again basing the cut on the absolute value of elements relative to the maximum absolute value. To be clear, I am not permanently dropping part of the stored, evolving $\mathbf{A}$, only the matrix used to compute $\mathbf{AQ'A}$. I apply another similar cut defined by $\epsilon_{\mathbf{A}'}$ to $\mathbf{A}' = \mathbf{AQ'A}$, before using it to update $\mathbf{A}$ in each λ step. In practice, for simplicity, I only use one of these two cuts at a time, finding the $\epsilon_{\mathbf{A}}$ cut to be slightly more efficient in my test problems.

E. Numerical demonstration

For numerical tests I use one dimensional scenarios similar to [1]. I use signal power spectrum $P(k) = A(k/k_p)^\gamma \exp(-k^2)$ with $\gamma = 0$ or -0.5, where k is measured in units of the data cell size. I add unit variance noise to each cell. I generate mock data with $A_0 = 1$ and calculate likelihoods as a function of A. I use pivot $k_p = 0.1$ so that the $\gamma = -0.5$ case has both signal and noise dominated ranges of scales. To be sure the test covers both fine structure and edges, I create statistically inhomogeneous data sets where the rms noise level in every fourth cell is multiplied by a factor of 10, and the noise in the last quarter of the data vector is similarly multiplied.

It is more difficult to make a non-trivial test with the innovations in this paper, because if I assume periodic data with homogeneous noise so that I can compute the exact likelihood to compare to using FFTs, the obvious choice of $\mathbf{A}_\star$ sets $\mathbf{A} \equiv 0$ so the RG evolution is almost trivial. If I also find the maximum likelihood field to use for ϕ_0, so that $\mathbf{b} \equiv 0$, it is completely trivial. For this reason I only do tests with inhomogeneous data in this paper, first on data sets small enough to compute the exact likelihood by brute force linear algebra, demonstrating that the RG method works precisely in the appropriate limit of the numerical parameters, then with large data sets where the truth is determined by using much better than necessary values for numerical parameters.

After some experimentation, my standard numerical parameter settings are as follows: $\mathbf{A}_\star$ is set to $0.47N_0^{-1}$, where N_0 is the noise power in the good part of the data—this sets the accumulated $\text{Tr}[\mathbf{AQ'}]$ term in Eq. (15) to

approximately zero (the results are insensitive to the exact value of $\mathbf{A}_\star$, as long as it is reasonable). I specify the number of mid-point method λ steps per factor of two coarse graining by a numerical parameter $N_{d\mathbf{Q}'}$. My standard setting is $N_{d\mathbf{Q}'} = 8$ (in an advanced version of the method, one could try to apply all the usual tricks for solving differential equations numerically). I set $\epsilon_{\mathbf{Q}'} = 0.02$, and $\epsilon_{\mathbf{A}} = 0.0005$.

