# The Kullback–Leibler Divergence and the Convergence Rate of Fast Covariance Matrix Estimators in Galaxy Clustering Analysis

Zhigang Li[1] , Zhejie Ding[2,3], Yu Yu[2,3] , and Pengjie Zhang[2,3,4]

[1] School of Physics and Electronic Engineering, Nanyang Normal University, 1638 Wolong Road, Nanyang, 473061, People's Republic of China; zhigang@nynu.edu.cn
[2] Department of Astronomy, School of Physics and Astronomy, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, 200240, People's Republic of China; zhejied@sjtu.edu.cn, yuyu22@sjtu.edu.cn
[3] Key Laboratory for Particle Astrophysics and Cosmology (MOE)/Shanghai Key Laboratory for Particle Physics and Cosmology, People's Republic of China
[4] Tsung-Dao Lee Institute, Shanghai Jiao Tong University, Shanghai, People's Republic of China

## Abstract

We present a method to quantify the convergence rate of the fast estimators of the covariance matrices in the large-scale structure analysis. Our method is based on the Kullback–Leibler (KL) divergence, which describes the relative entropy of two probability distributions. As a case study, we analyze the delete-$d$ jackknife estimator for the covariance matrix of the galaxy correlation function. We introduce the information factor or the normalized KL divergence with the help of a set of baseline covariance matrices to diagnose the information contained in the jackknife covariance matrix. Using a set of quick particle mesh mock catalogs designed for the Baryon Oscillation Spectroscopic Survey DR11 CMASS galaxy survey, we find that the jackknife resampling method succeeds in recovering the covariance matrix with 10 times fewer simulation mocks than that of the baseline method at small scales ($s \leqslant 40\,h^{-1}$ Mpc). However, the ability to reduce the number of mock catalogs is degraded at larger scales due to the increasing bias on the jackknife covariance matrix. Note that the analysis in this paper can be applied to any fast estimator of the covariance matrix for galaxy clustering measurements.

## 1. Introduction

The covariance matrix plays an important role in the data analysis of the galaxy large-scale structure and contains important information on the statistical and systematical errors on the data. An accurate covariance matrix is crucial to pass the errors on the data to the errors on the inferred cosmological parameters correctly (Hartlap et al. 2007; Dodelson & Schneider 2013; Taylor et al. 2013; Percival et al. 2014; Taylor & Joachimi 2014). However, we usually do not know the true covariance matrix from first principles. Instead, the standard way is to estimate the covariance matrix from the data themselves or from the artificial or mock catalogs (Reid et al. 2010; Manera et al. 2013, 2015; Anderson et al. 2014; Gil-Marín et al. 2016). The mock catalogs are created to follow the statistical properties of the data set as closely as possible and include the diverse observational effects (Manera et al. 2015). A large number of mock catalogs are required to reduce the statistical errors on the covariance matrix (Percival et al. 2014). The creation of mock catalogs and the analysis of them has become one of the most computationally consuming steps in the modern galaxy clustering analysis (Monaco et al. 2002, 2013; Scoccimarro & Sheth 2002; Manera et al. 2013, 2015; Tassev et al. 2013; Kitaura et al. 2014, 2015; White et al. 2014; Chuang et al. 2015; Feng et al. 2016; Balaguera-Antolínez et al. 2019), especially for the ongoing and upcoming next-generation galaxy surveys such as the Dark Energy Survey (Frieman & Dark Energy Survey Collaboration 2013), the Dark Energy Spectroscopic Instrument survey

(DESI; Schlegel et al. 2011), the Large Synoptic Survey Telescope survey (LSST Science Collaboration et al. 2009), and the Euclid satellite mission surveys (Laureijs et al. 2011).

Numerous efforts have been devoted to finding alternatives to obtain reliable estimates of the covariance matrix quickly and accurately. In real observations, the covariance matrix involves complex effects from the galaxy evolution, the scale-dependent and non-Poissonian shot noise, the stochastic bias, and the redshift space distortion (Takahashi et al. 2009; Zhang et al. 2013; Li et al. 2014; Blot et al. 2015; Shi et al. 2016; Zheng & Song 2016; Howlett & Percival 2017; Klypin & Prada 2018). Theoretical modeling of the covariance matrix has achieved great progress on the dark matter power spectrum (Neyrinck 2011; Mohammed & Seljak 2014; Carron et al. 2015; Bertolini et al. 2016; Grieb et al. 2016; Mohammed et al. 2017; Hikage et al. 2020; Taruya et al. 2021), the galaxy power spectrum (e.g., Lacasa & Kunz 2017; Sugiyama et al. 2020), and the galaxy correlation function (e.g., Philcox et al. 2020; Rashkovetskyi et al. 2023). Wadekar & Scoccimarro (2020) have proposed a promising analytical method to compute the covariance matrix of galaxy power spectrum multipoles including various theoretical and observational effects. Their results show that the analytic approach has the benefit of sampling noise free and saving computational resources to recompute covariances in the model fitting process (Wadekar et al. 2020).

Meanwhile, many methods have been proposed to reduce the number of mock catalogs or the size of the simulation boxes that are required to obtain a reliable and accurate covariance matrix estimate. Some of them are based on fitting the estimated covariance from a small number of mocks to an empirical model with several free parameters (Pope & Szapudi 2008; O'Connell et al. 2016; Pearson & Samushia 2016). O'Connell & Eisenstein (2019) extended the method, fitting a

jackknife covariance matrix from a single survey volume to obtain the fitting parameters without reference to any mocks in real analysis. Howlett & Percival (2017) proposed a method to reduce the size of the simulation box and to correct for the supersample covariance and the window function effect analytically. Other approaches aim to reduce the number of mock catalogs by the resampling method (Norberg et al. 2009; Schneider et al. 2011; Arnalte-Mur & Norberg 2014; Escoffier et al. 2016; Mohammad & Percival 2022) or tapering method (Paz & Sánchez 2015).

An important step in assessing the efficiency of the covariance matrix estimator of the large-scale structure is to calculate the convergence rate, i.e., the number of mock catalogs needed to obtain an equivalent covariance matrix compared to the brute-force sample variance from a given number of mock catalogs. Two factors are concerned in this case, the noise and the bias. The noise level is usually represented by the mean variance of the elements of the covariance matrix and commonly used to estimate the convergence rate. However, the different parts in the covariance matrix do not play equal roles in the parameter fitting process, typically the diagonal and off-diagonal terms. Compared to the dominant diagonal terms, the off-diagonal terms are usually much smaller but contain critical information on the mode coupling and the window function. The mean variance of the elements cannot distinguish the difference between them. The bias on the estimator of the covariance matrix also plays an important role in the parameter fitting process and should be recognized. Although the biases on the diagonal elements are easy to show, those on the off-diagonal elements are not due to the fact that they have small values and high noise levels. Furthermore, it is the precision matrix, i.e., the inverse of the covariance matrix, that appears in the likelihood function. The matrix inversion is a nonlinear process. It mixes the diagonal and off-diagonal elements in the covariance matrix, then makes the effect of the noise and bias much more complicated. So we need alternatives to quantify the performance of the estimator of the covariance matrix.

In this paper, we present a simple method to estimate the convergence rate of the covariance matrix estimators efficiently. Since we are not comparing two arbitrary matrices but instead two Gaussian likelihood functions characterized by the two covariance matrices, there is a prominent tool to accomplish our goal, the Kullback–Leibler (KL) divergence (Kullback & Leibler 1951). The KL divergence measures the relative entropy between two probability distributions, and it can describe how different two covariance matrices are in the sense of the Gaussian likelihood functions. It has been adopted in multiple literature to study the convergence of covariance matrices from fast methods (e.g., O'Connell et al. 2016; Lippich et al. 2019; Philcox et al. 2020). In our study, we apply the method on a recently proposed covariance matrix estimator of the galaxy correlation functions, which combines the delete-$d$ jackknife resampling and the mock catalogs (Escoffier et al. 2016).

This paper is organized as follows. In Section 2, we introduce the KL divergence and its application to the convergence of covariance matrix estimators. In Section 3, we describe the data set and the two methods to estimate the covariance matrix of the galaxy correlation function, i.e., the brute-force method and the jackknife resampling. In Section 5,

we first test the KL divergence using the brute-force covariance matrices from different numbers of mock catalogs. In addition, we apply the KL divergence test to the covariance matrices from the jackknife resampling method to estimate its convergence rate. We close the paper with a brief discussion and summary in Section 6.

