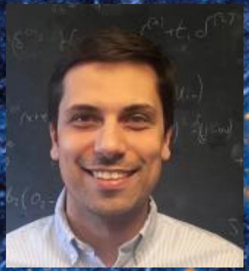
Fewer Mocks and Less Noise:

Reducing the Dimensionality of Cosmological Observables with Subspace Projections



Oliver Philcox (Princeton)

DESI Galaxy & Quasar Clustering Telecon

Based on: Philcox, Ivanov, Zaldarriaga, Simonovic, Schmittfull (2020, arXiv: 2009.03311)

The Curse of Dimensionality

- Cosmological observables are **high-dimensional**, e.g.;
  - BOSS had ~ 100 power spectrum bins
  - Tomographic analyses will have many more

but these are only used to measure a **few** parameters

- Inference proceeds via a **Gaussian** likelihood:

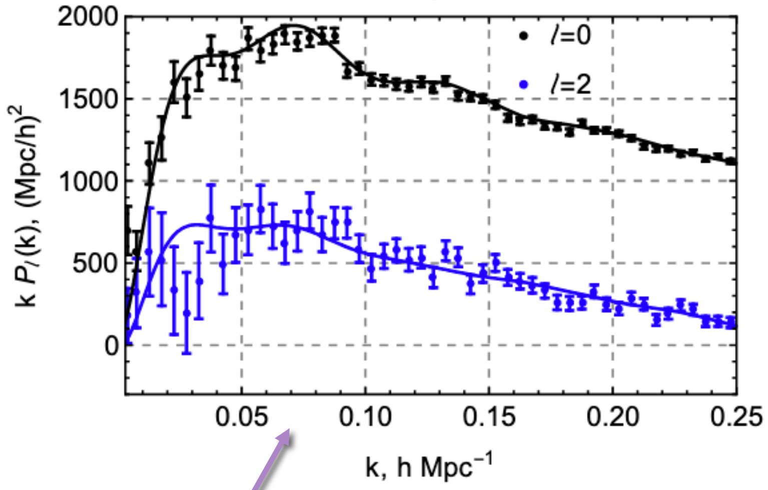
$$L(\theta) \propto \exp\left[-\frac{1}{2}(d - m(\theta)C^{-1}(d - m(\theta))\right]$$

which requires the **inverse** covariance matrix

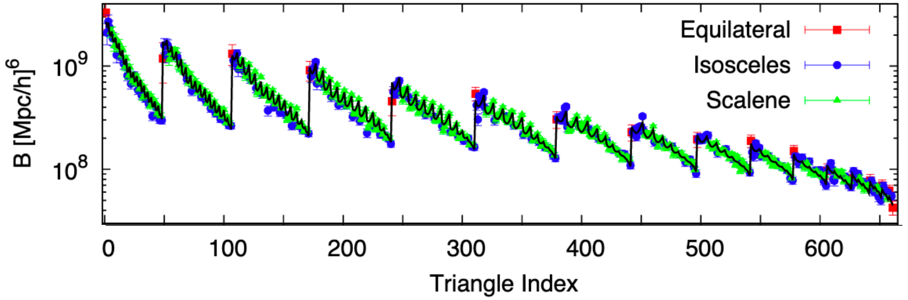
- Normally use a **sample** covariance:

$$\mathrm{cov}\,(P_a, P_b) \equiv \mathsf{C}_{D,ab} = \frac{1}{N_{\mathrm{mock}} - 1} \sum_{i=1}^{N_{\mathrm{mock}}} \left(\hat{P}_a^{(i)} - \overline{P}_a\right)\left(\hat{P}_b^{(i)} - \overline{P}_b\right)$$

with samples from **N-body** simulations



Power Spectrum Bispectrum



Gil-Marin+16, Ivanov+19

The Curse of Dimensionality

- Need $N_{mock} > N_{bin}$ to invert the sample covariance

- For finite N_{mock} this is a **biased** inverse: [Anderson'03/Hartlap+07]

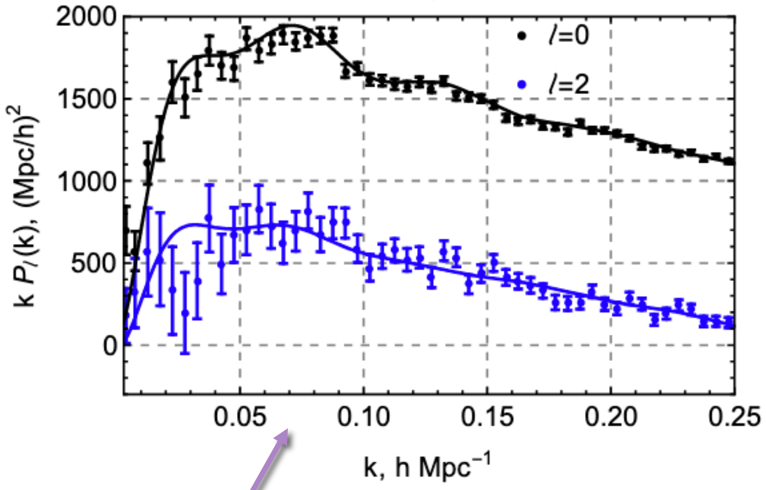
$$\Psi_D = f_{\rm H} \times {\sf C}_D^{-1}, \quad f_{\rm H} = \frac{N_{\rm mock} - N_{\rm bin} - 2}{N_{\rm mock} - 1}$$

$$\chi^2(\theta) \rightarrow f_H \, \chi^2(\theta)$$

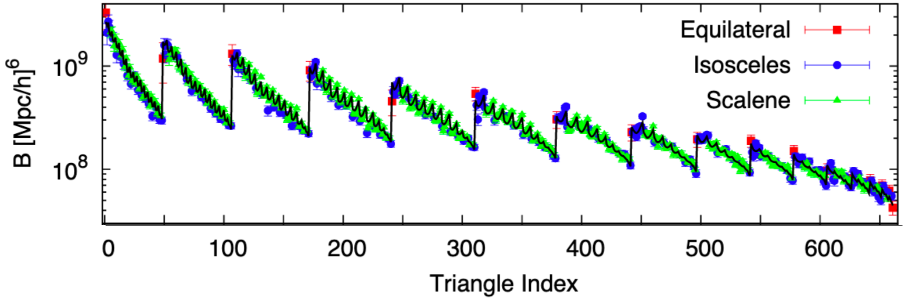
(cf. Sellentin & Heavens '15)

- **Noise** in the covariance matrix gives stochastic **shifts** in the best-fit parameters: [Percival+13]
  - Must **inflate** the output covariances by $\sim 1 + \frac{N_{bin} - N_{param}}{N_{mock}}$
  - This **loses** constraining power

- Can be **reduced** if we **compress** the data



Power Spectrum Bispectrum

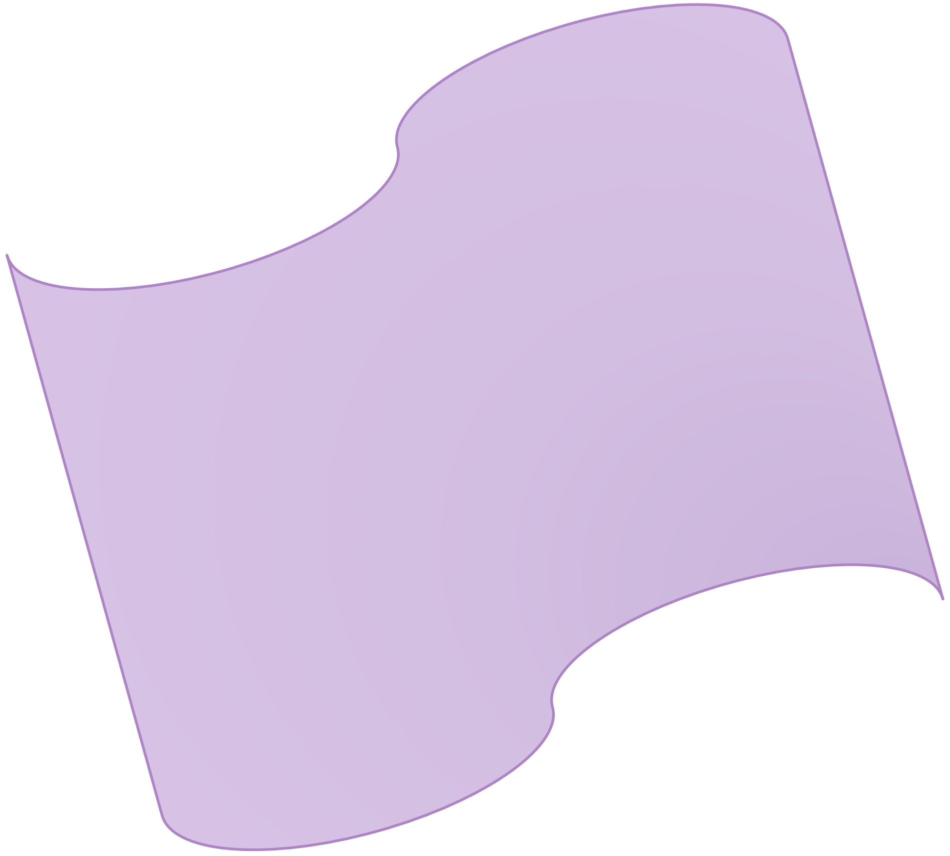


Gil-Marin+16, Ivanov+19

Creating a Metric in Parameter Space

- Consider an analysis measuring parameters $\vec{\theta}$

- The space of all physical models for the analysis are described by a **manifold** (with boundary)



