Reconstructing Cosmology
using
Principal Component Analysis

Ranbir Sharma

IISER Mohali, Punjab

Collaborators: *Ankan Mukherjee, H K Jassal*

Content:

- Cosmology
- Observational datasets and methods to probe the Universe
- Statistical ways to Interpret - Infer - Imply
- Principal Component Analysis (PCA)
- Reconstruction of late-time cosmology using PCA
- Inference of model parameter using PCA
- Summary and Conclusion

Cosmology

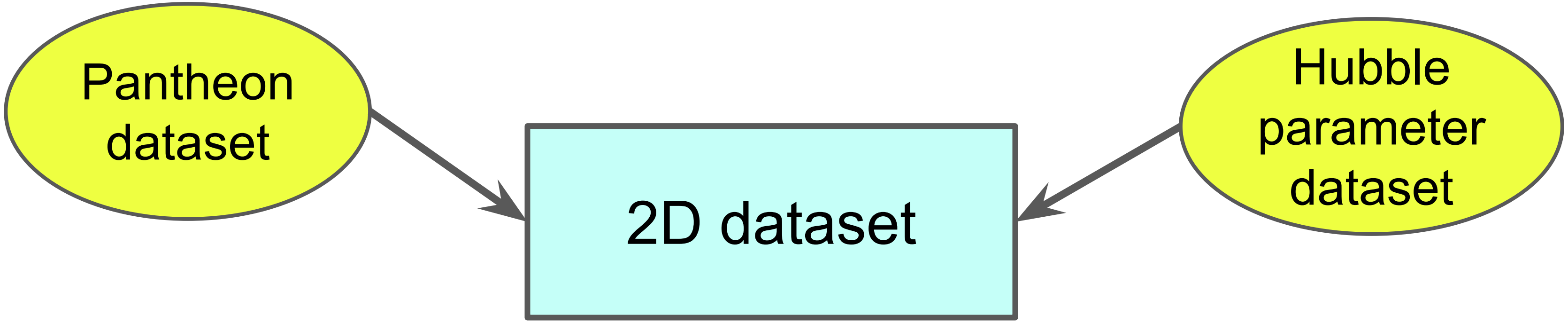
$$R^i_k - \frac{1}{2}\delta^i_k R = 8\pi G T^i_k$$

$$\frac{2\ddot{a}}{a} + \frac{\dot{a}^2 + k}{a^2} = -8\pi G p$$

$$\frac{\dot{a}^2 + k}{a^2} = \frac{8\pi G}{3}\rho$$

$$H^2(z) = H_0^2\left[\Omega_m(1+z)^3 + \Omega_x e^{3\int_0^z \frac{1+w(z')}{1+z'}dz'}\right]$$

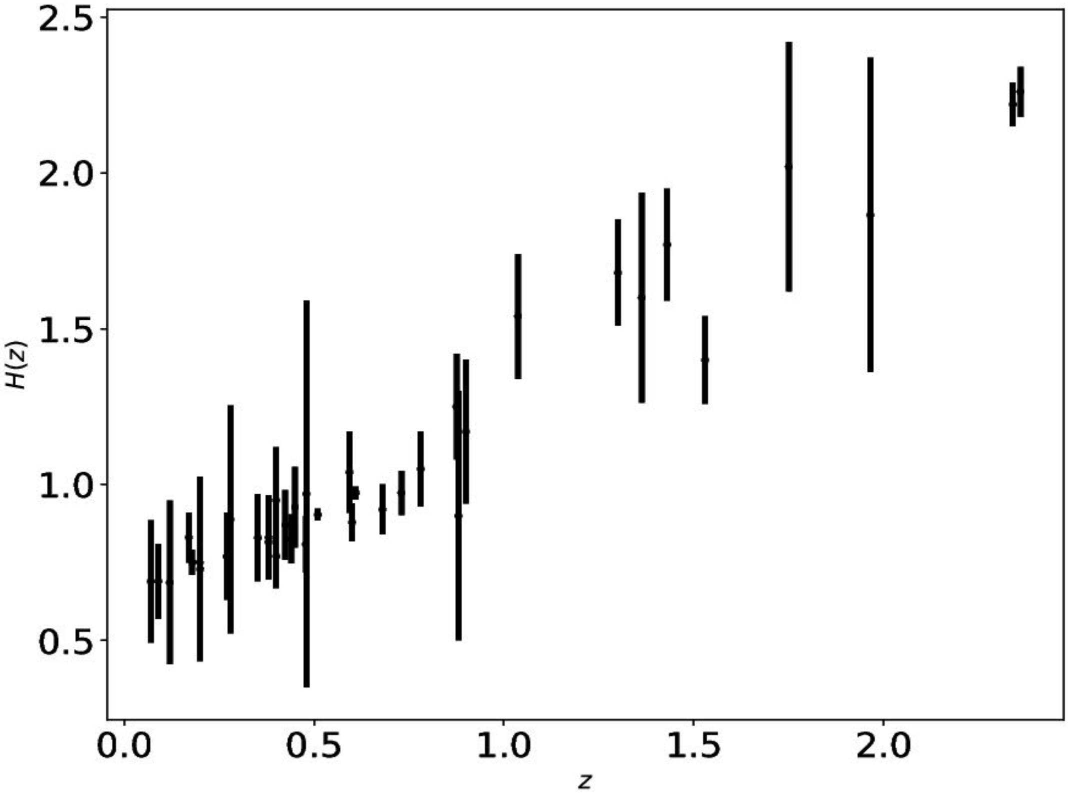
Observational datasets



z	m	σ_m
0.01	0.139	0.198
.	.	.
.	.	.

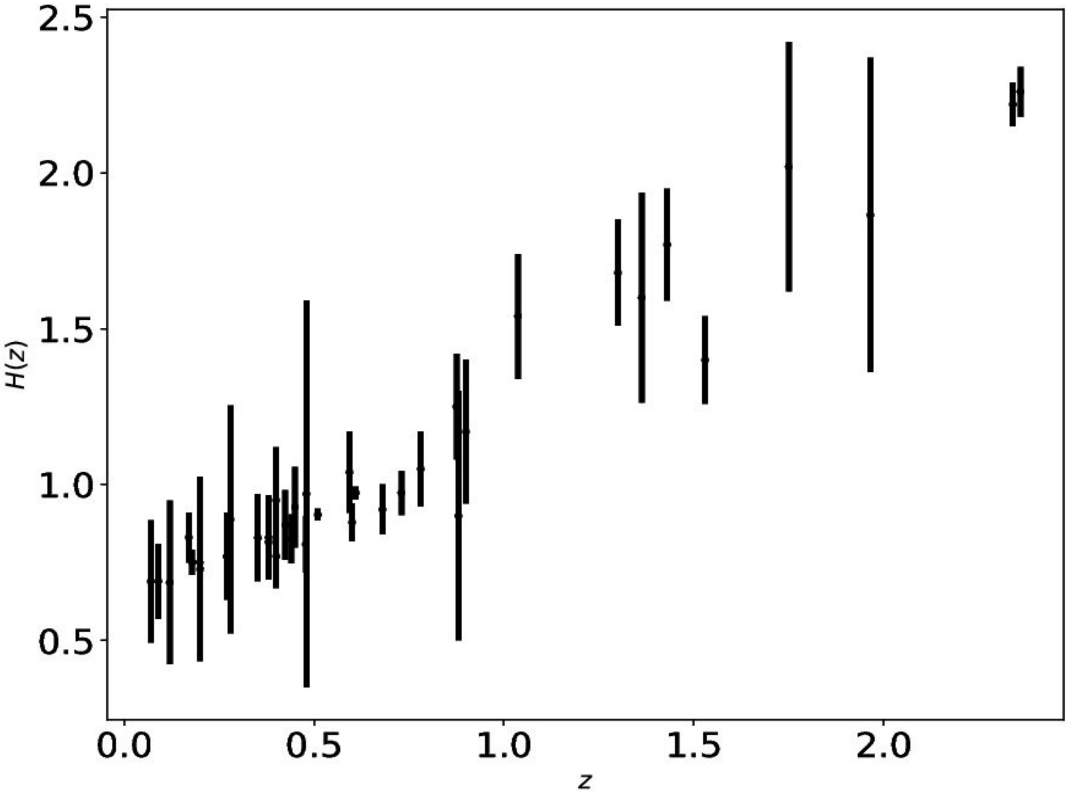
z	h	σ_h
0.07	0.69	0.196
.	.	.
.	.	.