1. Small problems

I first do some tests with $N = 16384$, where we can still pretty quickly compute the exact likelihood by brute force linear algebra, shown in Fig. 1. The results are good, by construction of course. Both using a maximum likelihood
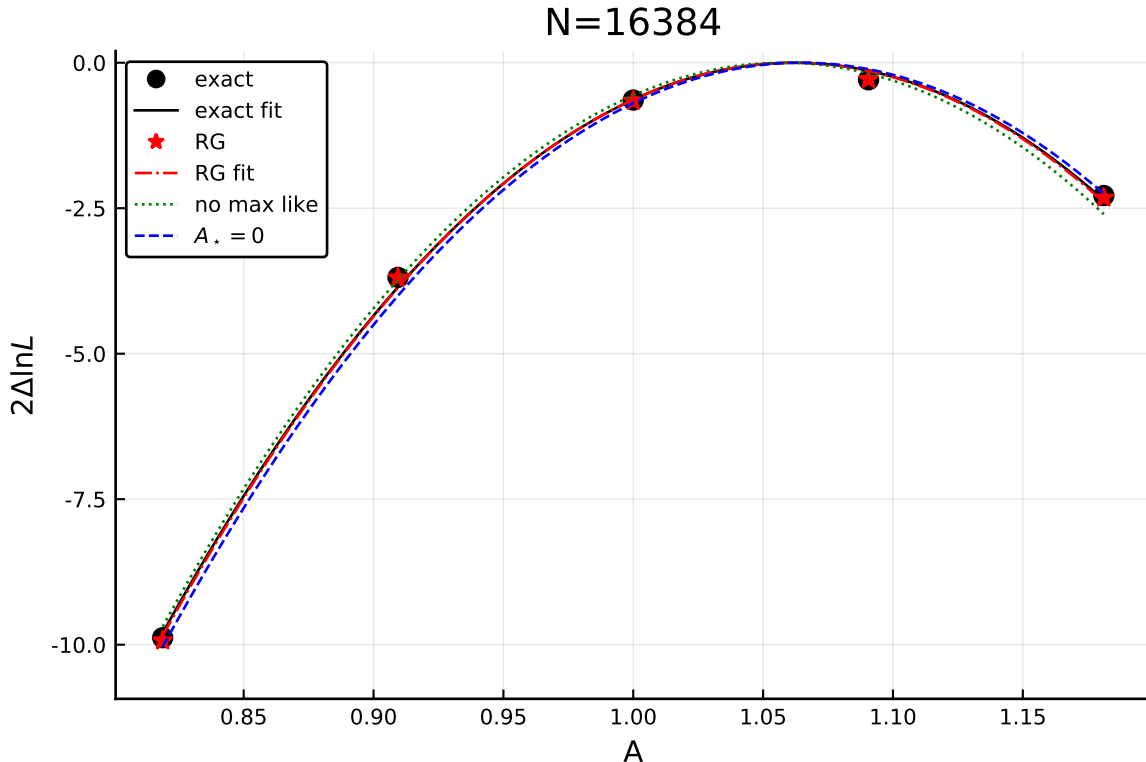

FIG. 1. $N = 16384$ test. The exact likelihood is computed by brute force linear algebra at five representative values of A. To guide the eye, I fit a quadratic polynomial to the points, using this to define the maximum. I use the RG method to compute the likelihood at the same five points, and similarly plot a quadratic fit representation—the results are essentially indistinguishable in this example. For the case with no maximum likelihood field, i.e., $\phi_0 = 0$, and the case with $\mathbf{A}_\star = 0$, I plot only the fitted quadratic, to reduce clutter.

ϕ_0 and using $\mathbf{A}_\star$ to remove the mean effect of $\mathbf{A}$ from the evolution improve the accuracy at fixed parameter settings, although for these settings (which were driven by larger data sets) the difference is not critical. This example has $\gamma = -0.5$, which is generally a little more difficult for the algorithm than $\gamma = 0$.