## 2. KL Divergence

The KL divergence from the probability distribution $Q$ to the reference probability distribution $P$ is a measure of $Q$ diverging from $P$, which is defined as

$$\mathrm{KL}(P|Q) = \int P(\boldsymbol{x}) \ln \frac{P(\boldsymbol{x})}{Q(\boldsymbol{x})} \, d\boldsymbol{x}. \tag{1}$$

The KL divergence is positive definite, that is, $\mathrm{KL}(P|Q) \geqslant 0$. The equality holds if and only if $P = Q$. Another important property of the KL divergence is that it is asymmetric in general, $\mathrm{KL}(P|Q) \neq \mathrm{KL}(Q|P)$. In Bayesian language, the KL divergence measures the information loss when one uses $Q$ (usually a model) to approximate $P$ ("true" distribution; Baez & Fritz 2014).

If $P$ and $Q$ are both multivariate normal distributions with the same mean, the KL divergence simplifies to

$$\mathrm{KL}(P|Q) = \frac{1}{2} \left[ \mathrm{Tr}(C_Q^{-1} C_P) + \ln\left(\frac{\det C_Q}{\det C_P}\right) - N \right], \tag{2}$$

where $C_P$ and $C_Q$ are the covariance matrices of $P$ and $Q$, respectively. $N$ is the dimension of the concerned random variables or the data vectors. $\mathrm{Tr}(A)$ represents the trace of matrix $A$, and $\det A$ is the determinant. In the following, the distributions $P$ and $Q$ are always assumed to be the multivariate normal distributions.

### 2.1. Sample Variance of Gaussian-distributed Data

In data analysis of the cosmological large-scale structure, we usually estimate the covariance matrix of measured data from the sample variance of a large number of independent simulation mock catalogs. If data follow a multivariate normal distribution, the estimated covariance matrix $\hat{C}$ follows a Wishart distribution (Wishart 1928),

$$\mathcal{P}(\hat{C}) = \frac{2^{-np/2}}{\Gamma_p(\frac{n}{2})} (n^p \det \hat{C})^{\frac{n-p-1}{2}} (\det \Sigma)^{-\frac{n}{2}}$$
$$\times \exp\left\{ \mathrm{Tr}\left( -\frac{n}{2} \Sigma^{-1} \hat{C} \right) \right\}, \tag{3}$$

where $N_m$ is the number of mock samples, $n = N_m - 1$ is the degrees of freedom, $\Sigma$ is the expectation value of the covariance matrix $E(\hat{C})$, and $p$ is the dimension of $\hat{C}$. $\Gamma_p(x)$ is the multivariate gamma function. Considering that $\hat{C}_P$ and $\hat{C}_Q$ are the sample variances of two subsets of the same parent simulation mock catalogs, the expectation value of the KL divergence from $Q$ (with the covariance matrix $\hat{C}_Q$) to $P$ (with

the covariance $\hat{C}_P$) can be calculated as

$$\langle \mathrm{KL}(P|Q) \rangle = \frac{1}{2} \left\{ \frac{p(p+1)}{N_Q - p - 1} + \Gamma_p\left(\frac{N_Q}{2}\right) \right.$$
$$\left. - \Gamma_p\left(\frac{N_P}{2}\right) + p \ln\left(\frac{N_P}{N_Q}\right) \right\}, \qquad (4)$$

where $N_P$ and $N_Q$ are the number of subset mock catalogs used to calculate $\hat{C}_P$ and $\hat{C}_Q$, respectively. To derive the above equation, we have used the fact that the expectation value of $\hat{C}_P$ and $\hat{C}_Q$ is the same.

### 2.2. Biased Covariance Matrix

Supposing that the expectation value of the estimated covariance matrix $\hat{C}_{\tilde{Q}}$ has a linear bias $\alpha$ with respect to the expectation value of $\hat{C}_Q$, that is, $\Sigma_{\tilde{Q}} = \alpha \Sigma_Q$, the expectation value of the KL divergence from the multivariate normal distribution $\tilde{Q}$ with the covariance matrix $\hat{C}_{\tilde{Q}}$ to $P$ will be

$$\langle \mathrm{KL}(P|\tilde{Q}) \rangle = \langle \mathrm{KL}(P|Q) \rangle + \Delta_{\mathrm{KL}}(\alpha), \qquad (5)$$

where $\Delta_{\mathrm{KL}}(\alpha)$ is defined as

$$\Delta_{\mathrm{KL}}(\alpha) = \{\ln \alpha + (1/\alpha - 1)(1 + a)\}p/2, \qquad (6)$$

with $a = (p+1)/(N_Q - p - 1)$. In most cases, $\Delta_{\mathrm{KL}}(\alpha)$ is positive except for a tiny range of $1 \leqslant \alpha \leqslant (1 + a)$, where it has a minimal negative value of $(\ln(1 + a) - a)p/2 \approx -pa^2/4$, which is close to 0 when the dimension of the covariance matrix is much smaller than the number of mock realizations, i.e., $p \ll N_Q$. When $N_Q \to \infty$, $a \to 0$. Then the bias on the covariance matrix sets a lower limit on $\Delta_{\mathrm{KL}}(\alpha)$,

$$\Delta_{\mathrm{KL,min}}(\alpha) = \{\ln \alpha + (1/\alpha - 1)\}p/2, \qquad (7)$$

which is positive definite when $\alpha \neq 1$.

## 3. Data Sets

We conduct the KL divergence on the covariance matrix calculated from the delete-$d$ jackknife resampling. We discuss the galaxy mock catalogs that we use to calculate the jackknife covariance matrix in Section 3.1. We show the calculated correlation functions and their covariance matrices in Sections 3.2 and 3.3, respectively.

### 3.1. Quick Particle Mesh Mock Samples

In this paper, we use the publicly released mock catalog by the Baryon Oscillation Spectroscopic Survey (BOSS) collaboration. These mocks are generated using the quick particle mesh (QPM) method (White et al. 2014) with low mass and force resolution. The simulations are run in a flat $\Lambda$CDM cosmology with parameters of $\Omega_m = 0.29$, $h = 0.7$, $\Omega_b = 0.0458$, $\sigma_8 = 0.8$, and $n_s = 0.97$. The cubic simulation box has a side length of $2560\,h^{-1}$ Mpc and contains $1280^3$ particles. Halos are assigned to a subset of the simulation particles, which are chosen based on their smoothed local density. The halo masses are then sampled with a density-dependent probability to match the halo mass function and the large-scale bias from the reference high-resolution $N$-body simulations. The galaxies are populated in the resolved halos using the halo occupation distribution (HOD) approach (e.g., Wechsler & Tinker 2018).

The HOD parameters are adjusted to fit the small-scale projected two-point correlation function of the BOSS DR11 CMASS galaxies. The galaxies are further downsampled based on the radial selection function and the angular mask of the BOSS DR11 CMASS survey on the north Galactic cap, which covers 6391 deg$^2$ and extends over a wide redshift range of $0.43 < z < 0.70$ (Beutler et al. 2014). For more information on the QPM galaxy mock catalogs, we refer the reader to White et al. (2014).

### 3.2. Two-point Correlation Function

We measure the galaxy two-point correlation function of the QPM mocks based on the Landy & Szalay (1993) method,

$$\xi(s, \mu) = \frac{DD(s, \mu) - 2DR(s, \mu) + RR(s, \mu)}{RR(s, \mu)}, \qquad (8)$$

where $s$ is the separation between two galaxies and $\mu$ is the cosine of the angle spanned by the galaxy separation vector and the line-of-sight vector. Here we define the line of sight for each pair of galaxies as the direction of the vector passing through the median point of the pair separation and the observer, $\boldsymbol{h} = (\boldsymbol{s}_1 + \boldsymbol{s}_2)/2$, with $\boldsymbol{s}_{1,2}$ being the position vector of galaxy 1 and galaxy 2. $DD(s, \mu)$ is the number of galaxy–galaxy pairs whose separation is located in the $(s, \mu)$ bin, normalized by the total number of pairs. $DR(s, \mu)$ and $RR(s, \mu)$ are the number of galaxy–random and random–random pairs, respectively. For each galaxy mock sample, we generate randomly distributed points with the consideration of the radial selection function and the angular mask of the BOSS DR11 CMASS survey. We set the number of random points to be 10 times that of the mock galaxies.