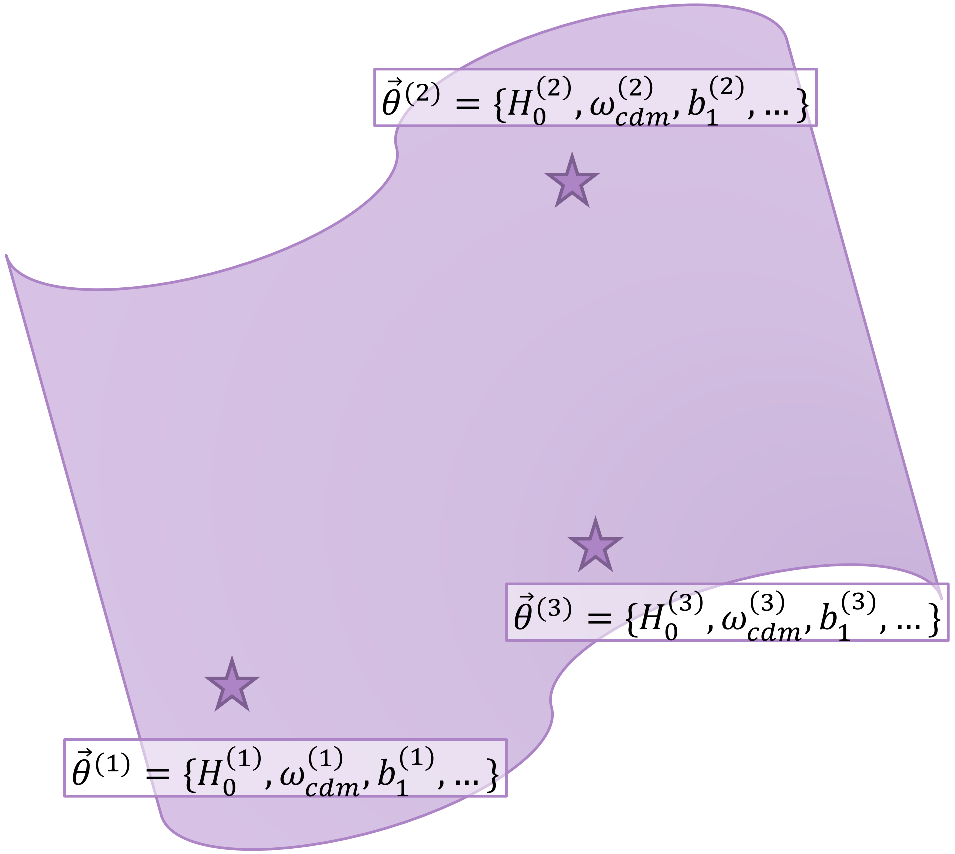
*somewhat inspired by gravitational wave analyses [e.g. Roulet+19]

Philcox+20

Creating a Metric in Parameter Space

- Consider an analysis measuring parameters $\vec{\theta}$
- The space of all physical models for the analysis are described by a **manifold** (with boundary)
- Co-ordinates on the manifold → cosmological + nuisance **parameters:**

$$\vec{\theta}^{(i)} = \{H_0^{(i)}, \omega_{cdm}^{(i)}, b_1^{(i)}, \dots\}$$



Philcox+20

Creating a Metric in Parameter Space

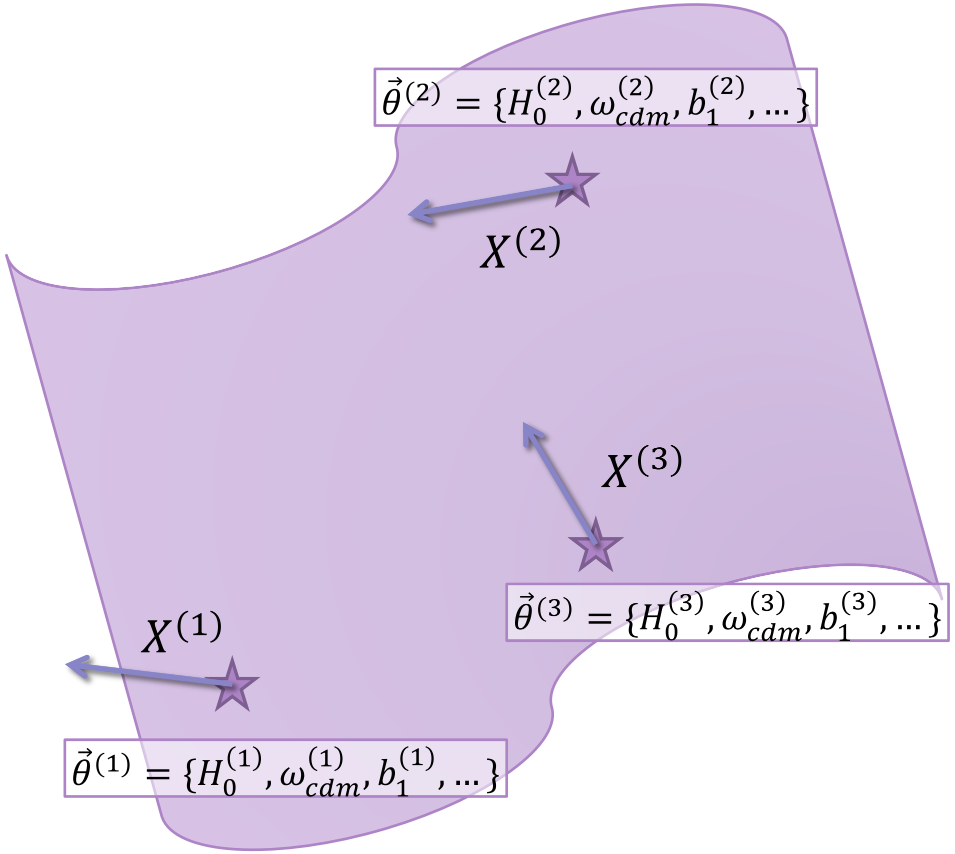
- Consider an analysis measuring parameters $\vec{\theta}$
- The space of all physical models for the analysis are described by a **manifold** (with boundary)
- Co-ordinates on the manifold → cosmological + nuisance **parameters:**

$$\vec{\theta}^{(i)} = \{H_0^{(i)}, \omega_{cdm}^{(i)}, b_1^{(i)}, \ldots\}$$

- The **tangent vector** to each point is the **theory model**

$$X_a(\theta) \equiv \sum_{ab} \mathsf{C}_b^{-1/2} \left[P_b(\theta) - \overline{P}_b\right]$$

noise-weighting for later use. C is a **fiducial** covariance.

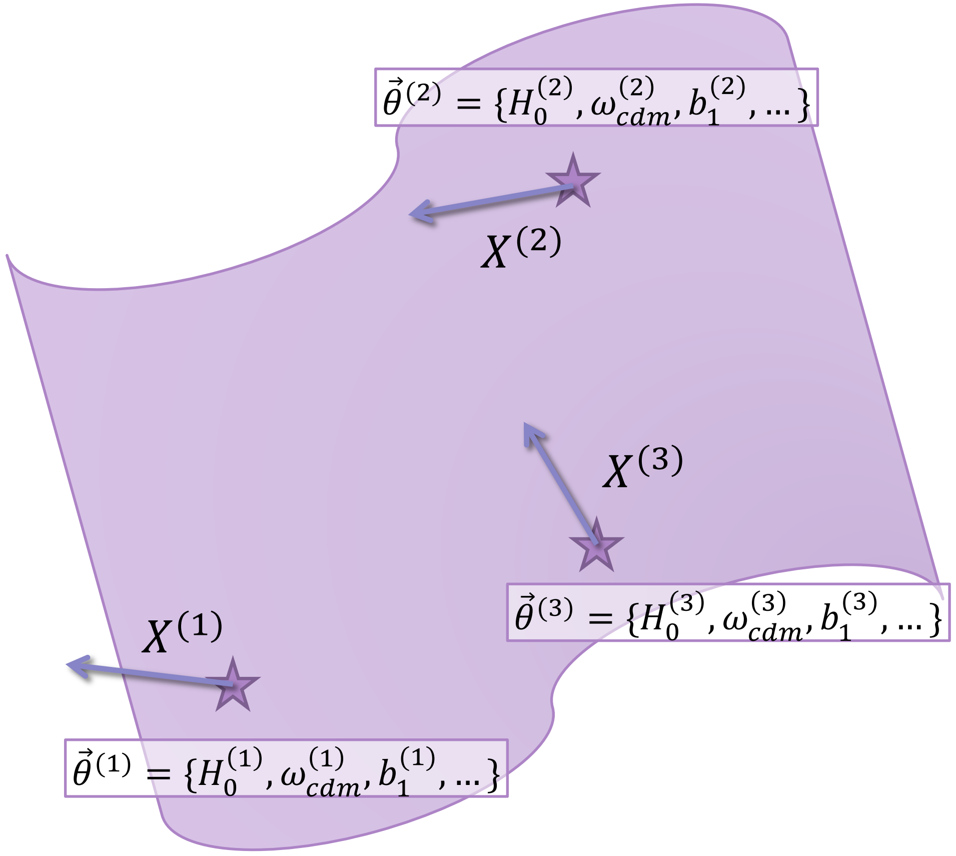


Philcox+20

Creating a Metric in Parameter Space

- We define an **inner product** on the manifold using the **tangent vectors**:

$$\langle X^{(i)} | X^{(j)} \rangle = \sum_a X_a^{(i)} X_a^{(j)}$$



Philcox+20

Creating a Metric in Parameter Space

- We define an **inner product** on the manifold using the **tangent vectors**:

$$\langle X^{(i)} | X^{(j)} \rangle = \sum_a X_a^{(i)} X_a^{(j)}$$

- This gives a notion of **distance** between two points*

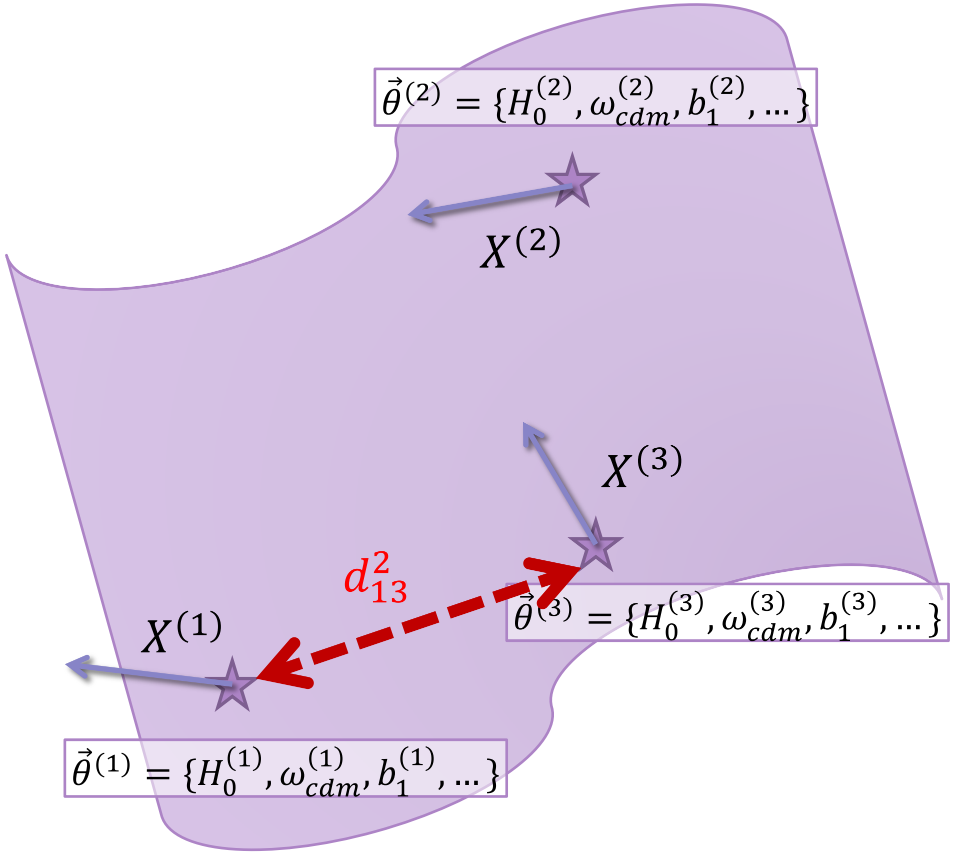
$$d_{ij}^2 = \langle X^{(i)} - X^{(j)} |\, X^{(i)} - X^{(j)} \rangle$$

This is just a **Euclidean** metric.

- In terms of **P**:

$$d_{ij}^2 = \sum_{ab} \left(P_a(\theta^{(i)}) - P_a(\theta^{(j)})\right) C_{ab}^{-1} \left(P_b(\theta^{(i)}) - P_b(\theta^{(j)})\right)$$

which is just the Gaussian χ^2.



*Assuming Riemannian geometry, *i.e.* a Gaussian likelihood

Philcox+20

The Subspace Projection

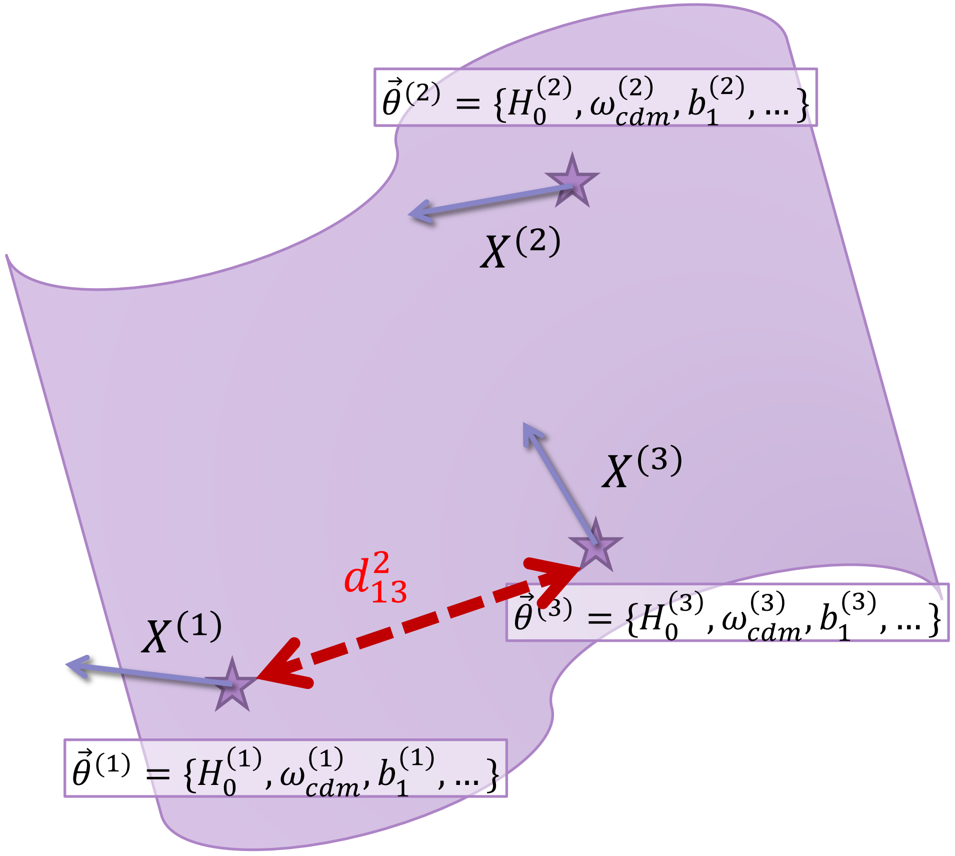
- The tangent vectors X are **high-dimensional** (size N_{bin})

- Can we identify a low-dimensional **subspace** that preserves the distance information?

- To do this:
  1. Draw **samples** from the manifold (*i.e.* $\{\theta^{(i)}\}$)
  2. Compute the **tangent-vectors** $X(\theta)$ at each point
  3. Perform a **S**ingular **V**alue **D**ecomposition