Ways of reconstruction from data



- ➢ **Reconstruction in model dependent parametric manner**
- ➢ **Reconstruction in model independent non-parametric manner**
- ➢ **Reconstruction through Principal Component Analysis(PCA)**

Ways of reconstruction using PCA



- **Using Fisher Matrix approach**
- **Using different realisation from dataset in hand**
- **Algorithm of PCA + Correlation coefficient calculation (CCC)**

Correlation coefficient calculation (CCC)

Pearson CC → L →

$$\rho = \frac{\mathrm{Cov}(A,B)}{\sigma_A \sigma_B}$$

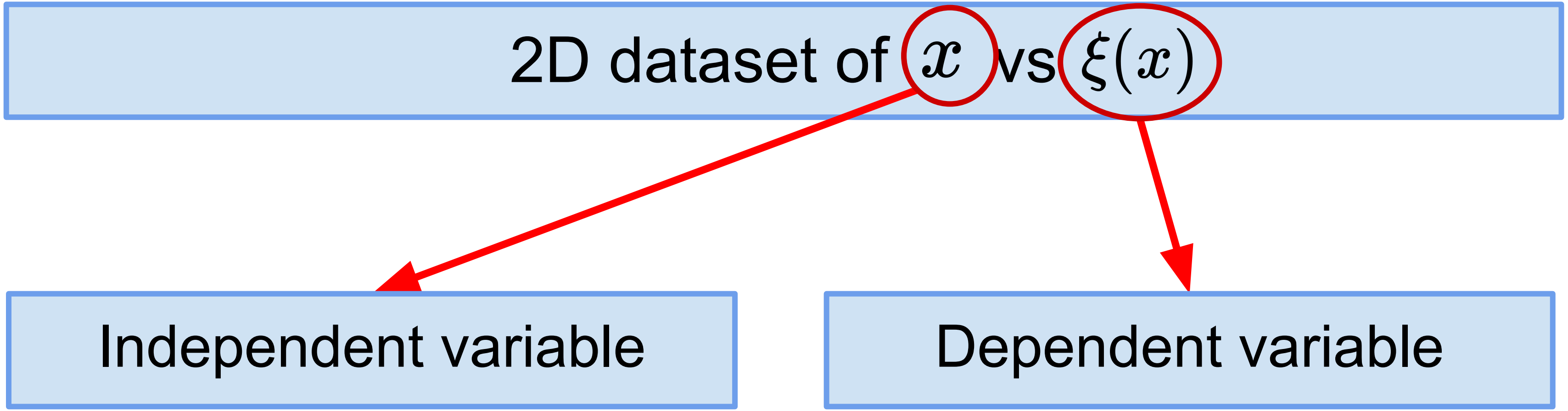
Spearman CC → NL →

$$r = \frac{\mathrm{Cov}(r_A, r_B)}{\sigma_{r_A} \sigma_{r_B}}$$

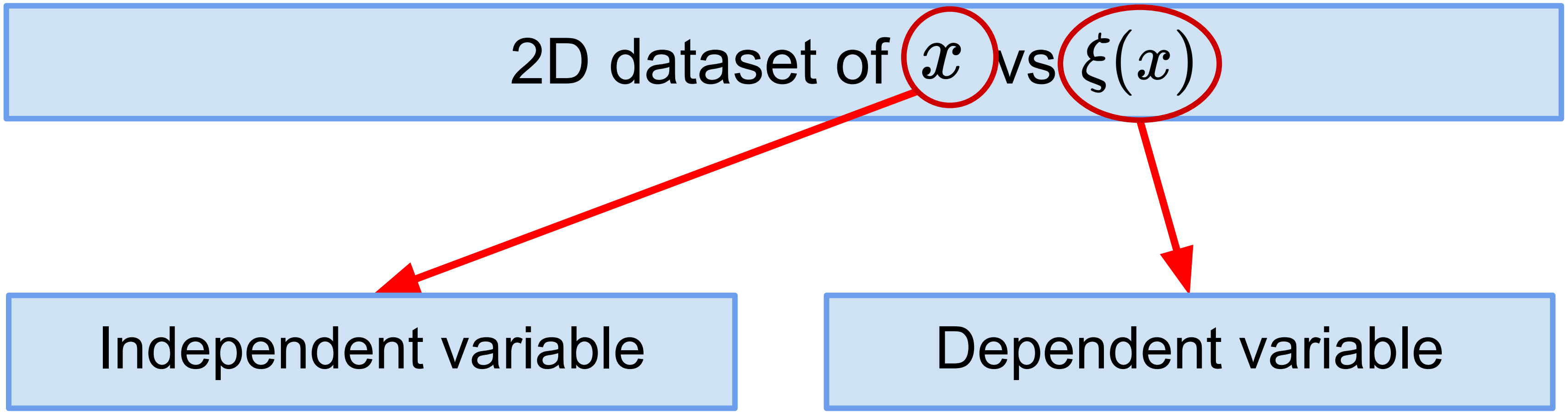
Kendall CC → NL →

$$\tau = \frac{\text{Actual Score}}{\text{Maximum Possible Score}}$$

Requirements of reconstruction



Requirements of reconstruction



Data should have **lesser** non-linear correlation than the linear

Algorithm of reconstruction from PCA

1) Expressing the dependent variable in a polynomial

$$\xi(x) = \sum_{i=1}^{N} \alpha_i f(x)^{(i-1)}$$

Independent Unit basis

Reconstructed quantity

Coefficients

Algorithm of reconstruction from PCA

1) Expressing the dependent variable in a polynomial

$$\xi(x) = \sum_{i=1}^{N} \alpha_i f(x)^{(i-1)}$$

Dimension of coefficient space

Independent Unit basis

Reconstructed quantity

Coefficients

Algorithm of reconstruction from PCA

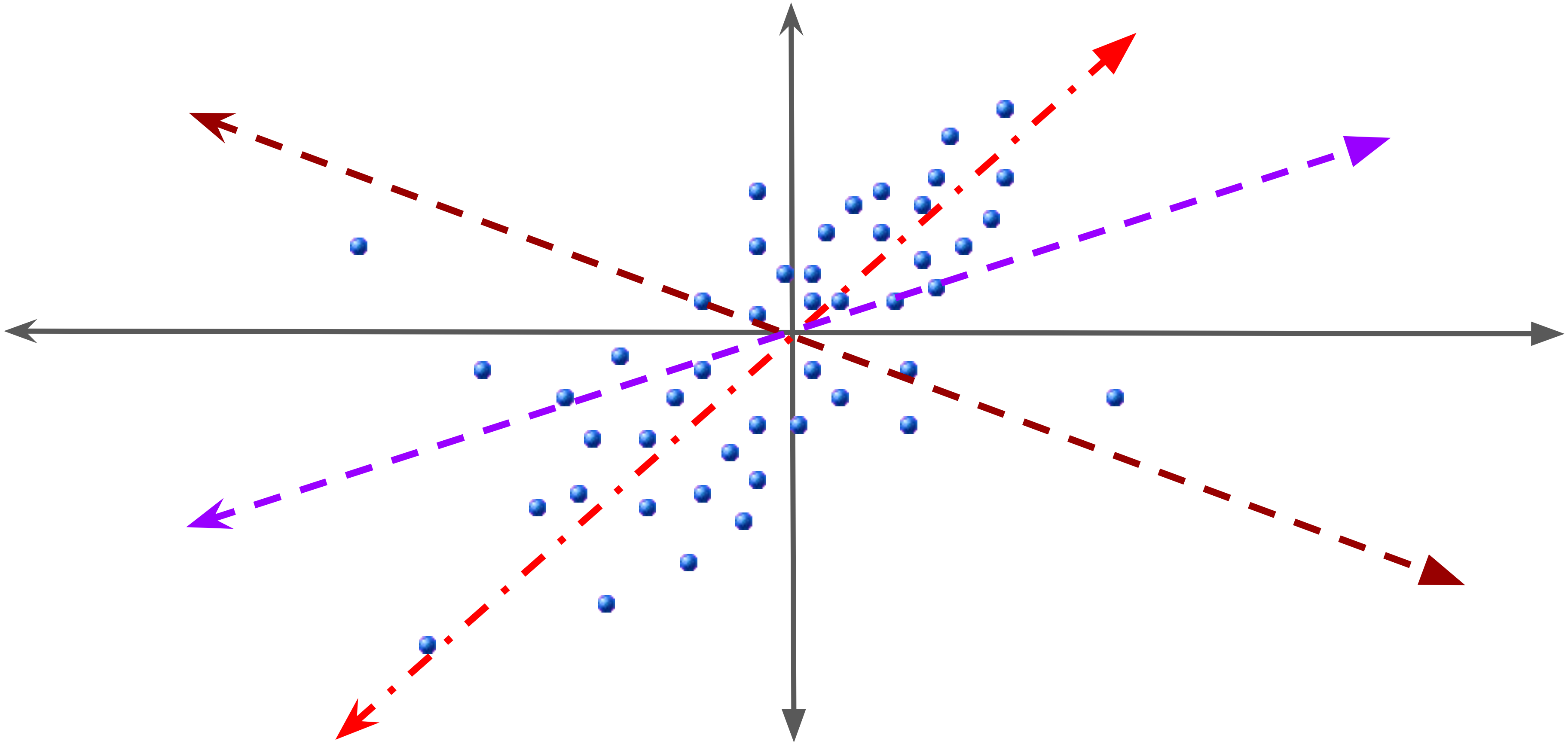
2) Division of N-dim coefficient space and the calculation of data-matrix of PCA

$$\chi^2 = \sum_{j=1}^{k} \frac{\left(\xi(x)_{data} - \xi(\{b_i\},x)\right)^2}{\sigma_j^2}$$

From tabulated dataset

From polynomial expression

Algorithm of reconstruction from PCA



Algorithm of reconstruction from PCA

3) PCA data-matrix

$$\mathbf{Y} = \begin{pmatrix} b_1^{(1)} & b_2^{(1)} & \cdots & b_n^{(1)} \\ b_1^{(2)} & b_2^{(2)} & \cdots & b_n^{(2)} \\ b_1^{(3)} & b_2^{(3)} & \cdots & b_n^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ b_1^{(N)} & b_2^{(N)} & \cdots & b_n^{(N)} \end{pmatrix}$$