2. Large problems

If we are convinced that the algorithm works in the sense of producing accurate results in the appropriate limit of numerical parameters, we can do non-trivial large-scale tests by simply looking for convergence as the numerical parameters are changed, i.e., we assume that if there is convergence it is to the correct result. Figure 2 shows an $N = 524288$ test, for $\gamma = -0.5$ again. The results are again excellent. One might guess based on these figures that my numerical parameter settings are too conservative, i.e., that I could loosen them to achieve better speed. This is not actually true—there seems to be some cancelation of errors that makes the results in these particular examples so perfect, and they go bad very quickly if the parameters are loosened.
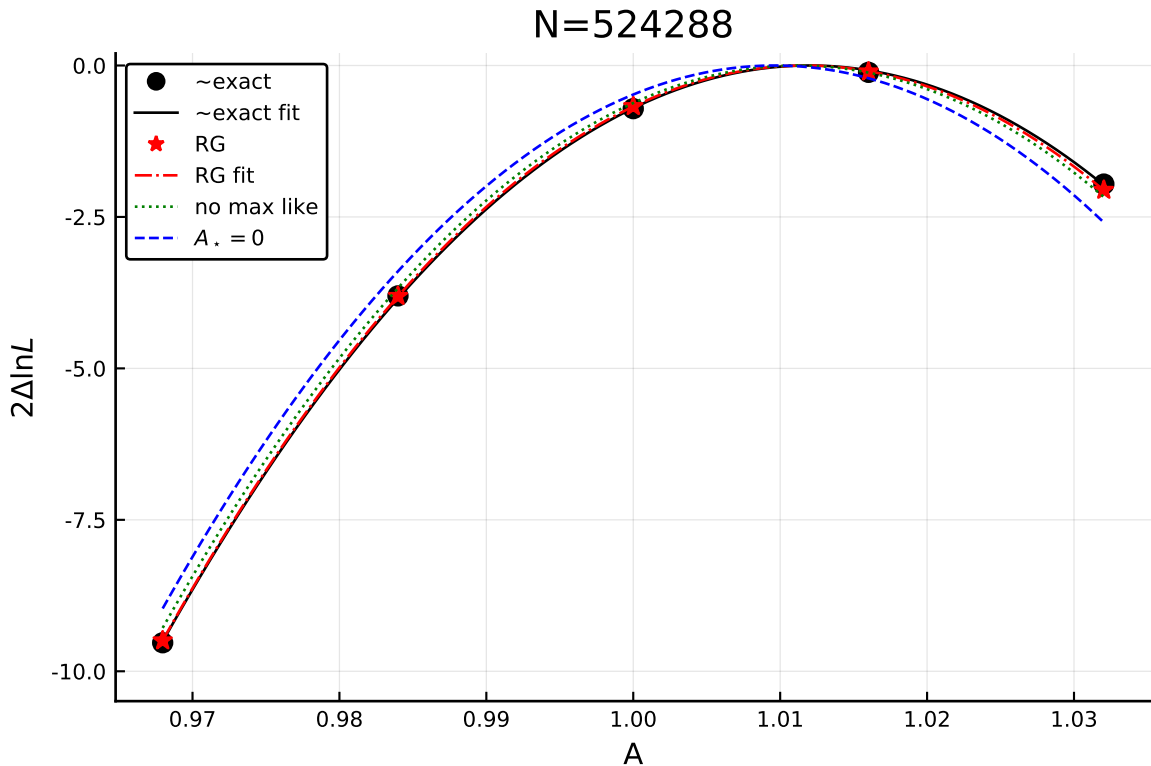
FIG. 2. $N = 524288$ test similar to Fig. 1. The "exact" likelihood is not strictly exact, but computed for $N_{d\mathbf{Q}} = 25$, $\epsilon_{\mathbf{Q}'} = 0.0005$, and $\epsilon_{\mathbf{A}} = 0.0002$, which is perfectly converged at the level of differences in this figure.

I stop at $N = 2^{19}$ for these examples because careful testing on my laptop becomes tedious beyond this, especially running with extremely conservative parameter settings to be certain of the exact result. I have run up to two million cells with good looking results. A one million cell example runs in two minutes. At four million I start to exhaust the memory on my laptop in my current Julia implementation, although it would be possible to go somewhat further with more optimization. In any case, it is clear that billion cell data sets could be done comfortably on a supercomputer.

I tried evolving using $Q(\lambda, k) = Q(0, k)e^{-k^2\lambda^2}$, more like in [1], but with a maximum likelihood ϕ_0, $\mathbf{A}_\star$, and element size cuts as introduced in this paper, but was unable to come within a factor of ten of the performance of the pairwise suppression approach of this paper.

III. DISCUSSION

To summarize, I have suggested the following improvements to the basic RG approach of [1]:

- Integrating out the difference between cells that are to be combined, rather than small-scale structure more generally, by defining $\mathbf{Q}'$ directly to be proportional to the difference between the current and target covariance.

- Shifting integration variables to integrate around a maximum likelihood signal field, if available, as ϕ_0.

- Subtracting a statistically homogeneous approximation out of the numerically evolving matrix $\mathbf{A}$, through the definition of $\mathbf{A}_\star$.

- Cuts on matrix element size, specified by $\epsilon_{\mathbf{Q}'}$, $\epsilon_{\mathbf{A}}$, etc., instead of a simple range cut.

The first of these is by far the most important. In the end it is clear that the algorithm is fast and straightforward enough for convenient practical data analysis.

