Constraints on local Primordial non-Gaussianities with large-scale structures

Laboratoire d'Annecy-le-Vieux de Physique Théorique

Marina S. Cagliari

Collaborators:
E. Castorina, M. Bonici, D. Bianchi, M. Barberi-Squarotti, K. Pardede, G. D'Amico, A. Bairagi, B. Wandelt

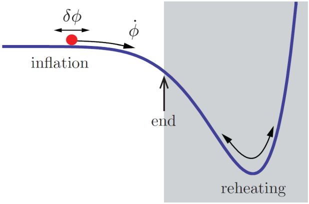
TUG - LAPTh - 11/07/2024

Inflation

Inflation is a **key ingredient of ΛCDM**.
It solves problems and makes predictions:

- large scales causally connected in the past,
- observed Universe is close to flat,
- spectral index,
- almost adiabatic fluctuations.

What is the dynamics of inflation?



Baumann (2009)

Inflation and $f_{\rm NL}$

Local PNG, $f_{\rm NL}$, give an insight into the inflation dynamics,

$$\lim_{k_1 \to 0} B_\zeta(k_1, k_2, k_3) = 4\, f_{\rm NL}\, P_\zeta(k_1)\, P_\zeta(k_3)\,.$$

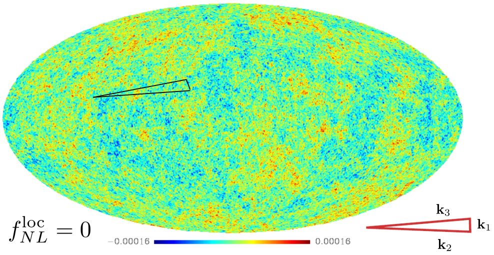
They are zero in **single-field inflation models**, and $f_{\rm NL} \sim 1$ in **multi-field models**:

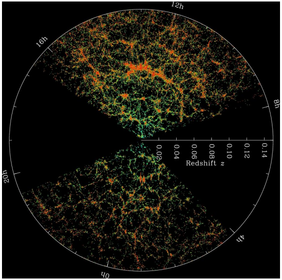
- detection $\to$ rule out single field models,
- non-detection of $f_{\rm NL} \sim 1 \to$ constraints on multi-field models.

To learn something we have to aim for $\sigma_{f_{\rm NL}} \lesssim 1$.

Maldacena (2003)
Creminelli and Zaldarriaga (2004)

$f_{\rm NL}$ probes





$$\sigma_{f_{\rm NL}}^{\rm CMB} \sim 5 \qquad \sigma_{f_{\rm NL}}^{\rm LSS} \sim 25$$

Planck+ (2019), SDSS, Castorina+ (2019), Cabass+ (2022), D'Amico+ (2022)

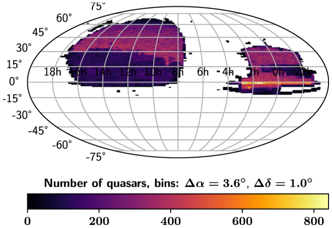
Old and new surveys

The new-generation surveys are here:

- DESI DR1 data public in five months,
- Euclid DR1 data public in one year and a half.

Use **old-generation surveys** to learn how to **extract information from the new data**.

- Results from eBOSS DR16 quasar sample:
  - 343 708 QSOs,
  - $0.8 < z < 2.2$



DESI Collaboration (2016)
Euclid Collaboration (2024)
Ross+ (2020)
Lyke+ (2020)

How does $f_{\rm NL}$ effect LSS?