N —-> Number of dimension of coefficient space

n —--> Number of patches created

Algorithm of reconstruction from PCA

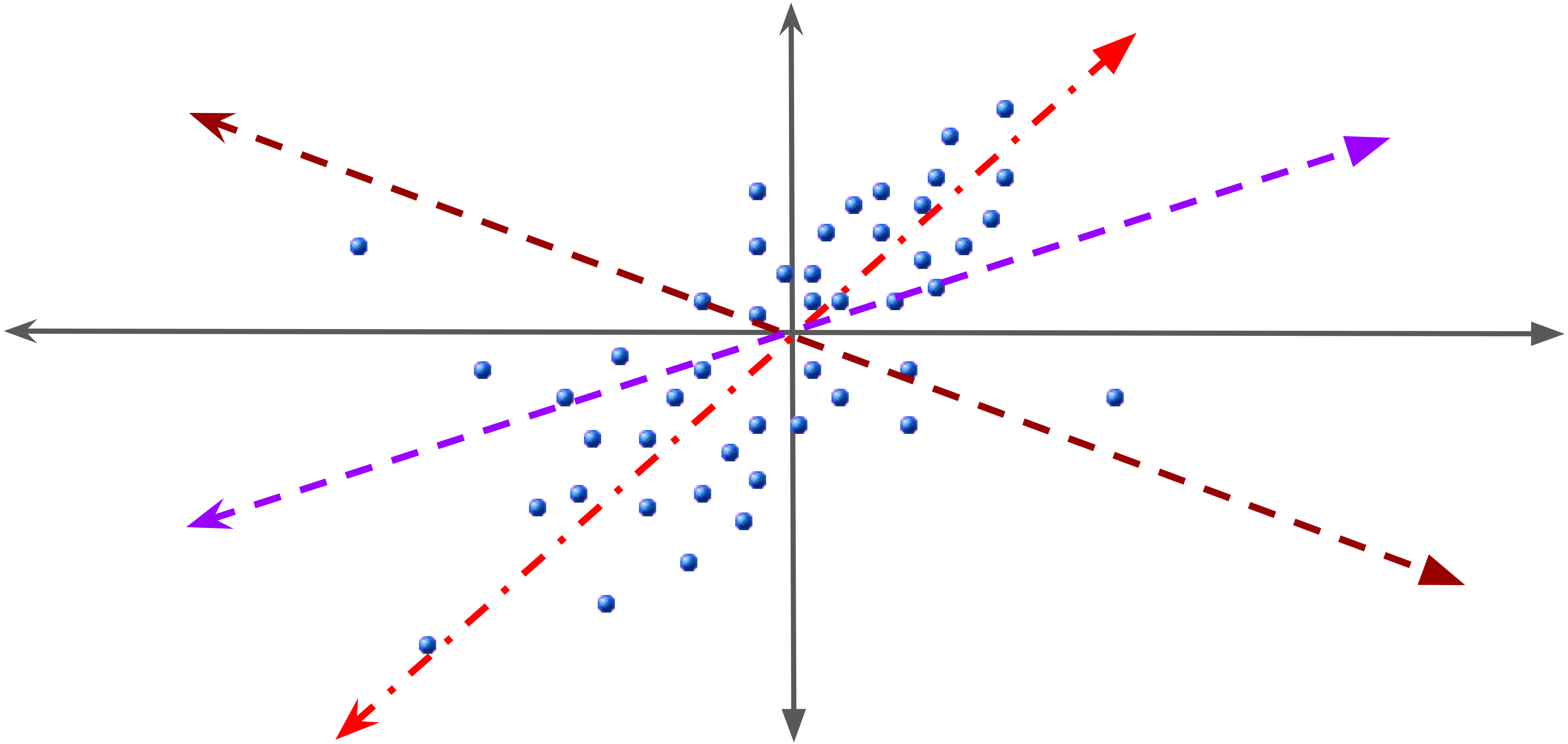
4) Covariance matrix $\mathbf{C} = \frac{1}{n}\mathbf{Y}\mathbf{Y}^T$

$$\mathcal{C} = \begin{matrix} \sigma_{b_1}^2 & Cov(b_1, b_2) & Cov(b_1, b_3) \\ Cov(b_2, b_1) & \sigma_{b_2}^2 & Cov(b_2, b_3) \\ Cov(b_3, b_1) & Cov(b_3, b_2) & \sigma_{b_3}^2 \end{matrix}$$