It was surprising to me that the pair-oriented definition of $\mathbf{Q}'$ made such a large (factor $\gtrsim 10$) difference in speed. While the the principle that if we know which cells we will combine we should focus on integrating out the difference between them seems good enough to expect some improvement, I would have been happy with a factor of two. It

may be that I do not have the best possible implementation of the smooth cutoff option. In any case though, it seems like the pair-oriented approach is the way to go.

Of course it is only useful to integrate around a maximum likelihood field if that field can be found more quickly than the RG analysis could be done without it. This was the case in my tests, where finding the maximum likelihood field by conjugate gradient (CG) takes about 5% of the time in each likelihood computation. This might not always be the ratio, as my CG solution was massively accelerated by being able to multiply by things like $\mathbf{P}$ in Fourier space, including for preconditioning (e.g., without preconditioning finding a maximum likelihood field takes longer than the RG integration without it). If, e.g., the CG had to be done using less efficient spherical harmonic transforms, it might be faster not to use it. An interesting possibility is to use the RG method itself to find the maximum likelihood field. [1] showed how to find the data-constrained mean of any function of ϕ, with $\langle\phi\rangle$ itself being the simplest possible version of this. For a Gaussian problem $\langle\phi\rangle$ is the maximum likelihood field, while for a non-Gaussian problem it is not but would probably be a better starting point than the maximum likelihood field in that case anyway. Finding $\langle\phi\rangle$ can be piggybacked on a standard likelihood computation with minimal extra cost, but to get a speedup in likelihood calculations you would need to feed the result back into a recalculation. This would only be effective if a useful estimate of $\langle\phi\rangle$ could be found with looser numerical settings than would be required to do the calculation with $\phi_0 = 0$, which seems quite possible. When, e.g., computing derivatives with respect to parameters, we would probably achieve most of the benefit by computing $\langle\phi\rangle$ only for the central model (remember that accurate results can be achieved for any ϕ_0, it is just a question of how tight numerical settings need to be to do it).

Note that it may not always be beneficial to use $\mathbf{A}_\star \neq 0$. There is no cost if all cells in a formal data vector have measurements, i.e., there are no zeros on the diagonal of $\mathbf{R}^t\mathbf{N}^{-1}\mathbf{R}$, but if a substantial number of cells represent large holes in the data set or zero padding, so that these elements of $\mathbf{A}(0)$ can be dropped from sparse storage, setting $\mathbf{A}_\star \neq 0$ will remove this possibility. This must be considered on a problem-by-problem basis.

While my prototype code is already quite fast, at two minutes per likelihood evaluation per million cells, there is clearly more room for optimization. Most obviously, I am not taking advantage of the fact that $\mathbf{A}$ and $\mathbf{Q}'$ are symmetric matrices at all, for no better reason than not knowing canned operations in Julia that will do this. Other simple improvements could be tuning of things like the cuts I've parameterized by $\epsilon_{\mathbf{Q}'}$, etc.. I kept these cuts constant for all iterations but this could be wasteful if the required cut value is set by coarser levels of the calculation that do not take much total time. A less obvious but I think promising optimization idea is the following: The effect of evolving Eq. (13) is non-linear in the $\mathbf{Q}'$ matrix as initial changes in $\mathbf{A}$ are multiplied back together to find the next step, i.e., we get products of $\mathbf{Q}'$ with itself. The required number of steps is surely set by the products of the largest elements of $\mathbf{Q}'$—the products of small elements are perturbatively much smaller. This suggests that $\mathbf{Q}'$ could be split into two or more pieces based on element size. The piece(s) with larger elements, which would be very short-range (i.e., few elements, i.e., fast to multiply), could be evolved first, then longer-range the pieces with smaller elements evolved with fewer steps, possibly even one, because their self-products are negligible. As long as our set of $\mathbf{Q}'$ steps integrates to $\mathbf{Q}_2 - \mathbf{Q}_1$, we are free to choose the details.

The next step is to implement this for realistic cosmological scenarios.