$$X_{ia} = \sum_{\alpha} U_{i\alpha} D_{\alpha} V_{\alpha a}$$

Set of samples *Basis vectors*



Philcox+20

The Subspace Projection

$$X_{ia} = \sum_{\alpha} U_{i\alpha} D_{\alpha} V_{\alpha a}$$

Set of samples *Basis vectors*

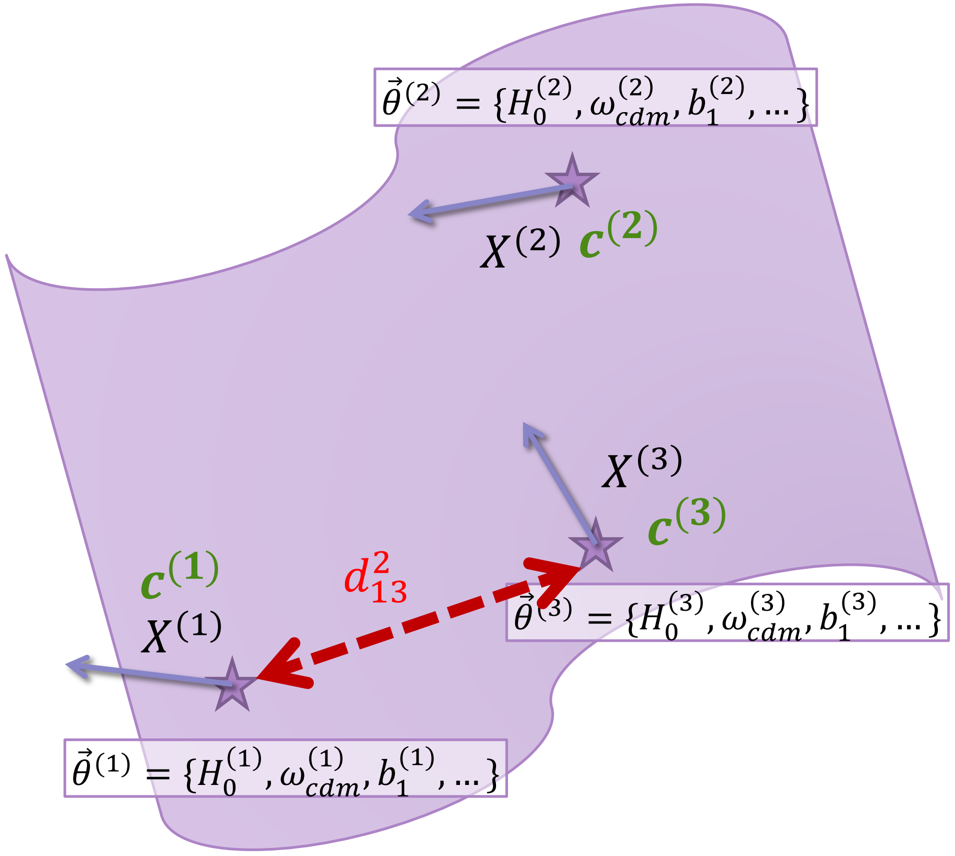
- This defines a set of **basis** vectors:

$$X_a^{(i)} \approx \sum_{\alpha=1}^{N_{\rm SV}} c_\alpha^{(i)} V_{\alpha a}$$

Subspace Coefficients

- All information is in the $c^{(i)}$ **subspace** coefficients

- If $N_{SV} = N_{bin}$ this is just a rotation
- If $N_{SV} < N_{bin}$ we have **compressed** the statistic



Philcox+20

Properties of the Decomposition

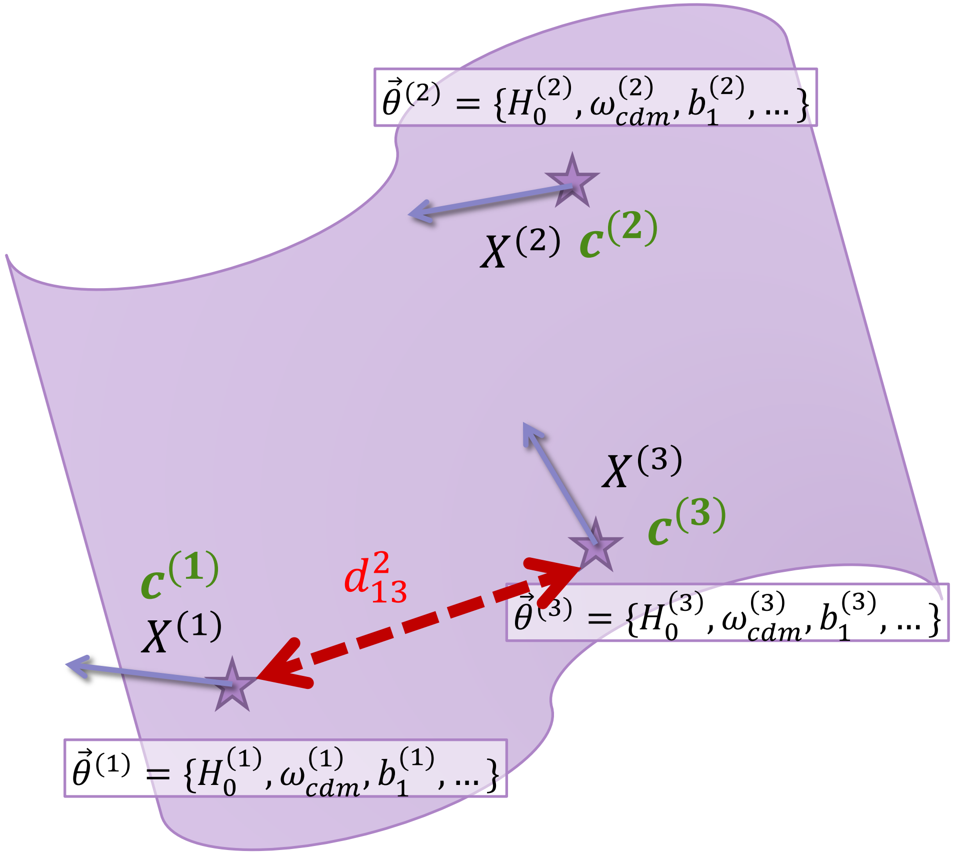
- The linear decomposition is **optimal** with respect to $d^2 \equiv \chi^2$

- We can **set** the size of the space robustly:

  - Choose N_{SV} by requiring that the **error** in χ^2 is **below** some threshold, **averaged** over the prior

  - If we need **higher** precision, just use more basis vectors!

- All the analysis is in terms of N_{SV} subspace coefficients



Philcox+20

Analysis in the Projected Subspace

- How do we apply this to data?

- Likelihood of statistic *P*:

Model ↓ Data ↓

$$-2\log\mathcal{L}(\theta) = \hat{\chi}^2(\theta) = \sum_{ab}\left(\hat{P}_a - P_a(\theta)\right)\mathsf{C}^{-1}_{D,ab}\left(\hat{P}_b - P_b(\theta)\right)$$

True Covariance

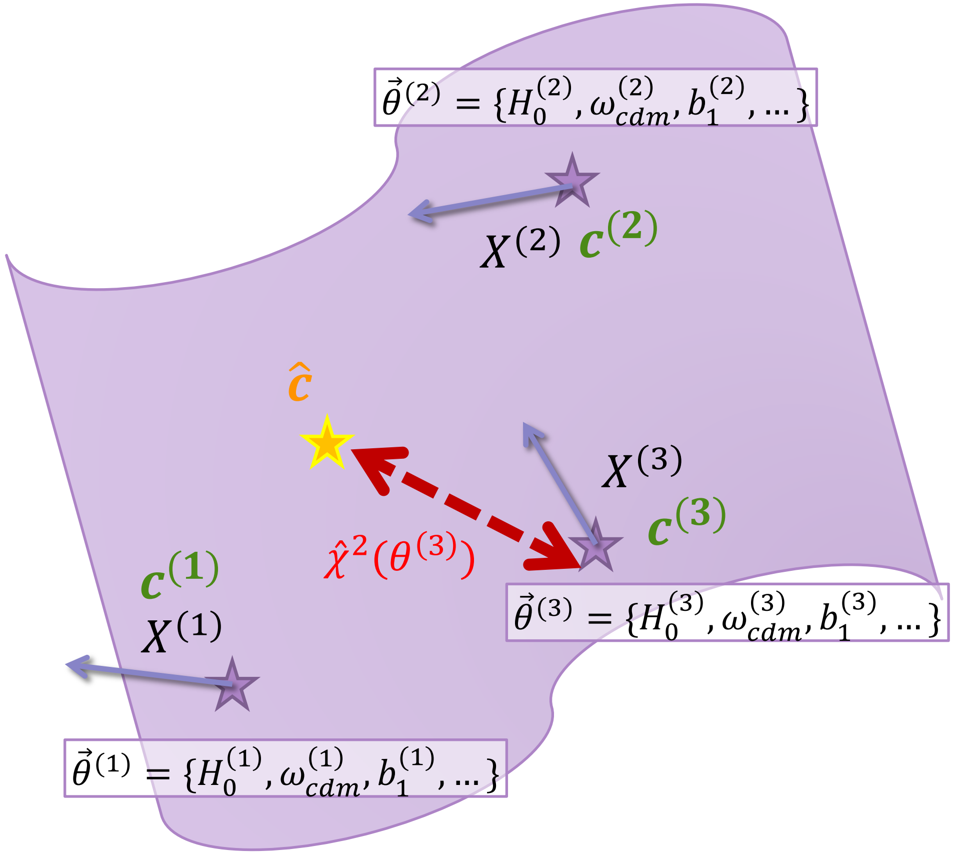
- Likelihood of *subspace* coefficients:

Model ↓ Data ↓

$$-2\log\mathcal{L}(\theta) = \sum_{\alpha=1}^{N_{\rm SV}}\sum_{\beta=1}^{N_{\rm SV}}\left(\hat{c}_\alpha - c_\alpha(\theta)\right)\mathcal{C}^{-1}_{D,\alpha\beta}\left(\hat{c}_\beta - c_\beta(\theta)\right)$$

True Covariance (almost diagonal)

where $\hat{c}$ are **observed** coefficients: $\hat{c}_\alpha = \sum_{ab} V_{a\alpha}\, \mathsf{C}_{ab}^{-\frac{1}{2}}\, \hat{P}_b$



Philcox+20

Overview of the Procedure

Generating the Basis Vectors

1. Draw a set ($\sim 10^4$) of cosmological + nuisance parameters from the **priors**
2. Compute the noise-weighted statistic at each point forming a **template bank**
3. Perform an **SVD** on these samples to identify **basis vectors**
4. **Restrict** to the first N_{SV} vectors, setting N_{SV} by constraining **error** in χ^2

