The Statistical Properties of Large Scale Structure

Alexander Szalay

Department of Physics and Astronomy

The Johns Hopkins University

Outline

- Lecture 1:
  - *Basic cosmological background*
  - *Growth of fluctuations*
  - *Parameters and observables*
- Lecture 2:
  - *Statistical concepts and definitions*
  - *Practical approaches*
  - *Statistical estimators*
- Lecture 3:
  - *Applications to the SDSS*
  - *Angular correlations with photometric redshifts*
  - *Real-space power spectrum*

Lecture #2

- Statistical concepts and definitions
  - *Correlation function*
  - *Power spectrum*
  - *Smoothing kernels*
  - *Window functions*
- Statistical estimators

Basic Statistical Tools

- Correlation functions
  - *N-point and N^{th}-order*
  - *Defined in real space*
  - *Easy to compute, direct physical meaning*
  - *Easy to generalize to higher order*

$\langle \rho_1 \rho_2 \rangle$ $\langle \rho_1 \rho_2 \rho_3 \rangle$ $\langle \rho_1^2 \rho_2^3 \rangle$

- Power spectrum
  - *Fourier space equivalent of correlation functions*
  - *Directly related to linear theory*
  - *Origins in the Big Bang*
  - *Connects the CMB physics to redshift surveys*
- Most common are the 2^{nd} order functions:
  - *Variance σ_8^2*
  - *Two-point correlation function $\xi(r)$*
  - *Power spectrum $P(k)$*

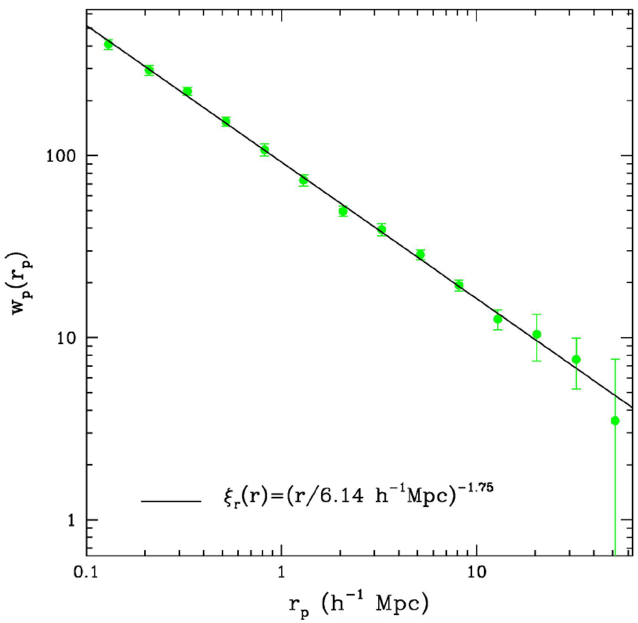
The Galaxy Correlation Function

- First measured by Totsuji and Kihara, then Peebles etal
- Mostly angular correlations in the beginning
- Later more and more redshift space
- Power law is a good approximation

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}$$

- Correlation length r_0=5.4 h^{-1} Mpc
- Exponent is around γ=1.8
- Corresponding angular correlations

$$w(r) = \left(\frac{\theta}{\theta_0}\right)^{1-\gamma}$$



The Overdensity

- We can observe galaxy counts $n(\mathbf{x})$, and compare to the expected counts $\langle n \rangle$

- Overdensity

$$\delta(\mathbf{x}) = \frac{n(\mathbf{x})}{\langle n \rangle} - 1$$

- Fourier transform

$$\delta(\mathbf{x}) = \frac{1}{(2\pi)^3} \int d^3\mathbf{k}\, e^{i\mathbf{kx}} \delta(\mathbf{k})$$

- Inverse

$$\delta(\mathbf{k}) = \int d^3\mathbf{x}\, e^{-i\mathbf{kx}} \delta(\mathbf{x})$$

