Higher-Order Statistics on Linear and Non-Linear Scales

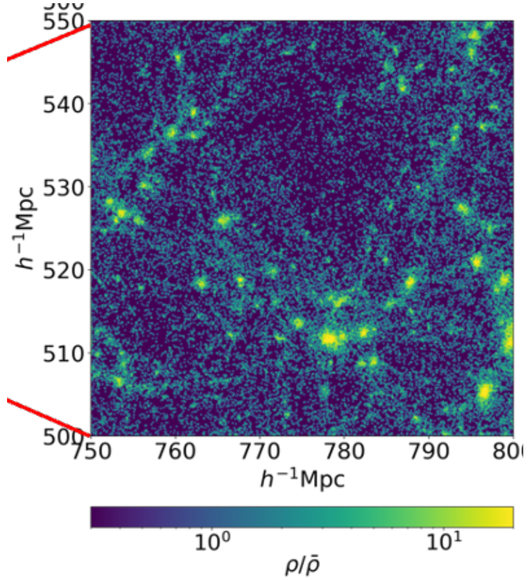
From Galaxies to Cosmology with Large Spectroscopic Surveys, Marseille, 4th July, 2022

Lado Samushia, Kansas State University & Abastumani Observatory

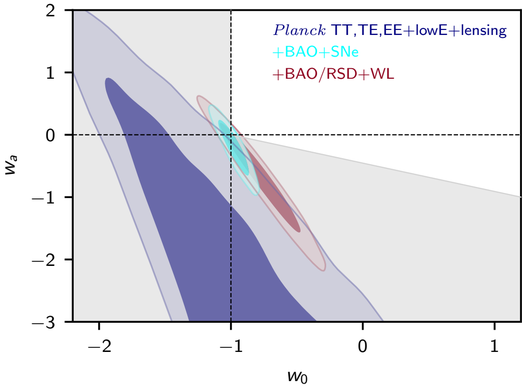
Sponsors:
Shota Rustaveli national science foundation of Georgia grant FR 19-498
NASA ROSES grants 12-EUCLID12-0004
US Department of Energy grants DE-SC0021165 and DE-SC0011840

From Galaxies to Cosmology

- The Ultimate goal of our field in a nutshell is to take a distribution of observed galaxies and to learn important things about the physics of the Universe (derive constraints on cosmological parameters).
- The direct field-level analysis approach tries to achieve it by running lots of high-fidelity simulations and directly comparing the galaxy distributions in the simulations to data.
- This is in principle the most complete solution but is currently not viable because the required computational costs are still prohibitive.