Performing the Analysis

1. **Project** the data onto the N_{SV} **subspace**-coefficients
2. Run MCMC with the Gaussian **subspace** likelihood:

$$-2\log\mathcal{L}(\theta) = \sum_{\alpha=1}^{N_{\rm SV}}\sum_{\beta=1}^{N_{\rm SV}} \left(\hat{c}_\alpha - c_\alpha(\theta)\right)\mathcal{C}^{-1}_{D,\alpha\beta}\left(\hat{c}_\beta - c_\beta(\theta)\right)$$

Requirements

- Gaussian Likelihood
- Theory Model
- Priors on parameters
- Approximate (smooth) fiducial covariance*

*only used to define basis vectors

Philcox+20

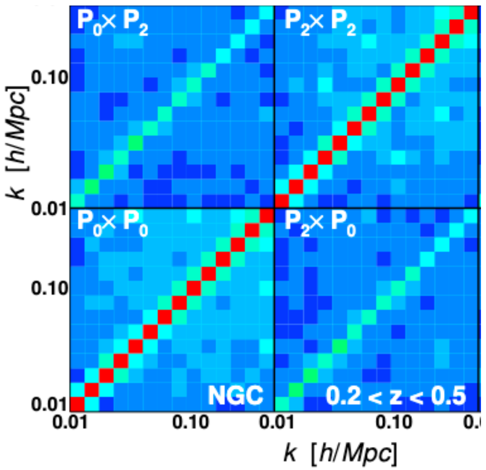
Comparison to Other Approaches

Covariance Matrix PCA [e.g. Scoccimarro 2000]

1. Form the observable **covariance matrix**
2. Perform a **P**rincipal **C**omponent **A**nalysis of this
3. **Restrict** to the first N basis vectors
4. **Project** the data onto these

- PCA finds directions that contribute most to **signal-to-noise**
  - Are these directions **useful**?
- Our SVD finds the directions that contribute most to the **log-likelihood**
  - Optimal for a **specific** analysis

Power Spectrum Covariance



PCA →

$$C = W\Lambda W^T$$

$$P(k) \approx \sum_i a_i W_i(k)$$

Coefficients

Basis Vectors

Beutler+16

Comparison to Other Approaches

<u>Covariance Matrix PCA</u> [e.g. Scoccimarro 2000]

1. Form the observable **covariance matrix**
2. Perform a **P**rincipal **C**omponent **A**nalysis of this
3. **Restrict** to the first N basis vectors
4. **Project** the data onto these

- PCA finds directions that contribute most to **signal-to-noise**
  - Are these directions **useful**?
- Our SVD finds the directions that contribute most to the **log-likelihood**
  - Optimal for a **specific** analysis

<u>MOPED</u> [e.g. Heavens 2000]

- Compresses to N_{param} numbers based on the **Fisher matrix**
- Technically only exact for **Gaussian** posterior [but often a good approximation]
- Decomposition centered around a **point** in space
  - May have to **iterate** the procedure
- Number of basis vectors is **fixed**
- Our SVD does **not** assume a Gaussian posterior
  - Invariant to reparametrizations of manifold
  - Non-Gaussianity and **multi-modality** allowed
  - Arbitrarily **accurate** given large enough N_{SV}

[see also Alsing+18 , KL: Tegmark+97]

Application: BOSS Power Spectra

Test Case: Full-Shape analysis of BOSS power spectra [Ivanov+19]

- 10-parameter analysis:

$$\theta = \{\omega_{\rm cdm}, A_s/A_{s,\rm fid}, h, ...\} \times \{b_1, b_2, b_{G_2}, b_4, c_{s,0}, c_{s,2}, P_{\rm shot}\}$$

- 96-bin power spectrum (high-z NGC sample, monopole + quadrupole)
- Covariance estimated from **MultiDark-Patchy** mocks [Kitaura+15]

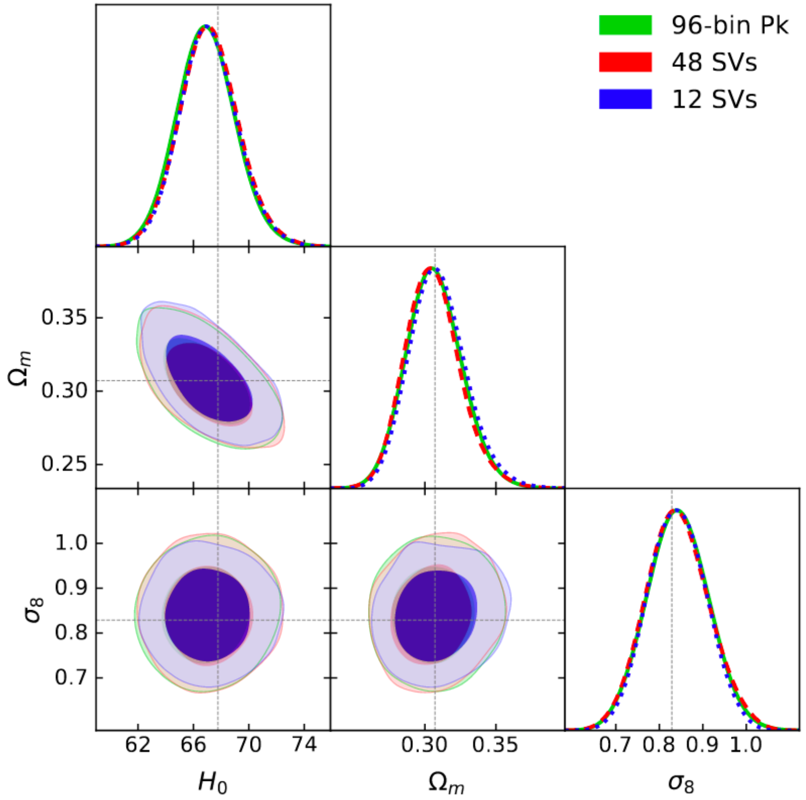
To generate basis vectors:

○ Compute theory model (1-loop **E**ffective **F**ield **T**heory) at 10^4 random draws in parameter space

○ Fiducial covariance is a **Gaussian** model [Wadekar+19]

○ Set $N_{SV} = 12$, by setting $\Delta\chi^2 < 0.1$ **averaged** across prior

BOSS mean-of-mocks analysis



No bias from the subspace decomposition!

Philcox+20

Application: BOSS Power Spectra

More realistic case: data-set is a **single** Patchy mock

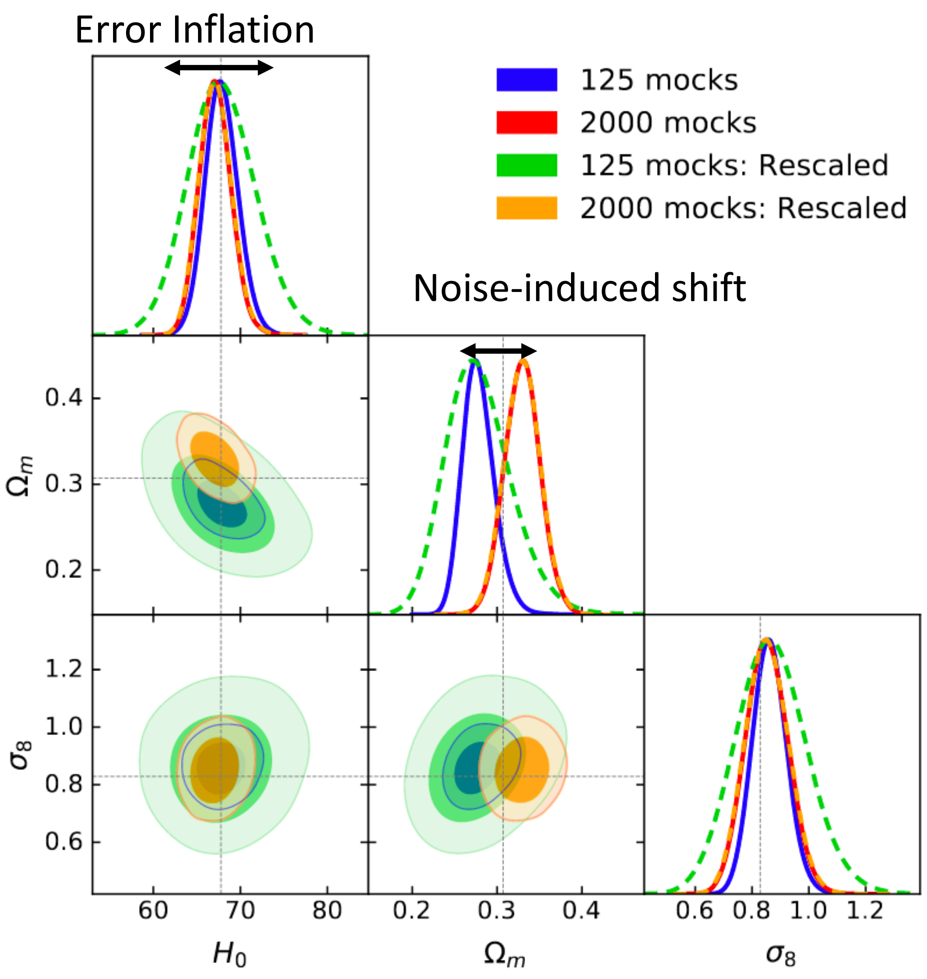
- Sample covariance from:
  a) 125 mocks
  b) 2000 mocks