- Wave number

$$k = \frac{2\pi}{\lambda}$$

The Power Spectrum

- Consider the ensemble average

$$\left\langle \delta(\mathbf{k}_1)\delta^*(\mathbf{k}_2)\right\rangle$$

- Change the origin by ***R***

$$\left\langle \tilde{\delta}(\mathbf{k}_1)\tilde{\delta}^*(\mathbf{k}_2)\right\rangle = e^{i(\mathbf{k}_1-\mathbf{k}_2)\mathbf{R}}\left\langle \delta(\mathbf{k}_1)\delta^*(\mathbf{k}_2)\right\rangle$$

- Translational invariance

$$\left\langle \delta(\mathbf{k}_1)\delta^*(\mathbf{k}_2)\right\rangle = (2\pi)^3\delta^{(3)}(\mathbf{k}_1-\mathbf{k}_2)\left|\delta(\mathbf{k}_1)\right|^2$$

- Power spectrum

$$P(\mathbf{k}) = \left|\delta(\mathbf{k})\right|^2$$

- Rotational invariance

$$P(\mathbf{k}) = P(k)$$

Correlation Function

- Defined through the ensemble average

$$\xi(\mathbf{x}_1, \mathbf{x}_2) = \left\langle \delta(\mathbf{x}_1)\delta(\mathbf{x}_2) \right\rangle$$

- Can be expressed through the Fourier transform

$$\xi(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{(2\pi)^6} \int d^3\mathbf{k}_1 d^3\mathbf{k}_2 \; e^{i(\mathbf{k}_1\mathbf{x}_1 - \mathbf{k}_2\mathbf{x}_2)} \left\langle \delta(\mathbf{k}_1)\delta^*(\mathbf{k}_2) \right\rangle$$

- Using the translational invariance in Fourier space

$$\xi(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{(2\pi)^3} \int d^3\mathbf{k}_1 d^3\mathbf{k}_2 \; e^{i(\mathbf{k}_1\mathbf{x}_1 - \mathbf{k}_2\mathbf{x}_2)} \delta^{(3)}(\mathbf{k}_1 - \mathbf{k}_2) P(k_1)$$

- The correlation function only depends on the distance

$$\xi(\mathbf{x}_1, \mathbf{x}_2) = \int d^3\mathbf{k} \; e^{i\mathbf{k}(\mathbf{x}_1 - \mathbf{x}_2)} P(k) = \xi(\mathbf{x}_1 - \mathbf{x}_2)$$

P(k) vs ξ(r)

- The Rayleigh expansion of a plane wave gives

$$e^{i\mathbf{kr}} = \sum_l i^l (2l+1)\, j_l(kr) P_l(\hat{\mathbf{k}}\hat{\mathbf{r}})$$

- Using the rotational invariance of P(k)

$$\xi(\mathbf{x}_1 - \mathbf{x}_2) = \xi(r) = \frac{1}{4\pi^2} \int dk\, k^2 j_0(kr) P(k)$$

- The power per logarithmic interval

$$\xi(r) = \frac{1}{4\pi^2} \int d \ln k \;\; j_0(kr) \left[k^3 P(k)\right]$$

- The power spectrum and correlation function form a Fourier transform pair

Filtering the Density

- Effect of a smoothing kernel K(x), where $\int d^3\mathbf{x}\, K(\mathbf{x}) = 1$

$$\delta_s(\mathbf{x}) = \delta * K = \int d^3\mathbf{x}'\, \delta(\mathbf{x}')K(\mathbf{x}-\mathbf{x}')$$

- Convolution theorem

$$\delta_s(\mathbf{k}) = \delta(\mathbf{k})K(\mathbf{k})$$

- Filtered power spectrum

$$P_s(\mathbf{k}) = |\delta(\mathbf{k})|^2\, |K(\mathbf{k})|^2 = P(\mathbf{k})\, |K(\mathbf{k})|^2$$

- Filtered correlation function

$$\xi_s(r) = \frac{1}{4\pi^2}\int d\ln k\; j_0(kr)\left[k^3 P(k)|K(k)|^2\right]$$