The multipoles of the correlation function are calculated by expanding the 3D correlation function using the Legendre polynomial $P_l(\mu)$, i.e.,

$$\xi_l(s) = \frac{(2l + 1)}{2} \int_{-1}^{1} d\mu\, \xi(s, \mu) P_l(\mu), \qquad (9)$$

where $P_0(\mu) = 1$ and $P_2(\mu) = (3\mu^2 - 1)/2$. We are focusing on the monopole ($l = 0$) and quadrupole ($l = 2$) in the following. To do the above integration, we use 20 $\mu$ bins with an equal bin width of 0.05. Our data vector is $\xi = (\xi_0, \xi_2)$, where $\xi_0 = (\xi_0(s_1), ..., \xi_0(s_N))$, $\xi_2 = (\xi_2(s_1), ..., \xi_2(s_N))$, and $N$ is the number of $s$ bins.

We mainly study the correlation function multipoles at the scale range of $0\,h^{-1}$ Mpc $< s < 40\,h^{-1}$ Mpc with the bin size $\Delta s = 2\,h^{-1}$ Mpc, where the jackknife method (Xu et al. 2023) can still give a satisfactory covariance matrix compared to the more sophisticated HOD-based (Yu et al. 2022) or emulator-based (Chapman et al. 2022; Yuan et al. 2022) methods. For the large-scale structure analysis, we choose two scale ranges, the intermediate scale of $20\,h^{-1}$ Mpc $< s < 80\,h^{-1}$ Mpc with $\Delta s = 4\,h^{-1}$ Mpc and the large scale of $24\,h^{-1}$ Mpc $< s < 160\,h^{-1}$ Mpc with $\Delta s = 8\,h^{-1}$ Mpc. On these large scales, the survey window effect becomes important. The large-scale clustering breaks the independence among different jackknife subregions, which will introduce an increasing bias on the jackknife covariance matrix along with the increasing scales.

### 3.3. The Covariance Matrix

Based on the QPM mocks, we can calculate the baseline covariance matrix from the brute-force method and take it as the true covariance. To show the convergence rate of the fast covariance matrix estimators based on the KL divergence, we choose the delete-$d$ jackknife resampling method as a case study.

#### 3.3.1. Brute-force Covariance Estimation

The baseline covariance matrix is estimated from the variance of independent mock samples drawn from the fiducial cosmological model, which we denote as the brute-force method. In our case, each QPM mock sample has the same observational effect as the real observation, including the complex survey geometry and the completeness effect.

The brute-force covariance estimation of the correlation function multipoles over mock catalogs is calculated using the following formula:

$$\hat{C}_{ij} = \frac{1}{N_m - 1}\sum_{k=1}^{N_m}(\xi_i^{(k)} - \bar{\xi}_i)(\xi_j^{(k)} - \bar{\xi}_j),$$
$$\bar{\xi}_i = \frac{1}{N_m}\sum_{k=1}^{N_m}\xi_i^{(k)}, \qquad (10)$$

where $i, j = 1, 2, \ldots, N, N+1, \ldots, 2N$ with the first (last) $N$ elements corresponding to the $N$ radial bins of the monopole (quadrupole) correlation functions. The superscript $k$ enclosed in brackets denotes the mock index, and $N_m$ is the total number of mocks.

The above-obtained brute-force covariance matrix, as random variables, will follow a Wishart distribution if the correlation functions follow a multivariate Gaussian distribution. The Wishart distribution can be characterized by the degrees of freedom $N_m - 1$, the dimension of the data vector or the number of bins $p$, and the true or expected covariance matrix $\Sigma$.

#### 3.3.2. Jackknife Resampling

Unlike the brute-force covariance matrix, which needs a large number of mock samples, the jackknife covariance matrix can be calculated directly from the observational data and so can save the computational cost greatly. The practical analysis shows that the jackknife covariance matrix suffers from a large noise level and cannot meet the requirement of future large-scale structure analysis. It has been proposed that applying the jackknife technique on individual mock samples and averaging over them can greatly enhance the precision of the jackknife covariance matrix estimation (Escoffier et al. 2016). The performance of the jackknife covariance matrix can be further improved by using the delete-$d$ jackknife technique.

The traditional jackknife covariance matrix is calculated by dividing the observational data into $N_s$ subregions. The jackknife samples are constructed by deleting one subregion at a time. We calculate the correlation function for each jackknife sample. The covariance matrix is calculated by

$$\hat{C}_{ij,\text{J1}} = \frac{N_s - 1}{N_s}\sum_{k=1}^{N_s}(\xi_{i,\text{J1}}^{(k)} - \bar{\xi}_{i,\text{J1}})(\xi_{j,\text{J1}}^{(k)} - \bar{\xi}_{j,\text{J1}}), \qquad (11)$$

where $\xi_{\text{J1}}^{(k)}$ is the correlation function for the $k$th jackknife sample, and $\bar{\xi}_{\text{J1}}$ is the mean correlation function averaged over all the jackknife samples.

The above covariance matrix is shown to be not optimal for nonsmooth or nonlinear statistics (Wu 1986), which tends to be true for our case due to the effects from the window function and the redshift-dependent galaxy selection function. Shao & Wu (1989) proposed two delete-$d$ subsamples, instead of one, at a time to construct the jackknife samples and proved that it can give an asymptotically unbiased covariance matrix for the case with nonsmooth statistics when $N_s - d \to \infty$ and $\sqrt{N_s}/d \to \infty$. This is called the delete-$d$ jackknife resampling. The delete-$d$ jackknife covariance matrix is calculated by

$$\tilde{C}_{ij,\text{JK}} = \frac{N_s - d}{d \times N_{\text{JK}}}\sum_{k=1}^{N_{\text{JK}}}(\xi_{i,\text{JK}}^{(k)} - \bar{\xi}_{i,\text{JK}})(\xi_{j,\text{JK}}^{(k)} - \bar{\xi}_{j,\text{JK}}), \qquad (12)$$

where $\xi_{\text{JK}}^{(k)}$ is the correlation function of the $k$th delete-$d$ jackknife sample, and $\bar{\xi}_{\text{JK}}$ is the mean correlation function over the total number of delete-$d$ jackknife samples given by

$$N_{\text{JK}} = \frac{N_s!}{d!(N_s - d)!}. \qquad (13)$$

In addition, if a jackknife covariance matrix is calculated from a mock, we can further reduce its sample variance by averaging over the jackknife covariance matrices from multiple mocks (Escoffier et al. 2016), i.e.,

$$\hat{C}_{ij,\text{JK}} = \frac{1}{N_m}\sum_{m=1}^{N_m}{}^{(m)}\tilde{C}_{ij,\text{JK}}, \qquad (14)$$

where ${}^{(m)}\tilde{C}_{\text{JK}}$ is the covariance matrix obtained by applying Equation (12) on the $m$th mock.

The delete-$d$ jackknife variance estimator is asymptotically unbiased when $N_s$ and $d$ go to infinity. Increasing the number of subsamples will reduce the minimal transverse size of each subsample and then the number of independent modes. It will also increase the number of jackknife resamples. Escoffier et al. (2016) show that the delete-$d$ jackknife covariance matrix of the galaxy correlation function converges when $N_s \geqslant 9$ and the choice of $d$ has a small effect. In this analysis, we choose $N_s = 12$ and $d = 6$. The total number of jackknife samples for each mock catalog is calculated from the combination formula, i.e., $N_{\text{JK}} = \binom{12}{6} = \frac{12!}{6! \times 6!} = 924$.