- Should **inflate** posterior contours to account for **stochastic shifts** from noise in the covariance matrix* [Percival+13]

$$(\Delta\theta)^2 \approx (N_{bin} - N_{param})/N_{mock}$$

- Significant **shifts** from using 125 mocks with 96-bin P(k)

- Inflation factor is **large**



(a) 96-bin Power Spectrum

Philcox+20

*Assuming Gaussian likelihoods, cf. Sellentin & Heavens '15

Application: BOSS Power Spectra

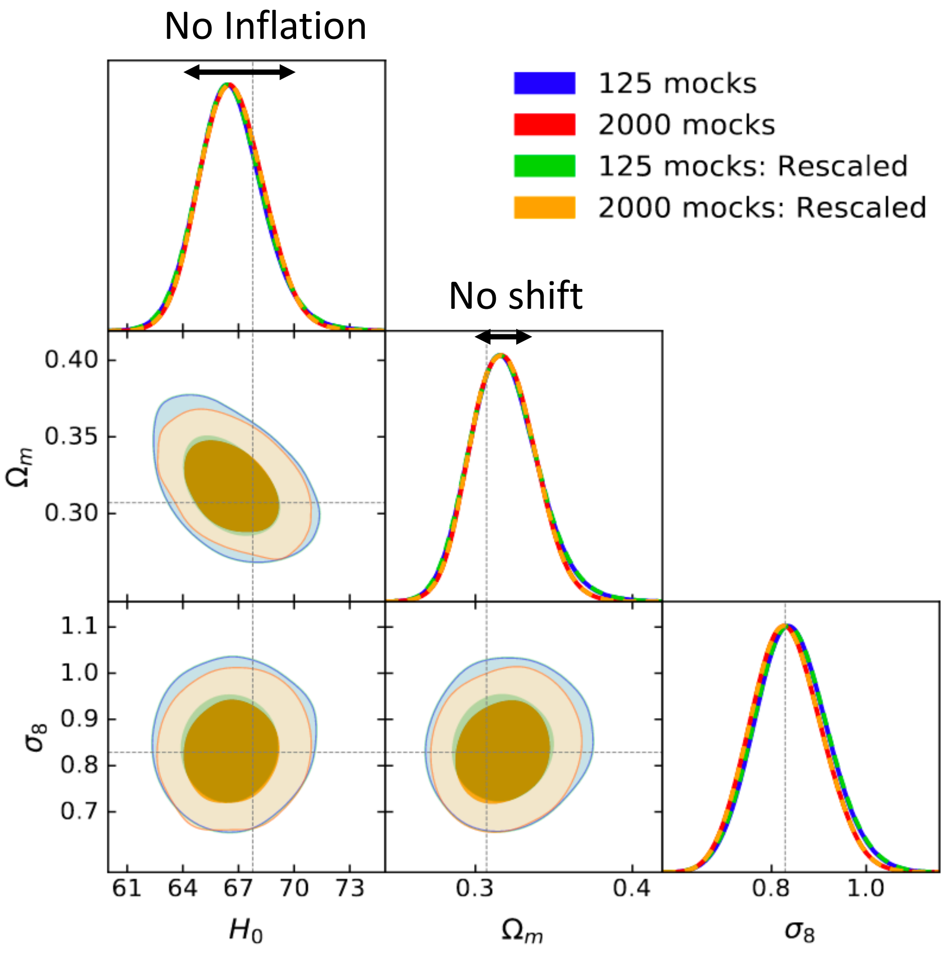
More realistic case: data-set is a **single** Patchy mock

- Sample covariance from:
  a) 2000 mocks
  b) 125 mocks

- Should **inflate** posterior contours to account for **stochastic shifts** from noise in the covariance matrix* [Percival+13]

$$(\Delta\theta)^2 \approx (N_{bin} - N_{param})/N_{mock}$$

- **No** significant **shifts** from using 125 mocks with 12 SVs
- Inflation factor is **small**



(c) 12 Subspace Coefficients

Philcox+20

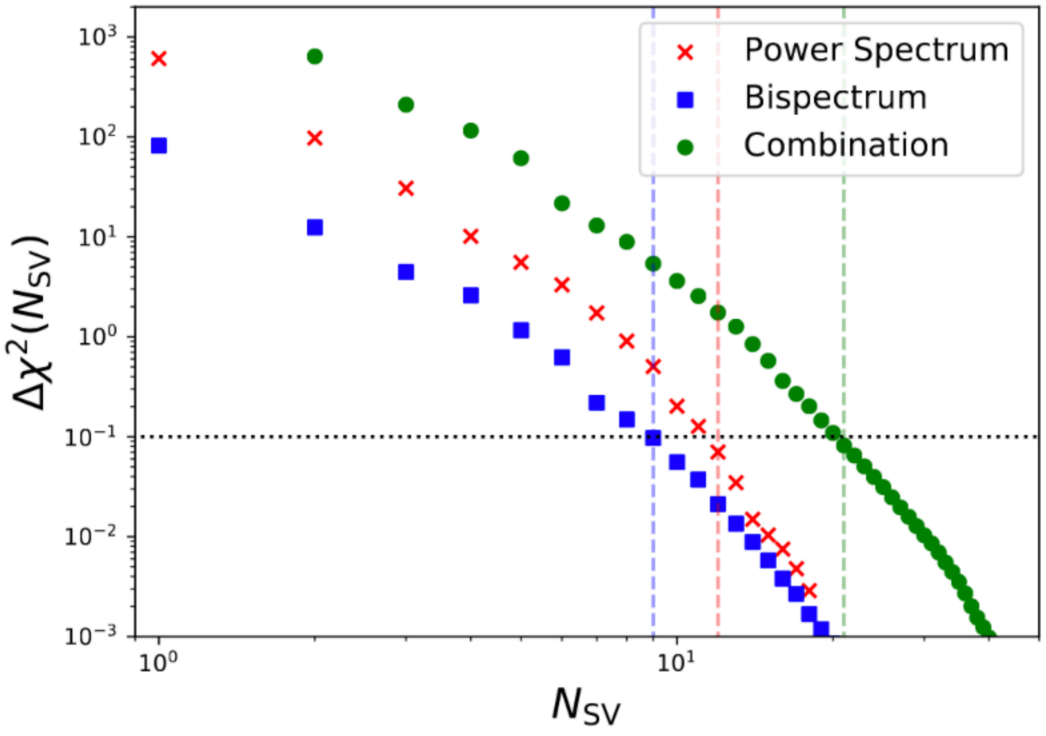
*Assuming Gaussian likelihoods, cf. Sellentin & Heavens '15

Beyond Power Spectra

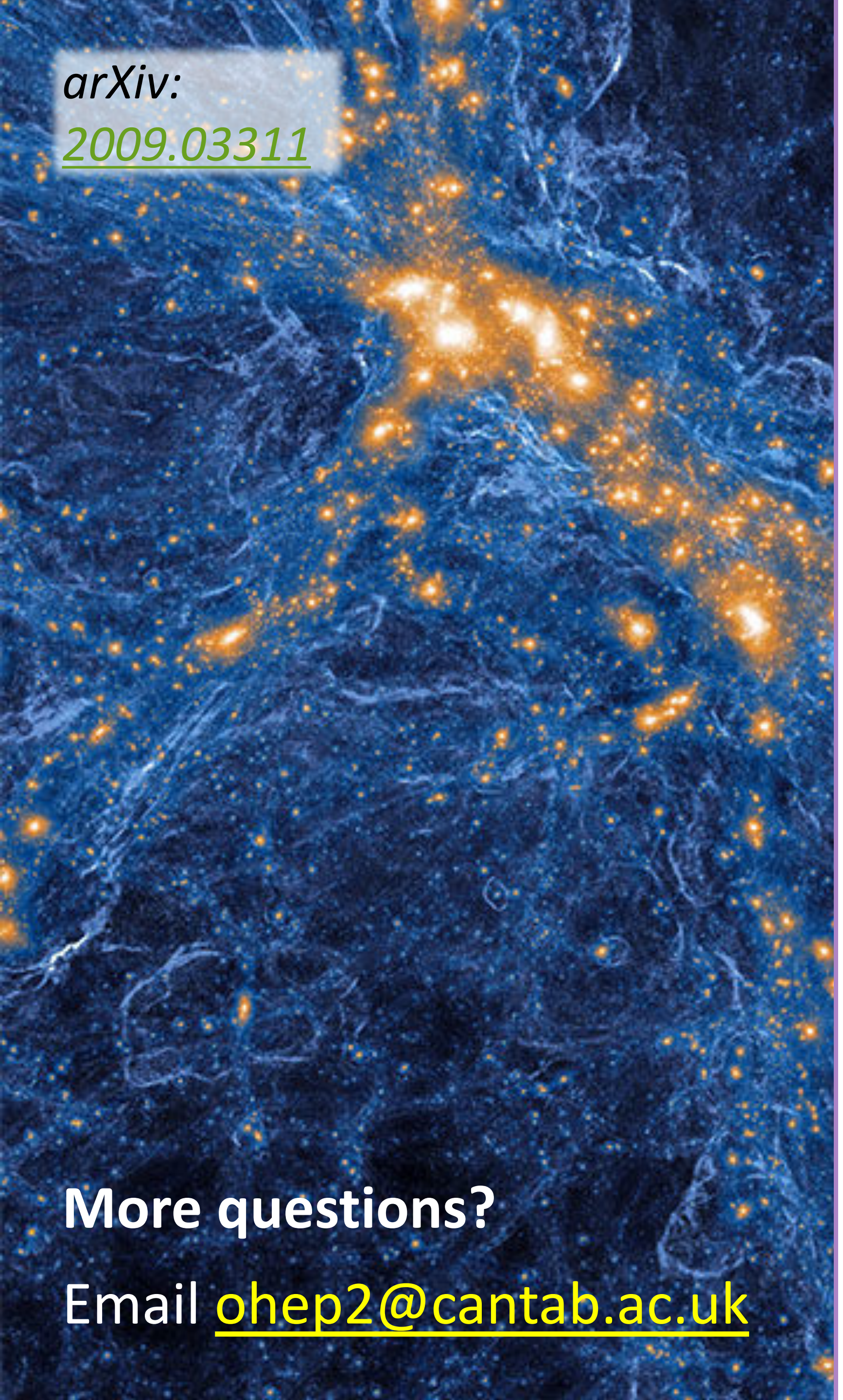
- This applies to **any** Gaussian-likelihood analysis, given a **theory** model, parameter **priors** and a **fiducial** covariance.