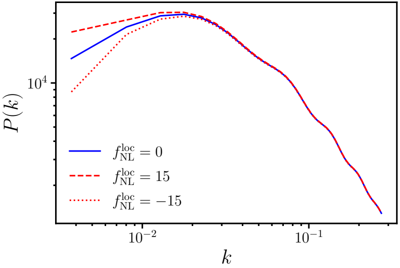
For $f_{\rm NL} \neq 0$ the variance of the short-scale density modes is affected by the large scales,

$$\phi = \phi_s + \phi_l$$
$$\langle \delta_s^2 \rangle^{1/2} = \sigma(1 + 2\, f_{\rm NL}\, \phi_l)\,.$$

The coupling of ϕ_l and ϕ_s produce a **scale dependent** term in the **bias**:

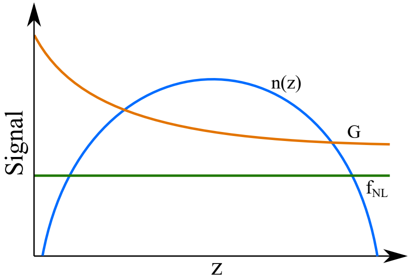
$$b_1 \rightarrow b_1 + f_{\rm NL}\, b_\phi k^{-2}\,,$$

with $b_\phi = b_1 - p$.



Dalal+ (2008)
Slosar+ (2008)

Optimal Weights



$$\delta_g \approx b_\phi \phi_\mathrm{p} + b_1 \delta_m(z)$$

Tegmark+ (1998), Castorina+ (2019)

$$P(k) \propto \langle \delta^\mathrm{w}(k)\, \delta^\mathrm{w}(-k) \rangle \,.$$

We use a **unique optimal galaxy weighting scheme** for constraining f_NL from LSS observations:

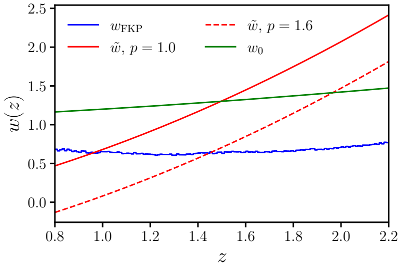
$$P(k) \propto \langle \tilde{\delta}(k)\, \delta_0(-k) \rangle \,,$$

where z-dependent weights were applied to the galaxy field,

$$\tilde{w}(z) = (b(z) - p)\,,$$

$$w_0(z) = D(z) \left(b(z) + \frac{f(z)}{3} \right) \,.$$

Optimal Weights



$$\delta_g \approx b_\phi \phi_p + b_1 \delta_m(z)$$

Tegmark+ (1998), Castorina+ (2019)

$$P(k) \propto \langle \delta^{w}(k)\, \delta^{w}(-k) \rangle \,.$$

We use a **unique optimal galaxy weighting scheme** for constraining $f_{\rm NL}$ from LSS observations:

$$P(k) \propto \langle \tilde{\delta}(k)\, \delta_0(-k) \rangle \,,$$

where z-dependent weights were applied to the galaxy field,

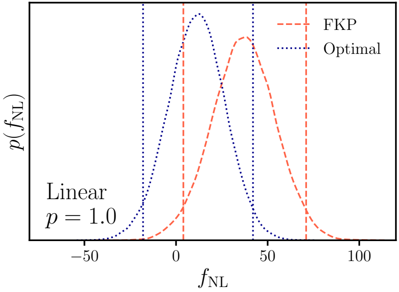
$$\tilde{w}(z) = (b(z) - p)\,,$$

$$w_0(z) = D(z)\, \left(b(z) + \frac{f(z)}{3} \right) \,.$$

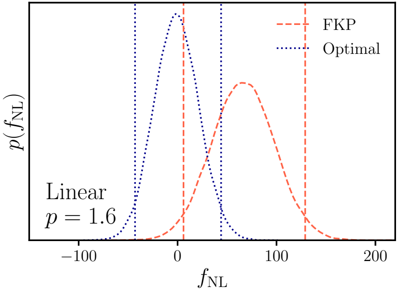
$f_{\rm NL}$ posterior

Cagliari+ (2023)

$p = 1.0$



$p = 1.6$



FKP: $\sigma_{f_{\rm NL}} \sim 17$ | $\sigma_{f_{\rm NL}} \sim 31$

Optimal: $\sigma_{f_{\rm NL}} \sim 15$ | $\sigma_{f_{\rm NL}} \sim 21$

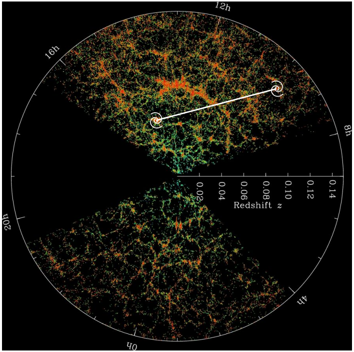
11% improvement! | 29% improvement!