#### 3.3.3. Covariance Matrix from QPM Mocks

In this section, we use the QPM mock catalog to calculate the delete-$d$ jackknife covariance matrices and to compare them with the brute-force ones. We calculate the baseline brute-force covariance matrix from 1000 mock samples and take it as the benchmark. For the jackknife covariance matrix, we calculate the mean over 100 mocks based on Equation (14). In Figure 1, we compare the two types of covariance matrices at the scale $0\,h^{-1}\,\text{Mpc} < s < 40\,h^{-1}\,\text{Mpc}$ with a bin width of $2\,h^{-1}\,\text{Mpc}$. Overall, the jackknife covariance matrix with 100 mock samples has good consistency with the baseline brute-force one. In the upper panel of Figure 1, we compare the diagonal terms of the covariance matrices. The data points are from the jackknife one, and the solid lines are from the baseline brute-force one. For comparison, we also show the brute-force covariance matrix with fewer mock samples, $N_m = 100$. As
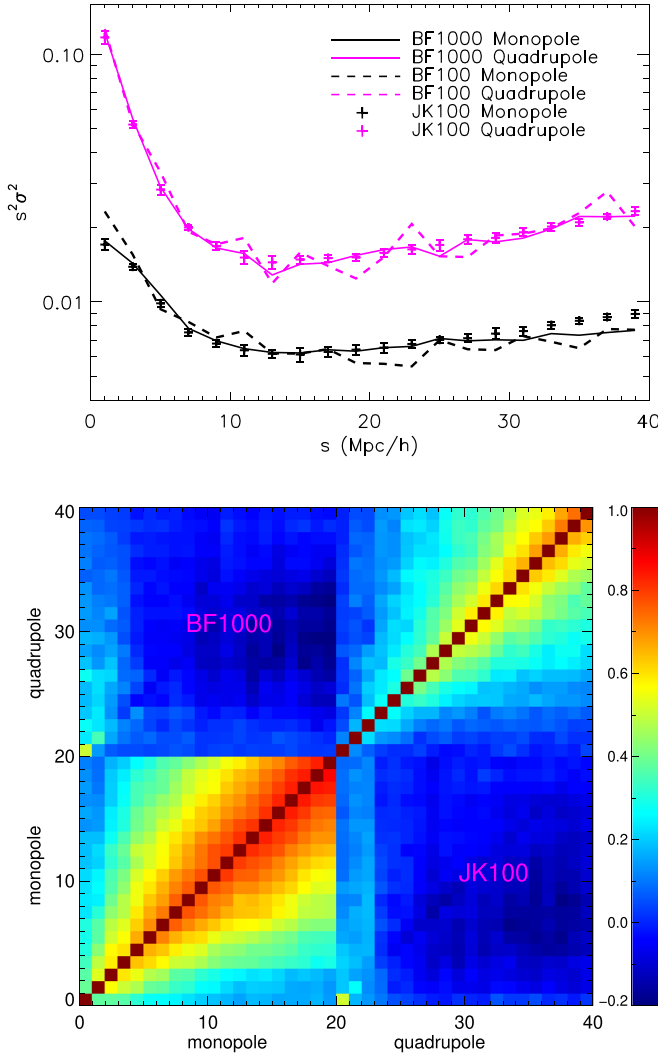
**Figure 1.** The brute-force and delete-$d$ jackknife covariance matrices of the correlation function monopoles and quadrupoles of the QPM mock catalogs on scales $0\ h^{-1}\ \mathrm{Mpc} < s < 40\ h^{-1}\ \mathrm{Mpc}$. The upper panel shows the diagonal elements of the covariance matrices: solid lines for the brute-force method with 1000 mock samples (black for the monopole and magenta for the quadrupole), dashed lines for the brute-force method using 100 mock samples, and plus signs for the delete-$d$ jackknife method with 100 mock samples. The lower panel shows the cross-correlation matrices from the brute-force (upper left corner) and delete-$d$ (lower right corner) methods, respectively.

expected, it has a larger fluctuation compared to the baseline for both the correlation function monopole (black dashed line) and quadrupole (magenta dashed line). For the monopole, the diagonal terms of the jackknife covariance matrix show a slightly increasing bias compared to the baseline as $s > 30\ h^{-1}\ \mathrm{Mpc}$.

We also show the cross-correlation matrix for the jackknife covariance matrix and the baseline brute-force one (in the lower panel in Figure 1), which is defined as

$$R_{ij} = \frac{C_{ij}}{\sqrt{C_{ii} C_{jj}}}. \qquad (15)$$

The cross-correlation matrix is symmetric and has unity diagonal elements. So we show the baseline cross-correlation matrix in the upper left corner (as indicated by the text BF1000) and the jackknife one in the lower right corner

(denoted as JK100). The symmetric feature on small scales indicates that there is good agreement between the two correlation matrices. Again, the discrepancy increases as the scale becomes larger.

There are two important requirements in the delete-$d$ jackknife resampling method. One is that the number of subsamples $N_s$ and the omitted subsamples $d$ should be large enough to satisfy the relation of $N_s - d \gg 1$ and $\sqrt{N_s}/d \gg 1$. The success on small scales demonstrates that our choice of $N_s = 12$ and $d = 6$ is reasonable for the correlation function analysis. Another requirement is that the data in each subsample should be identically and independently distributed. Violating this could break the robustness of the jackknife variance estimator and introduce bias. For galaxy two-point correlation function measurement, subsamples are correlated with each other on large scales due to the large-scale modes of galaxy clustering. This may be the cause for the growing bias with scales on the jackknife covariance matrices.

## 4. KL Divergence: Measurements

In this section, we show the results of the KL divergence for the baseline covariance matrices coming from the same or different sets of mock samples (Section 4.1), as well as the KL divergence for the jackknife covariance matrices and the baseline covariance matrices (Section 4.2). Then we introduce the information factor to estimate the convergence rate of the jackknife resampling method in Section 4.3.

### 4.1. KL Divergence for Brute-force Covariance Matrices

We partition the full 1000 QPM mock samples exclusively into multiple groups. For the setting of $n$-partition, there are $N_S^{(n)} = 1000/n$ groups, each of which contains $n$ mock samples. The groups are arranged in this way: we pack the first $n$ samples of the full catalog into the first group $S_n^1$, the second $n$ samples into $S_n^2$, and so on. We denote the $i$th group as $S_n^i$. We introduce the set $\boldsymbol{S}_n$ to represent the $n$-partition, i.e., $\boldsymbol{S}_n = \{S_n^i | i = 1, \ldots, N_S^{(n)}\}$. We can calculate the brute-force covariance matrix $\hat{C}_n^i$ for each member $S_n^i$ in $\boldsymbol{S}_n$.

Given two multivariate normal distributions, $Q_n^j$ with the covariance matrix $\hat{C}_n^j$ and $P_m^i$ with the covariance matrix $\hat{C}_m^i$, the KL divergence from $Q_n^j$ to $P_m^i$, $\mathrm{KL}(P_m^i | Q_n^j)$, can be calculated based on Equation (2). To avoid any possible correlation between $\hat{C}_n^j$ and $\hat{C}_m^i$, we require that $S_n^j$ and $S_m^i$ do not contain any common mock samples. We can calculate the mean and variance of $\mathrm{KL}(P_m^i | Q_n^j)$ from all the available pairs of $Q_n^j$ and $P_m^i$,

$$\mathrm{KL}(P_m | Q_n) = \frac{1}{N_p} \sum_{i,j} \mathrm{KL}(P_m^i | Q_n^j), \qquad (16)$$

$$\sigma_{\mathrm{KL}}^2(P_m, Q_n) = \frac{1}{N_p - 1} \sum_{i,j} (\mathrm{KL}(P_m^i | Q_n^j) - \mathrm{KL}(P_m | Q_n))^2, \qquad (17)$$

where $N_p$ is the total number of available pairs of $Q_n^j$ and $P_m^i$.

We show the mean and variance of $\mathrm{KL}(P_m^i | Q_n^j)$ calculated using the above formula for the various combinations of $n$-partition and $m$-partition in Figure 2. First, the KL divergence is asymmetric under the exchange of $P_m$ and $Q_n$, that is $\mathrm{KL}(P_m | Q_n) \neq \mathrm{KL}(Q_n | P_m)$. Second, given a fixed $m$ (the number of mock samples to calculate the brute-force covariance matrix of $P_m$), $\mathrm{KL}(P_m | Q_n)$ decreases along with increasing $n$ (the
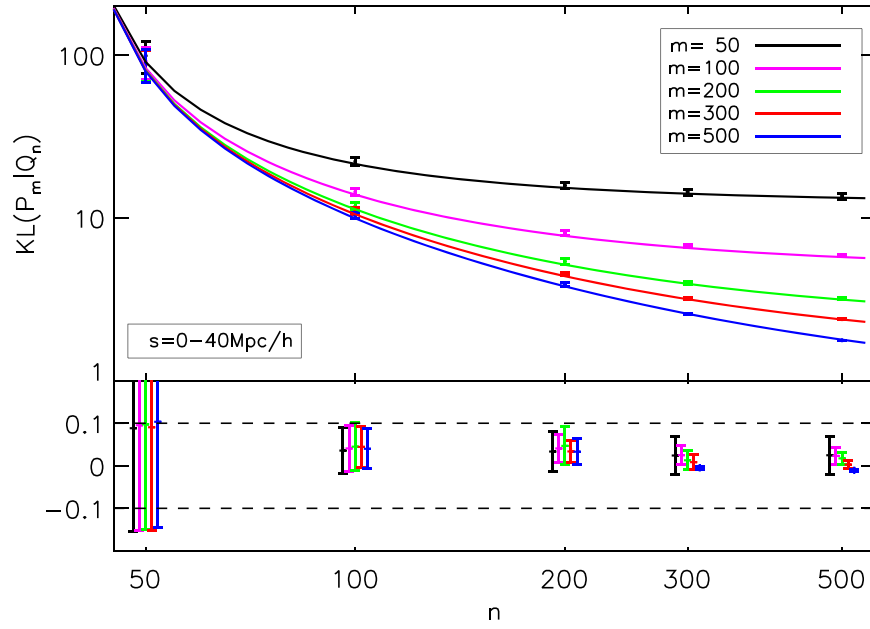
**Figure 2.** The mean and variance of KL($P_m^i|Q_n^j$) as given in Equations (16) and (17), respectively. The upper panel shows the mean KL divergence KL($P_m|Q_n$) (colored symbols) and their model prediction (solid lines). Different colors denote different values of $m$. The error bars are from the standard deviations $\sigma_{\mathrm{KL}}(P_m, Q_n)$. The lower panel shows the fractional difference between the measurements and the model prediction at different $n$. For clarity, we slightly shift the results from different $m$ at a given $n$.

number of mock samples to calculate the brute-force covariance matrix of $Q_n$) and saturates at some large value of $n$. Finally, the value of KL($P_m|Q_n$) is dominated by the covariance matrix, which is calculated from fewer mock samples and so has a higher noise level.