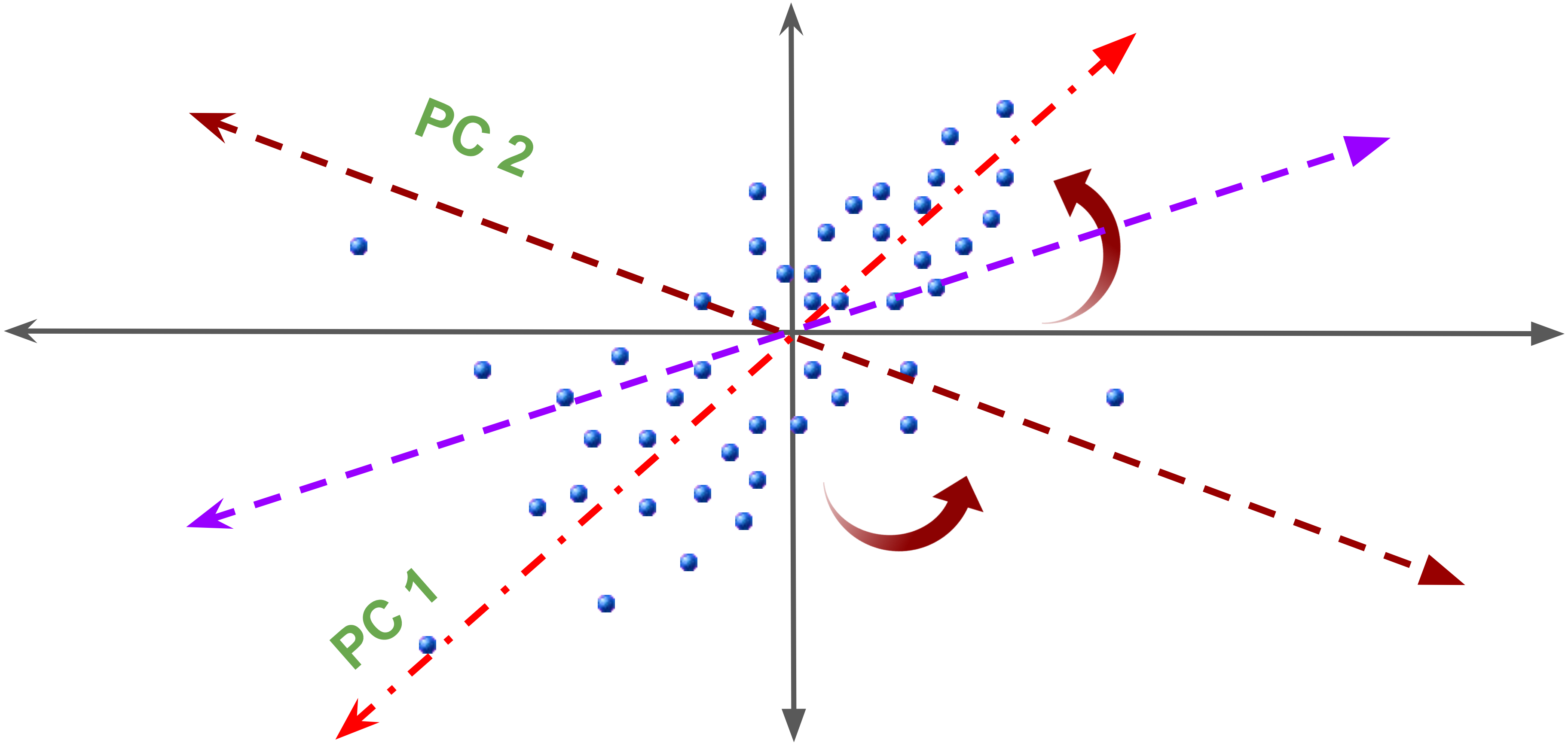
From noise

From signal

Algorithm of reconstruction from PCA

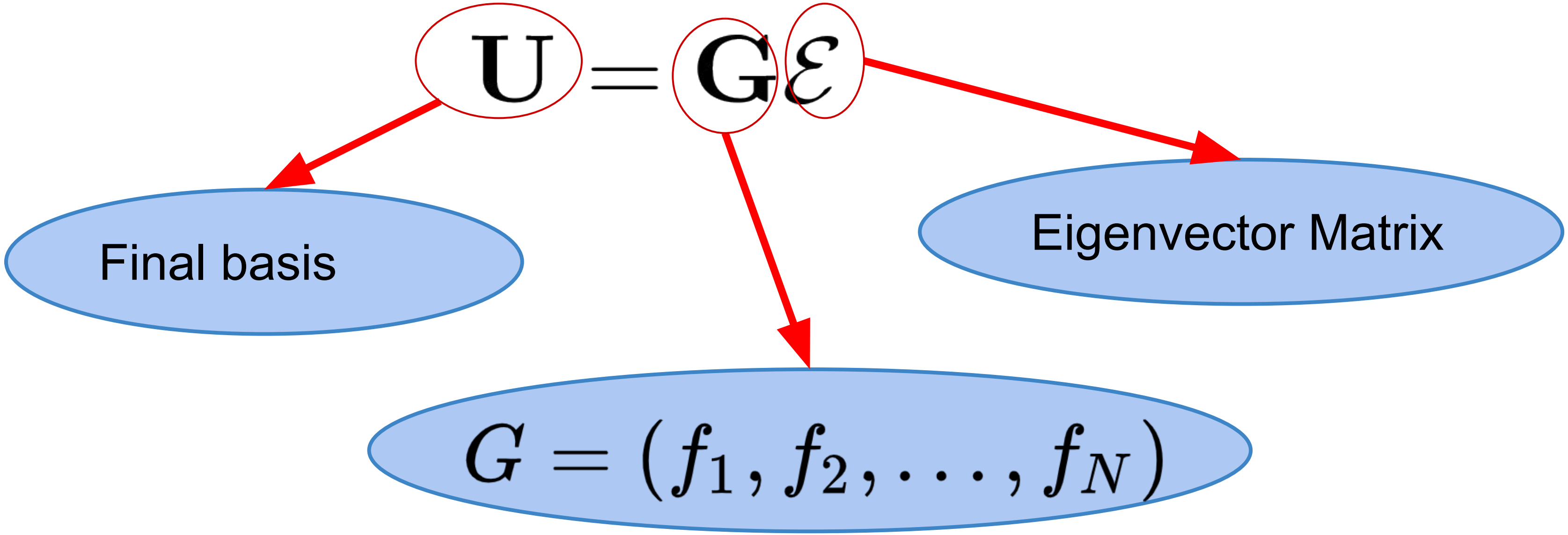


Algorithm of reconstruction from PCA



Algorithm of reconstruction from PCA

5) Diagonalisation of the covariance Matrix and finding of Eigenfunctions



Algorithm of reconstruction from PCA

6) Reduction of dimension and the final form

Reduced final dimension

Final basis

$$\xi(x) = \sum_{i=1}^{M} \beta_i u(x)^{(i-1)}$$

Final coefficients

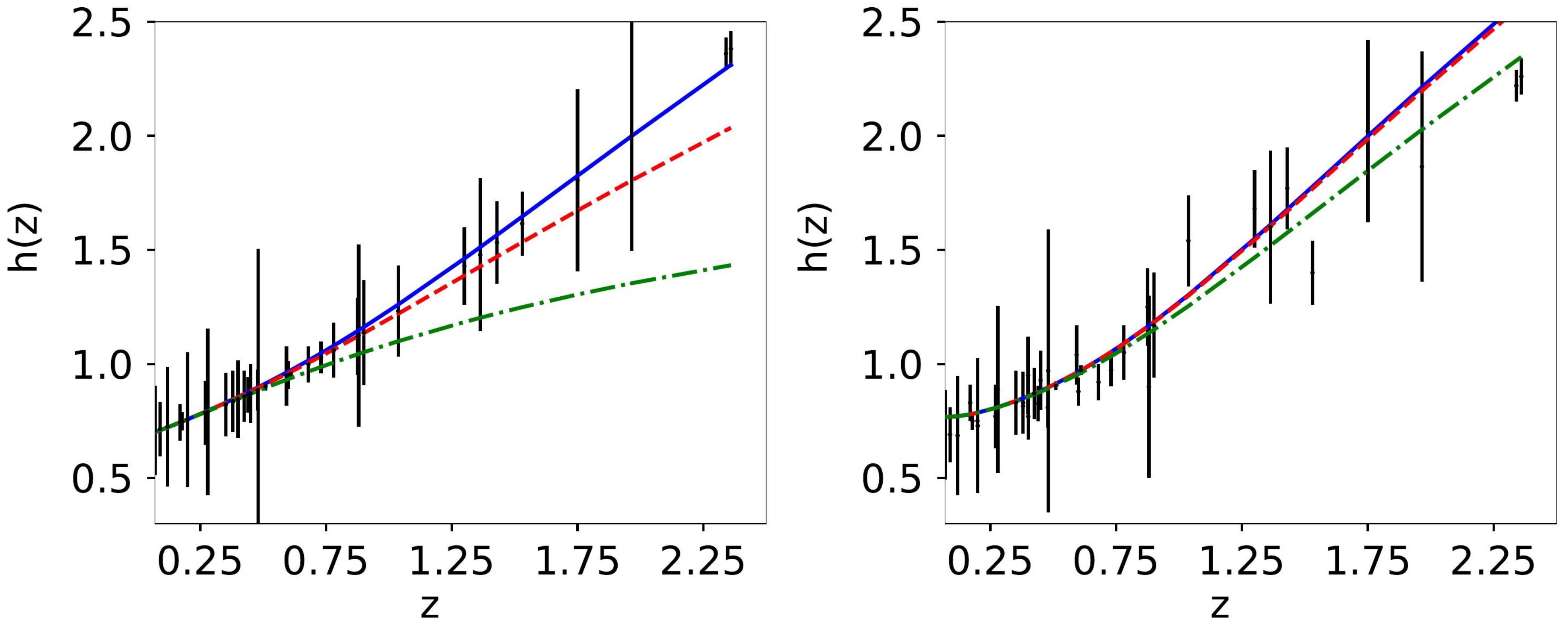
Reconstruction of w(z) derived approach

$$H^2(z) = H_0^2\left[\Omega_m(1+z)^3 + \Omega_x e^{3\int_0^z \frac{1+w(z')}{1+z'}dz'}\right]$$

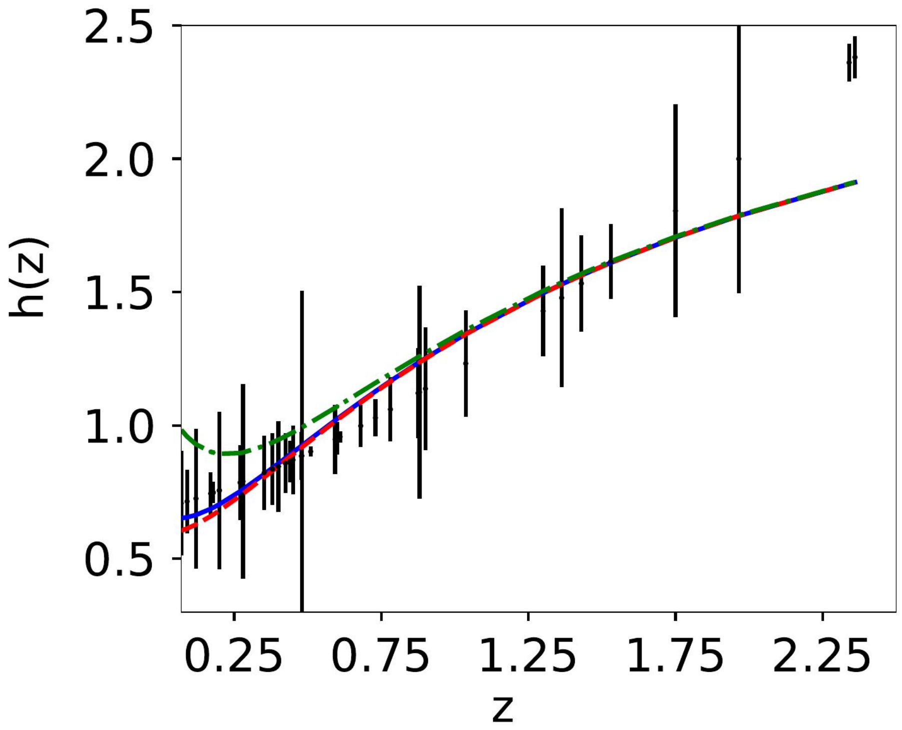
From PCA

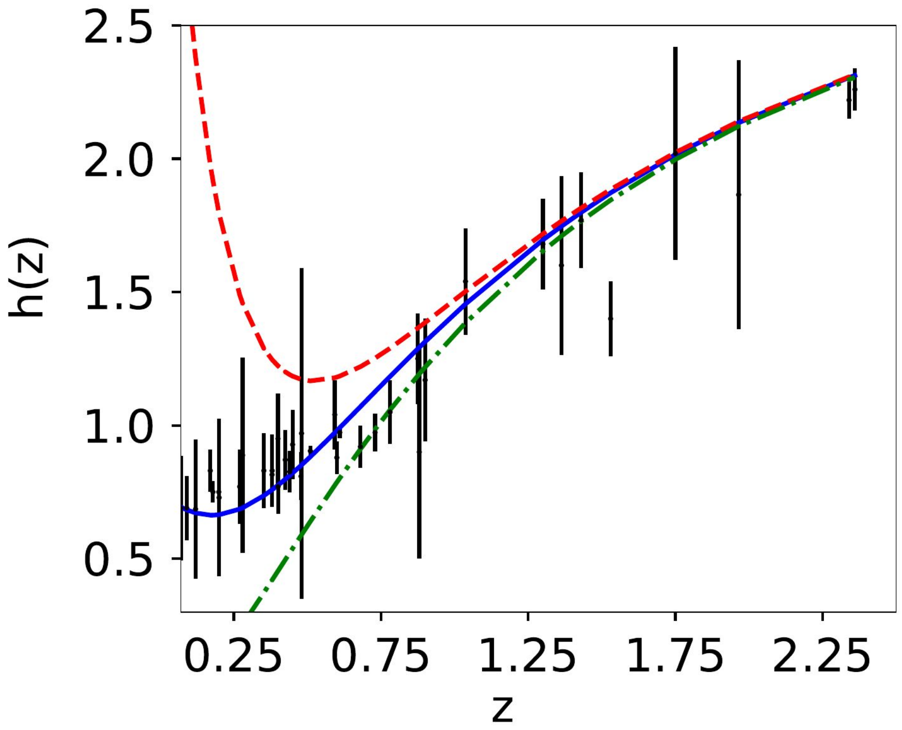
$$w(z) = \frac{3H^2 - 2(1+z)HH'}{3H_0^2(1+z)^3\Omega_M - 3H^2}.$$