Simulation of a Universe
From Kaushal et al. 2020, 2111.02441

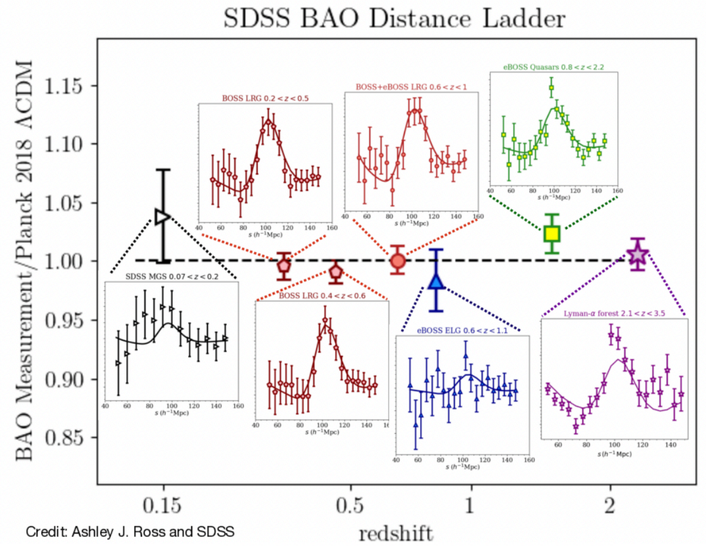


Cosmological constraints
From Planck 2018, 1807.06209

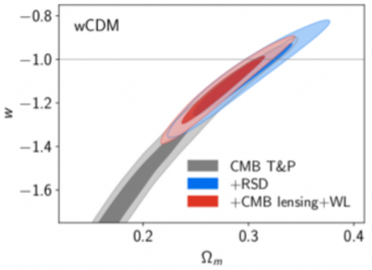
Standard analysis

BAO/RSD

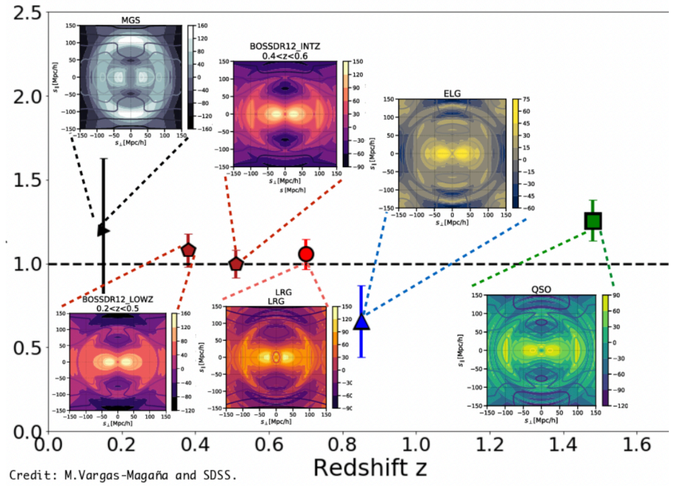
- Compute a 2pt statistics of the galaxies (configuration space - correlation function, Fourier space - power spectrum)
- BAO feature is sensitive to the Alcock-Paczyński effect. Can be used to derive measurements of distance.
- RSD feature is sensitive to the growth of structure. Can be used to constrain the properties of gravity.
- The measurements are well-understood. The theory can be computed both with analytical tools and simulations.
- Robust with respect to observational systematics and small-scale physics.
- Can we do more?



Credit: Ashley J. Ross and SDSS



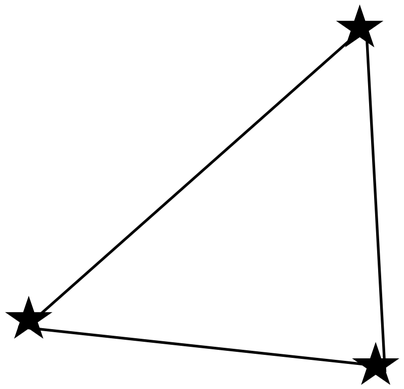
Cosmology from eBOSS standard analysis



Credit: M.Vargas-Magaña and SDSS.

Higher-Order Statistics

3pt - Bispectrum



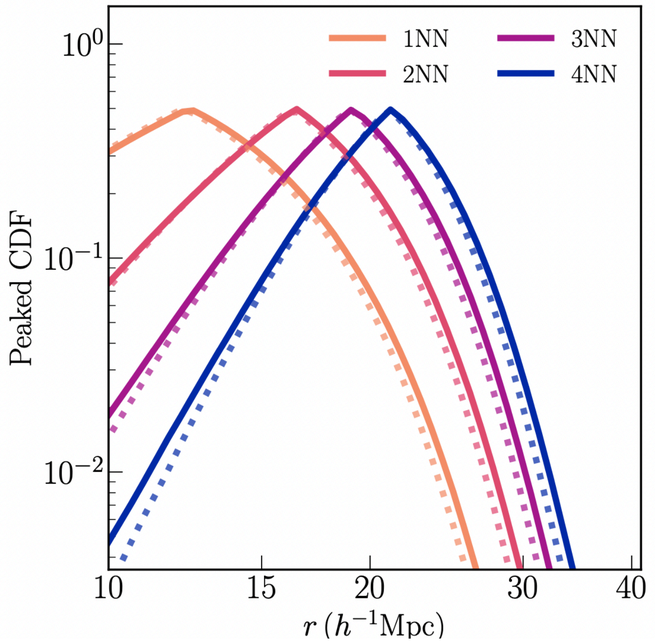
3pt function - the probability of finding three galaxies in a specific configuration

- The hierarchy of the n-point correlation functions contains all of the non-stochastic information about the field (although see Caron 2011, APJ, **738**, 86, for the problems with convergence)
- They are difficult and time-consuming to compute (although see Philcox & Slepian, 2022, 2106.10278, for new developments)
- 3pt functions are the most natural step beyond the standard analysis.
- In principle, a 5D function of three triangle side-lengths and two orientation angles.
- They can now be computed very efficiently (see e.g. https://github.com/ladosamushia/Bispectrum)
- On large scales can be modeled both theoretically and with simulations, and may contain most of the leftover information from the 2pt.