In addition, we show the model prediction of KL($P_m|Q_n$) by assuming that the covariance matrices of $P_m$ and $Q_n$ follow the same Wishart distribution as the solid lines in Figure 2. There is quite good agreement between the measurement and the model prediction, which indicates that the correlation functions measured from the mock samples follow a multivariate normal distribution closely.

### 4.2. KL Divergence for Jackknife Covariance Matrices

Following Section 4.1, for each member in set $S_n$, we calculate the mean delete-$d$ jackknife covariance matrix using Equation (14) and denote it as $\hat{C}_{n,\mathrm{JK}}^j$. Supposing $\hat{C}_{n,\mathrm{JK}}^j$ is the covariance matrix of the multivariate normal distribution $Q_{n,\mathrm{JK}}^j$, the KL divergence from $Q_{n,\mathrm{JK}}^j$ to $P_m^i$, KL($P_m^i|Q_{n,\mathrm{JK}}^i$), can be calculated using Equation (2). Then we calculate the mean and variance of KL($P_m^j|Q_{n,\mathrm{JK}}^i$) using the nonoverlapped pairs of $S_n^j$ and $S_m^i$, similar to Equations (16) and (17), i.e.,

$$\mathrm{KL}(P_m|Q_{n,\mathrm{JK}}) = \frac{1}{N_p}\sum_{i,j}\mathrm{KL}(P_m^j|Q_{n,\mathrm{JK}}^i),$$

$$\sigma_{\mathrm{KL}}^2(P_m, Q_{n,\mathrm{JK}}) = \frac{1}{N_p - 1}\sum_{i,j}(\mathrm{KL}(P_m^j|Q_{n,\mathrm{JK}}^i) - \mathrm{KL}(P_m|Q_{n,\mathrm{JK}}))^2. \quad (18)$$

As discussed in the previous sections, if $\hat{C}_P$ and $\hat{C}_Q$ contain the same signal or expectation value, the KL divergence from the multivariate normal distribution $Q$ with the covariance matrix $\hat{C}_Q$ to the multivariate normal distribution $P$ with the covariance matrix $\hat{C}_P$, KL($P|Q$), is determined by the noise level of $\hat{C}_Q$ and $\hat{C}_P$.

KL($P|Q$) can be larger when $\hat{C}_Q$ contains less information (or a higher noise level) than $\hat{C}_P$, and vice versa; hence, we expect that KL($P|Q$) can measure the relative amount of information contained in $\hat{C}_P$ and $\hat{C}_Q$. As shown in Figure 1, due to a limited number of mock samples, the sample variance can cause large fluctuation (noise) on the baseline covariance matrix from the brute-force method. To reduce such a noise effect, we introduce the information factor defined as

$$\eta(Q_m|Q_{n,\mathrm{JK}}) = \frac{\mathrm{KL}(P_m|Q_{n,\mathrm{JK}})}{\mathrm{KL}(P_m|Q_m)}, \quad (19)$$

whose denominator is calculated from Equation (16) with $m = n$. The information factor compares the statistical information contained in the jackknife covariance matrix $\hat{C}_{n,\mathrm{JK}}$ and the baseline covariance matrix $\hat{C}_m$. If $\hat{C}_{n,\mathrm{JK}}$ contains the same information $\hat{C}_m$, then $\eta = 1$. Otherwise, $\eta < 1$ or $\eta > 1$, if $\hat{C}_{n,\mathrm{JK}}$ contains more or less information, respectively. Therefore, the intersections between the solid curves (linked to the data points) and the horizontal dashed line (showing $\eta = 1$) in Figure 3 give the estimated number of mock samples required for the jackknife covariance matrices to be equivalent to the baseline covariance matrices in the sense of KL divergence.

The variance of the information factor can be roughly estimated using the variance of the numerator in Equation (19),

$$\sigma_\eta^2(Q_m, Q_{n,\mathrm{JK}}) \approx \frac{\sigma_{\mathrm{KL}}^2(P_m, Q_{n,\mathrm{JK}})}{[\mathrm{KL}(P_m|Q_m)]^2}. \quad (20)$$

We do not account for the contribution from the variance of KL divergence between the baseline covariance matrices (the denominator), so it will underestimate the true variance. As can be seen in Figure 2, the proportional variance of the denominator is about 10% at most. In Figure 3, we show the error bars of $\eta$ based on the approximated variance. The dashed curves show the
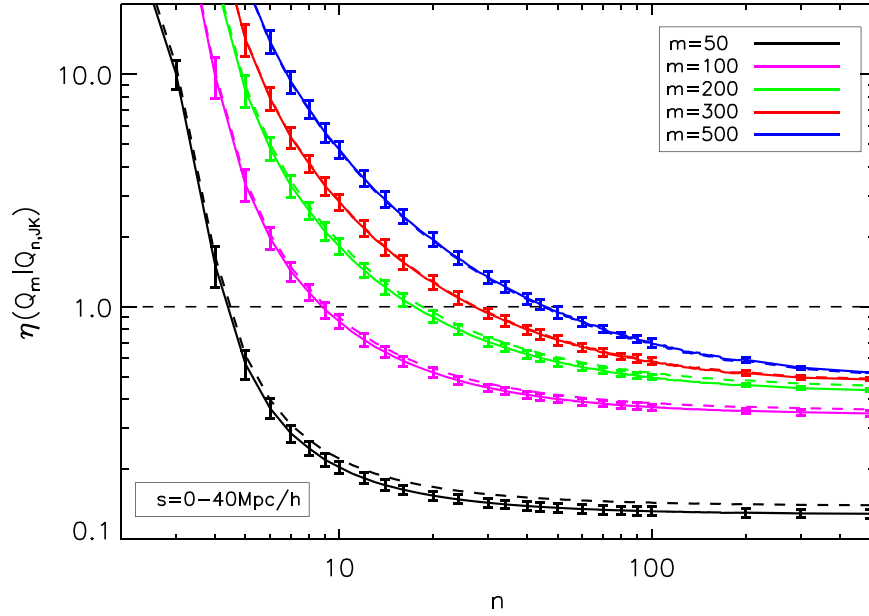
**Figure 3.** Information factor of QPM mocks, $\eta(Q_m|Q_{n,\mathrm{JK}})$, defined in Equation (19). $n$ is the number of mock samples used to calculate the jackknife covariance matrices as in Equation (14), and $m$ is the number of mock samples to calculate the baseline covariance matrices. Different colors denote different $m$. We also show the results with the denominator in Equation (19) replaced by the model prediction as the dashed curves.
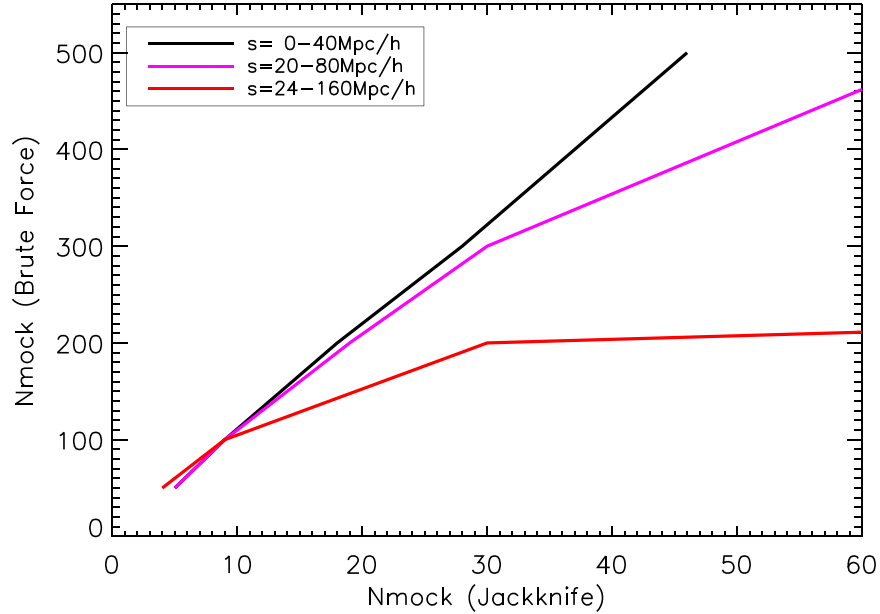


**Figure 4.** Scaling law of the number of mock samples needed by the jackknife resampling method and the baseline method to give statistically equivalent covariance matrices.

information factors that replace the denominator in Equation (19) by the model prediction for Gaussian-distributed data. They have a good agreement with the solid curves.