How to improve the $P(k)$ results?

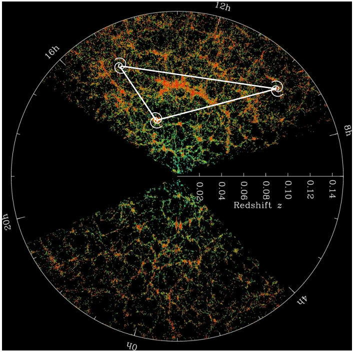
- Use larger surveys not shot-noise-dominated, like DESI and Euclid. However, larger surveys pose challenges in managing the **large-scale systematic effects**.
- Combine the 2-point statistics (power spectrum) with higher-order statistics, e.g. the bispectrum. In this case, we extract **more information from data we already have** and will be available in the future.

DISCLAIMER: We still need larger surveys!

Bispectrum: 3-point information



+



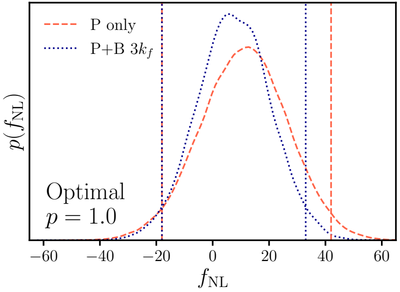
$$P_g(k) \propto \langle \delta_g(k)\, \delta_g(-k) \rangle\,,$$

$$B_g(k_1, k_2, k_3) \propto \langle \delta_g(k_1)\, \delta_g(k_2)\, \delta_g(k_3) \rangle\,.$$

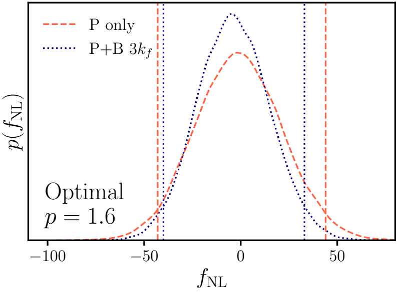
$f_{\rm NL}$ posterior

Cagliari+ (in prep.)

$p = 1.0$



$p = 1.6$



P only: $\sigma_{f_{\rm NL}} \sim 15$ $\sigma_{f_{\rm NL}} \sim 21$

P+B: $\sigma_{f_{\rm NL}} \sim 13$ $\sigma_{f_{\rm NL}} \sim 19$

13% improvement! 10% improvement!

Can field-level information help?

Memory problems plague ML-based field-level inference.

Divide the survey in patches

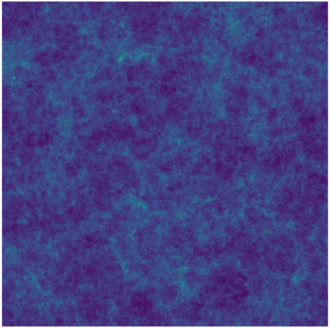
$\downarrow$ CNN

Small scales: patch summary

Large scales: $P(k)$ and $B(k)$ of the box

$\downarrow$ Combine

Get better constraints?



Can field-level information help?

Memory problems plague ML-based field-level inference.

Divide the survey in patches

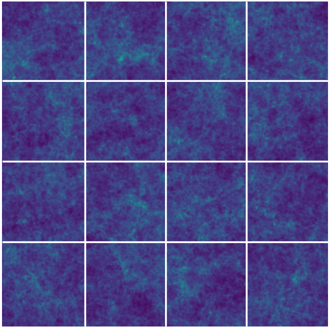
$\downarrow$CNN

Small scales: patch summary

Large scales: $P(k)$ and $B(k)$ of the box

$\downarrow$Combine

Get better constraints?



