

Benchmarking field-level cosmological inference from galaxy redshift surveys

Hugo Simon ^a, François Lanusse ^b and Arnaud de Mattia ^a

^a Université Paris-Saclay, CEA, Irfu,
91191, Gif-sur-Yvette, France

^b Université Paris-Saclay, Université Paris Cité, CEA, CNRS, AIM,
91191, Gif-sur-Yvette, France

E-mail: hugo.simon@cea.fr, francois.lanusse@cnrs.fr,
arnaud.de-mattia@cea.fr

ABSTRACT: Field-level inference has emerged as a promising framework to fully harness the cosmological information encoded in next-generation galaxy surveys. It involves performing Bayesian inference to jointly estimate the cosmological parameters and the initial conditions of the cosmic field, directly from the observed galaxy density field. Yet, the scalability and efficiency of sampling algorithms for field-level inference of large-scale surveys remain unclear. To address this, we introduce a standardized benchmark using a fast and differentiable simulator for the galaxy density field based on JaxPM. We evaluate a range of sampling methods, including standard Hamiltonian Monte Carlo (HMC), No-U-Turn Sampler (NUTS) without and within a Gibbs scheme, and both adjusted and unadjusted microcanonical samplers (MAMS and MCLMC). These methods are compared based on their efficiency, in particular the number of model evaluations required per effective posterior sample. Our findings emphasize the importance of carefully preconditioning latent variables and demonstrate the significant advantage of (unadjusted) MCLMC for scaling to $\geq 10^6$ -dimensional problems. We find that MCLMC outperforms adjusted samplers by over an order-of-magnitude, with a mild scaling with the dimension of our inference problem. This benchmark, along with the associated publicly available code, is intended to serve as a basis for the evaluation of future field-level sampling strategies. The code is openly available at <https://github.com/hsimonfroy/benchmark-field-level>.

KEYWORDS: cosmological parameters from LSS, cosmological simulations, Statistical sampling techniques, redshift surveys

ARXIV EPRINT: [2504.20130](https://arxiv.org/abs/2504.20130)

Contents

1	Introduction	1
2	Methodology	4
2.1	Model	4
2.2	MCMC samplers	6
2.3	Conditioning strategies	10
2.4	Metrics	12
3	Numerical Results	13
3.1	Initialization	13
3.2	MCMC samplers	14
3.3	Conditioning strategies	17
3.4	Scaling effects	18
4	Conclusion	20
A	Additional comparisons	21
A.1	Non-trivial selection	21
A.2	Scaling of adjusted and unadjusted samplers	22
B	Hamiltonian formulation of the isokinetic Langevin dynamics	23
C	Handling bias in unadjusted Langevin dynamics	24

1 Introduction

In the current cosmological model the evolution of the Universe is governed by a small set of cosmological parameters. Inflation gives rise to an initial field of primordial fluctuations, which evolve under gravity to form the large scale structure (LSS) of the Universe. Galaxy surveys such as DESI [1], Euclid [2], and Vera Rubin’s Legacy Survey of Space and Time [3] serve as powerful probes of LSS by mapping the spatial distribution of galaxies, which traces the underlying matter density field and provides a direct observational signature of the evolved primordial fluctuations.

A common approach to extracting information from galaxy surveys is to compress the observable galaxy density field δ_g into a summary statistics $s(\delta_g)$. As illustrated in figure 1 (top panel), a marginal likelihood $s | \Omega, b$, is then constructed to relate cosmological parameters Ω and a reduced set of nuisance parameters b to the summary statistics s , effectively marginalizing over the stochasticity of the fields δ_L and δ_g . This marginalization can be performed with a Gaussian model, e.g. with analytical or simulation-based mean and covariance [4], or with more general techniques of Neural Density Estimation, e.g. normalizing flows, also fitted on simulations [5]. Finally, the posterior distribution $\Omega, b | s$ is inferred,

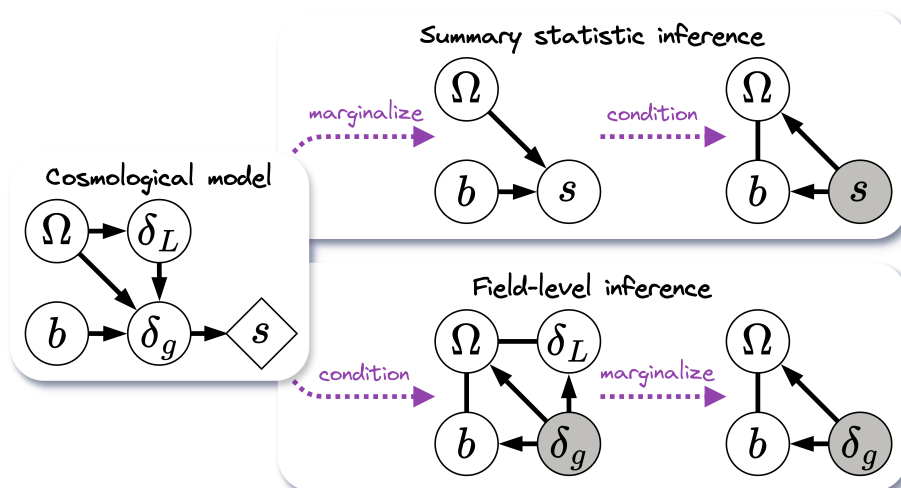


Figure 1. Two approaches to inferring cosmology Ω from the galaxy density field δ_g , that depends on Ω , initial field δ_L , and nuisance parameters b via a cosmological model (left panel). Top panel: inference based on summary statistics $s(\delta_g)$ marginalizes over δ_L and δ_g (e.g. by fitting a Gaussian model or a Neural Density Estimator on simulations) to obtain marginal model $p(\Omega) p(b) p(s | \Omega, b)$, then conditions on s (e.g. by MCMC sampling) to obtain final posterior $p(\Omega, b | s)$. Bottom panel: field-level inference conditions on δ_g (e.g. by MCMC sampling) to obtain full posterior $p(\Omega, b, \delta_L | \delta_g)$, then marginalizes over δ_L (by discarding δ_L samples) to obtain final posterior $p(\Omega, b | \delta_g)$. In this Bayesian network formalism, white (resp. grey, diamond) nodes denote latent (resp. observed, deterministic) variables, and directed (resp. undirected) edges denote computed (resp. uncomputed) dependencies.

often using Markov Chain Monte Carlo (MCMC) sampling. The choice of summary statistics is guided by a trade-off between analytical or variational tractability of its likelihood, and the amount of cosmological information retained.

A well-motivated choice for $s(\delta_g)$ is the power spectrum P of δ_g . On large scales, where the field remains approximately Gaussian, the power spectrum is close to being sufficient statistics, i.e. $\Omega | P \simeq \Omega | \delta_g$. Additionally, the power spectrum likelihood is relatively tractable, as it can often be well approximated by a Gaussian distribution with an analytical mean. Yet, gravity-driven structure formation is non-linear, making the galaxy density field non-Gaussian. In this case, the power spectrum is not sufficient, hence the development of analyses relying on alternative statistics [6], such as higher order statistics (3PCF [7], bispectrum [8]), object correlations (peaks [9], voids [10, 11], density splits [12, 13],...), multiscale counts (wavelet scattering transform [14], 1D-PDFs [15], Minkowski functionals [16, 17], k-nearest neighbors [18], Betti numbers [19]...), or automatically-learned statistics (CNN compression [5] or GNN compression [20]). Nonetheless, striking a balance between the interpretability of the compressed statistic, the tractability of its likelihood, and its ability to probe the cosmological signal exhaustively, remains a fundamental challenge.

An alternative approach, known as field-level inference, bypasses the need for summary statistics by directly conditioning on the observation δ_g . It proceeds so by jointly sampling the initial linear field δ_L with cosmology Ω and nuisance parameters b , as illustrated in figure 1 (bottom panel). The marginal posterior $\Omega, b | \delta_g$ is then readily obtained by discarding δ_L

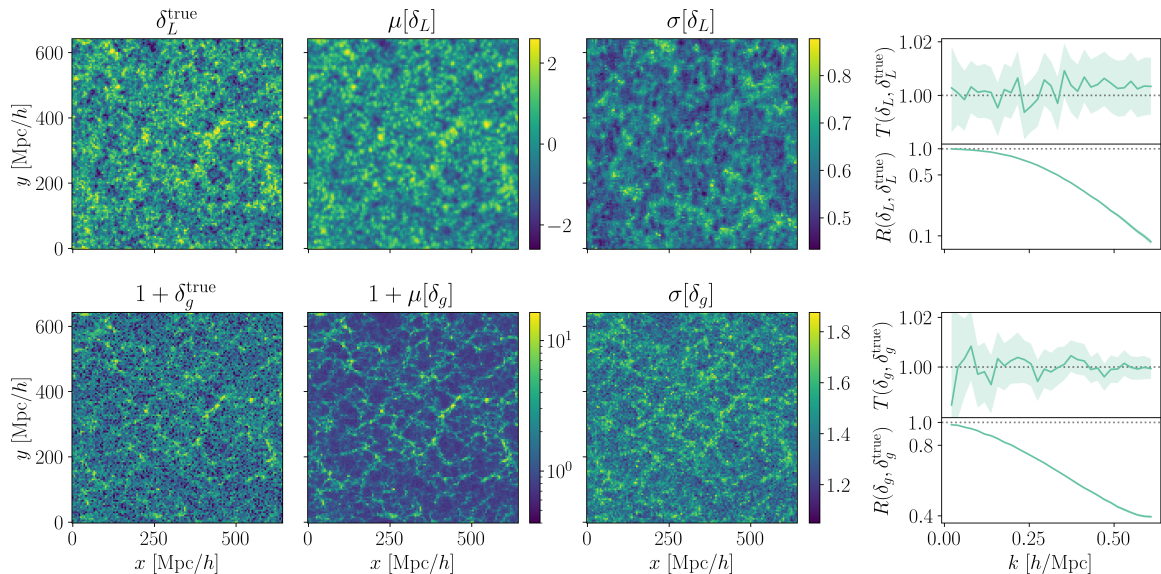


Figure 2. Posterior of the initial linear matter field δ_L (top row) and the observable galaxy density field δ_g (bottom row), obtained by field-level inference from the observed galaxy density field δ_g^{true} . Panels from left to right display respectively the voxel true value, posterior mean, and posterior standard deviations, for a thin slice of these fields. Rightmost panels show the posterior median and 95%-credible intervals of (top) the transfer function $T(\delta_i, \delta_j) = \sqrt{P_{ii}/P_{jj}}$ and (bottom) the spectra correlation $R(\delta_i, \delta_j) = P_{ij}/\sqrt{P_{ii}P_{jj}}$ (with $P_{ij}(k) = \langle \delta_i(k)\delta_j(k)^* \rangle$) between the sampled field and the true field associated with the observation. Both the linear matter and galaxy density power spectra are well recovered, while the fields themselves are well reconstructed at large scale (spectra correlation close to 1).

samples. These samples also enable a posterior reconstruction of the Universe’s history at the field level, as illustrated by figure 2 or by [21]. By construction, this approach extracts all the information encoded in the (observed and discretized) galaxy density field and is therefore quasi lossless. However, applying it to the volume of a large-scale survey like DESI ($V_{\text{eff}} \approx 20 \text{ (Gpc/h)}^3$), discretized at a typical non-linear scale of 5 Mpc/h , results in inferring a latent field with $\dim(\delta_L) \approx 10^9$ parameters. Field-level inference thus shifts the challenge from performing analytical or variational marginalization to solving a high-dimensional sampling problem.