### 4.3. Convergence Rate of Jackknife Covariance Matrix

As shown in Section 4.2, the information factor defined in Equation (19) can quantify the relative information contained in the jackknife covariance matrices $\hat{C}_{n,\mathrm{JK}}$ and the baseline covariance matrices $\hat{C}_m$. Given $m$, increasing the number of mock samples for $\hat{C}_{n,\mathrm{JK}}$ decreases the information factor $\eta(Q_m|Q_{n,\mathrm{JK}})$. At the point $\eta = 1$, we consider that the jackknife covariance matrices converge to the baseline covariance matrices, since they contain

the same information statistically. In Figure 4, we show the convergence rate of the covariance matrix calculated from he jackknife resampling and the brute-force methods based on the QPM mock samples. For the distance scales of $0\ h^{-1}\,\mathrm{Mpc} < s < 40\ h^{-1}\,\mathrm{Mpc}$ (black solid line), there is a linear scaling law for the number of mocks required to obtain the statistically equivalent covariance matrices from the two methods.

## 5. Convergence Rate: Tests

### 5.1. The Reference Covariance Matrices

When we compare the statistical information of the jackknife and baseline covariance matrices using the information factor
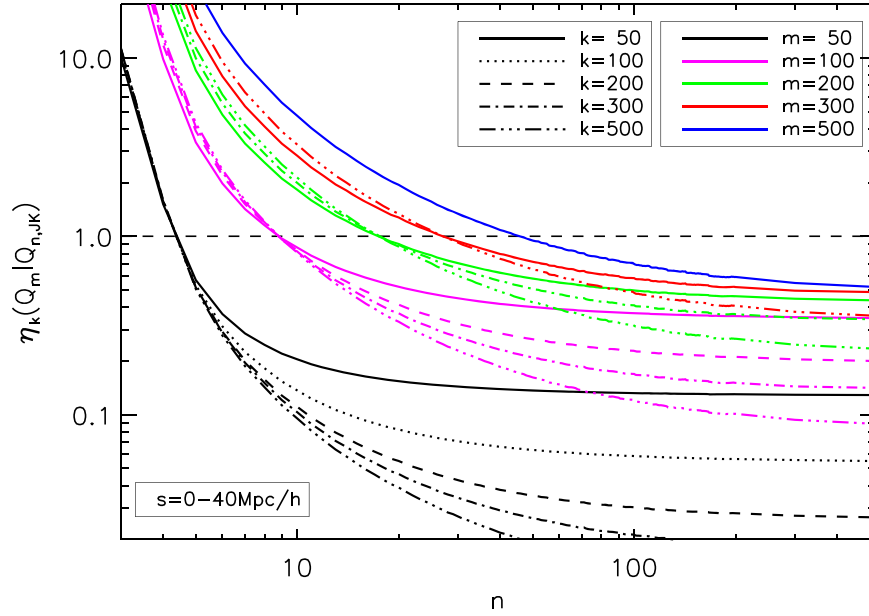
**Figure 5.** Extended information factor, $\eta_k(Q_m|Q_{n,\mathrm{JK}})$. The different colors of the lines represent different numbers of mock samples used in $\hat{C}_m$, and different line types denote different numbers of mocks used in the reference covariance matrices $\hat{C}_k$.

defined in Equation (19), we use $P_m$ as the reference distribution function. Since the covariance matrices of $P_m$ and $Q_m$ are calculated using the same number of mock samples, they contain roughly the same statistical information. In the following, we break this limitation and replace $P_m$ with $P_k$, where $k$ is not necessarily equal to $m$, to test the robustness of the information factor in quantifying the information contained in two covariance matrices. We introduce the extended information factor,

$$\eta_k(Q_m|Q_{n,\mathrm{JK}}) = \frac{\mathrm{KL}(P_k|Q_{n,\mathrm{JK}})}{\mathrm{KL}(P_k|Q_m)}. \quad (21)$$

If $Q_{n,\mathrm{JK}}$ and $Q_m$ are the same distribution function, i.e., their covariance matrices $\hat{C}_{n,\mathrm{JK}}$ and $\hat{C}_m$ contain the same information, then $\eta_k(Q_m|Q_{n,\mathrm{JK}}) = 1$ for any $k$.

In Figure 5, we show the extended information factor $\eta_k(Q_m|Q_{n,\mathrm{JK}})$ as a function of $n$ for difference combinations of $k$ and $m$. As shown by the different types of lines with the same color, the shape of the extended information factor as a function of $n$ (the number of mock samples used to calculate the jackknife covariance matrices) varies as $k$ changes; i.e., it is steeper for larger $k$. However, they converge to the value of 1 at almost the same position (at the same $n$) when the two covariance matrices contain almost the same amount of information. This clearly shows the robustness of the information factor defined by Equation (19) to quantify the relative information between the two covariance matrices with respect to the reference covariance matrices.

### 5.2. When Bias Is Present: Intermediate and Large Scales

As discussed in Section 2.2, a linear bias on $\hat{C}_Q$ with respect to $\hat{C}_P$ will introduce an additional term, $\Delta_{\mathrm{KL}}$ (Equation (6)), on the KL divergence from the multivariate normal distribution $Q$ with covariance matrix $\hat{C}_Q$ to the multivariate normal distribution $P$ with covariance matrix $\hat{C}_P$, KL($P|Q$). And $\Delta_{\mathrm{KL}}$ is almost positive definite. So any bias on the estimate of the covariance
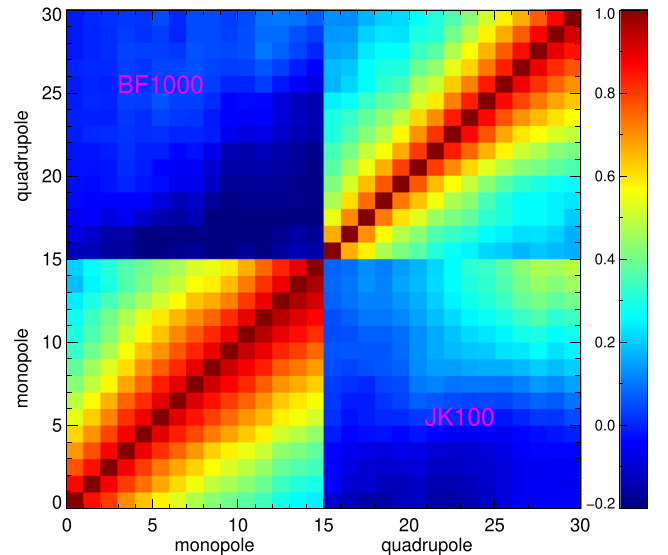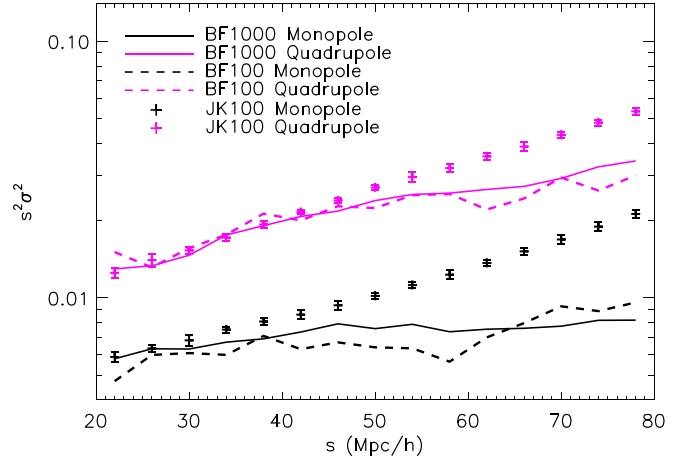