Results(derived approach) [basis: (1-a)] Hz

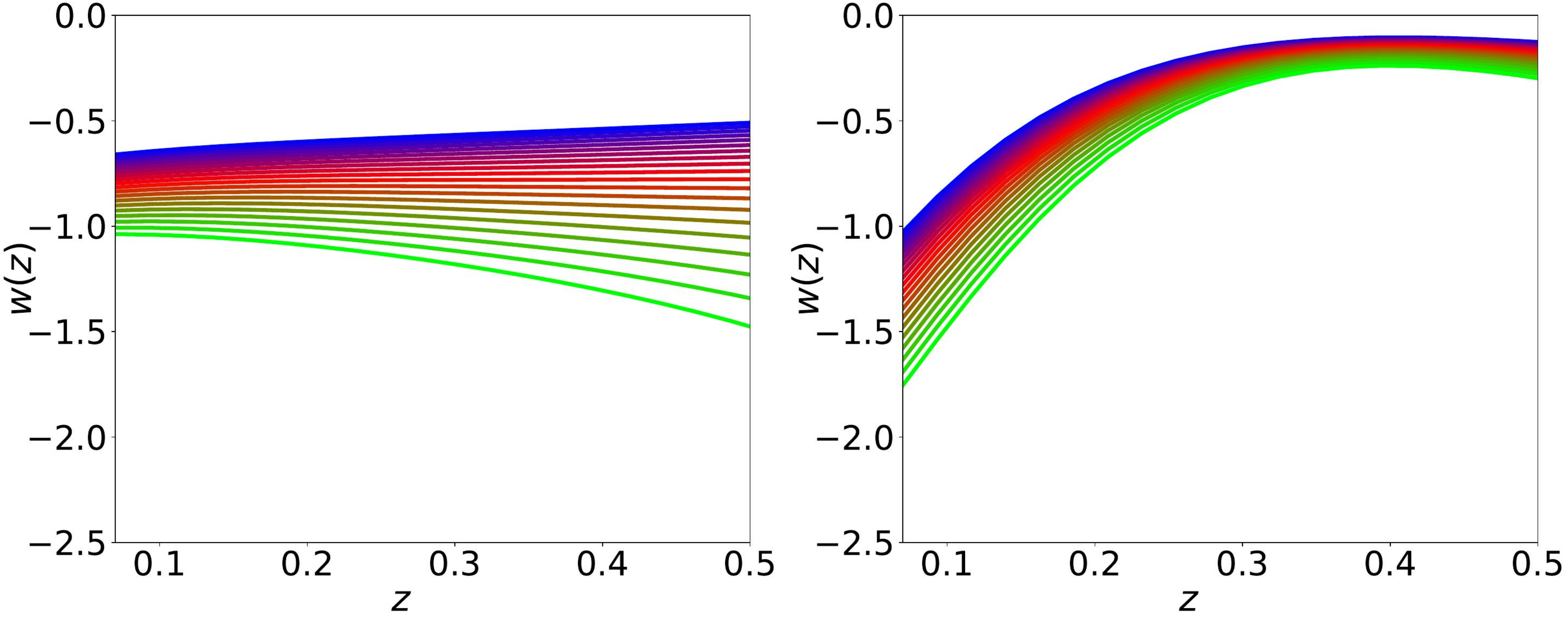


Results(derived approach) [basis: a] Hz

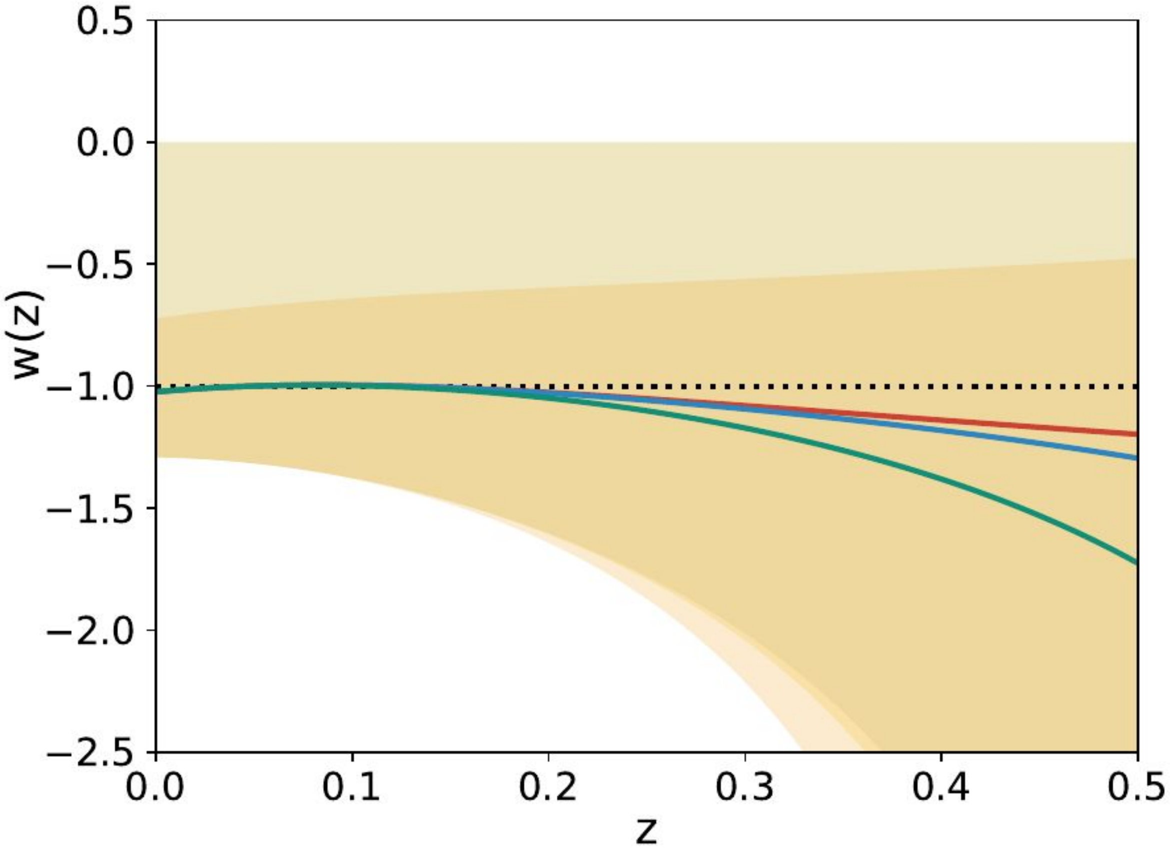




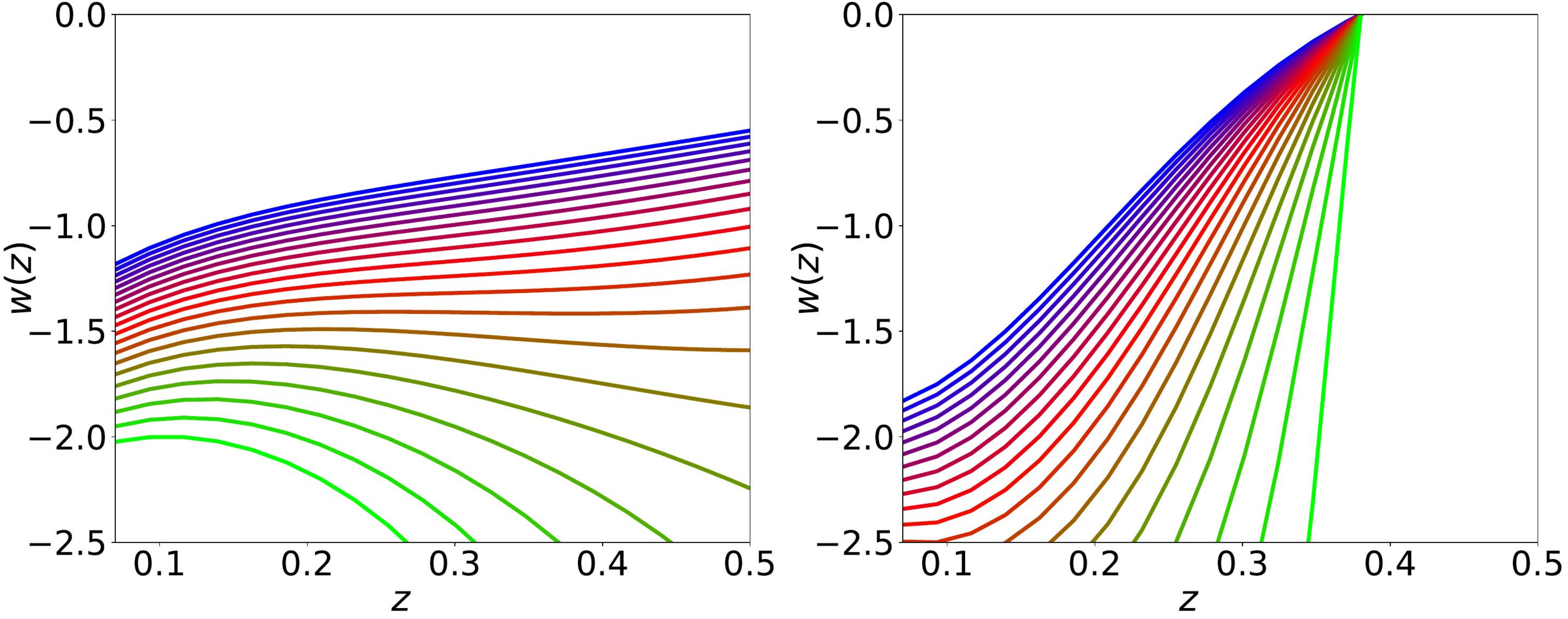
Results: w(z) from Hz simulated data



Results: w(z) from Hz simulated data



Results: w(z) from Hz real data



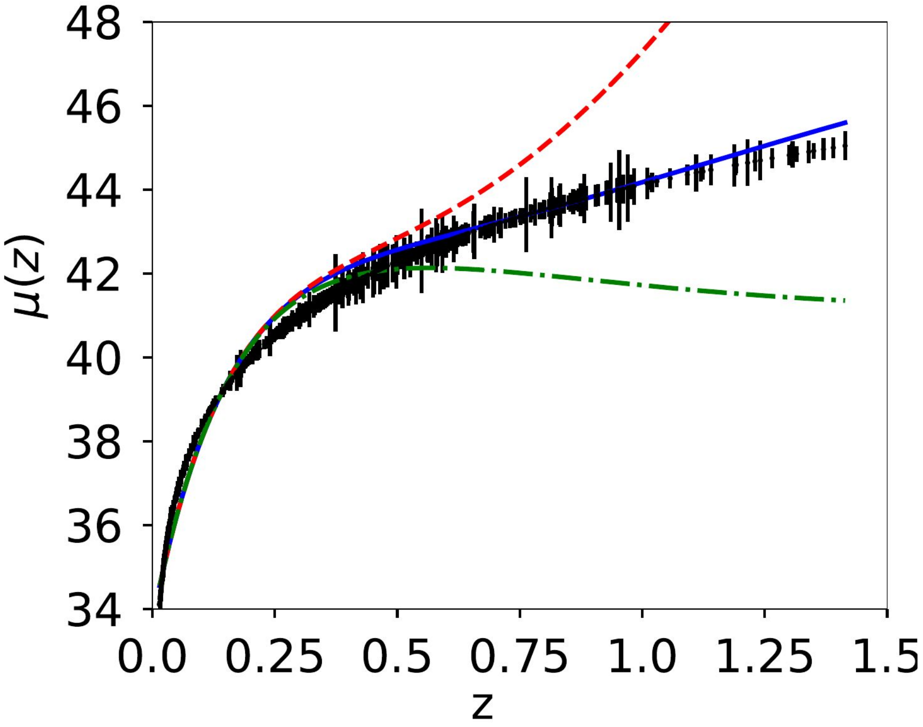