To tackle this, state-of-the-art sampling methods, such as the Hamiltonian Monte Carlo (HMC) algorithm [22, 23], leverage the gradient information to efficiently explore high-dimensional posteriors. In practice, the requirement for the gradient is a mild constraint on the model implementation, as modern computational frameworks like JAX¹ [24] enable automatic differentiation, eliminating the need for manual derivations. HMC is used in most field-level inference applications [21, 25–30]. The No U-Turn Sampler (NUTS) [31], a variation of the standard HMC algorithm with automatically tuned trajectory length, has been used by [32, 33]. [34] recently proposed a different sampler dubbed MicroCanonical Langevin Monte Carlo (MCLMC) to perform cosmological field-level inference, showing it can

¹<https://github.com/jax-ml/jax>

require over an order-of-magnitude less model evaluations than HMC. Moreover, the choice of reparametrization, i.e. model preconditioning strategies, generally affects the efficiency of these sampling methods. A comprehensive comparison of these strategies is required to identify the most efficient and robust approaches, to assess their scalability to next-generation galaxy surveys, and to guide future algorithmic developments.

In this work, we conduct a comprehensive benchmark of MCMC samplers and preconditionings, focusing on field-level modeling and inference for galaxy clustering data. Our goal is to identify strategies that enable reliable posterior sampling while minimizing the number of computationally expensive simulations. Additionally, we investigate how modeling choices, such as resolution, volume, and complexity in large-scale structure formation modeling impact the high-dimensional sampling process.

The structure of the paper is as follows. In section 2 we present the model, samplers, conditioning strategies and metrics for the benchmark. Comparison of the samplers, conditioning strategies, and impact of modeling complexity are presented and discussed in section 3. We conclude in section 4.

2 Methodology

2.1 Model

To benchmark various sampling strategies under the most relevant conditions, we implement a forward model that encapsulates the key physical processes that could impact high-dimensional field-level sampling. These include the initial matter field generation, its non-linear gravitational evolution, the galaxy-matter connection, redshift-space distortions (RSD), and observational noise. The analysis is performed in a self-specified context, where the observations are generated from the same model used for inference. This idealized setting allows us to isolate and study the performance of sampling algorithms without the confounding effects of model misspecification [35, 36]. Questions related to the convergence of the density and velocity fields at the probed scales, as well as survey-specific modeling choices, such as observational systematics, are deferred to future work. By focusing on the fundamental components of galaxy clustering models while remaining agnostic to specific survey characteristics, this modeling is designed to be broadly informative for upcoming applications of field-level inference to galaxy surveys.

In practice, we make use of **JAX**, a **Python** library providing GPU acceleration, Just-In-Time (JIT) compilation, automatic vectorization/parallelization, and differentiation. We implement our model via **NumPyro**² [37], a Probabilistic Programming Language (PPL) powered by **JAX**, which facilitates model definition, differentiation, and sampling. Our hierarchical model is thus highly modular, and contains the following stages:

1. Initial matter field: cosmological parameters Ω include the matter density Ω_m , and the amplitude of linear fluctuations (smoothed at a scale of $8 \text{ Mpc}/h$) σ_8 . The linear matter field δ_L is then sampled as a Gaussian field, with power spectrum determined by these parameters via the fitting formula of [38] implemented in **jax-cosmo**³ [39].

²<https://github.com/pyro-ppl/numpyro>

³https://github.com/DifferentiableUniverseInitiative/jax_cosmo

2. LSS formation: matter particles are initialized on a regular grid and displaced following first/second order Lagrangian Perturbation Theory (1LPT/2LPT), or through a Particle Mesh (PM) N-body scheme. We rely on `JaxPM`,⁴ which is a JAX version of `FlowPM` [40]. To compute N-body displacement in particular, we use 5 steps of the BullFrog solver, recently proposed by [41], and implemented within `JaxPM` framework. In addition to faster convergence, BullFrog does not require prior LPT initialization, eliminating the sensitivity to the starting redshift of other PM solvers, like FastPM [42].
3. Redshift-Space Distortions: particles are advected once more according to their velocities $\dot{\mathbf{q}}$ projected along line-of-sight to account for contribution of galaxy peculiar velocity to redshift measurement. The RSD displacement depends on the Hubble function H , evaluated at the simulation redshift z , and is given by

$$\Delta \mathbf{q} = H(z)^{-1}(\dot{\mathbf{q}} \cdot \hat{\boldsymbol{\eta}}) \hat{\boldsymbol{\eta}} \quad (2.1)$$

where line-of-sight $\hat{\boldsymbol{\eta}}$ is fixed and taken to be one axis of the grid.

4. Galaxy bias: the galaxy density field is linked to the matter field via a second order Lagrangian bias expansion [43, 44]. Formally, a particle at position \mathbf{q} and observed at redshift z is assigned the weight

$$w_g(\mathbf{q}, z) := (1 + b_1 \delta_L + b_2(\delta_L^2 - \langle \delta_L^2 \rangle) + b_{s^2}(s^2 - \langle s^2 \rangle) + b_{\nabla^2} \nabla^2 \delta_L)(\mathbf{q}, z) \quad (2.2)$$

where $\delta_L(\mathbf{q}, z) = D(z)\delta_L(\mathbf{q})$ with D the linear growth factor and $s^2 = s^{ij}s_{ij}$ with $s_{ij} = [\partial_i \partial_j \nabla^{-2} - \delta_{ij}^K/3]\delta_L$ the tidal tensor.

5. Observational noise: the biased field is then converted into the observable galaxy density field δ_g , by adding Gaussian noise with variance \bar{N}_g^{-1} to each voxel, where \bar{N}_g is the average number of galaxies per voxel.

In total, the latent parameters include 2 cosmological parameters $\Omega := (\Omega_m, \sigma_8)$; 4 bias parameters from the second order Lagrangian bias expansion $b := (b_1, b_2, b_{s^2}, b_{\nabla^2})$; and the linear matter field δ_L , whose dimensionality depends on discretization of the simulated volume. These latent parameters are preconditioned to improve sampling efficiency, which means that the base variables (Ω, b, δ_L) are reparametrized into sample variables $(\tilde{\Omega}, \tilde{b}, \tilde{\delta}_L)$ whose posteriors are closer to a standard Gaussian. We refer to section 2.3 for details on the different conditioning strategies. The true values of Ω and b used in experiments (to generate synthetic data), their priors, as well as the fiducial values taken to compute the initialization and the static preconditioning (see section 2.3) are referenced in table 1. The prior ranges for bias parameters are typical of values measured from N-body simulations populated with Halo Occupation Distribution scheme [45] and we ensure in the following that computed posteriors are likelihood-dominated.

As a baseline in this work, we simulate a cosmological volume of $V_{\text{sim}} = (320 \text{ Mpc}/h)^3$ discretized into a 64^3 mesh, resulting in a voxel resolution of $5 \text{ Mpc}/h$. Also, $\bar{N}_g := \bar{n}_g V_{\text{cell}}$ is fixed assuming a DESI-like mean galaxy density of $\bar{n}_g = 10^{-3} (\text{Mpc}/h)^{-3}$.

⁴<https://github.com/DifferentiableUniverseInitiative/JaxPM>

Name	Symbol	Truth	Prior	Fiducial
Matter density	Ω_m	0.3	$\mathcal{N}(0.3111, 0.5^2, [0.05, 1])$	0.3111
Amplitude of linear fluctuations	σ_8	0.8	$\mathcal{N}(0.8102, 0.5^2, [0, +\infty])$	0.8102
	b_1	1	$\mathcal{N}(1, 0.5^2)$	1
Lagrangian galaxy bias	b_2	0	$\mathcal{N}(0, 2^2)$	0
	b_{s^2}	0	$\mathcal{N}(0, 2^2)$	0
	b_{∇^2}	0	$\mathcal{N}(0, 2^2)$	0
Linear matter power spectrum	P_L	Given by cosmology (Ω_m, σ_8)		
Initial linear matter field	δ_L	$\mathcal{N}(\mathbf{0}, P_L V_{\text{cell}}^{-1})$		
Observable galaxy density field	δ_g	Evolved from initial field δ_L		

Table 1. Summary of model cosmological and bias parameters, and fields, with their true values used to generate the observations, their priors, and their fiducial values used for inference initialization. $\mathcal{N}(\boldsymbol{\mu}, C, A)$ denotes a Gaussian variable of mean $\boldsymbol{\mu}$ and covariance C , truncated to the set A .

2.2 MCMC samplers

To perform the high-dimensional sampling required in field-level inference, we cannot afford to employ a classical Metropolis-Hastings (MH) algorithm. This algorithm samples a target density via a simple Markov chain where each step is agnostic to local information about the density. This leads to two major issues: (i) as the dimensionality increases, more directions lead away from high-density regions, so that the step size must scale as $\mathcal{O}(d^{-1/2})$ to maintain a given acceptance rate; (ii) once the dynamics has reached high-density regions, motion becomes locally Brownian (a random walk), meaning n^2 steps of size ε are required to move by $\mathcal{O}(n\varepsilon)$, which is an inherently slow way to explore the sampling space.