**Figure 6.** Same as Figure 1 but for the scales $20\ h^{-1}$ Mpc $< s < 80\ h^{-1}$ Mpc.
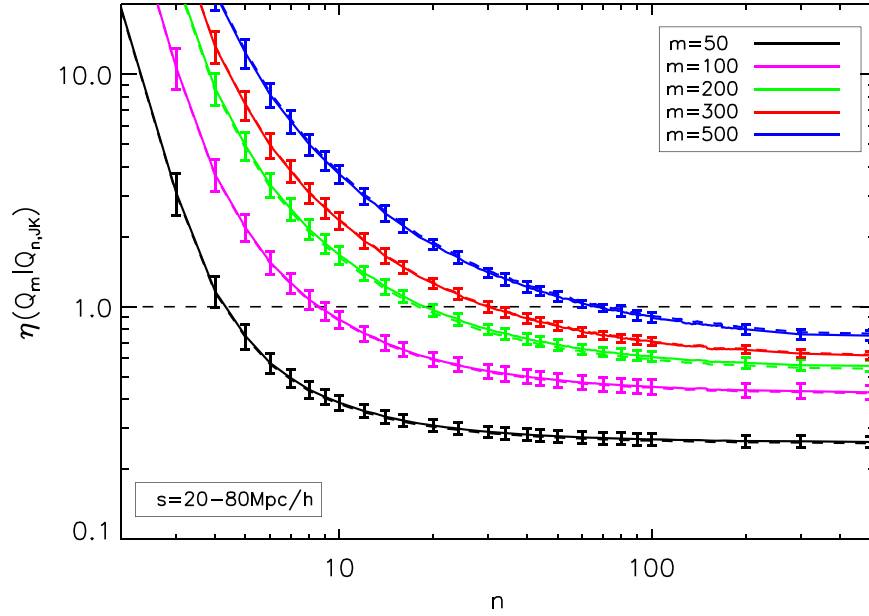
**Figure 7.** Same as Figure 3 but for the scales $20\,h^{-1}\,\mathrm{Mpc} < s < 80\,h^{-1}\,\mathrm{Mpc}$.

matrix will affect the power of the information factor to quantify the relative information between the estimated covariance matrix and the true one.

The jackknife resampling method tends to produce biased estimates for the covariance matrix of the galaxy correlation function. This is partially due to the nonlinear feature of the two-point correlation function. Furthermore, the possible correlation between different subregions in the galaxy sample will break the independence of the jackknife observations and introduce bias on the jackknife covariance matrix.

In this section, we study the correlation function at intermediate scales ($20\,h^{-1}\,\mathrm{Mpc} < s < 80\,h^{-1}\,\mathrm{Mpc}$) to test the effect of the covariance matrix bias on the information factor. Following Section 3.3, we calculate the baseline covariance matrix by the brute-force method and the jackknife covariance matrix using the delete-$d$ method. Similar to Figure 1, we show the resulting baseline covariance matrix from 1000 QPM mocks and one of the jackknife covariance matrices from 100 QPM mocks in Figure 6. As shown in the upper panel, there are significant biases on the diagonal terms of the jackknife covariance matrix that are larger at larger scales. The off-diagonal terms also show clear biases in the lower panel.

Similarly, the information factors for the distance scales $20\,h^{-1}\,\mathrm{Mpc} < s < 80\,h^{-1}\,\mathrm{Mpc}$ are shown in Figure 7. Compared with the small-scale one, the information factor as a function of $n$ (the number of mocks for the jackknife covariance matrix) becomes flatter and converges to a larger value for a given $m$ (the number of mocks for the brute-force covariance matrix) on intermediate scales. As a result, more mock samples are needed for the jackknife covariance matrix to contain the same information as the brute-force covariance matrix. The solid magenta line in Figure 4 shows the scaling law of $n$ as a function of $m$ on intermediate scales, which is clearly flatter than the small-scale one (solid black line).

It can be expected that the information factor will converge to larger and larger values when $m$ increases and eventually will stay above 1 forever, where the bias on the jackknife covariance matrix dominates the statistical noise. Limited by the number of mock samples available, this phenomenon is not

explored further in our study. Instead, we observe a similar phenomenon using data on larger scales where the bias on the jackknife covariance matrices becomes even larger. The information factors for the distance scales of $24\,h^{-1}\,\mathrm{Mpc} < s < 160\,h^{-1}\,\mathrm{Mpc}$ are shown in Figure 8. As expected, the information factors with $m = 300$ converge to a value close to 1. When $m$ goes up to 500, the information factors stay above 1 all the time. The scaling law of the number of mock samples needed by the jackknife resampling method and the baseline method for the distance scales of $24\,h^{-1}\,\mathrm{Mpc} < s < 160\,h^{-1}\,\mathrm{Mpc}$ is shown in Figure 4 as the red solid line, which is much flatter than those from the smaller distance scales.

## 6. Discussion and Summary

In this paper, we have proposed a simple method to diagnose the equality or similarity of two covariance matrices and then to calculate the convergence rate of the fast covariance matrix estimators for galaxy clustering measurements. The essence of the method is based on the fact that we are only interested in the Gaussian likelihood function characterized by the covariance matrix, rather than the covariance matrix itself. The KL divergence is a perfect tool to do this job.

As a case study, we explore the delete-$d$ jackknife covariance matrix estimator, which is one of the fast covariance matrix estimators in galaxy clustering analysis. In general, the jackknife covariance matrix contains both bias and noise with respect to the true covariance matrix, which contributes to the KL divergence. So we introduce the information factor (Equation (19)) to study the statistical information in the jackknife covariance matrix.

In this work, we focus on the anisotropic two-point galaxy correlation function and study its covariance matrix based on the QPM mock samples. We first test the KL divergence for the brute-force covariance matrices coming from different numbers of mock samples on the scale range $0\,h^{-1}\,\mathrm{Mpc} < s < 40\,h^{-1}\,\mathrm{Mpc}$. We find that they are consistent with the Gaussian predictions. Then we calculate the information factor using the jackknife and the brute-force covariance matrices and
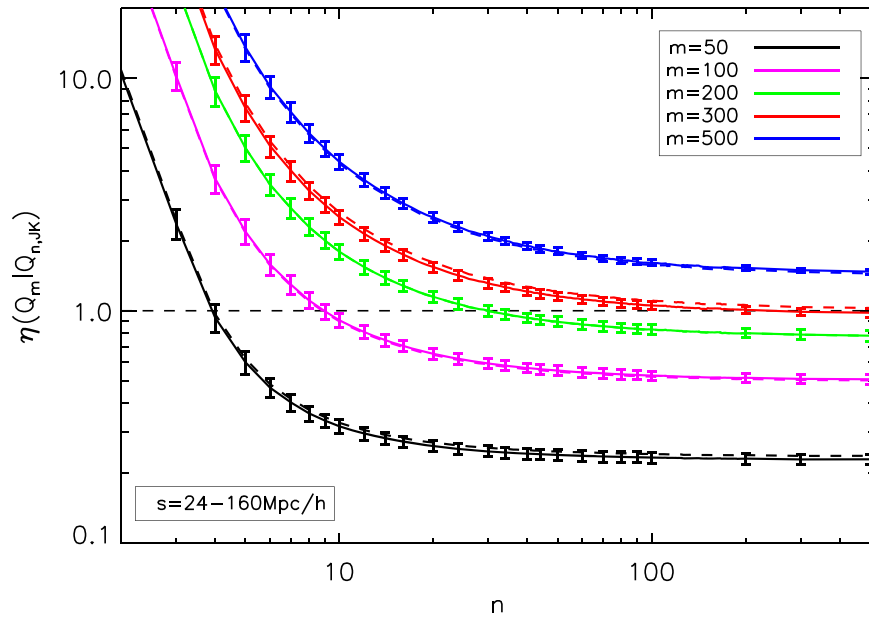
**Figure 8.** Same as Figure 3 but for the scales 24 $h^{-1}$ Mpc $< s < 160$ $h^{-1}$ Mpc.

estimate the convergence rate for the jackknife resampling method. We find that the jackknife resampling can recover the brute-force covariance matrices statistically by using about 10 times fewer mock samples. This can be supportive for the study of galaxy clustering at small scales with a small number of mocks.