Variance

- At r=0 separation we get the variance:

$$\sigma_R^{\ 2} = \frac{1}{4\pi^2} \int d\ln k \left[k^3 P(k) \left| K_R(k) \right|^2 \right]$$

- Usual kernel is a 'top-hat' with an R=8h^{-1} Mpc radius

$$K_R(r) = \begin{cases} 1, & \text{if } r < R \\ 0, & \text{if } r \geq R \end{cases} \qquad K_R(k) = \left[\frac{j_1(kr)}{kr} \right]$$

- The usual normalization of the power spectrum is using this window

$$\sigma_8^{\ 2} = \frac{1}{4\pi^2} \int d\ln k \left[k^3 P(k) \left| K_8(k) \right|^2 \right]$$

Selection Window

- We always have an anisotropic selection window, both over the sky and along the redshift direction

$$W(\mathbf{r}) = \begin{cases} > 0, & \text{if } \mathit{inside} \\ 0, & \text{if } \mathit{outside} \end{cases}$$

- The observed overdensity is

$$\delta_w(\mathbf{x}) = \delta(\mathbf{x}) W(\mathbf{x})$$

- Using the convolution theorem

$$\delta_w(\mathbf{k}) = \int d^3\mathbf{k}' \, \delta(\mathbf{k}') W(\mathbf{k} - \mathbf{k}')$$

The Effect of Window Shape

- The lines in the spectrum are at least as broad as the window – the PSF of measuring the power spectrum!

$$\left\langle \delta_w(\mathbf{k})\delta_w^*(\mathbf{k}')\right\rangle = \int d^3\mathbf{k}'' d^3\mathbf{k}''' \left\langle \delta_w(\mathbf{k}'')\delta_w^*(\mathbf{k}''')\right\rangle W(\mathbf{k}-\mathbf{k}'')W^*(\mathbf{k}'-\mathbf{k}''')$$

$$\left\langle \delta_w(\mathbf{k})\delta_w^*(\mathbf{k}')\right\rangle = (2\pi)^3 \int d^3\mathbf{k}'' P(\mathbf{k}'')\, W(\mathbf{k}-\mathbf{k}'')W^*(\mathbf{k}'-\mathbf{k}'')$$

- The shape of the window in Fourier space is the conjugate to the shape in real space
- The larger the survey volume, the sharper the k-space window => survey design

1-Point Probability Distribution

- Overdensity is a superposition from Fourier space

$$\delta(\mathbf{x}) = \frac{1}{(2\pi)^3} \int d^3\mathbf{k}\; e^{i\mathbf{kx}} \delta(\mathbf{k})$$

- Each δ depends on a large number of modes
- Variance (usually filtered at some scale R)

$$\sigma^2 = \left\langle |\delta|^2 \right\rangle$$

- Central limit theorem: Gaussian distribution

$$P(\delta) = \frac{e^{-\frac{\delta^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$

N-point Probability Distribution

- Many ‘soft’ pixels, smoothed with a kernel
- Raw dataset **x**
- Parameter vector Θ
- Joint probability distribution is a multivariate Gaussian

$$f(\mathbf{x},\mathbf{\Theta}) = (2\pi)^{-n/2}\left|\mathbf{C}\right|^{-1/2}\exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})\right)$$

- M is the mean, C is the correlation matrix

$$\mathbf{m} = \left\langle \mathbf{x} \right\rangle \qquad C_{ij} = \left\langle x_i x_j \right\rangle - m_i m_j$$

- C depends on the parameters Θ

Fisher Information Matrix

- Measures the sensitivity of the probability distribution to the parameters

$$\mathbf{F}_{ij} = -\left\langle \frac{\partial^2 \ln f}{\partial \Theta_i \partial \Theta_j} \right\rangle$$