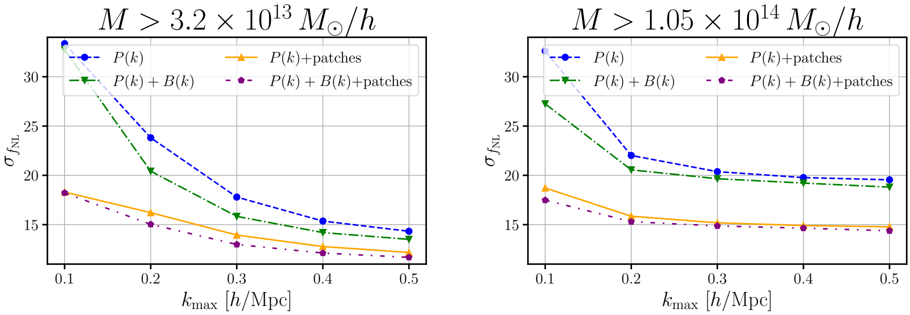
A Fisher information study

- $$F_{ij} = \left(\frac{\partial \mathbf{x}}{\partial \theta_i}\right)^T \mathbf{C}^{-1} \left(\frac{\partial \mathbf{x}}{\partial \theta_j}\right) \longrightarrow \sigma_i = \sqrt{(F^{-1})_{ii}}\,,$$
  where $\mathbf{x}$ is the estimator and $\mathbf{C}$ its covariance matrix.
- I compute the Fisher information content of $\mathbf{x} = \{P, B, \text{patches}\}$ in **simulated boxes** at $z = 1$.
- I use the halo catalogues of the Quijote-PNG simulations.
- I divide the $1^3\,(\text{Gpc}/h)^3$ simulation into $125^3\,(\text{Mpc}/h)^3$ patches with resolution $8\,\text{Mpc}/h$.

Villaescusa-Navarro+ (2020)
Coulton+ (2023)

Constraints on $f_{\rm NL}$

Cagliari+ (in prep.)



Conclusions

- I presented different ways to improve our current LSS analyses for f_{NL}:
  - Optimal $P(k)$: 29% improvement w.r.t. FKP $\rightarrow$ survey with double volume,
  - $P+B$: 38% improvement w.r.t. FKP $P(k)$ $\rightarrow$ almost a survey with three time the current volume.
- We obtained $\sigma_{f_{\mathrm{NL}}} \sim 13$, the best constraint on f_{NL} from a consistent LSS analysis.
- Combining standard summary statistics with ML field-level data compression can improve constraints and reduce the modelling for the analyses.

Thanks for your attention!

I also work in Euclid...

2109.07303

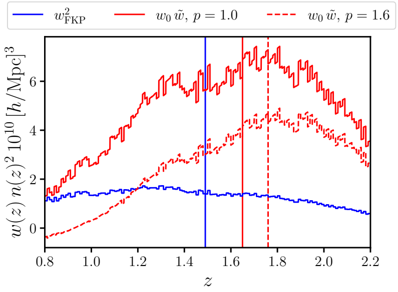
2403.08726

and with ML for photo-zs...