ACKNOWLEDGMENTS

[1] Patrick McDonald, "Renormalization group computation of likelihood functions for cosmological data sets," *Phys. Rev.* **D 99**, 043538 (2019).
[2] K. G. Wilson and J. Kogut, "The renormalization group and the ϵ expansion," *Phys. Rept.* **12**, 75–199 (1974).
[3] T. Banks, *Modern Quantum Field Theory, by Tom Banks, Cambridge, UK: Cambridge University Press, 2008* (2008).
[4] P. J. E. Peebles, *Principles of Physical Cosmology by P.J.E. Peebles. Princeton University Press, 1993. ISBN: 978-0-691-01933-8* (1993).
[5] P. McDonald and A. Roy, "Clustering of dark matter tracers: generalizing bias for the coming era of precision LSS," JCAP **8**, 020 (2009), arXiv:0902.0991 [astro-ph.CO].
[6] Patrick McDonald and Zvonimir Vlah, "Large-scale structure perturbation theory without losing stream crossing," *Phys. Rev.* **D 97**, 023508 (2018).
[7] J. R. Bond, A. H. Jaffe, and L. Knox, "Estimating the power spectrum of the cosmic microwave background," *Phys. Rev.* **D 57**, 2117–2137 (1998).

[8] Benjamin D. Wandelt, David L. Larson, and Arun Lakshminarayanan, "Global, exact cosmic microwave background data analysis using Gibbs sampling," *Phys. Rev.* **D 70**, 083511 (2004), arXiv:astro-ph/0310080 [astro-ph].

[9] F. S. Kitaura and T. A. Enßlin, "Bayesian reconstruction of the cosmological large-scale structure: methodology, inverse algorithms and numerical optimization," *Mon. Not. Roy. Astron. Soc.* **389**, 497–544 (2008), arXiv:0705.0429 [astro-ph].

[10] T. A. Enßlin, M. Frommert, and F. S. Kitaura, "Information field theory for cosmological perturbation reconstruction and nonlinear signal analysis," *Phys. Rev.* **D 80**, 105005 (2009), arXiv:0806.3474.

[11] Torsten A. Enßlin, "Information Theory for Fields," Annalen der Physik **531**, 1800127 (2019), arXiv:1804.03350 [astro-ph.CO].

[12] P. McDonald, "Primordial non-Gaussianity: Large-scale structure signature in the perturbative bias model," *Phys. Rev.* **D 78**, 123519–+ (2008), arXiv:0806.1061.

[13] N. Bartolo, J. P. Beltrán Almeida, S. Matarrese, M. Pietroni, and A. Riotto, "Signatures of primordial non-Gaussianities in the matter power-spectrum and bispectrum: the time-RG approach," JCAP **3**, 011 (2010), arXiv:0912.4276 [astro-ph.CO].

[14] T. Giannantonio and C. Porciani, "Structure formation from non-Gaussian initial conditions: Multivariate biasing, statistics, and comparison with N-body simulations," *Phys. Rev.* **D 81**, 063530 (2010), arXiv:0911.0017 [astro-ph.CO].

[15] J.-O. Gong and S. Yokoyama, "Scale-dependent bias from primordial non-Gaussianity with trispectrum," *Mon. Not. Roy. Astron. Soc.* **417**, L79–L82 (2011), arXiv:1106.4404 [astro-ph.CO].

[16] M. Alvarez, T. Baldauf, J. R. Bond, N. Dalal, R. de Putter, O. Doré, D. Green, C. Hirata, Z. Huang, D. Huterer, D. Jeong, M. C. Johnson, E. Krause, M. Loverde, J. Meyers, P. D. Meerburg, L. Senatore, S. Shandera, E. Silverstein, A. Slosar, K. Smith, M. Zaldarriaga, V. Assassi, J. Braden, A. Hajian, T. Kobayashi, G. Stein, and A. van Engelen, "Testing Inflation with Large Scale Structure: Connecting Hopes with Reality," ArXiv e-prints (2014), arXiv:1412.4671.

[17] Azadeh Moradinezhad Dizgah and Cora Dvorkin, "Scale-dependent galaxy bias from massive particles with spin during inflation," Journal of Cosmology and Astro-Particle Physics **2018**, 010 (2018).