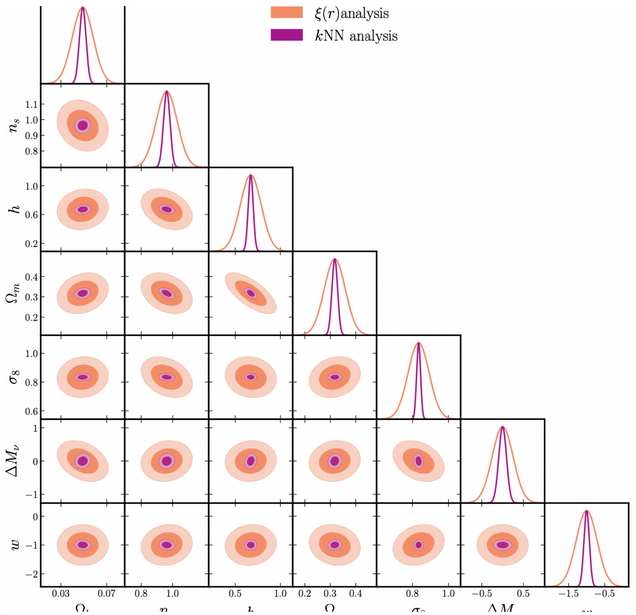
Beyond Two-Point

Integrated statistics

- In principle, one does not have to measure the hierarchy of the n-point functions.
- We can pick any statistics of the galaxy field that is easy to measure both in the simulations and in the data.
- Something that we believe can be made free from systematics (observational, simulation-based).
- These measurements are under the hood some kind of effective integrated combination of n-point functions, but we don't really need to know what they are.
- E.g. k-nearest neighbor cumulative distribution functions.



Peaked CDF at z=0 (solid) and z=0.5 (dotted)

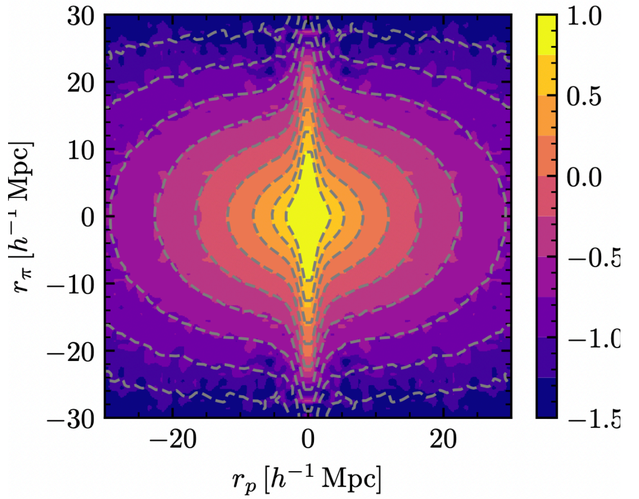


Cosmological constraints from kNN-CDF
From Banerjee & Abel, 2020, 2007.13342

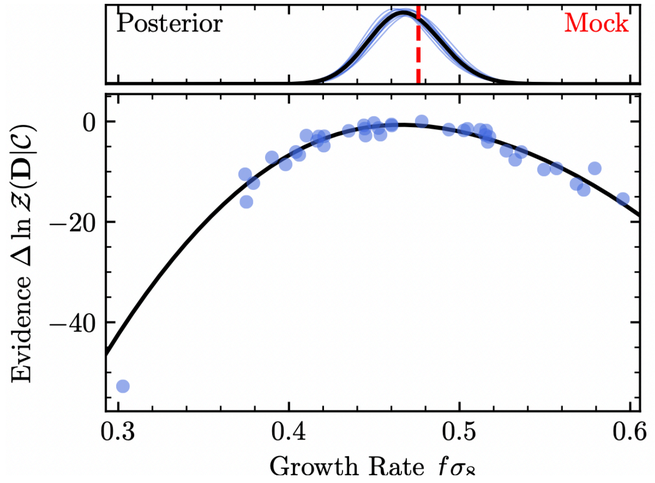
Small-Scale Clustering

HOD and cosmology

- Clustering on Small Scales in very highly nonlinear
- It is reasonable to expect that the higher-order statistics could be very useful in this regime
- The main problem with the small-scale clustering is that it is very sensitive to the physics of galaxy formation which is difficult to account for in simulations
- The Halo Occupation Distribution: Populate halos (that we have in simulations) with galaxies (that we do not have in simulations) based on the local properties of the halo and its neighborhood (mass, local density)
- Many groups working on the hydrodynamical simulations have reported noticeable disagreement with N-body simulations at tens of Megaparsec scales! The HOD approach may have non-resolvable problems at those scales.
- Multiple groups are working on cosmology fits to the small-scale clustering