2211.01901

Contacts:
marina.cagliari@lapth.cnrs.fr
github/mcagliari

Effective redshift

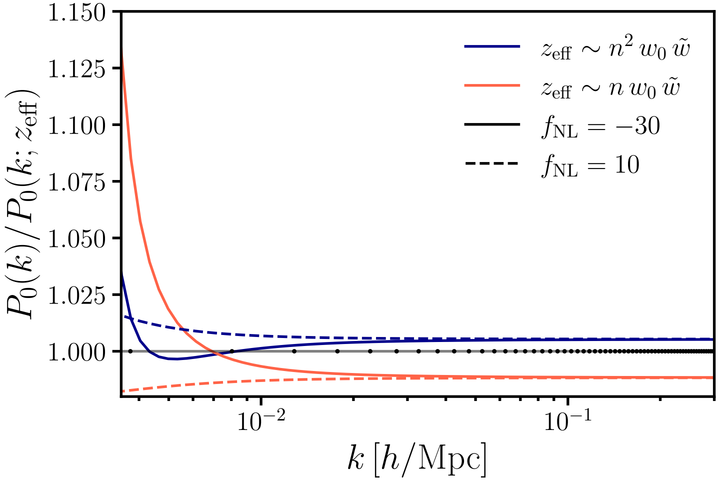


To not lose large-scale information along the line-of-sight we make a **full volume analysis**.

$$z_{\rm eff} = \frac{\sum_{i=1}^{N_{\rm QSO}} z_i\ n_i\ w^2_{{\rm c},\,i}\ w^2_{{\rm FKP},\,i}\ \tilde{w}_i\ w_{0,\,i}}{\sum_{i=1}^{N_{\rm QSO}} n_i\ w^2_{{\rm c},\,i}\ w^2_{{\rm FKP},\,i}\ \tilde{w}_i\ w_{0,\,i}}\,.$$

$$z_{\rm eff} = \begin{cases} 1.49 & {\rm FKP}\,, \\ 1.65 & p = 1.0\,, \\ 1.76 & p = 1.6\,. \end{cases}$$

How good is the z_{eff} approximation?



Power spectrum

The signal is at low k and at high k dominated by the z error $\rightarrow$ linear theory is enough.

$$P_{\mathrm{QSO}}(k, \mu, z_{\mathrm{eff}}) = G(k, \mu; \sigma_{\mathrm{FoG}})^2 \left(b_{tot}(k) + f\,\mu^2\right)^2 P_m(k) + N\,,$$

$$b_{tot}(k) = b_{\mathrm{QSO}} + f_{\mathrm{NL}}(b_{\mathrm{QSO}} - p)\tilde{\alpha}(k)\,.$$

$P_m(k)$ is computed in Planck2018 cosmology.

Window & Integral Constraint: the recipe

We are only interested in the monopole:

$$P^{(0)}_{\mathrm{QSO}}(k) = Q^{(0)}_{k\,k_p}\,P^{(0)}_{k_p} - \frac{1}{5}\,Q^{(2)}_{k\,k_p}\,P^{(2)}_{k_p} + \frac{1}{9}Q^{(4)}_{k\,k_p}\,P^{(4)}_{k_p} + \dots\,.$$

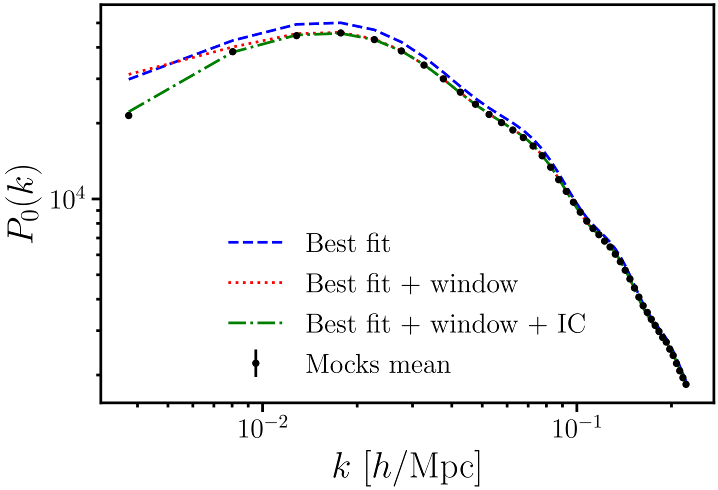
This needs to be corrected due to integral constraint effects:

$$P^{(0),\,\mathrm{IC}}_{\mathrm{QSO}}(k) = P^{(0)}_{\mathrm{QSO}}(k) - P^{(0)}_{\mathrm{QSO}}(0)\,|W_0(k)|^2 - P^{(0)}_{\mathrm{QSO}}(k)\,W_{\mathrm{RIC}}(k)\,.$$