Addressing these issues requires integrating local information about the target density into the dynamics. Moreover, significant advances in numerical sampling and optimization have been achieved by exploiting their connection to statistical physics: many randomized iterative methods exhibit particle-like dynamics, which can be analyzed statistically and leveraged to inform new developments. The following subsections provide an overview of canonical and microcanonical Langevin MCMC methods to introduce the different samplers that we consider in our benchmark.

2.2.1 Canonical sampling with Langevin dynamics

Our goal is to sample from the d -variate target density $p \propto e^{-U}$. In the potential U evolves a particle of position \mathbf{q} , momentum \mathbf{p} , and mass (matrix) M . The Hamiltonian of this particle is then given by:

$$\mathcal{H}(\mathbf{q}, \mathbf{p}) = U(\mathbf{q}) + \frac{1}{2} \mathbf{p}^\top M^{-1} \mathbf{p} \quad (2.3)$$

As a result, its motion follows Hamilton equations

$$\begin{cases} \dot{\mathbf{q}} = \partial_{\mathbf{p}} \mathcal{H} = M^{-1} \mathbf{p} \\ \dot{\mathbf{p}} = -\partial_{\mathbf{q}} \mathcal{H} = -\nabla U(\mathbf{q}) \end{cases} \quad (2.4)$$

which constrain the particle to the constant-energy surface $\{\mathbf{q}, \mathbf{p} \mid \mathcal{H}(\mathbf{q}, \mathbf{p}) = E\}$, where $E := \mathcal{H}(\mathbf{q}_0, \mathbf{p}_0)$ denotes the initial particle energy. Confined to this surface, the particle cannot sample from a general target density p without energy exchange.

Energy exchange can be provided by the particle's surrounding medium, a heat bath, where it collides with other particles at a frequency γ . The resulting dynamics in the continuous limit of interaction is described by the Langevin equations

$$\begin{cases} \dot{\mathbf{q}} = M^{-1}\mathbf{p} \\ \dot{\mathbf{p}} = -\nabla U(\mathbf{q}) - \gamma\mathbf{p} + (2\gamma M)^{1/2}\dot{\mathbf{w}} \end{cases} \quad (2.5)$$

where \mathbf{w} is a Wiener process (so $\dot{\mathbf{w}}$ is a Gaussian white noise). These equations add to the Hamiltonian dynamics a dissipation term $-\gamma\mathbf{p}$ and a fluctuation term $(2\gamma M)^{1/2}\dot{\mathbf{w}}$, which account respectively for the particle dissipating coherent kinetic energy as heat and gaining dispersed kinetic energy from thermal fluctuations.

Energy exchange allows the particle system to reach thermal equilibrium, where the distribution of particle states is given by the canonical ensemble $p_C(\mathbf{q}, \mathbf{p}) \propto e^{-\mathcal{H}(\mathbf{q}, \mathbf{p})}$. In particular and as desired, the particle position \mathbf{q} is distributed as the target p :

$$\int p_C(\mathbf{q}, \mathbf{p}) d\mathbf{p} = \int p(\mathbf{q}) \mathcal{N}(\mathbf{p} \mid \mathbf{0}, M) d\mathbf{p} = p(\mathbf{q}). \quad (2.6)$$

2.2.2 Microcanonical sampling with isokinetic Langevin dynamics

A limitation of introducing energy exchange into the dynamics is that the target distribution is only recovered by averaging particle trajectories across multiple energy levels, potentially slowing convergence. Hopefully, energy exchange is not necessary if, instead of a classical particle, we consider an isokinetic particle, i.e. a particle of constant kinetic energy $\dot{\mathbf{q}}^\top M \dot{\mathbf{q}}/2 = |\mathbf{u}|^2/2 = 1/2$, where $\mathbf{u} := M^{1/2}\dot{\mathbf{q}}$ denotes an auxiliary momentum.

Physically, this isokineticity can be enforced by coupling the particle to a Gaussian IsoKinetic (GIK) thermostat [46], which constantly extracts or provides energy to the particle so that its temperature, as a kinetic energy, is a constant of motion (see appendix B). When placed in the rescaled potential $U/(d-1)$ (recalling $d \geq 2$ is dimension), such particle admits the GIK Hamiltonian⁵

$$\mathcal{H}(\mathbf{q}, \mathbf{p}) = \frac{\mathbf{p}^\top M^{-1}\mathbf{p}}{2m(\mathbf{q})} - \frac{m(\mathbf{q})}{2} \quad (2.7)$$

where $m := e^{-U/(d-1)}$ and $\mathbf{p} := mM^{1/2}\mathbf{u} = mM\dot{\mathbf{q}}$ is the canonical momentum. Remarkably, this Hamiltonian's microcanonical ensemble $p_{MC}(\mathbf{q}, \mathbf{p}) \propto \delta(\mathcal{H}(\mathbf{q}, \mathbf{p}))$ coincides exactly with its isokinetic ensemble $p_{IK}(\mathbf{q}, \mathbf{u}) \propto e^{-U(\mathbf{q})}\delta(|\mathbf{u}|^2 - 1)$. This means that under ergodicity, the particle is sampling these ensembles, and therefore has position \mathbf{q} distributed as target p :

$$\int p_{MC}(\mathbf{q}, \mathbf{p}) d\mathbf{p} = \int p_{IK}(\mathbf{q}, \mathbf{u}) d\mathbf{u} \propto \int p(\mathbf{q}) \delta(|\mathbf{u}|^2 - 1) d\mathbf{u} \propto p(\mathbf{q}). \quad (2.8)$$

⁵In contrast to [47–49] we adopt the GIK Hamiltonian formulation that does not require any momentum-dependent time rescaling. This preserves p as a stationary density of the dynamics, and therefore reconciles the microcanonical formulation from [47, 48] and the isokinetic formulation from [49]. In appendix B, we show that we can rederive from this Hamiltonian the MCLMC equations of motion.

However, merely following the induced Hamiltonian dynamics does not ensure the ergodicity required for sampling the microcanonical/isokinetic ensemble. In fact, if the system exhibits additional symmetries beyond time translation (which already enforces energy conservation), Noether’s theorem guarantees the opposite: the particle remains confined to a subset of the constant-energy surface defined by an additional conserved quantity.

To ensure ergodicity, [48] reintroduces the particle into a heat bath, where it collides with other particles at a frequency γ . However, due to the GIK thermostat, these interactions always conserve kinetic energy. For numerical efficiency, it is preferred to compute the dynamics for state (\mathbf{q}, \mathbf{u}) , rather than (\mathbf{q}, \mathbf{p}) , which follows in the continuous limit of interaction the isokinetic Langevin equations [49] (in Stratonovich form)

$$\begin{cases} \dot{\mathbf{q}} = M^{-1/2}\mathbf{u} \\ \dot{\mathbf{u}} = P(\mathbf{u})(-M^{-1/2}\nabla U(\mathbf{q})/(d-1) + (2\gamma)^{1/2}\dot{\mathbf{w}}) \end{cases} \quad (2.9)$$

where $P(\mathbf{u}) := I - \mathbf{u}\mathbf{u}^\top$ is the orthogonal projection onto $\text{span}(\mathbf{u})^\perp$. Therefore thermal fluctuations only impact the momentum direction \mathbf{u} while both Hamiltonian \mathcal{H} and kinetic energy $|\mathbf{u}|^2/2$ are conserved.

2.2.3 Derived samplers

The Langevin dynamics modeled by equations (2.5) and (2.9) are the basis of several sampling algorithms, involving different regimes of interaction with medium γ and inertia M . In practice, the dynamics are solved numerically, often using symplectic integrators like KDK-Verlet/Leapfrog or McLachlan/Minimum-Norm integrator [50]. Numerical integration comes with discretization errors that can induce a bias, which can be either eliminated via MH adjustment (accept-reject step), or controlled via the integration step size.

In this benchmark, we consider three canonical and adjusted samplers, and two microcanonical samplers. Our selection is guided by both the existing literature on field-level inference and the availability of the implementations. It is designed to highlight key methodological distinctions, particularly between canonical vs. microcanonical, and adjusted vs. unadjusted methods.

HMC. The celebrated Hamiltonian Monte Carlo algorithm [23] simulates a classical particle evolving according to equations (2.5) in a sparse medium ($\gamma = 0$) for a tunable trajectory time length L , before colliding with a heavy particle, which refreshes its momentum: $\mathbf{p} \sim \mathcal{N}(\mathbf{0}, M)$. This arrival state is then used as a proposal for a MH adjustment step, after which the dynamics repeats.

NUTS. No-U-Turn Sampler [31] is a variant of HMC that automatically and dynamically adapts the trajectory time length L . It constructs sets of proposals along the Hamiltonian trajectories, which are stopped whenever they start to reverse direction, thereby avoiding redundant exploration and the need for manual tuning of L .

NUTSwG. Any MCMC sampler can be implemented within a Gibbs sampling scheme [51], where MCMC sampling alternates over subsets of parameters. For instance, assume one

wants to sample from three blocks of variables q_1, q_2, q_3 . Then a MCMC within Gibbs scheme consists in sampling alternatively with the given MCMC from $q_1 | q_2, q_3$, then $q_2 | q_1, q_3$, then $q_3 | q_1, q_2$, using the sample of one block to condition the others. It can be seen as a sampling generalization of alternating gradient descent. We consider a NUTS within Gibbs implementation, referred to as NUTSwG in the following.

MAMS. Metropolis-Adjusted Microcanonical Sampler [52] is a microcanonical counterpart to canonical HMC. It simulates an isokinetic particle evolving according to equations (2.9) in a sparse medium ($\gamma = 0$) for a tunable average trajectory time length L , before colliding with a heavy particle, which refreshes its momentum: $\mathbf{u} \leftarrow \mathbf{z}/|\mathbf{z}|$ where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, I)$. This arrival state is then used as a proposal for a MH adjustment step, after which the dynamics repeats.