Reconstruction of w(z) derived approach

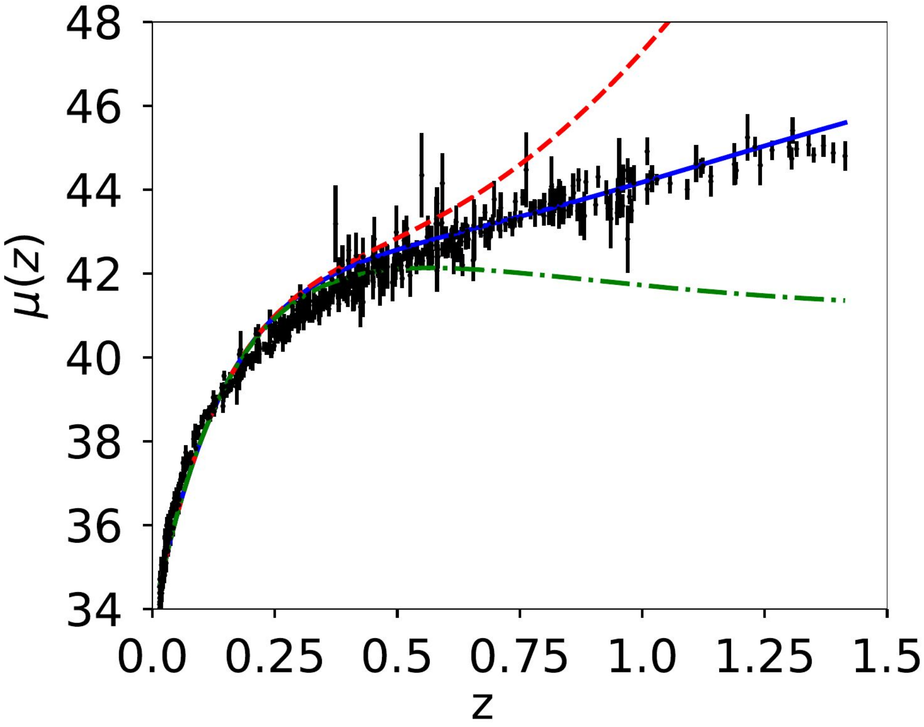
$$\mu(z) = 5\log\left(\frac{d_L}{1Mpc}\right) + 25$$

$$\mu \longrightarrow d_L \longrightarrow w$$

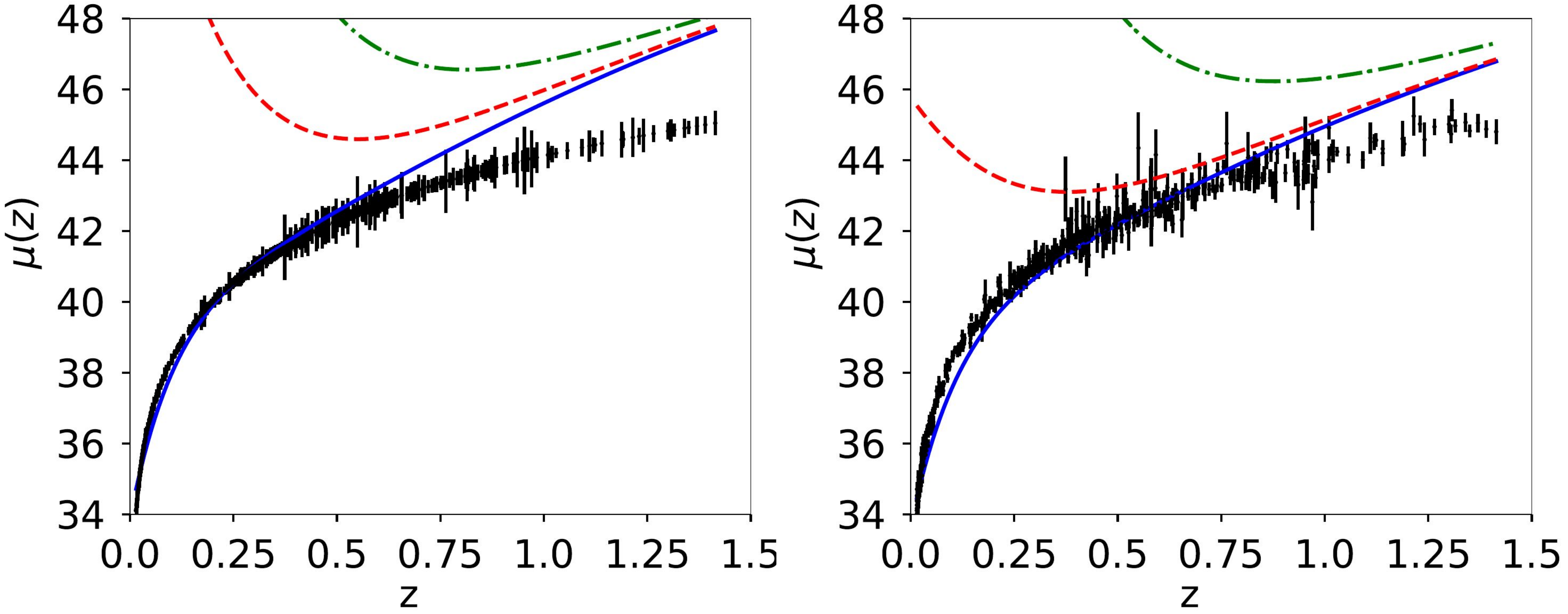
$$d_L(z) = \frac{c}{H_0}(1+z)\int_0^z \left(\Omega_m(1+z)^3 + \Omega_x e^{3\int_0^z \frac{(1+w(z'))dz'}{(1+z')}}\right)^{-1/2} dz$$