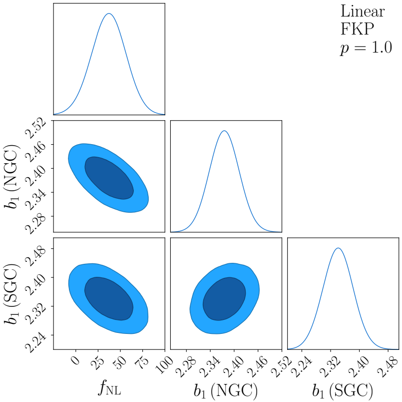
$P^{(0),\,\mathrm{IC}}_{\mathrm{QSO}}(k)$ is what enters in the likelihood.

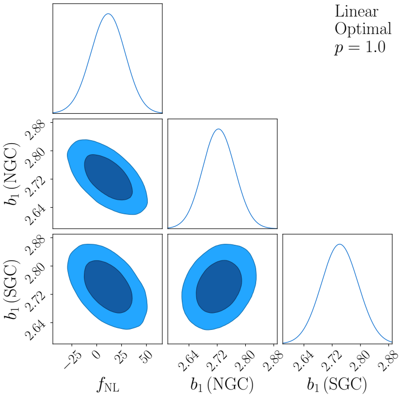
Wilson+ (2017)
De Mattia and Ruhlmann-Kleider (2019)

Everything together

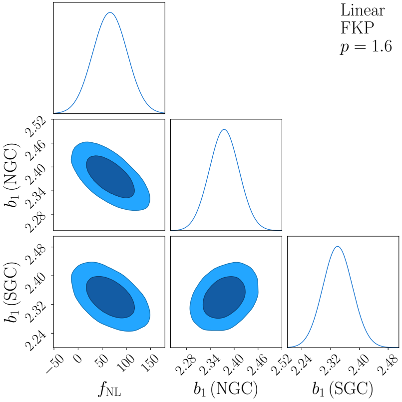


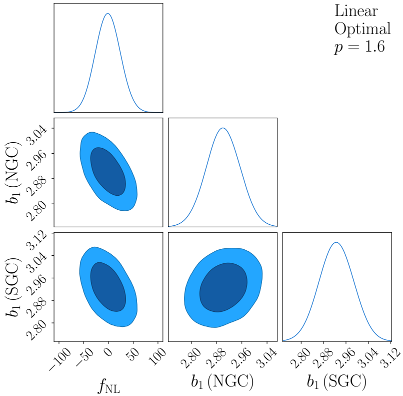
Bias and f_{NL} correlation: $p = 1.0$





Bias and $f_{\rm NL}$ correlation: $p = 1.6$





Three-level bispectrum

$$B_{\rm QSO}(k_1,k_2,k_3) = \mathcal{Z}_1(k_1)\mathcal{Z}_1(k_2)\mathcal{Z}_1(k_3)B(k_1,k_2,k_3) + \\ + 2\mathcal{Z}_1(k_1)\mathcal{Z}_1(k_2)\mathcal{Z}_2(k_1,k_2)P(k_1)P(k_2) + 2\,\mathrm{perm}\,,$$

where also $B(k_1,k_2,k_3)$ depends on the matter $P(k)$.

The window function and GIC are applied to the matter $P(k)$s that enter the bispectrum model. The RIC is applied similarly to the $P(k)$ model:

$$B^{(0),\,\rm IC}_{\rm QSO}(k_1,k_2,k_3) = B^{(0),\rm GIC}_{\rm QSO}(k_1,k_2,k_3) - \\ - B^{(0)}_{\rm QSO}(k_1,k_2,k_3)\,W^{B}_{\rm RIC}(k_1,k_2,k_3)\,.$$

Window approximation