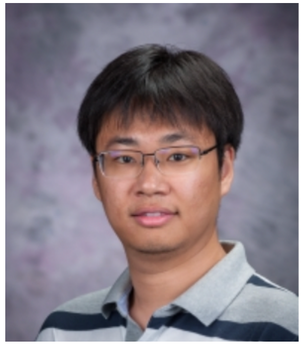


BOSS small scale 2pt function



5% growth rate measurement from BOSS small scale 2pt function
From Lange et al. 2020, 2101.12261

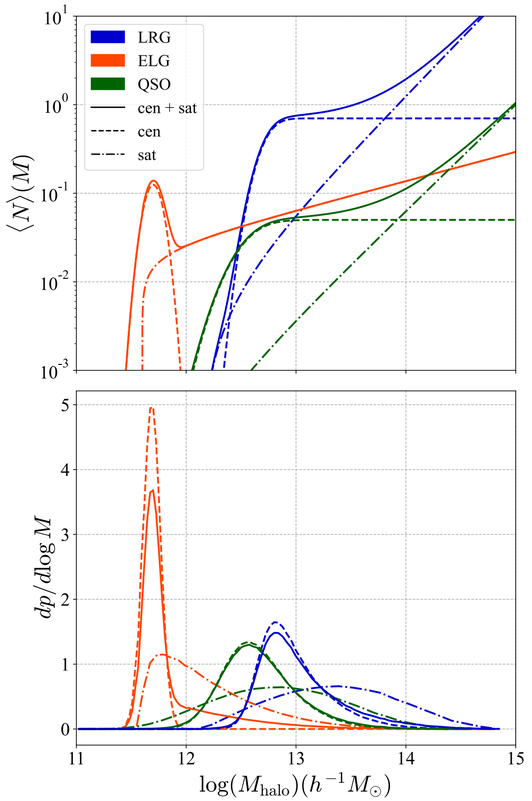
HOD constraints from projected 3pt correlation function



Hanyu Zhang

Preliminaries

- The project was led by a current KSU student Hanyu Zhang
- On small scales, where the clustering is highly nonlinear, can higher-order statistics enhance constraints on galaxy-halo connection?
- HOD - put galaxies in halos based on probabilistic prescriptions about position and velocity (usually based on halo mass and other local parameters).
- Constrain HOD parameters by comparing projected 2pt/3pt functions measured from the Abacus simulations with the measurements.



The average number of central/satellite galaxies in a halo.
Probability of a halo hosting a galaxy.
From Zhang et al. (2022, arXiv, 2203.17214)

A simple parameterization for the average Number of central/satellite galaxies in a halo containing adjustable HOD parameters

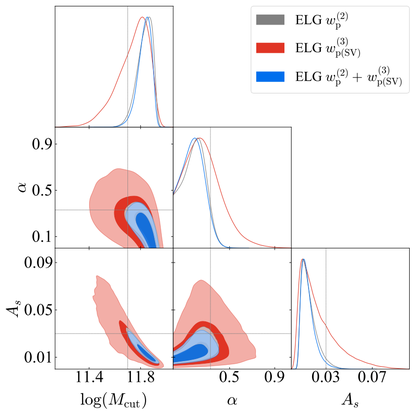
$$\langle N^c_{\rm lrg} \rangle(M) = \frac{A_c}{2}\left(1 + {\rm erf}\left[\frac{\log(M) - \log(M_{\rm cut})}{\sigma}\right]\right).$$

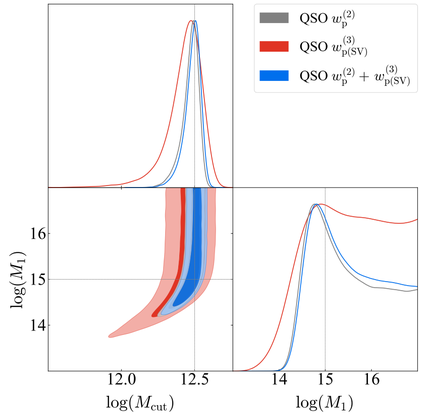
$$\langle N^s_{\rm lrg} \rangle(M) = A_s \left(\frac{M - M_0}{M1}\right)^{\alpha} H(M - M_0).$$