[18] Ue-Li Pen, "Fast power spectrum estimation," *Mon. Not. Roy. Astron. Soc.* **346**, 619–626 (2003), arXiv:astro-ph/0304513 [astro-ph].

[19] N. Padmanabhan, U. Seljak, and U. L. Pen, "Mining weak lensing surveys." New Astronomy **8**, 581–603 (2003).

[20] Kendrick M. Smith, Oliver Zahn, and Olivier Doré, "Detection of gravitational lensing in the cosmic microwave background," *Phys. Rev.* **D 76**, 043510 (2007), arXiv:0705.3980 [astro-ph].

[21] Uroš Seljak, Grigor Aslanyan, Yu Feng, and Chirag Modi, "Towards optimal extraction of cosmological information from nonlinear data," Journal of Cosmology and Astro-Particle Physics **2017**, 009 (2017).

[22] A. Font-Ribera, P. McDonald, and A. Slosar, "How to estimate the 3D power spectrum of the Lyman-α forest," JCAP **1**, 003 (2018), arXiv:1710.11036.

[23] N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, N. Bartolo, E. Battaner, K. Benabed, A. Benoît, A. Benoit-Lévy, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, A. Bonaldi, L. Bonavera, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, A. Catalano, A. Challinor, H. C. Chiang, P. R. Christensen, D. L. Clements, L. P. L. Colombo, C. Combet, A. Coulais, B. P. Crill, A. Curto, F. Cuttaia, L. Danese, R. D. Davies, R. J. Davis, P. de Bernardis, A. de Rosa, G. de Zotti, J. Delabrouille, F. X. Désert, E. Di Valentino, C. Dickinson, J. M. Diego, K. Dolag, H. Dole, S. Donzelli, O. Doré, M. Douspis, A. Ducout, J. Dunkley, X. Dupac, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, J. Fergusson, F. Finelli, O. Forni, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frejsel, S. Galeotta, S. Galli, K. Ganga, C. Gauthier, M. Gerbino, M. Giard, E. Gjerløw, J. Gonzá lez-Nuevo, K. M. Górski, S. Gratton, A. Gregorio, A. Gruppuso, J. E. Gudmundsson, J. Hamann, F. K. Hansen, D. L. Harrison, G. Helou, S. Henrot-Versillé, C. Hernández- Monteagudo, D. Herranz, S. R. Hildebrandt, E. Hivon, W. A. Holmes, A. Hornstrup, K. M. Huffenberger, G. Hurier, A. H. Jaffe, W. C. Jones, M. Juvela, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Knoche, L. Knox, M. Kunz, H. Kurki-Suonio, G. Lagache, A. Lähteenmäki, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, R. Leonardi, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Lilley, M. Linden-Vørnle, V. Lindholm, M. López- Caniego, J. F. Macías-Pérez, B. Maffei, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, M. Maris, P. G. Martin, E. Martínez-González, S. Masi, S. Matarrese, P. R. Meinhold, A. Melchiorri, M. Migliaccio, M. Millea, S. Mitra, M. A. Miville-Deschênes, A. Moneti, L. Montier, G. Morgante, D. Mortlock, S. Mottet, D. Munshi, J. A. Murphy, A. Narimani, P. Naselsky, F. Nati, P. Natoli, F. Noviello, D. Novikov, I. Novikov, C. A. Oxborrow, F. Paci, L. Pagano, F. Pajot, D. Paoletti, B. Partridge, F. Pasian, G. Patanchon, T. J. Pearson, O. Perdereau, L. Perotto, V. Pettorino, F. Piacentini, M. Piat, E. Pierpaoli, D. Pietrobon, S. Plaszczynski, E. Pointecouteau, G. Polenta, N. Ponthieu, G. W. Pratt, S. Prunet, J. L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, C. Renault, A. Renzi, I. Ristorcelli, G. Rocha, M. Rossetti, G. Roudier, B. Rouillé d'Orfeuil, J. A. Rubiño-Martín, B. Rusholme, L. Salvati, M. Sandri, D. Santos, M. Savelainen, G. Savini, D. Scott, P. Serra, L. D. Spencer, M. Spinelli, V. Stolyarov, R. Stompor, R. Sunyaev, D. Sutton, A. S. Suur-Uski, J. F. Sygnet, J. A. Tauber, L. Terenzi, L. Toffolatti, M. Tomasi, M. Tristram, T. Trombetti, M. Tucci, J. Tuovinen, G. Umana, L. Valenziano, J. Valiviita, F. Van Tent, P. Vielva, F. Villa, L. A. Wade, B. D. Wandelt, I. K. Wehus, D. Yvon, A. Zacchei, and A. Zonca, "Planck 2015 results. XI. CMB power spectra, likelihoods, and robustness of parameters," *Astron. Astrophys.* **594**, A11 (2016).