MCLMC. MicroCanonical Langevin Monte Carlo [48] simulates an isokinetic particle evolving according to equations (2.9) in a dense medium. There, the particle undergoes continuous collisions with lightweight particles, causing its momentum to be partially refreshed at each integration step: $\mathbf{u} \leftarrow (\mathbf{u} + \sqrt{2\gamma}\mathbf{z})/|\mathbf{u} + \sqrt{2\gamma}\mathbf{z}|$, where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, I)$,⁶ and the tunable momentum decoherence length L is introduced such that $\gamma := (e^{2\varepsilon/L} - 1)/(2d)$. Crucially, MCLMC is run without MH adjustment step, meaning that every integration state is directly accepted as a sample. Any potential bias introduced by this unadjusted scheme is controlled by the integration step size ε , as discussed below.

On the one hand, adjusted algorithms are asymptotically unbiased, but at the cost of significantly reducing the integration step size as dimension increases in order to maintain a reasonable acceptance rate. Formally, [23] show that for a target product density, the number of model evaluations needed to reach a new effective sample grows as d for the MH algorithm, as $d^{1/3}$ for adjusted overdamped ($M \ll \gamma$) Langevin dynamics, and as $d^{1/4}$ for adjusted underdamped ($\gamma \ll M$) Langevin dynamics.

On the other hand, unadjusted algorithms have no unbiased guarantees, but the relation between step size and bias has been theoretically studied for both canonical and microcanonical samplers [54, 55]. In particular, [55] show, both formally for a target Gaussian and empirically, that the bias can be controlled via $\langle \Delta E^2 \rangle / d$, the Energy Error Variance Per Dimension (EEVPD) across integration steps, which itself depends on the step size in a dimension-independent manner. Therefore, by tuning the step size to keep the EEVPD below some threshold, one ensures that the total sampling error (in terms of Wasserstein distance or Mean Square Error) is dominated by the Monte Carlo error. Since the Effectively, this EEVPD threshold governs the trade-off between the sampling efficiency (the number of model evaluations needed to reach a new effective sample) and the bias, independently of the dimension. This is a really promising result compared to the scaling of adjusted methods. Admittedly, the sampled posterior is rarely an isotropic Gaussian in practice, so we may expect some dependency of the sampling efficiency to dimension. However, as explained in section 2.3, we can reparameterize the model such that the posterior approaches an isotropic Gaussian.

⁶This refreshment step replaces the less direct-to-sample Brownian motion on the hypersphere \mathbb{S}^{d-1} [53], corresponding to the diffusion term $\sqrt{2\gamma}P(u)\dot{W}$ in equations (2.9). Both preserve p as a stationary density and provide the desired ergodicity [49], and they even closely agree when $\varepsilon \ll L$.

2.3 Conditioning strategies

For any sampling problem, one is free to choose any bijective reparametrization of the base variables. Formally, sampling from a target density in a variable space \mathcal{X} is analytically equivalent to sampling from the image density in the transformed space $\tilde{\mathcal{X}}$, and then transforming back the samples to \mathcal{X} . However, these are not algorithmically equivalent. The density in \mathcal{X} can be challenging to sample (due to degeneracies, discontinuities, boundaries, etc.), but a transformed space $\tilde{\mathcal{X}}$ can often be defined where the density is smoother and easier to sample. This section focuses on preconditioning the latent space so that the posterior of latent variables approximates an isotropic Gaussian.

2.3.1 Model preconditioning

With full knowledge of the posterior, finding a suitable reparameterization would be simplified. But in practice, when sampling is precisely the way we access posterior information, we can only rely on partial posterior knowledge. A simple preconditioning is prior-conditioning, where (*a priori*) sample variables are made isotropic Gaussian. This is a relevant choice of conditioning if the likelihood constraining power is weak or aligned with the prior, as the resulting prior-conditioned posterior would also be approximately isotropic Gaussian. However, if the likelihood constraining power is strong and varies significantly across variables relatively to their priors, prior-conditioning can exacerbate degeneracies in the posterior [56].

In the context of field-level inference, model preconditioning is exemplified by the choice of parametrization for the linear field δ_L . The linear field can easily be prior-conditioned in real space by considering a Gaussian white noise, which is Fourier-transformed and multiplied by the square root of the power spectrum. However, the likelihood typically induces correlations between voxels, leading to a posterior that is ill-conditioned. Alternatively, the linear field can also easily be prior-conditioned in Fourier space with a complex Hermitian Gaussian noise, multiplied by the square root of the power spectrum. As a result of the large-scale isometry-invariance and linear growth of δ_L , the posterior wavevector amplitudes are approximately independent and Gaussian. However, since different scales are constrained with varying power, the posterior can still remain poorly conditioned.

In order to make the δ_L posterior more isotropic, we approximate it assuming a linear model for the observable galaxy density field, as described hereafter. In Fourier space, for any wavevector \mathbf{k} , the (discretized) linear matter field $\delta_L(\mathbf{k}) | \Omega$ is distributed as a complex-Gaussian $\mathcal{CN}(0, P_L(\mathbf{k})V_{\text{cell}}^{-1})$, recalling that $\mathcal{CN}(\boldsymbol{\mu}, C) = \mathcal{N}(\text{Re}(\boldsymbol{\mu}), C/2) + i\mathcal{N}(\text{Im}(\boldsymbol{\mu}), C/2)$ for any complex vector $\boldsymbol{\mu}$ and covariance matrix C . In the flat-sky approximation, this field is then evolved linearly by a factor $B_K := (b_E + (\hat{\mathbf{k}} \cdot \hat{\boldsymbol{\eta}})^2 f)D$ known as the Kaiser boost [57], where D and f are the linear growth factor and growth rate at the observed redshift and cosmology, $b_E := 1 + b_1$ is the linear Eulerian bias associated to the linear Lagrangian bias b_1 , and $\hat{\mathbf{k}} \cdot \hat{\boldsymbol{\eta}}$ is the cosine of the angle between wavevector \mathbf{k} and line-of-sight $\hat{\boldsymbol{\eta}}$. Finally, as in the reference forward model, we include observational noise as a Gaussian distribution of variance \bar{N}_g^{-1} , the mean number of galaxies per voxel. Put together and applying Bayes

formula for this Gaussian linear model, we have

$$\underbrace{\mathcal{CN}(\delta_g | B_K \delta_L, \bar{N}_g^{-1})}_{\text{likelihood}} \underbrace{\mathcal{CN}(\delta_L | 0, P_L V_{\text{cell}}^{-1})}_{\text{prior}} = \underbrace{\mathcal{CN}(\delta_L | \mu, \sigma^2)}_{\text{posterior}} \underbrace{\mathcal{CN}(\delta_g | 0, \bar{N}_g^{-1} + B_K^2 P_L V_{\text{cell}}^{-1})}_{\text{evidence}} \quad (2.10)$$

with

$$\begin{aligned} \sigma &:= (\bar{N}_g B_K^2 + P_L^{-1} V_{\text{cell}})^{-1/2} \\ \mu &:= \sigma^2 \bar{N}_g B_K \delta_g \end{aligned}$$

meaning that we *a posteriori* have $\delta_L | \delta_g, \Omega, b \sim \mathcal{CN}(\mu, \sigma^2)$. Hence we define the preconditioned linear field:

$$\tilde{\delta}_L := (\delta_L - \mu)/\sigma \sim \mathcal{CN}(-\mu/\sigma, P_L V_{\text{cell}}^{-1}/\sigma^2) \quad (2.11)$$

such that its estimated posterior is as desired $\tilde{\delta}_L | \delta_g, \Omega, b \sim \mathcal{CN}(0, I)$. This reparametrization obviously depends on cosmology and bias parameters, which we specify following these two options (that we compare in section 3.3):

1. Static Kaiser conditioning: we assume fixed fiducial values $\Omega^{\text{fid}}, b_1^{\text{fid}}$, as given in table 1.
2. Dynamic Kaiser conditioning: we use the values of Ω, b_1 of the current sample, thereby adapting the conditioning to the local shape of the posterior.⁷

Cosmological and bias parameters Ω, b are also transformed into $\tilde{\Omega}, \tilde{b}$ based on a rough estimate of their posterior standard deviation $\sigma[\Omega, b | \delta_g]$, such that $\tilde{\Omega}, \tilde{b} | \delta_g$ is approximately a standard Gaussian. To estimate $\sigma[\Omega, b | \delta_g]$, we observe that it scales as the inverse square root of the galaxy sample volume and extrapolate the constraints from a fast (low-dimensional) inference performed within a smaller volume.

The preconditioning strategy described above leverages approximate knowledge of the posterior to enhance sampling efficiency. However, at smaller scales, the Kaiser model approximation no longer holds. In addition, when considering observational data, the spatially-varying survey selection function will introduce correlations between wavevector amplitudes. We thus provide in appendix A.1 an approximate generalization of the preconditioning for curved-sky and non-trivial selection, as well as a comparison of sampling performance against the periodic box case. More generally, to further improve the model conditioning where analytic prescriptions break down, we can rely on an adaptive mass matrix.

2.3.2 Mass matrix conditioning

The MCMC samplers we introduced in section 2.2 all admit a mass matrix M as a tunable parameter. Formally, a particle of position \mathbf{q} and mass M will follow the same dynamics (up to a time and initial momentum rescaling) as a particle of position $\tilde{\mathbf{p}} := M^{1/2} \mathbf{q}$ and unit mass I , i.e. the mass matrix M is equivalent to a linear conditioning of the sample positions. Therefore, a good candidate for the mass matrix is the inverse covariance of the sample variables [59].

⁷This is equivalent to an analytical and spatially-varying Riemannian metric, as leveraged in Riemannian Manifold Hamiltonian Monte Carlo (RMHMC) [58].

The adopted strategy is to leverage prior knowledge to analytically precondition \mathbf{q} into $\tilde{\mathbf{q}}$ as described in section 2.3.1, then rely on adaptive methods to set $M = \tilde{\Sigma}^{-1}$, where $\tilde{\Sigma}$ is an online estimate of the covariance of $\tilde{\mathbf{q}}$. In practice, this estimation is performed during a warmup phase using Welford’s algorithm for numerical stability, and the mass matrix is kept diagonal for computational efficiency.