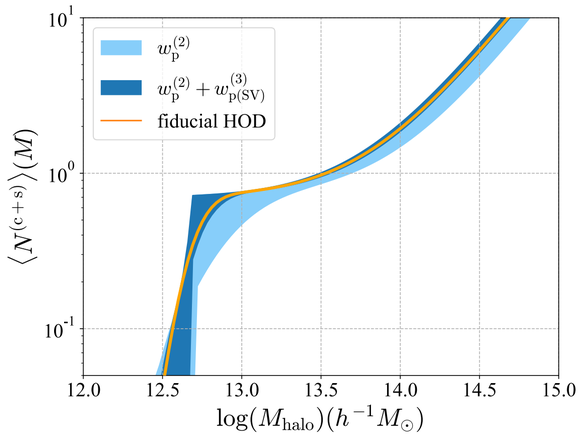
HOD constraints from projected 3pt correlation function

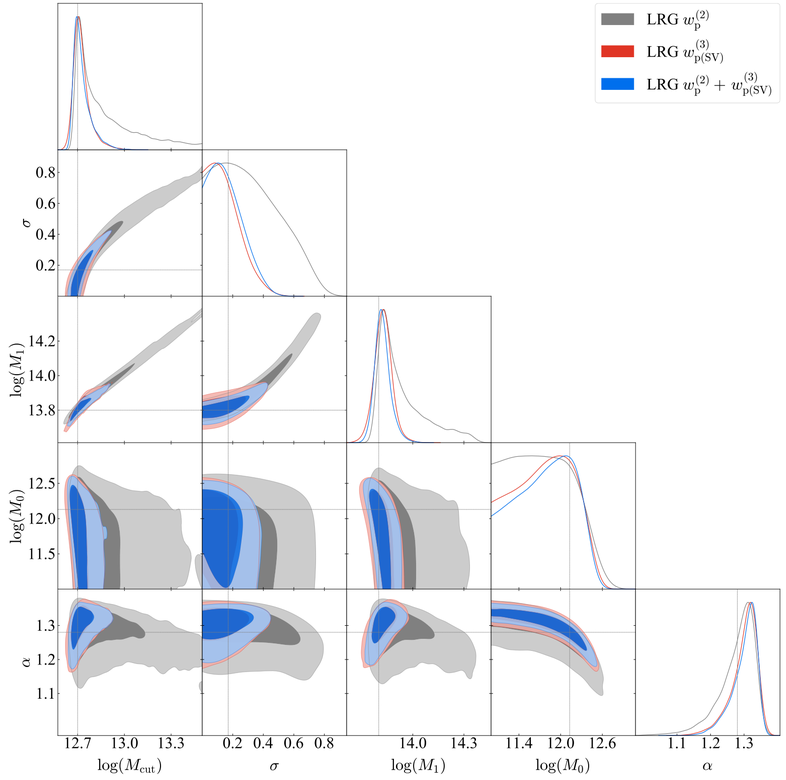
DESI LRG-like sample from Abacus Simulations

- 3pt signal completely dominates the 2pt signal on the HOD parameters for the LRGs
- In other words, small changes in HOD result in large changes in the 3pt correlation (compared to the 2pt)
- This is not the case for the ELGs, where the contribution of the 3pt to the total constraining power is negligible.
- For this, accelerating triplet counts, binning pitfalls, and many other things see Zhang et al. 2022 (arXiv, 2203.17214)