- **More** precise data will require **more** coefficients (fixing $\Delta\chi^2 < 0.1$)
  - Adding **reconstructed** BAO information: [cf. Philcox+20a]
    - $N_{SV} = 14$
  - Increasing volume by 10x [DESI-like]:
    - $N_{SV} = 16$
  - 2135-bin BOSS **bispectrum**
    - $N_{SV} = 9$
  - Power spectrum + **bispectrum**
    - $N_{SV} = 21$

Average χ^2 error across prior



Philcox+20



arXiv:
2009.03311

More questions?

Email ohep2@cantab.ac.uk

Conclusions

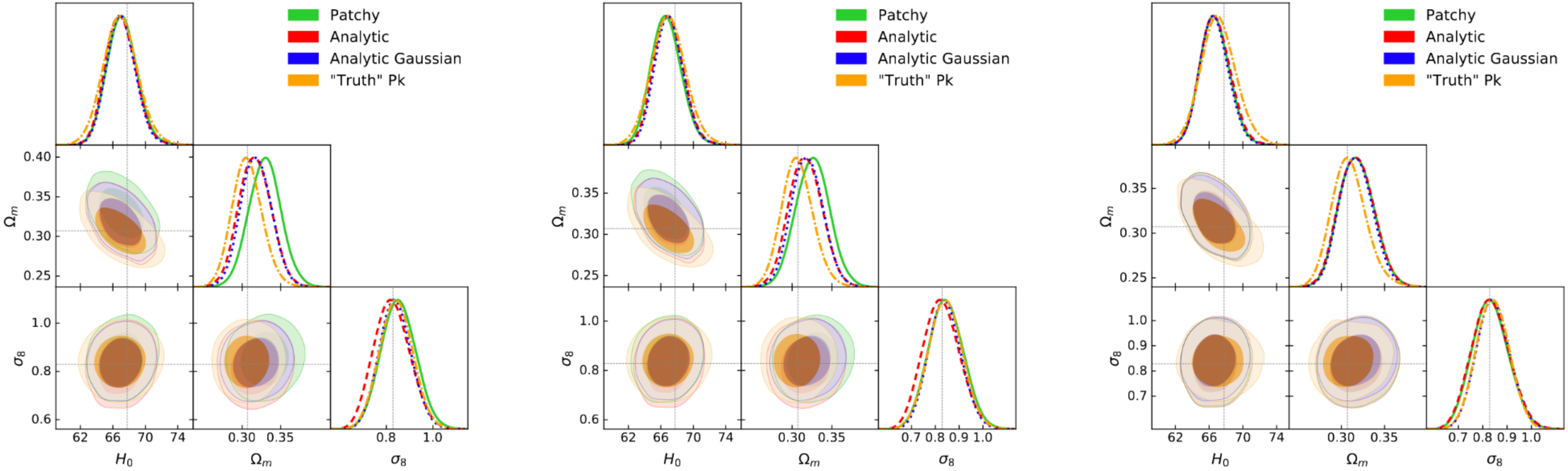
- Using **model-specific** subspace projections we can **heavily** compress cosmological data-sets

- The decomposition is
  1. **Robust** and **accurate**
  2. Widely **applicable**
  3. **Fast** and simple to use

- **Reduce** impact of **covariance** matrix noise:
  - **Sharpening** parameter constraints
  - Allows **fewer** mocks to be computed