Results(derived approach) [basis: (1-a)] SNIa





Results(derived approach) [basis: a] SNIa



Error in PCA

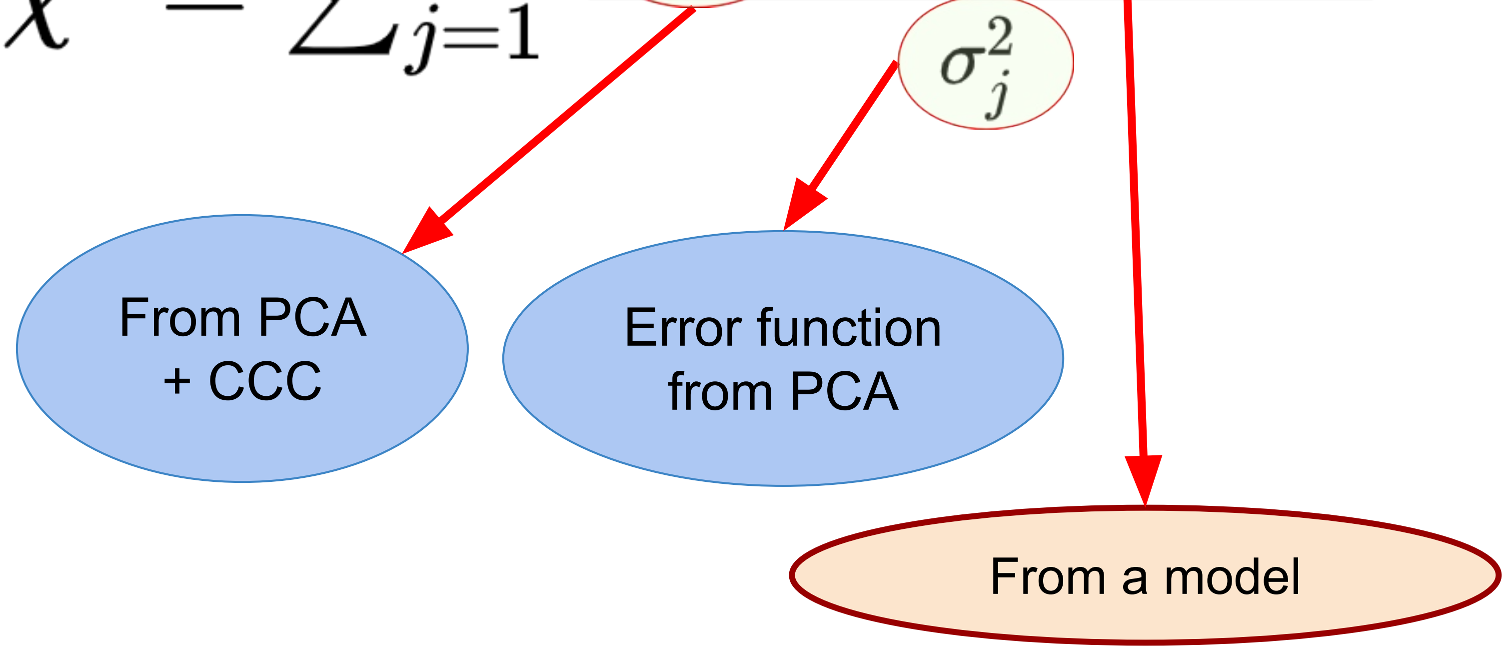
- By masking the under-fitting points of the PCA-data-matrix we re-create the cov-matrix

- EigenValues of the cov-matrix gives the error information of the reconstruction

$$\sigma(\xi(x)) = \left[\sum_{i=1}^{M} \sigma^2(\beta_i) e_i^2(x)\right]^{\frac{1}{2}}$$

Inference of model's parameters from PCA

$$\chi^2 = \sum_{j=1}^{k} \frac{\left((\xi(x))_{PCA} - \xi(\{b_i\},x)\right)^2}{\sigma_j^2}$$



Inference of model’s parameters from PCA

$$\chi^2 = \sum_{j=1}^{k} \frac{\left(\xi(x)_{PCA} - \xi(\{b_i\},x)\right)^2}{\sigma_j^2}$$

From PCA + CCC

Error function from PCA

$$H^2(z) = H_0^2\left[\Omega_m(1+z)^3 + \Omega_x e^{3\int_0^z \frac{1+w(z')}{1+z'}dz'}\right]$$

Inference of model's parameters from PCA

$$\chi^2 = \sum_{j=1}^{k} \frac{\left((\xi(x))_{PCA} - \xi(\{b_i\}, x) \right)^2}{\sigma_j^2}$$

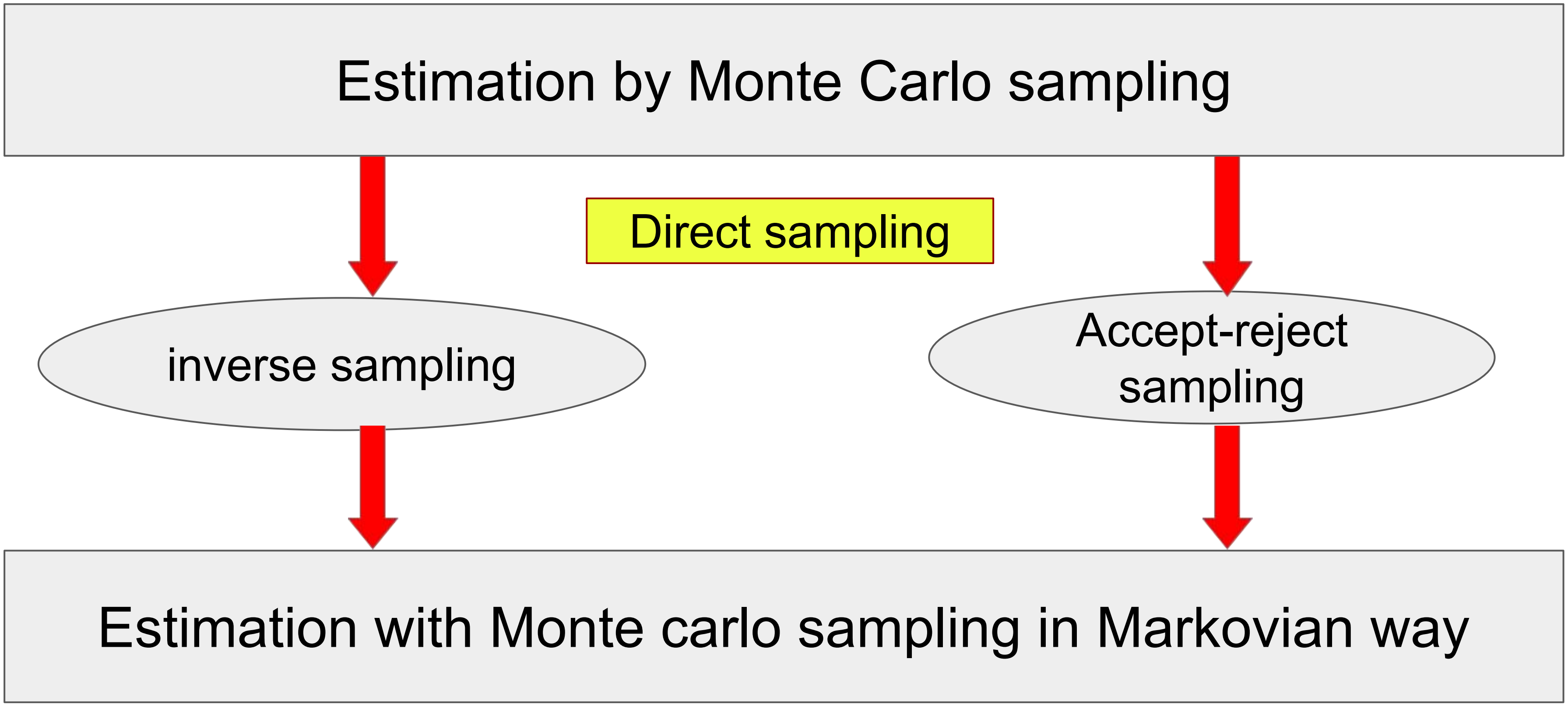
From PCA
+ CCC

Er

$$w(z) = \sum_{i=1}^{m} \beta_i \Gamma(z)^{i-1}$$

$$H^2(z) = H_0^2 \left[\Omega_m (1+z)^3 + \Omega_x e^{3 \int_0^z \frac{1+w(z')}{1+z'} dz'} \right]$$

Markov Chain Monte Carlo (MCMC)