By introducing the extended information factor (Equation (21)), we test the robustness of the information factor in Section 5.1. Although a general choice of $k$ in the extended information factor can give us more information, the simplification by taking $k = m$ can still catch the point where $Q_{n,\text{JK}}$ and $Q_m$ contain the same information. In addition, we study the influence of the bias of the covariance matrix on the KL divergence based on the correlation functions on larger scales. We find that the bias on the jackknife covariance matrix reduces the power of the jackknife resampling method in recovering the brute-force covariance matrix.

The analysis presented in this paper can be applied to other fast estimators of the galaxy clustering covariance matrix. The findings on the limitation of the jackknife resampling method, especially the significant bias of the covariance matrix on large scales, are generic. Favole et al. (2021) recently studied the constraint on the baryon acoustic oscillation scale from the jackknife covariance based on the CMASS data or mocks and found no significant bias compared to that from the brute-force covariance matrix. However, the jackknife covariance bias is still worthy of being carefully investigated for the next-generation redshift surveys, such as DESI and Euclid, as their statistical error will be subdominant compared to the systematics. A larger set of simulation mock catalogs is required to investigate such cases with small bias.

## Acknowledgments

## ORCID iDs

Zhigang Li ⓘ https://orcid.org/0000-0002-1324-0893
Yu Yu ⓘ https://orcid.org/0000-0002-9359-7170

## References

Anderson, L., Aubourg, É., Bailey, S., et al. 2014, MNRAS, 441, 24
Arnalte-Mur, P., & Norberg, P. 2014, in IAU Symp. 306, Statistical Challenges in 21st Century Cosmology, ed. A. Heavens, J.-L. Starck, & A. Krone-Martins (Cambridge: Cambridge Univ. Press), 247
Baez, J. C., & Fritz, T. 2014, arXiv:1402.3067
Balaguera-Antolínez, A., Kitaura, F.-S., Pellejero-Ibáñez, M., Zhao, C., & Abel, T. 2019, MNRAS, 483, L58
Bertolini, D., Schutz, K., Solon, M. P., Walsh, J. R., & Zurek, K. M. 2016, PhRvD, 93, 123505
Beutler, F., Saito, S., Seo, H.-J., et al. 2014, MNRAS, 443, 1065
Blot, L., Corasaniti, P. S., Alimi, J.-M., Reverdy, V., & Rasera, Y. 2015, MNRAS, 446, 1756
Carron, J., Wolk, M., & Szapudi, I. 2015, MNRAS, 453, 450
Chapman, M. J., Mohammad, F. G., Zhai, Z., et al. 2022, MNRAS, 516, 617
Chuang, C.-H., Zhao, C., Prada, F., et al. 2015, MNRAS, 452, 686
Dodelson, S., & Schneider, M. D. 2013, PhRvD, 88, 063537
Escoffier, S., Cousinou, M.-C., Tilquin, A., et al. 2016, arXiv:1606.00233
Favole, G., Granett, B. R., Silva Lafaurie, J., & Sapone, D. 2021, MNRAS, 505, 5833
Feng, Y., Chu, M.-Y., Seljak, U., & McDonald, P. 2016, MNRAS, 463, 2273
Frieman, J. & Dark Energy Survey Collaboration 2013, AAS Meeting Abstracts, 221, 335.01
Gil-Marín, H., Percival, W. J., Cuesta, A. J., et al. 2016, MNRAS, 460, 4210
Grieb, J. N., Sánchez, A. G., Salazar-Albornoz, S., & Dalla Vecchia, C. 2016, MNRAS, 457, 1577
Hartlap, J., Simon, P., & Schneider, P. 2007, A&A, 464, 399
Hikage, C., Takahashi, R., & Koyama, K. 2020, PhRvD, 102, 083514
Howlett, C., & Percival, W. J. 2017, MNRAS, 472, 4935
Kitaura, F.-S., Gil-Marín, H., Scóccola, C. G., et al. 2015, MNRAS, 450, 1836
Kitaura, F.-S., Yepes, G., & Prada, F. 2014, MNRAS, 439, L21
Klypin, A., & Prada, F. 2018, MNRAS, 478, 4602
Kullback, S., & Leibler, R. A. 1951, Ann. Math. Stat., 22, 79
Lacasa, F., & Kunz, M. 2017, A&A, 604, A104
Landy, S. D., & Szalay, A. S. 1993, ApJ, 412, 64

Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv:1110.3193
Li, Y., Hu, W., & Takada, M. 2014, PhRvD, 89, 083519
Lippich, M., Sánchez, A. G., Colavincenzo, M., et al. 2019, MNRAS, 482, 1786
LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv:0912.0201
Manera, M., Samushia, L., Tojeiro, R., et al. 2015, MNRAS, 447, 437
Manera, M., Scoccimarro, R., Percival, W. J., et al. 2013, MNRAS, 428, 1036
Mohammad, F. G., & Percival, W. J. 2022, MNRAS, 514, 1289
Mohammed, I., & Seljak, U. 2014, MNRAS, 445, 3382
Mohammed, I., Seljak, U., & Vlah, Z. 2017, MNRAS, 466, 780
Monaco, P., Sefusatti, E., Borgani, S., et al. 2013, MNRAS, 433, 2389
Monaco, P., Theuns, T., & Taffoni, G. 2002, MNRAS, 331, 587
Neyrinck, M. C. 2011, ApJ, 736, 8
Norberg, P., Baugh, C. M., Gaztañaga, E., & Croton, D. J. 2009, MNRAS, 396, 19
O'Connell, R., Eisenstein, D., Vargas, M., Ho, S., & Padmanabhan, N. 2016, MNRAS, 462, 2681
O'Connell, R., & Eisenstein, D. J. 2019, MNRAS, 487, 2701
Paz, D. J., & Sánchez, A. G. 2015, MNRAS, 454, 4326
Pearson, D. W., & Samushia, L. 2016, MNRAS, 457, 993
Percival, W. J., Ross, A. J., Sánchez, A. G., et al. 2014, MNRAS, 439, 2531
Philcox, O. H. E., Eisenstein, D. J., O'Connell, R., & Wiegand, A. 2020, MNRAS, 491, 3290
Pope, A. C., & Szapudi, I. 2008, MNRAS, 389, 766
Rashkovetskyi, M., Eisenstein, D. J., Aguilar, J. N., et al. 2023, MNRAS, 524, 3894
Reid, B. A., Percival, W. J., Eisenstein, D. J., et al. 2010, MNRAS, 404, 60
Schlegel, D., Abdalla, F., Abraham, T., et al. 2011, arXiv:1106.1706
Schneider, M. D., Cole, S., Frenk, C. S., & Szapudi, I. 2011, ApJ, 737, 11
Scoccimarro, R., & Sheth, R. K. 2002, MNRAS, 329, 629
Shao, J., & Wu, C. F. J. 1989, AnSta, 17, 1176
Shi, F., Yang, X., Wang, H., et al. 2016, ApJ, 833, 241
Sugiyama, N. S., Saito, S., Beutler, F., & Seo, H.-J. 2020, MNRAS, 497, 1684
Takahashi, R., Yoshida, N., Takada, M., et al. 2009, ApJ, 700, 479
Taruya, A., Nishimichi, T., & Jeong, D. 2021, PhRvD, 103, 023501
Tassev, S., Zaldarriaga, M., & Eisenstein, D. J. 2013, JCAP, 6, 036
Taylor, A., & Joachimi, B. 2014, MNRAS, 442, 2728
Taylor, A., Joachimi, B., & Kitching, T. 2013, MNRAS, 432, 1928
Wadekar, D., Ivanov, M. M., & Scoccimarro, R. 2020, PhRvD, 102, 123521
Wadekar, D., & Scoccimarro, R. 2020, PhRvD, 102, 123517
Wechsler, R. H., & Tinker, J. L. 2018, ARA&A, 56, 435
White, M., Tinker, J. L., & McBride, C. K. 2014, MNRAS, 437, 2594
Wishart, J. 1928, Biometrika, 20A, 32
Wu, C. F. J. 1986, AnSta, 14, 1261
Xu, H., Li, H., Zhang, J., et al. 2023, SCPMA, 66, 129811
Yu, J., Zhao, C., Chuang, C.-H., et al. 2022, MNRAS, 516, 57
Yuan, S., Garrison, L. H., Eisenstein, D. J., & Wechsler, R. H. 2022, MNRAS, 515, 871
Zhang, P., Pan, J., & Zheng, Y. 2013, PhRvD, 87, 063526
Zheng, Y., & Song, Y.-S. 2016, JCAP, 8, 050