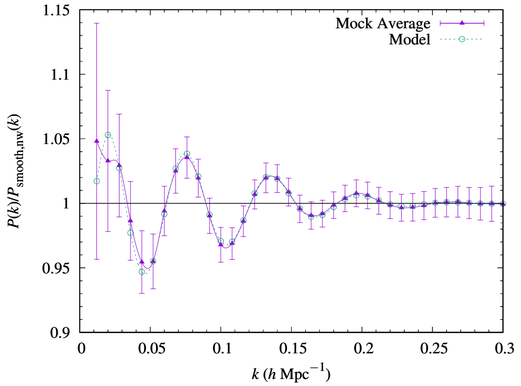




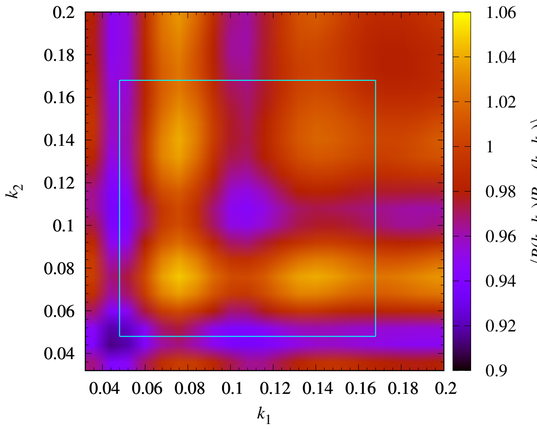
Large-Scale Clustering

BAO in the 3pt function

- A harmonic feature at very large scales (100 Mpc). Usually measured in the 2pt function.
- An effective spherical shell of over-density around each galaxy.
- Sensitive to distortions by the Alcock-Paczyński effect.
- Robust with respect to observational systematic effects.
- Robust with respect to small-scale galaxy physics.
- Should also be present in all the higher-order functions (correlated with the 2pt function).
- The constraining power of the data set scales as $\sqrt{V}$, where V is the survey volume.



BAO signal in the power spectrum



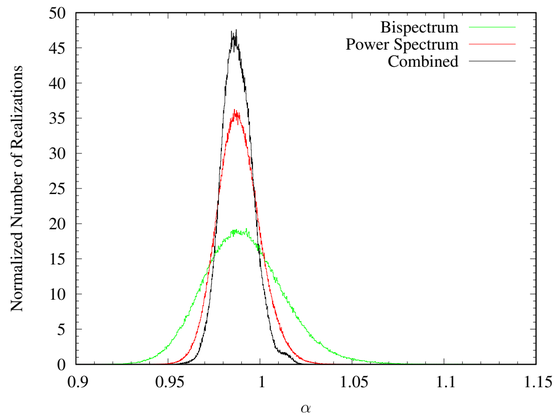
BAO signal in the integrated bispectrum
From Pearson & Samushia 2018

BAO from Galaxy Bispectrum

BOSS DR12

- The project was led by then a KSU Postdoc David Pearson (now a faculty at Pittsburgh State University)
- BAO signal is also present in the large-scale bispectrum as a 3D harmonic signal.
- By aligning the frequency of the harmonic signal in 3D we can constrain the distance redshift relationship.
- BOSS DR12 bispectrum results in ~ 20% improvement in the BAO parameters when combined with the power spectrum.
- 20% improvement is equivalent to the 40% effective increase in survey volume.
- Check out the methodology, systematics study, and how better covariances would improve things in Pearson & Samushia 2018, MNRAS, **478**, 4500

David Pearson

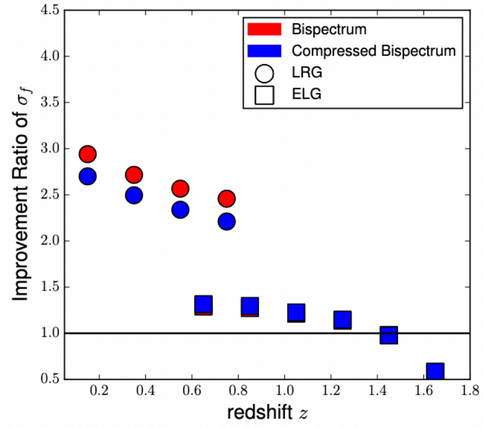


BAO parameter likelihood from the power spectrum (red) bispectrum (green), and the combined analysis (black) from Pearson & Samushia 2018.

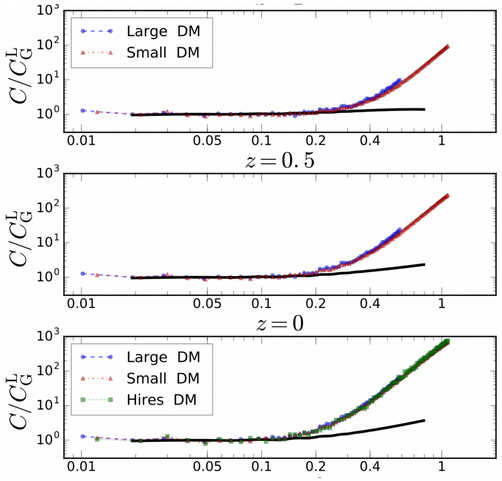
Bispectrum Information Content

Forecasts

- Forecasting the potential performance of higher-order statistics is not easy.
- Leading order computations with Gaussian approximation predict a big gain (see. e.g. Gagrani & Samushia, 2017, MNRAS, **467**, 928)
- Studies have shown that the nonlinear evolution and non-Gaussian effects become significant for the bispectrum on surprisingly large scales (Chan & Blot, 2017, PRD, **96**, 023528)
- Forecasts depend on the assumptions about the galaxy sample (higher-order biases, the number density)
- Higher number density allows you to go to higher wavenumber, where gains on the bispectrum are larger than on the power spectrum.