2.4 Metrics

The performance of an estimation procedure $\hat{\theta}$ can be quantified via its associated risk $\mathcal{R}[\hat{\theta}]$, such as the Mean Square Error (MSE). For a given number n of data samples, the efficiency $e_n := \mathcal{R}[\tilde{\theta}_n]/\mathcal{R}[\hat{\theta}_n]$ compares the risk of a particular estimator $\hat{\theta}_n$ to the risk of a reference estimator $\tilde{\theta}_n$. Alternatively, the number of samples N_{eff} needed for the reference risk to equate the particular risk, $\mathcal{R}[\tilde{\theta}_{N_{\text{eff}}}] = \mathcal{R}[\hat{\theta}_n]$, is commonly referred to as the Effective Sample Size (ESS).

In our case, we obtain from MCMC sampling a stationary and identically distributed, but possibly correlated, sequence of n samples, from which we are interested to estimate some target quantity θ (e.g. mean, median, covariance, quantile, credible region...). We can then compare the MSE based on the correlated sequence with the MSE that we would have if the samples were independent. The induced ESS denoted $N_{\text{eff}}[\theta]$ is then understood as “the number of independent samples that would yield the same statistical power as the MCMC samples for estimating the quantity θ ”. Typically, it is the sample size at which the (Markov Chain) Central Limit Theorem operates for this quantity, i.e. $\sqrt{N_{\text{eff}}[\theta]}(\hat{\theta}_n - \theta) \simeq \mathcal{N}(0, \sigma^2[\theta])$.

It is common and simpler to focus on the estimation of the mean μ , so in the following, the ESS denoted by N_{eff} refers to $N_{\text{eff}}[\mu]$ unless otherwise specified. Formally, for a stationary sequence of size n and autocorrelation ρ_t , the ESS is asymptotically

$$N_{\text{eff}} \simeq \frac{n}{\sum_{t=-\infty}^{+\infty} \rho_t} = \frac{n}{1 + 2 \sum_{t=1}^{+\infty} \rho_t} \quad (2.12)$$

and can be estimated based on sample autocorrelation. We see that in the stationary case, the more correlated the samples are, the smaller the ESS becomes. This aligns with the common (though not always true) intuition that highly correlated samples share redundant information, thus reducing their overall effectiveness.

The ESS is a relevant metric for quantifying the quality of a sample sequence, serving as a proxy for estimation error without requiring knowledge of the true mean value. On the other hand, if their true value is known, we can directly compute the error for any quantity of interest, typically the mean μ and standard deviation σ . Given n_c chains, we estimate their mean $\hat{\mu}_c$ and standard deviation $\hat{\sigma}_c$, and the (empirical and normalized) MSE is obtained as:

$$\text{MSE}[\theta] := \frac{1}{n_c} \sum_c \varepsilon_c^2[\theta] \quad \text{where in particular } \varepsilon_c[\theta] := \frac{\hat{\theta}_c - \theta}{\sigma \sqrt{n_c}} \times \begin{cases} 1 & \text{if } \theta \text{ is } \mu \\ \sqrt{2} & \text{if } \theta \text{ is } \sigma \end{cases}. \quad (2.13)$$

Normalization of errors ensures that for an asymptotically unbiased estimator $\text{MSE}[\theta] \simeq N_{\text{eff}}[\theta]^{-1} \chi^2(n_c)/n_c$, making it commensurable with the inverse ESS.

Given the high dimensionality of our problem, it is often necessary to aggregate metrics across sets of variables, such as cosmological, bias, and linear field parameters. The

multivariate generalization of MSE implies that we respectively take the arithmetic mean of empirical MSEs and the harmonic mean of ESSs.

Model evaluation (especially its gradient) is our main computational bottleneck for cosmological inference. We therefore characterize the efficiency of a MCMC sampler by the number of model evaluations required to yield one effective sample $N_{\text{eval}}/N_{\text{eff}}[\theta]$, and commensurably, to yield a unit error $N_{\text{eval}} \times \text{MSE}[\theta]$. These metrics are expected to be relatively stable over evaluations, so we check their convergence before reporting them.

Finally, unadjusted samplers may introduce an asymptotic bias in the sampled posterior (see appendix C), that is not directly quantified by the ESS or the MSE individually.⁸ However, since in the unbiased case, $\text{MSE}[\mu]$ is of the order of the inverse ESS, finding $\text{MSE}[\mu] \gg N_{\text{eff}}^{-1}$ would reveal a non-negligible bias.

3 Numerical Results

In this section we benchmark a comprehensive range of sampling methods and conditioning strategies according to the metrics specified in section 2.4 for the model presented in section 2.1. The task is to sample from the posterior $\Omega, b, \delta_L \mid \delta_g^{\text{true}}$ given an observation δ_g^{true} generated from the same model. In sections 3.2 and 3.3, the forward model performs 1LPT displacement in a $V_{\text{sim}} = (320 \text{ Mpc}/h)^3$ volume discretized into a 64^3 mesh (voxel resolution of $5 \text{ Mpc}/h$). In section 3.4, we vary these settings.

We run for each inference task 8 parallel chains, on $8 \times 80\text{GB}$ NVIDIA A100 GPU node. Our sampling pipeline is built upon JAX-powered packages NumPyro and BlackJAX⁹ [60].

3.1 Initialization

Simply drawing δ_L samples from the prior would produce an initial field with the fiducial power spectrum, but its amplitudes and phases would be distributed between the different wavevectors regardless of the observable field δ_g . Instead, we seed the chains with samples from the Kaiser posterior, given by equations (2.10) and fiducial parameters. This approximation results in an inaccurate initial linear power spectrum but aligns amplitudes and phases closer to their high-density regions. This is especially the case for the largest scales as reflected by the spectra correlation¹⁰ on figure 3. We observe that this initialization strategy indeed accelerates convergence.

We then perform a first warmup phase where exclusively δ_L is inferred while maintaining other (cosmological and bias) parameters fixed at their fiducial values. This allows the high-dimensional δ_L to converge towards its high-density region while preventing its dynamics to drag the other parameters away. This warmup phase is assumed successful when the sample power spectrum fluctuates around the fiducial one, as shown by the transfer function on figure 3. Note that if we had initialized the linear field from the start with the fiducial power spectrum, we would not be able to rapidly distinguish the chain convergence from the chain being stuck.

⁸A high MSE can be caused by both an inefficient sampling (low ESS) or a bias.

⁹<https://blackjax-devs.github.io/blackjax>

¹⁰Because spectra correlation is complex-valued, we only show its norm, though we may miss some potential phase shifts.

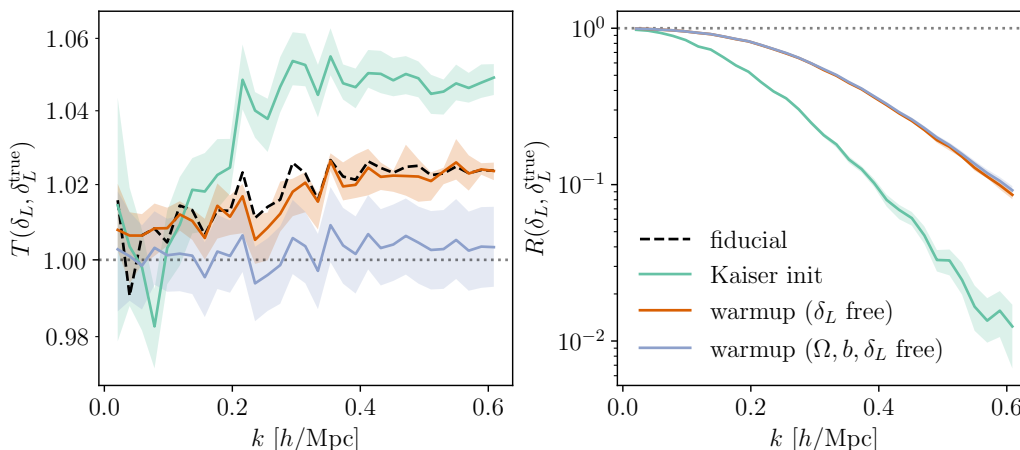


Figure 3. State of the chains at the initialization and at the end of the warmup phases, showed via the median and the 95%-credible region of the transfer function (left) and spectra correlation (right) between the sampled linear field and the true linear field (as defined in figure 2). The initialization based on the Kaiser posterior approximation (turquoise) provides a relevant reconstruction of the linear field δ_L at large scales (spectra correlation close to 1). From there, the subsequent warmup phases achieve convergence, first (orange) towards a fiducial power spectrum (assuming a fixed fiducial cosmology Ω^{fid} and bias parameters b^{fid}), then (indigo) towards the true power spectrum (jointly inferring Ω , b and δ_L).

Next, a second warmup phase is performed where all parameters are jointly sampled. This is also the phase where sampler hyperparameters are tuned, especially step size and mass matrix. This second warmup is assumed successful when the sampler hyperparameters of the different chains are in qualitative agreement.

Finally, the chains are run until reaching at least 500 ESS and a Gelman-Rubin statistic of at most 1.01 for each parameter (which is roughly equivalent, see [61]).

In total, about one fifth of the total number model evaluations is spent for warmup (to be conservative), and we do not include these evaluations in the benchmark results. We observe that a short warmup time is, as expected, correlated with a high sampling efficiency in the stationary regime.

3.2 MCMC samplers

We start by comparing the performance of the different samplers mentioned in section 2.2. For all samplers, we use the Kaiser dynamic preconditioning and tune a (diagonal) mass matrix. Their specific configurations are described hereafter:

HMC, NUTS, NUTSwG. We solve the equations of motion 2.4 with the KDK-Verlet integrator and tune the step size ε to target a 65% acceptance rate, theoretically optimal for isotropic Gaussian densities [23], which we corroborated in our tests.¹¹

¹¹The relevant range of the target acceptance rate for HMC-like algorithms is usually [0.6, 0.9], with higher values better suited to more complex densities [31, 62].

HMC. The trajectory time length L is based on the average trajectory length of NUTS, following the heuristic $L_{\text{HMC}} = \langle L_{\text{NUTS}} \rangle / 2$.¹²

NUTSwG. We split the sampled variables into two blocks: (i) the linear field δ_L and (ii) cosmological and bias parameters (Ω, b) , in a similar fashion to [26, 29].¹³ A NUTS within Gibbs scheme for these blocks means that we alternatively sample with NUTS from $\delta_L \mid \Omega, b, \delta_g$ then $\Omega, b \mid \delta_L, \delta_g$, using the sample of one block to condition the others. We also tested splitting the parameter space between (δ_L, Ω) and b , based on the idea that sampling $b \mid \delta_L, \Omega, \delta_g$ is computationally cheap since it only involves evaluating the bias model and does not require simulating structure formation. However, even under the optimistic counting of zero model evaluation for a recomputation of the bias model, this approach performed worse than the other sampling methods.