- Kramer-Rao theorem:
  one cannot measure a parameter more accurately than

$$\frac{1}{\sqrt{F_{ii}}}$$

Higher Order Correlations

- One can define higher order correlation functions

$$\zeta = \langle \delta_1 \delta_2 \delta_3 \rangle$$

$$\eta = \langle \delta_1 \delta_2 \delta_3 \delta_4 \rangle$$

- Irreducible correlations represented by connected graphs
- For Gaussian fields only 2-point, all other =0
- Peebles conjecture: only tree graphs are present
- Hierarchical expansion

$$\xi^N = Q_N \sum \xi_{ij} \xi_{jk} \ldots \xi_{ni} \text{ (N - 1 terms)}$$

- Important on small scales

Correlation Estimators

- Expectation value

$$\xi(\mathbf{x}_1, \mathbf{x}_2) = \langle \delta(\mathbf{x}_1)\delta(\mathbf{x}_2) \rangle$$

- Rewritten with the density as

$$\xi_{12} = \frac{\langle (\rho_1 - \langle \rho_1 \rangle)(\rho_2 - \langle \rho_2 \rangle) \rangle}{\langle \rho_1 \rangle \langle \rho_2 \rangle}$$

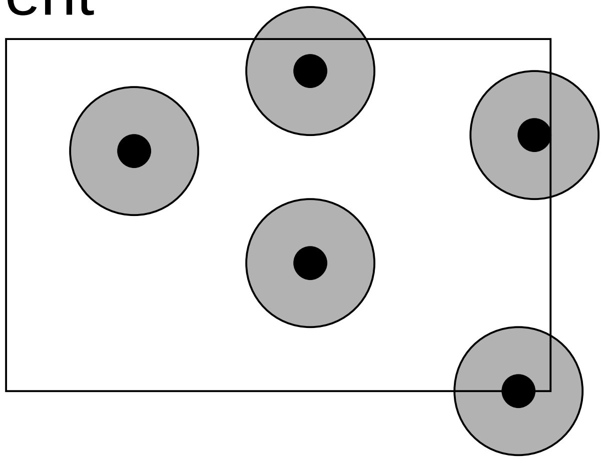
- Often also written as

$$\xi_{12} = \frac{\langle \rho_1 \rho_2 \rangle}{\langle \rho_1 \rangle \langle \rho_2 \rangle} - 1 = \frac{DD}{RR} - 1$$

- Probability of finding objects in excess of random
- The two estimators above are **NOT EQUIVALENT**

Edge Effects

- The objects close to the edge are different
- The estimator has an excess variance (Ripley)
- If one is using the first estimator, these cancel in first order (Landy and Szalay 1996)



$$\xi_{12} = \frac{DD - 2DR + RR}{RR}$$

$$\xi_{12} = \frac{\langle (\rho_1 - \langle \rho_1 \rangle)(\rho_2 - \langle \rho_2 \rangle) \rangle}{\langle \rho_1 \rangle \langle \rho_2 \rangle} = \frac{\langle (D_1 - R_1)(D_2 - R_2) \rangle}{R_1 R_2}$$

Discrete Counts

- We can measure discrete galaxy counts

$$n(\mathbf{r}) = \sum_{\alpha} \delta^{D}(\mathbf{r} - \mathbf{r}_{\alpha})$$

- The expected density is a known <n>, fractional
- The overdensity is

$$\delta(\mathbf{r}) = \frac{n - \langle n \rangle}{\langle n \rangle}$$

- If we define cells with counts N_i

$$\delta_i(\mathbf{r}) = \frac{N_i - \langle N_i \rangle}{\langle N_i \rangle}$$

Power Spectrum

- Naïve estimator for a discrete density field is

$$\hat{f}(\mathbf{k}) = \frac{1}{N}\sum_n e^{i\mathbf{k}\mathbf{r}_n}$$

$$\hat{P}(k) = \left|\hat{f}(\mathbf{k})\right| = \frac{1}{N^2}\sum_{n,n'} e^{i\mathbf{k}(\mathbf{r}_n - \mathbf{r}_{n'})} = \frac{1}{N^2}\sum_{n\neq n'} e^{i\mathbf{k}(\mathbf{r}_n - \mathbf{r}_{n'})} + \frac{1}{N}$$