Backup Slides

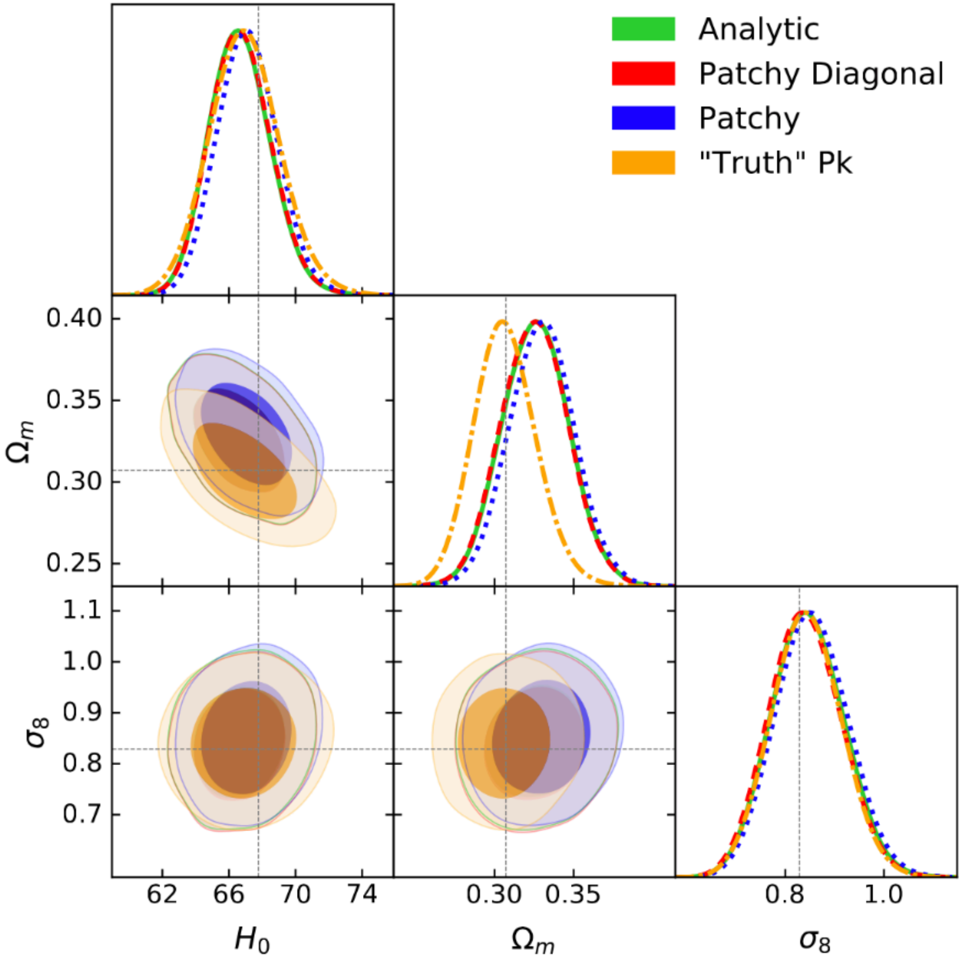
Altering the Data Covariance Matrix



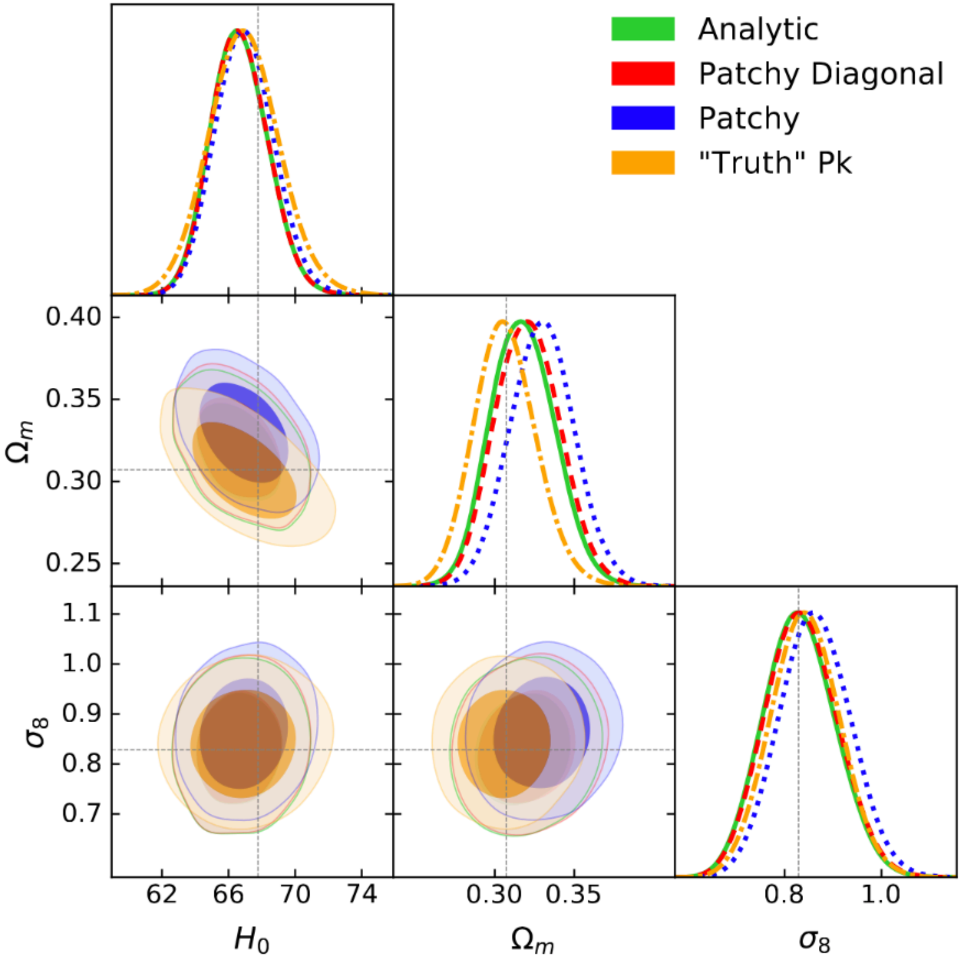
(a) 96-bin Power Spectrum (b) 48 Subspace Coefficients (c) 12 Subspace Coefficients

Philcox+20

Altering the Fiducial Covariance Matrix



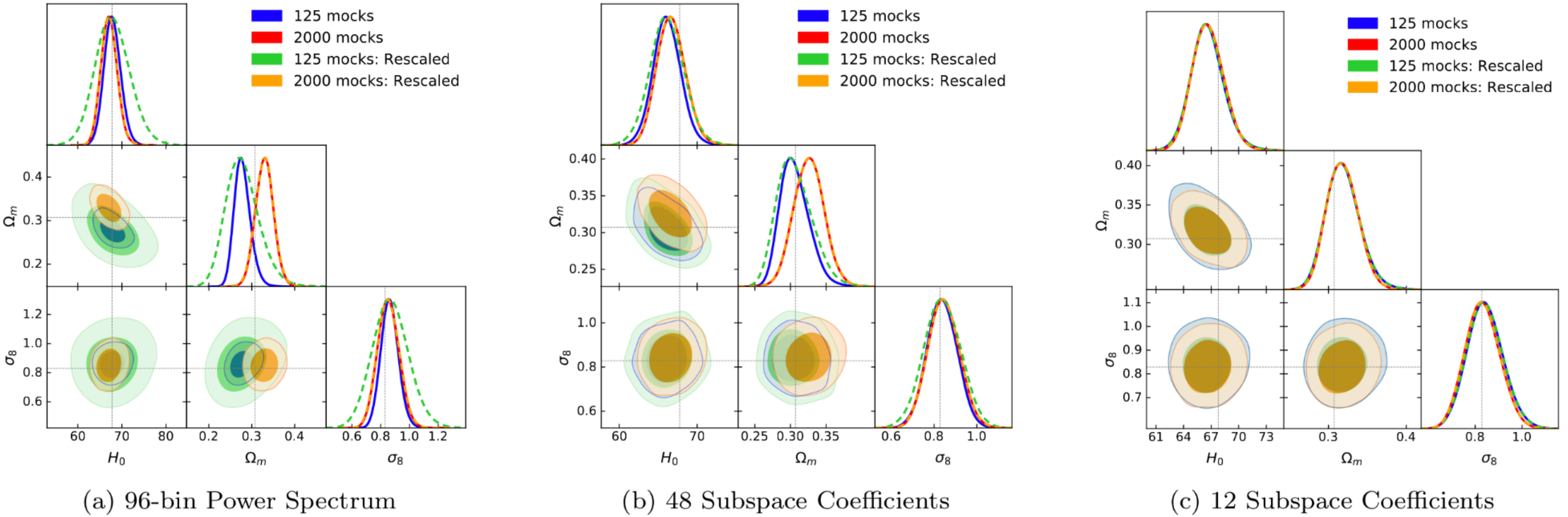
(a) 48 Subspace Coefficients



(b) 12 Subspace Coefficients

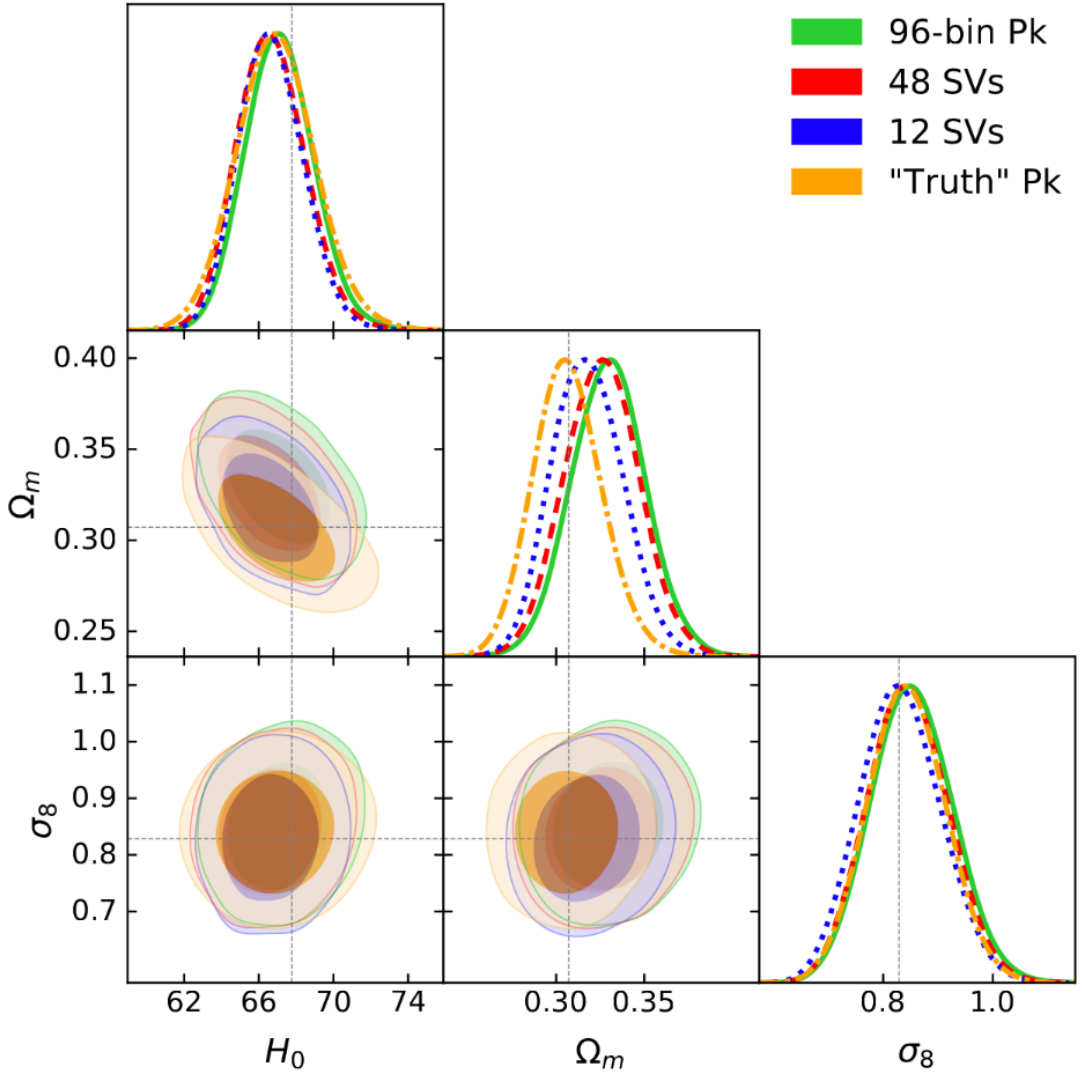
Philcox+20

Noise in the Covariance Matrix



Philcox+20

Single Mock Comparison



Philcox+20