MAMS, MCLMC. We tested a variety of symplectic integrators¹⁴ to solve the equations of motion 2.9. This includes the widely used KDK-Verlet integrator (2nd-order, 1 model evaluation per step), McLachlan integrator (2nd-order, 2 evaluations; see eq. 4.2 in [50]), Blanes integrator (2nd-order, 3 evaluations; see eq. 44 in [64]), and Omelyan integrator (4th-order, 5 evaluations; see eq. 77 in [65]). The McLachlan and Omelyan schemes are known to achieve minimum norm within their respective orders [66], while the Blanes scheme offers best stability among 3-evaluation methods. In our tests, the McLachlan integrator consistently achieved the best performance, albeit by a modest margin, and is therefore adopted in the remainder of this work. As expected, more stable or higher-order integrators allowed for larger step sizes, but the gain was insufficient to offset their increased number of model evaluations. Additional results on the scaling of microcanonical samplers and the impact of integrators on sampling efficiency can be found in appendix A.2.

MAMS. The step size ε is tuned to target a 65% acceptance rate, theoretically optimal for isotropic Gaussian densities and corroborated in our tests.¹⁵ The average trajectory length is set to heuristic $L = 0.3 \frac{N_{\text{step}}}{N_{\text{eff}}[\mu]} \varepsilon$ (with $N_{\text{step}} = N_{\text{eval}}/2$ for McLachlan integrator) [52], where $N_{\text{eff}}[\mu]$ is a rough estimate of the (multivariate) ESS.

MCLMC. The step size ε is tuned to obtain an EEVPD below 10^{-6} , ensuring negligible bias in the posterior distribution. A detailed discussion on the followed strategy to control the potential bias of MCLMC is provided in appendix C. We set the momentum decoherence length according to a worst-case heuristic $L = 0.4 \frac{N_{\text{step}}}{\min_i N_{\text{eff}}[\mu_i]} \varepsilon$, where $N_{\text{eff}}[\mu_i]$ is a rough estimate of the ESS of the i -th sampled variable. Contrary to adjusted samplers that yield one sample after a certain number of integration steps L/ε , MCLMC yields one sample at each step. However, saving all these correlated samples is memory inefficient, so we

¹²Hamiltonian trajectory length is fixed here, although randomizing it could improve certain performance metrics in some cases [63].

¹³This choice can be partly motivated by a lack of differentiable computation of cosmological quantities. In the referenced works, each cosmological and bias parameter is itself treated in a separate block.

¹⁴The specific interest of symplectic integrators for solving equations (2.9) is not direct since it is not formulated in canonical phase space (\mathbf{q}, \mathbf{p}) .

¹⁵[52] report higher values working well in their tests, which may be better suited for more complex densities.

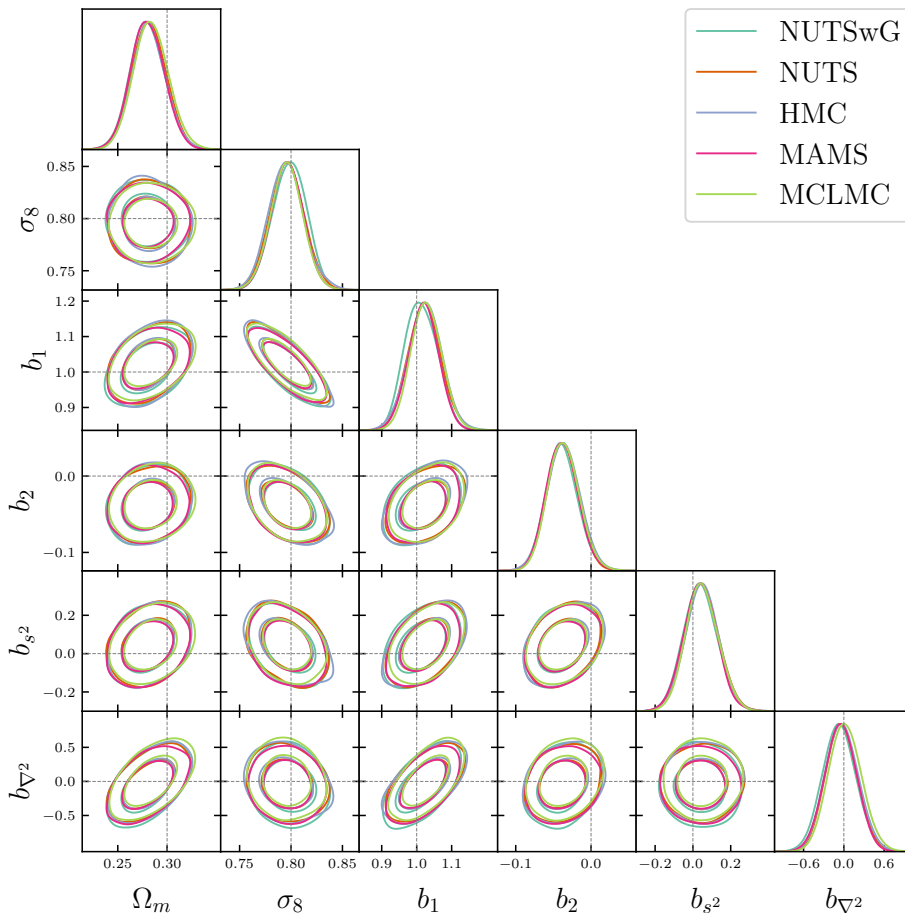


Figure 4. 68% and 95%-credible regions of the cosmological and bias parameters posterior $\Omega, b \mid \delta_g$, obtained with the five considered samplers (NUTSwG in turquoise, NUTS in orange, HMC in indigo, MAMS in magenta, MCLMC in lime). The dashed lines denote the parameter true values used to generate the (synthetic) observation.

apply chain thinning, retaining only every 16th step. We verify that this does not impact the performance metrics.

The posterior $\Omega, b \mid \delta_g$ obtained for different samplers are in good agreement, as illustrated in figure 4. Moreover, figure 5 shows the number of model evaluations per effective sample $N_{\text{eval}}/N_{\text{eff}}$ for each sampler and each set of parameters. Both microcanonical samplers, MAMS and MCLMC, outperform their canonical counterparts. Notably, MCLMC surpasses all adjusted samplers, achieving a tenfold speedup in this case. In particular, when compared to MAMS, this suggests that for such high-dimensional settings, the efficiency gain stems primarily from the unadjusted nature of MCLMC, with its microcanonical nature playing a secondary role.

Also, NUTSwG performs worse than standard NUTS, which is expected since MCMC within Gibbs sampling is theoretically suboptimal, particularly for correlated densities. While preconditioning and mass matrix tuning help mitigate correlations (see section 3.3), residual correlations may still contribute to this performance gap. Finally, HMC performs slightly

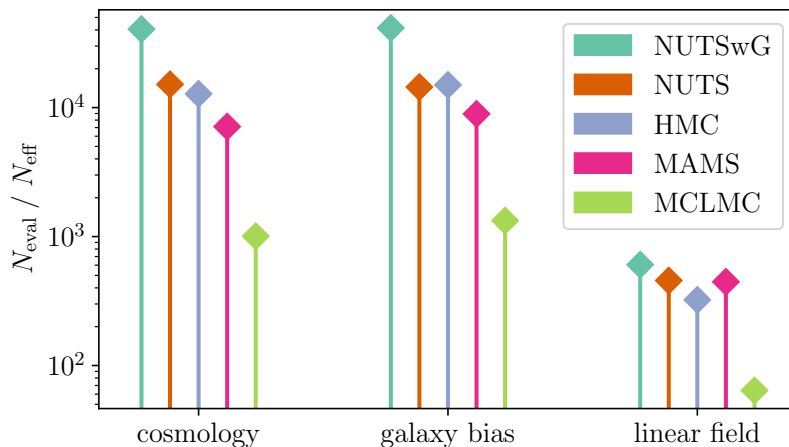


Figure 5. Number of model evaluations per effective sample for each sampler and each set of parameters (cosmology, galaxy bias and linear field).

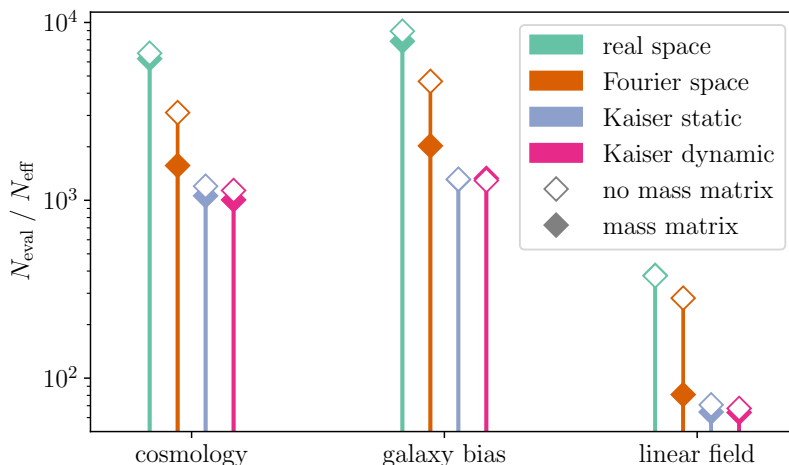


Figure 6. Number of model evaluations per effective sample for each set of parameters and each conditioning strategy: prior-conditioning in real space, prior-conditioning in Fourier space, static and dynamic Kaiser conditioning, all with (filled markers) and without (empty markers) mass matrix tuning.

better than NUTS, but this comparison benefits from the fact that HMC’s trajectory length was tuned using prior NUTS runs. In practice, NUTS is generally preferable to HMC, as it eliminates the need for manual tuning of this sensitive parameter.

3.3 Conditioning strategies

From this point onward, we focus on the MCLMC sampler, using it to benchmark four preconditioning strategies outlined in section 2.3, both with and without diagonal mass matrix tuning. As in the previous section, we verify that all resulting posteriors are consistent before reporting performance metrics.