- FKP (Feldman, Kaiser and Peacock) estimator
  The Fourier space equivalent to LS

$$\hat{f}(\mathbf{k}) = \sum_n \phi(\mathbf{r}_n) e^{i\mathbf{k}\mathbf{r}_n} - w(\mathbf{k})$$

$$\phi(\mathbf{r}) = \frac{\overline{n}(r)}{1+\overline{n}(r)P(k)}$$

Wish List

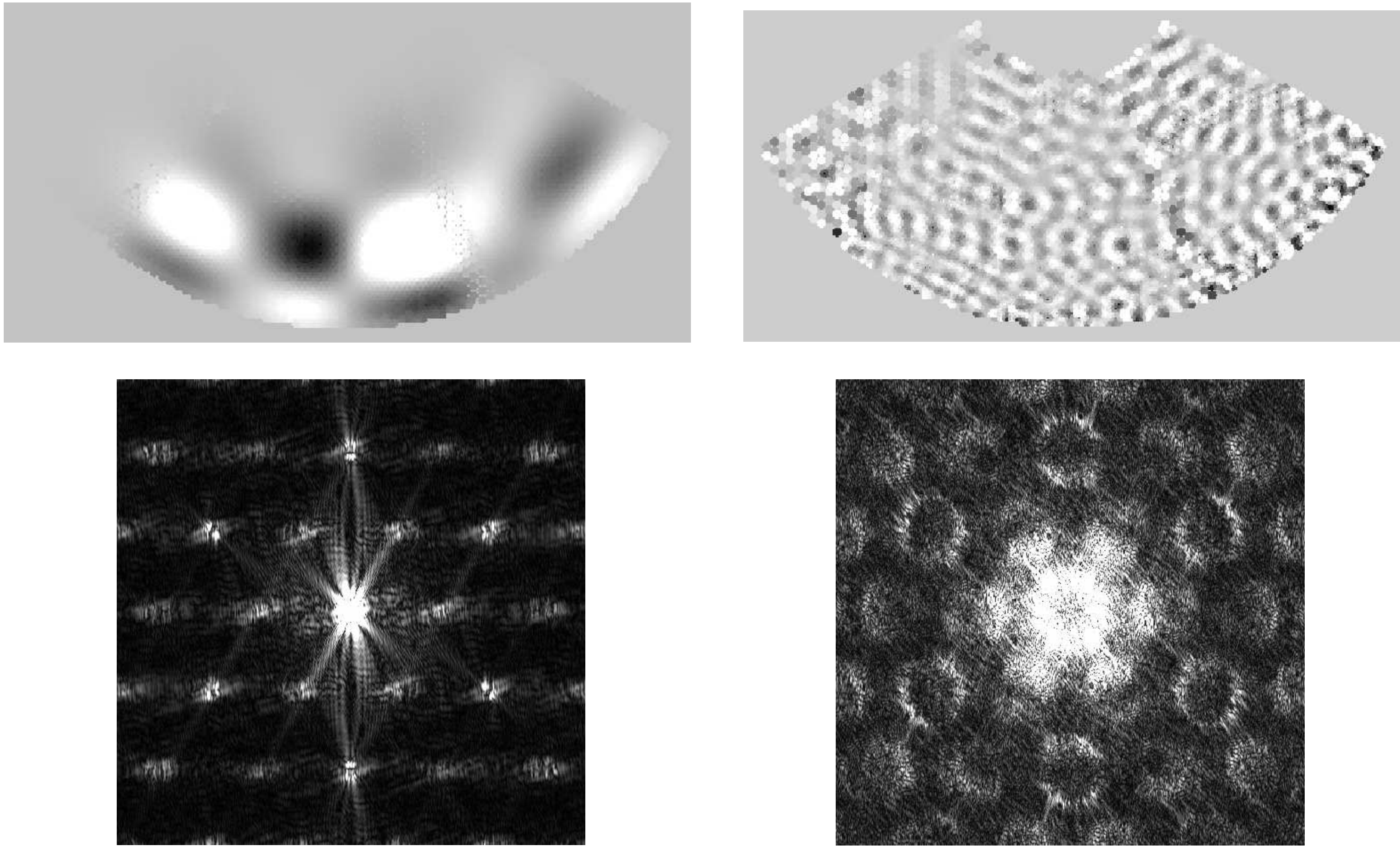
- Almost lossless for the parameters of interest
- Easy to compute, and test hypotheses (uncorrelated errors)
- Be computationally feasible
- Be able to include systematic effects