HMC and NUTS

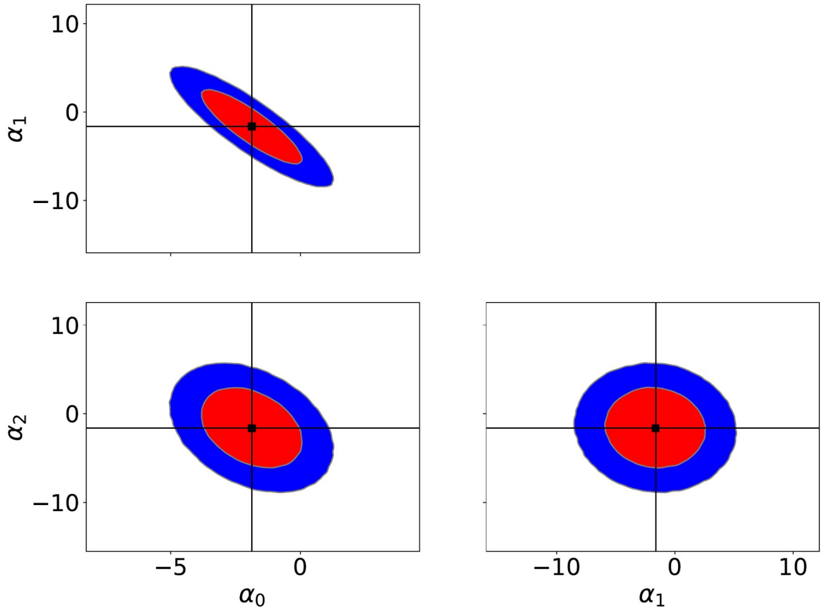
HMC →

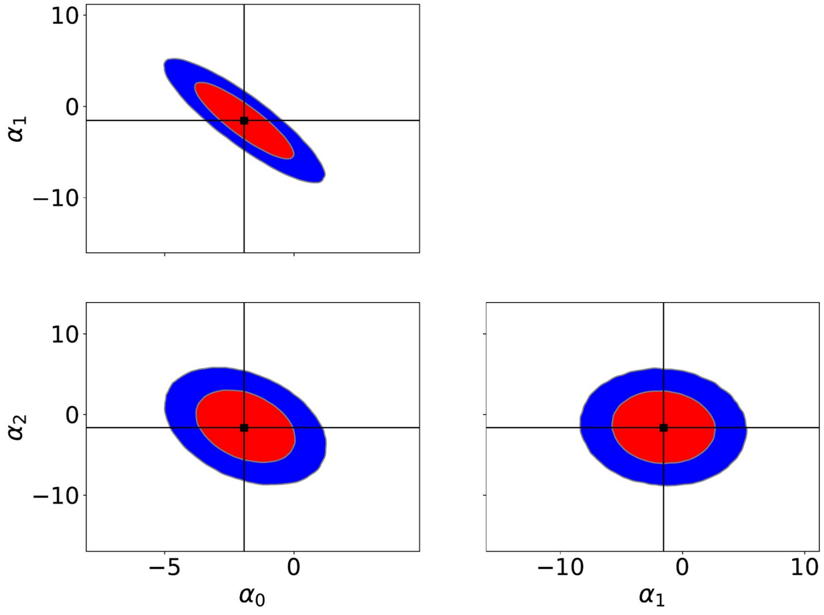
- Modified MH, with an introduction of an auxiliary variable (called momentum)
- With the restriction of following the Hamiltonian dynamics

NUTS →

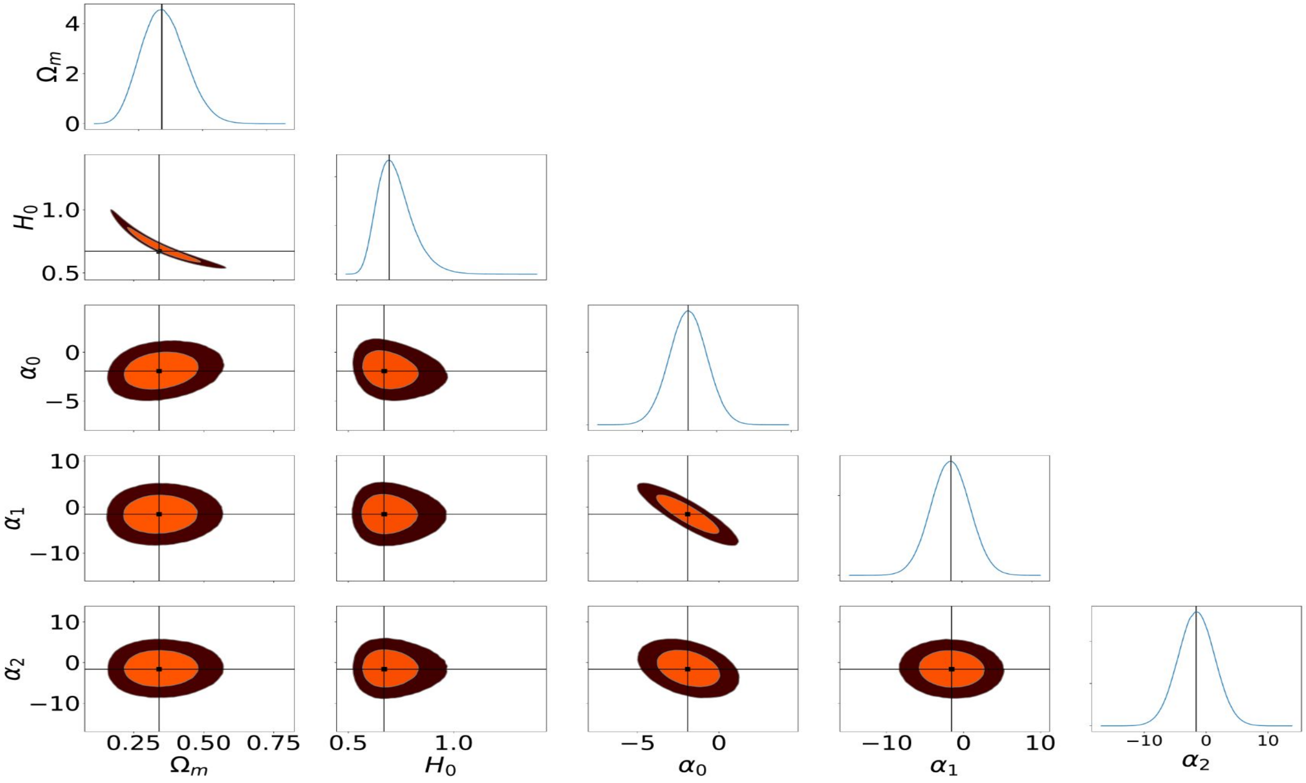
- Modified HMC, which can choose the "Leapfrog" steps and the step size automatically. This is done by imposing the "condition of Turning".

Inference of model's parameters from PCA

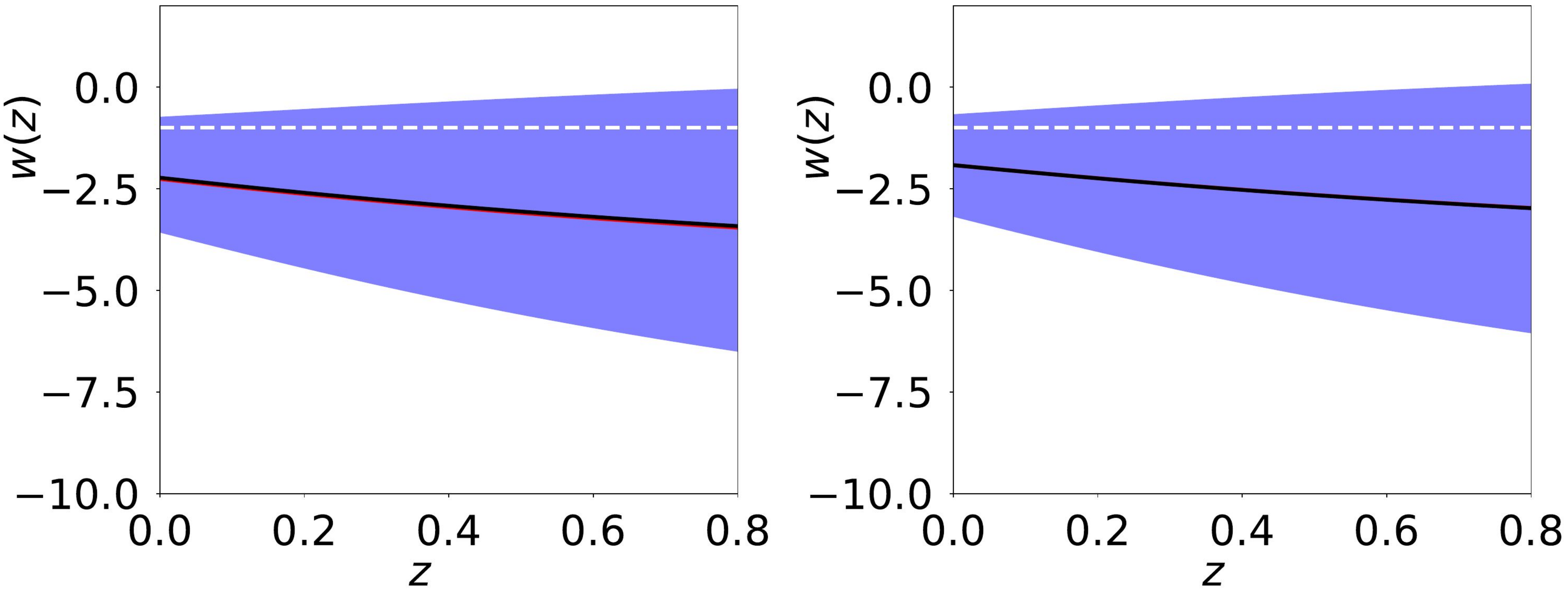




Inference of model's parameters from PCA



Inference of model's parameters from PCA



Summary and Conclusion :

- **PCA + CCC** gives a script to reconstruct a quantity in a non-parametric way.
- We need only the 2D dataset as input to reconstruct the quantity. **Choice of basis function and number of PCs comes intrinsically from the algorithm.**
- **PCA + CCC** implies a slowly varying dark-energy equation of state parameter.
- **PCA + CCC + MCMC** gives us a script to infer and constrain the parameters of a model without much biases.
- In the **PCA + CCC + MCMC** we need the error-function of the dataset.

Present and Future works :

- Comparison of different Kernel functions in the **Gaussian processes** reconstruction.
- Constraining **the modified gravity models** from PCA inference.
- Constraining Cosmological parameters from **Cluster and CMB data**.
- Principal Component Analysis in Maximum Likelihood estimation through **different error function**
- Maximum Likelihood Estimation using Principal Component Analysis for the **Pantheon dataset**

References :

- "**Reconstruction of late-time cosmology using Principal Component Analysis**", Ranbir Sharma, Anakan Mukherjee, H K Jassal, *EPJP (2022)137:219*
- "**Inference of model parameter from Principal Component Analysis**", Ranbir Sharma, H K Jassal, *arxiv::2211.13608*

Email ID: **ranbirsharma0313@gmail.com**