Strategy	$N_{\text{eval}}/N_{\text{eff}} \downarrow$	$N_{\text{eval}} \times \text{MSE}[\mu] \downarrow$	$N_{\text{eval}} \times \text{MSE}[\sigma] \downarrow$
Best	1005	879	390
w/o mass matrix	1136 $\times 1.1$	883 $\times 1.0$	518 $\times 1.3$
w/o dynamic	1200 $\times 1.1$	2256 $\times 2.6$	579 $\times 1.1$
w/o Kaiser	3112 $\times 2.6$	3314 $\times 1.5$	1395 $\times 2.4$
w/o Fourier	6714 $\times 2.2$	5504 $\times 3.5$	1391 $\times 1.0$

Table 2. Cumulative ablation of the different conditioning strategies from our best result with dynamic Kaiser preconditioning and mass matrix tuning. The three metrics $N_{\text{eval}}/N_{\text{eff}}$, $N_{\text{eval}} \times \text{MSE}[\mu]$ and $N_{\text{eval}} \times \text{MSE}[\sigma]$ are computed for the cosmological parameters, and the true value used for computing MSEs are obtained from a long NUTS sampling run. Removing conditioning strategies consistently and cumulatively degrades performance.

Figure 6 presents the number of model evaluations per effective sample for each conditioning strategy and set of parameters. As expected, sampling in real space performs the worse, since voxels can be highly correlated. This strong correlation also limits the potential efficiency gains from diagonal mass matrix tuning, as reflected in the small improvement observed in that case.

In contrast, sampling in Fourier space is expected to largely decorrelate the sampled δ_L parameters. Although naive prior-conditioning in Fourier space can still lead to poor conditioning, since different scales are not equally constrained, it enables significant improvement when combined with diagonal mass matrix tuning. Even better performance is achieved through preconditioning in Fourier space using a static Kaiser model, which appears to capture nearly all the diagonal conditioning benefits of Fourier space, as indicated by the minimal additional gain from mass matrix tuning. Finally, preconditioning with a dynamic Kaiser model performs only slightly better, likely because cosmology and bias parameters are sufficiently constrained in our setup, making the conditioning of the linear field weakly sensitive to their variations. For MCLMC in particular, better conditioning correlates with greater stability in numerical integration, allowing larger step sizes for the same nominal EEVPD, thereby enhancing sampling efficiency.

All metrics for the cosmological parameters are reported in table 2. The ground-truth values used for computing MSEs are obtained from a long NUTS sampling run, which is asymptotically unbiased. The metrics are consistent, with both $N_{\text{eval}}/N_{\text{eff}}$ and $N_{\text{eval}} \times \text{MSE}[\mu]$ values of the same order, indicating no detectable bias in MCLMC sampling. Ultimately, we demonstrate that effective conditioning can improve sampling efficiency by approximately a factor of ten.

3.4 Scaling effects

Now that we have more insights into the performance of MCMC samplers and conditioning strategies for field-level inference, we aim to fully assess their applicability to upcoming galaxy surveys. In particular, we are interested in the scalability of MCLMC sampling with respect to dimensionality and modeling complexity.

Figure 7 (left panel) shows the number of model evaluations per effective sample for various LSS formation models (1LPT, 2LPT and PM) at a fixed volume of $V_{\text{sim}} = (640 \text{ Mpc}/h)^3$

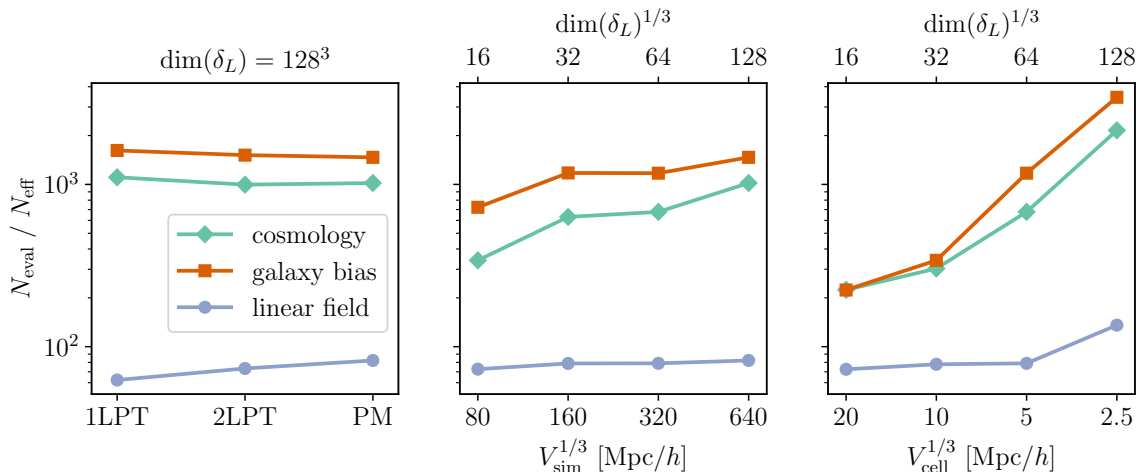


Figure 7. Number of model evaluations per effective sample for different model settings. Left: scaling with the complexity of the formation model at fixed resolution of $5 \text{ Mpc}/h$. Middle: scaling with the simulated volume for PM model. Right: scaling with the voxel resolution for PM model.

and voxel resolution of $5 \text{ Mpc}/h$ (hence $\dim(\delta_L) = 128^3$). We observe that in this setting, the sampling efficiency is largely independent of the complexity of the LSS formation model. Notably, we recover a similar magnitude in sampling efficiency as presented in [34] (below 10^2 and $10^3 N_{\text{eval}}/N_{\text{eff}}$ for linear field and cosmology respectively). One may notice a slight increase in the sampling cost of the linear field, which could be attributed to the increased non-linearity of the forward model for LSS formation.

Figure 7 (middle panel) shows the same metric for different volumes, at a fixed voxel resolution of $5 \text{ Mpc}/h$ with the PM formation model. We observe almost no increase in the sampling cost of the linear field with respect to dimensionality. Since increasing the volume at fixed resolution primarily adds independent wavevectors to infer from, this corroborates the theoretical expectation that MCLMC efficiency remains independent of dimension for product densities. However, for cosmological and bias parameters, we note a slight increase in sampling cost with dimensionality, approximately scaling as $d^{1/5}$. This is possibly related to the growing constraining power on these parameters, which reveals non-trivial correlations.

This effect is more pronounced when increasing resolution instead of volume. Figure 7 (right panel) shows sampling efficiency at a fixed volume of $V_{\text{sim}} = (320 \text{ Mpc}/h)^3$ and varying resolutions. We observe a clear increase in sampling cost with dimensionality, approximately scaling as d for cosmological and bias parameters. This indicates that field-level inference can become intrinsically more challenging at non-linear scales.

We must note that all these results are obtained for a mean galaxy density of $\bar{n}_g = 5 \times 10^{-4} (\text{Mpc}/h)^{-3}$, which is half the value used in previous sections and therefore yields twice the observational variance. This choice was made because, with the previously considered density of $\bar{n}_g = 10^{-3} (\text{Mpc}/h)^{-3}$ we found that for large volumes $V_{\text{sim}} \geq (640 \text{ Mpc}/h)^3$ or high resolution $V_{\text{cell}} \leq (2.5 \text{ Mpc}/h)^3$ and for the PM model only, the posterior exhibited some multimodality, with the chains getting stuck in different local modes¹⁶ (local in the sense

¹⁶We further checked this with a NUTS sampling run in case this was due to a MCLMC bias.

that some modes admit visibly lower density values than others). In practice, this problem could be addressed in several ways: (i) adopting an annealing or tempering strategy (as performed in [27, 67]) to help the chains converge to or navigate between global modes; (ii) smoothing or noising the observations (at the cost of reducing constraining power). We think the first option would make fair comparisons more difficult in this benchmark, and that it is worth being benchmarked on its own. Therefore, we opted for introducing more observational noise (remaining within the typical galaxy density range of current spectroscopic surveys) to focus on the general scalability aspects of the problem.

On an indicative basis, with our best sampler and conditioning strategy, in the most computational intensive setting ($\dim(\delta_L) = 128^3$ with 5-step PM displacement), a complete posterior can be obtained in less than 4 hours on a $8 \times 80\text{GB}$ GPU node. For the same setting and conditioning strategy, NUTS took more than 80 hours.

4 Conclusion

We have benchmarked various state-of-the-art samplers for field-level cosmological inference from galaxy redshift surveys, emphasizing the critical role of preconditioning latent variables, particularly the initial density field, δ_L . We find that Fourier-space sampling of δ_L , combined with a linear rescaling which assumes a simple Kaiser model, improves sampling efficiency by an order-of-magnitude compared to naive real-space sampling. In this setup, tuning the mass matrix provides only marginal additional gains. However, we note that for real survey applications where observational noise varies across the footprint, this preconditioning may need refinement, and mass matrix tuning could become more relevant.

Among the samplers tested (HMC, NUTS, NUTS within Gibbs, MAMS, and MCLMC), the recently proposed MCLMC achieves the highest efficiency, outperforming other methods by approximately an order-of-magnitude. This gain mostly arises from MCLMC being run unadjusted, at the cost of introducing a potential bias in the cosmological posterior. However, this bias can be effectively controlled and reduced to negligible levels with a modest number of additional sampling steps. For a relevant modeling at reasonable scales ($\geq 5 \text{ Mpc}/h$ with a PM scheme and 2nd order Lagrangian bias), MCLMC typically requires $\approx 10^3$ model evaluations per effective sample, mildly dependent of problem dimensionality, making it highly scalable for real-data applications. Nonetheless, we see that extending the inference to smaller physical scales may degrade efficiency, potentially requiring the use of tempering strategies to maintain robust performance. Similar complication may arise when considering more accurate small-scale modeling.

The benchmark code is publicly available¹⁷ and is intended to serve as a basis for future comparisons of alternative sampling methods. The forward model will be regularly augmented, in line with JaxPM developments, including the forthcoming implementation of massively parallel FFT calculations¹⁸ and improvements in the redshift-space galaxy density modeling.