Forecasted improvement on the growth rate (wrt 2pt only analysis) for the high-density DESI BGS sample.
From Gagrani & Samushia, 2017

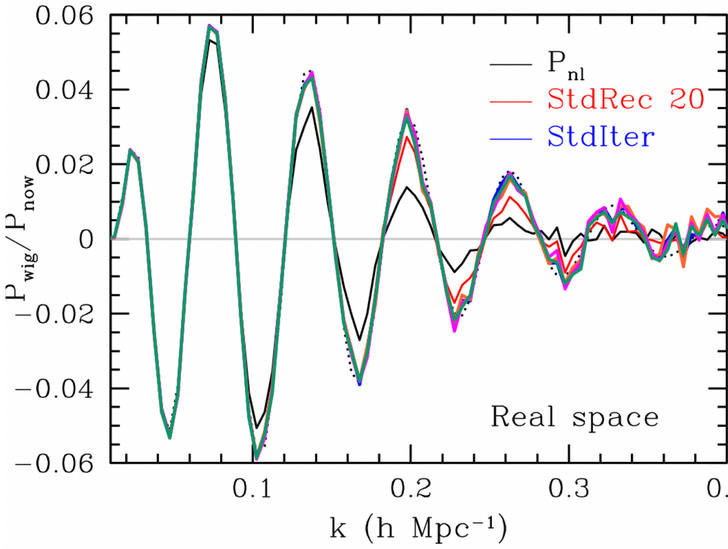


The ratio of the bispectrum variance wrt the Gaussian computation. The black line is the PT computation.
From Chan & Bolt, 2017

Gravitational Evolution

Conserving Information

- The matter field initially (shortly after the inflation) is Gaussian to a very high degree.
- Therefore it is fully described by its power spectrum.
- The field evolves to be non-linear under gravity, halo, and galaxy formation.
- The nonlinear field can only be fully described by all not functions.
- The mapping of the initial Gaussian field to the low redshift non-linear field is invertible in principle on large scales.
- Widely held conjecture: the information content of the Gaussian power spectrum is equal to the information content of the all-not functions of the nonlinearly evolved field.
- Corollary: If you reconstruct the galaxy field back to its linear version most of the higher-order information will go back to the 2pt function.
- We claim that the above statement is not true when the clustering statistics is used as an Alcock-Paczyński standard ruler.

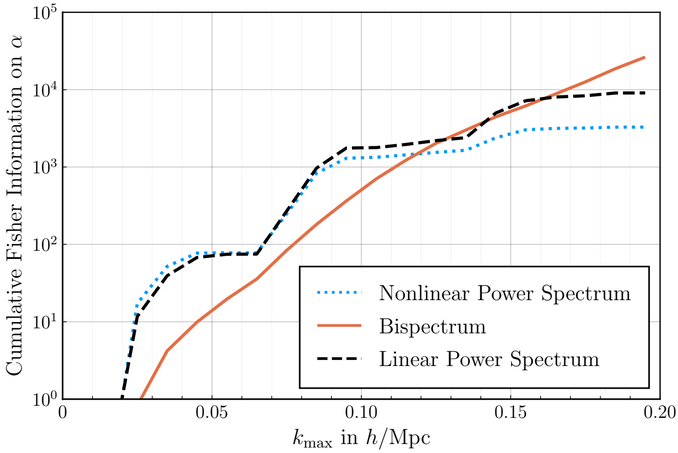


BAO feature in the non-linear 2pt function before (black), and after reconstruction (color). Reconstruction enhances the standard ruler nature of the 2pt function by removing some of the n-point signal.
From Seo et al. 2022, 2106.00530