The Karhunen-Loeve Transform

- Subdivide survey volume into thousands of cells
- Compute correlation matrix of galaxy counts among cells from fiducial P(k)+ noise model
- Diagonalize matrix
  - *Eigenvalues*
  - *Eigenmodes*
- Expand data over KL basis
- Iterate over parameter values:
  - *Compute new correlation matrix*
  - *Invert, then compute log likelihood*

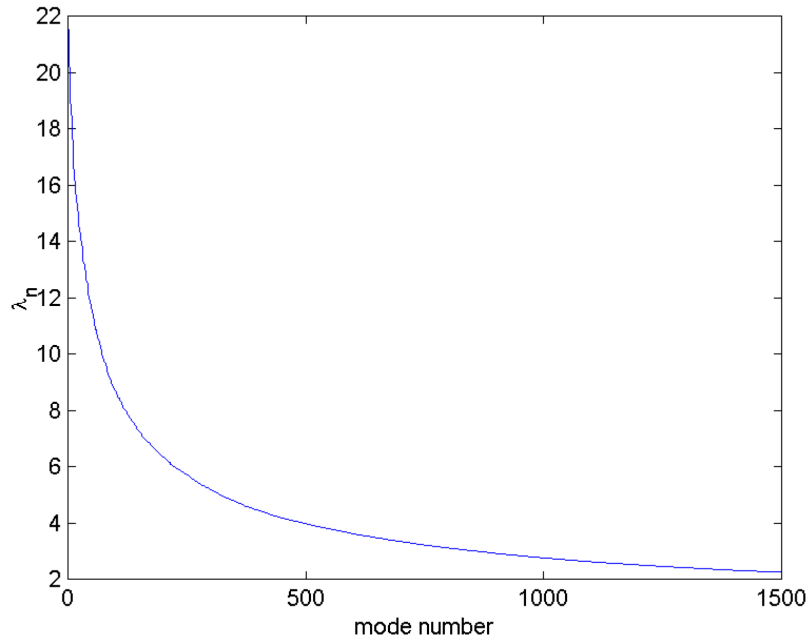
Vogeley and Szalay (1996)

Eigenmodes



Eigenmodes

- Optimal weighting of cells to extract signal represented by the mode
- Eigenvalues measure S/N
- Eigenmodes orthogonal
- In k-space their shape is close to window function
- Orthogonality = repulsion
- Dense packing of k-space
  => filling a ‘Fermi sphere’



Truncated expansion

- Use less than all the modes: truncation

$$\hat{f} = \sum_{i=1}^{M} b_i \Psi_i, \quad M < N$$

- Best representation in the *rms* sense

$$(f - \hat{f})^2 = \sum_{i=M+1}^{N} \lambda_i$$

- Optimal subspace filtering, throw away modes which contain only noise

Correlation Matrix

- The mean correlation between cells

$$\xi_{ij} = \iint d^3 r_1 d^3 r_2 \xi^{(s)}(r_1, r_2) W_i(r_1) W_j(r_2)$$

- Uses a fiducial power spectrum
- Iterate during the analysis

Whitening Transform

- Remove expected count n_i
- 'Whitened' counts

$$d_i = \left(\frac{g_i - n_i}{n_i} \right) \qquad R_{ij} = \xi_{ij} + \frac{\delta_{ij}}{n_i} + \frac{\varepsilon_{ij}}{n_i n_j}$$

- Can be extended to other types of noise
  => systematic effects
- Diagonalization: overdensity eigenmodes
- Truncation optimizes the overdensity

Truncation

- Truncate at 30 Mpc/h
  - *avoid most non-linear effects*
  - *keep decent number of modes*