[24] F. Beutler, H.-J. Seo, A. J. Ross, P. McDonald, S. Saito, A. S. Bolton, J. R. Brownstein, C.-H. Chuang, A. J. Cuesta, D. J. Eisenstein, A. Font-Ribera, J. N. Grieb, N. Hand, F.-S. Kitaura, C. Modi, R. C. Nichol, W. J. Percival, F. Prada, S. Rodriguez-Torres, N. A. Roe, N. P. Ross, S. Salazar-Albornoz, A. G. Sánchez, D. P. Schneider, A. Slosar, J. Tinker, R. Tojeiro, M. Vargas-Magaña, and J. A. Vazquez, "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Fourier space," *Mon. Not. Roy. Astron. Soc.* **464**, 3409–3430 (2017), arXiv:1607.03149.

[25] K. M. Górski, A. J. Banday, E. Hivon, and B. D. Wandelt, "HEALPix — a Framework for High Resolution, Fast Analysis on the Sphere," in *ASP Conf. Ser. 281: Astronomical Data Analysis Software and Systems XI* (2002) pp. 107–+.

Appendix A: Alternative approach to integrating out differences between cells

Before realizing I could define $\mathbf{Q}'$ by simply differencing the current and target $\mathbf{Q}$s, I worked out a method for integrating out the difference between cells closer to the original approach in [1]. I include it here to promote broader understanding of the possibilities.

The RG integration will be controlled by a parameter α which starts at zero and is taken to ∞. $\mathbf{Q}$ and S_I become functions of this parameter, i.e.,

$$\mathbf{Q}^{-1}(\alpha) \equiv \mathbf{P}^{-1} + \alpha \mathbf{K} \, , \tag{A1}$$

with $\mathbf{K}$ a fixed matrix to be specified. Obviously we can suppress fluctuations between cells 1 and 2 by adding a term to $S(\phi)$ proportional to $(\phi_1 - \phi_2)^2$. Repeating this over and over (e.g., $(\phi_3 - \phi_4)^2$, etc.) is equivalent to making $\mathbf{K}$ the following block diagonal matrix:

$$\mathbf{K} = \begin{bmatrix} \mathbf{k} & 0 & ... \\ 0 & \mathbf{k} & ... \\ ... & ... & ... \end{bmatrix} \, , \tag{A2}$$

where

$$\mathbf{k} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \, . \tag{A3}$$

I.e., by dialing α from 0 to ∞ in $\mathbf{Q}^{-1} = \mathbf{P}^{-1} + \alpha \mathbf{K}$, we will have effectively integrated out the differences between adjacent pairs of cells. We now have

$$\mathbf{Q}' = -\mathbf{Q}\mathbf{K}\mathbf{Q} = -\left(\mathbf{P}^{-1} + \alpha \mathbf{K}\right)^{-1} \mathbf{K} \left(\mathbf{P}^{-1} + \alpha \mathbf{K}\right)^{-1} \, . \tag{A4}$$

Unlike in [1], $\mathbf{K}$ is not exactly translation invariant, so we can't simply compute $\left(\mathbf{P}^{-1} + \alpha \mathbf{K}\right)^{-1}$ in Fourier space. The structure of $\mathbf{Q}'$ is the same everywhere, however, up to a distinction between odd and even cells, and it is limited to short range, so we can compute it by brute force inversion for a limited representative stretch of cells and then translate it everywhere.

This approach worked in preliminary tests, but not as efficiently as the one in the paper.