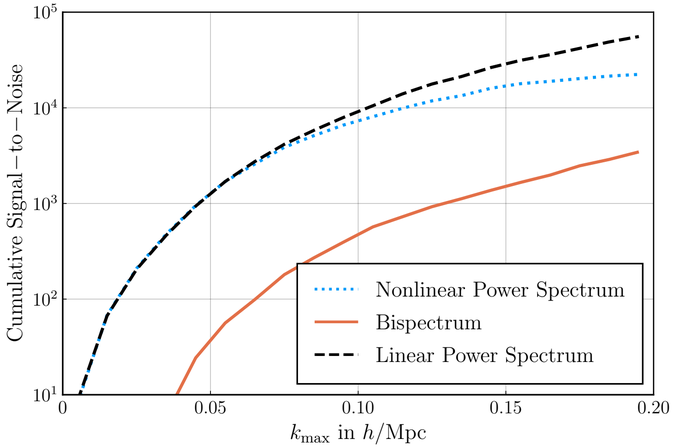
Information Content of the Bispectrum

Matter field used as a standard ruler

- The covariance matrices and parameter dependencies come from 16,000 Quijote simulations. They are free from the problems that come with theoretical modeling (non-linear effects).
- This is not a forecast! This is a deliberately artificial construction to demonstrate a point.
- The standard ruler power is determined by how sensitive the measurement is with respect to “stretching”.
- When used as a standard ruler the large-scale galaxy bispectrum can have a better constraining power than the linear power spectrum.
- Seems to be contradicting previous claims in the literature on the surface level.
- This only works when the bispectrum is used as a standard ruler.
- For the amplitude-like parameters, the information is “conserved” as the field evolves gravitationally from the linear Gaussian progenitor.



Signal-to-noise of the AP parameter (sensitivity to stretching)



Signal-to-noise of the amplitude parameter (e.g. σ_8). From Samushia, Slepian & Villaescusa-Navarro 2021, **505**, 628

Why does this feel counterintuitive?

Standard rulers are different

- The language we use to describe the BAO measurements is sometimes misleading. BAO dark energy constraints don't actually come from the dependence of the peak position on cosmology (e.g. 108 vs 110 Mpc). That the peak position changes with cosmology is actually a nuisance for the BAO distance measurements.
- If the BAO feature was already in the initial conditions (and didn't get enhanced by the gravitational evolution) how can there be more of it in a non-linear field? The information extracted by the standard ruler is not the information encoded in the initial conditions by baryonic physics (although that needed to happen for the standard ruler tests to become possible)
- Standard candles (Supernova) are a good analogy. Since they exist and happen to have standardizable luminosities we can use them to derive distance measurements. This information does not come from anywhere in particular in the initial Gaussian field. The amount of information we can extract from the standard candles is not limited by the properties of the initial Gaussian field.

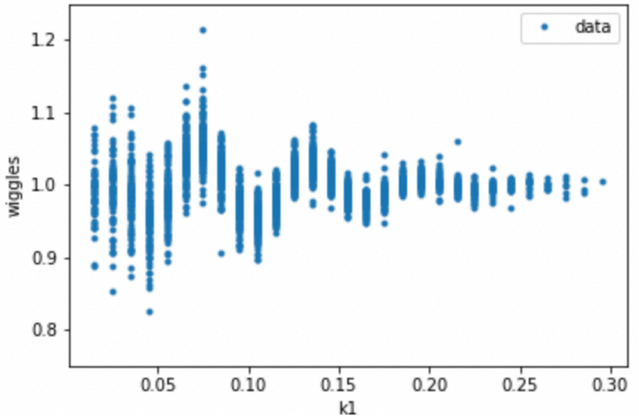
Improved BAO Constraints from Bispectrum

DESI LRG mocks

- The project is led by the KSU student Jayashree Behere and a postdoc Mehdi Rezaie.
- The goal is to perform the BAO/RSD fitting of the DESI bispectrum (jointly with the power spectrum).
- Properly marginalize broadband terms in the bispectrum so that the information comes only from the BAO feature and is not contaminated by anything else (see the figure on the right).
- Compute reliable covariance matrices that can be used for a thinly-binned bispectrum measurement up to a k ~ 0.2 Mpc/h.
- Make sure the procedure is robust with respect to DESI observational systematics
- Make sure the measurements recover cosmology without a systematic bias.

Jayashree Behere

Mehdi Rezaie



Preliminary

BAO only bispectrum of DESI LRGs
Work in progress.

Summary

Higher-Order Statistics on Linear and Non-Linear Scales

- Three-point correlation functions are relatively easy to compute in both configuration and Fourier spaces. Fast public parallel codes are available.
- We see that for some galaxy samples the small-scale 3pt functions can result in extremely tight constraints on the galaxy-halo connection. In some cases completely overwhelming the results obtainable with the 2pt analysis
- The large-scale 3pt functions can be used as standard rulers and their information content may turn out to be much higher than one would naively expect
- BAO measurements from the BOSS data enhance the standard 2pt measurements by up to 20%.
- We are working towards a DESI bispectrum analysis that would be robust with respect to theoretical and observational systematics