Applying these methods to real survey data will indeed require additional developments, including a proper light-cone model, refined galaxy-matter connection at the relevant scales

¹⁷<https://github.com/hsimonfroy/benchmark-field-level>

¹⁸<https://github.com/DifferentiableUniverseInitiative/jaxDecomp>

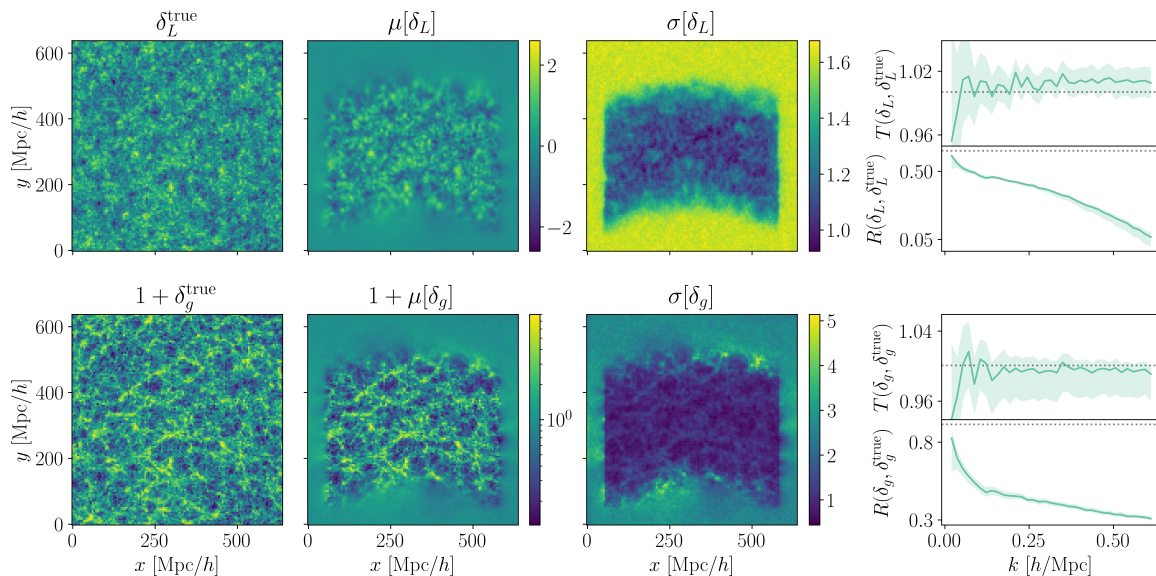


Figure 8. Posterior of (top row) the initial linear matter field δ_L and (bottom row) the evolved galaxy density field δ_g , obtained from field-level inference as in figure 2, but for a non-trivial selection function. Both the linear matter and galaxy density fields are well reconstructed on large scales within the selection (lower standard deviation, transfer function and spectra correlation close to 1), while they are less constrained outside of the selection (higher standard deviation).

in redshift space, and the incorporation of survey selection and observational systematics. These enhancements will be the focus of future work.

A Additional comparisons

A.1 Non-trivial selection

In this work, we benchmarked field-level inference in an idealized periodic box setting, selection and flat-sky setting. However, realistic surveys have significant masking and variable depth, which break translation invariance and introduce long-range coupling of wavevector amplitudes. Both our MCMC initialization and preconditioning rely on a linear diagonal approximation in Fourier space of our model given by 2.10, that is challenged in such more realistic setting. To prospect further developments, we briefly explore in this section the impact of curved-sky geometry and non-trivial selection functions on sampling efficiency.

We take a selection window W (normalized to unit mean value over voxels, $\bar{W} = 1$) consisting in a shell cut around the observer, and RSD are applied along the varying line-of-sight. Our model outputs an observable galaxy field N_g (in number of galaxies per voxel), from which we can define the fluctuation field $F_g := N_g - \bar{N}_g W$, such that equation (2.10) becomes

$$\begin{aligned}
 & \underbrace{\mathcal{N}(F_g | \bar{N}_g W B_K \delta_L, \bar{N}_g W)}_{\text{likelihood}} \underbrace{\mathcal{CN}(\delta_L | 0, P_L V_{\text{cell}}^{-1})}_{\text{prior}} \\
 &= \underbrace{\mathcal{CN}(\delta_L | \mu, \Sigma)}_{\text{posterior}} \underbrace{\mathcal{N}(F_g | 0, \bar{N}_g W + \bar{N}_g^2 V_{\text{cell}}^{-1} W B_K P_L B_K W)}_{\text{evidence}} \quad (\text{A.1})
 \end{aligned}$$

with

$$\begin{aligned}\Sigma &:= (\bar{N}_g B_K W B_K + P_L^{-1} V_{\text{cell}})^{-1} \\ \mu &:= \Sigma B_K F_g\end{aligned}$$

The main computational obstacle in applying this formula is that Σ is non-diagonal in both real and Fourier space, because W is diagonal in real space, whereas P_L is diagonal in Fourier space. To alleviate this, we approximate W by the constant $\bar{W} := \bar{W}^2{}^{1/2} = (1 + \sigma_W^2)^{1/2} \geq 1$, the root-mean-square of the window over voxels. Also, in curved-sky, $B_K := (b_E + (\hat{\mathbf{k}} \cdot \hat{\boldsymbol{\eta}})^2 f) D$ is neither diagonal in real space nor in Fourier space, but we can make the approximation that the line-of-sight $\hat{\boldsymbol{\eta}}$ is fixed to the line-of-sight towards the center of the survey volume, so that we keep B_K diagonal in Fourier space. The rest of the implementation is left unchanged.

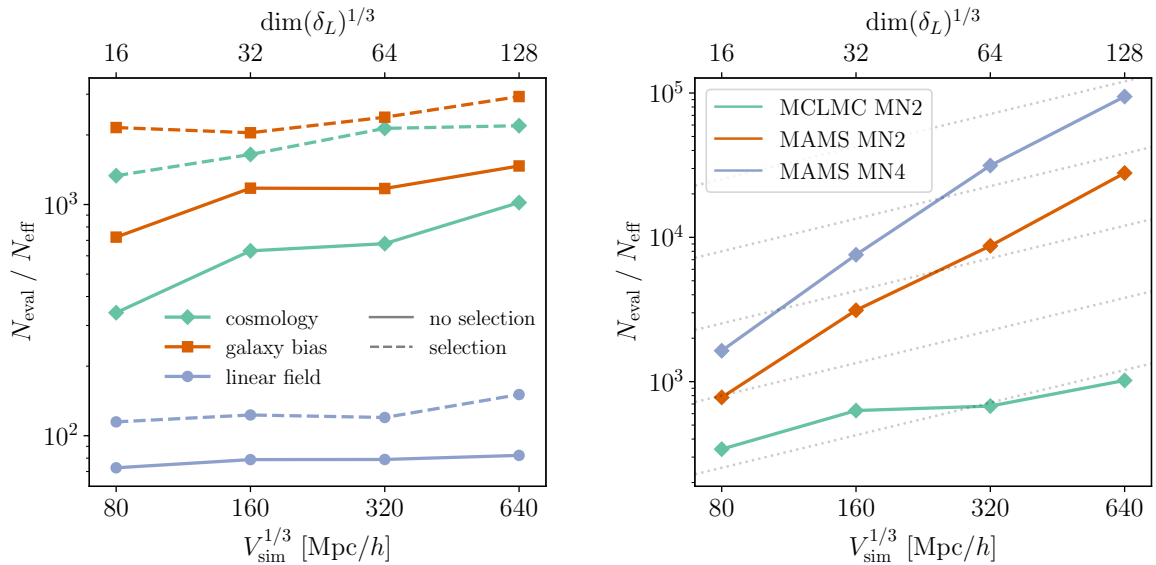
Figure 8 illustrates field-level inference for this more realistic setting. As we expect, the effect of the selection is clearly seen in the posterior standard deviation of the initial field and the galaxy density field: regions within the selection are better constrained than regions outside, which are more prior-dominated. More generally, the constraints are broader on all the metrics when compared to the no selection case in figure 2 (larger standard deviation at pixel level, larger variability in transfer function, faster decay of spectra correlation), which is consistent with a loss of information due to selection.

Finally, figure 9(a) shows the number of model evaluations per effective sample with respect to the dimension and simulated volume, for a PM model with and without selection and a fixed resolution of 5 Mpc/h. We observe a similar scaling between selections, but the non-trivial selection seems to degrade a bit the efficiency. This indicates that, while field-level inference may be more computationally demanding in realistic survey settings, the overall scaling of sampling performance with dimensionality remains largely unaffected.

A.2 Scaling of adjusted and unadjusted samplers

To illustrate the relevance of unadjusted sampling for field-level inference, we investigate whether higher-order integrators could help adjusted samplers to achieve efficiencies comparable to unadjusted ones. For this, we compare the unadjusted MCLMC to the adjusted MAMS, with McLachlan (alias 2nd-order Minium Norm, MN2) or Omelyan (4th-order Minimum Norm, MN4) integrators. The step size ε is tuned to target 65% and 80% acceptance rates for MN2 and MN4 respectively, which are theoretically optimal for isotropic Gaussian densities [52]. The average trajectory length is set following the heuristic $L \propto \frac{N_{\text{step}}}{N_{\text{eff}}} \varepsilon$, as mentioned in 3.2. We also tested a simpler $L \propto \varepsilon$ heuristic which happened to be worse for both integrators.

Figure 9(b) shows the number of model evaluations per effective sample with respect to the simulated volume, for a PM model (without survey selection) at fixed resolution of 5 Mpc/h, and for MCLMC and MAMS samplers with MN2 and MN4 integrators. As expected, unadjusted sampling exhibits strikingly better scaling with dimensionality than adjusted one. Also, both MCLMC and MAMS scale a bit worse than their respective theoretical rates for independent dimensions, respectively constant and $d^{1/4}$, which suggests potential non-trivial correlations appearing with growing constraining power. Moreover, MN4 integrator seems to perform consistently worse than MN2 integrator. This suggests that while higher-order integrators can be theoretically appealing by allowing for larger



(a) Scaling with the dimension and simulated volume, for a PM model with and without survey selection, at fixed resolution of 5 Mpc/h.

(b) Scaling for MCLMC and MAMS with McLachlan (MN2) and Omelyan (MN4) integrators. Dotted lines represent the $d^{1/4}$ scaling, expected for MAMS and HMC with 2nd order integrators.

Figure 9. Number of model evaluations per effective sample with respect to dimension and simulated volume, for different models and MCMC sampler settings.

step sizes, they can be numerically less stable and may not provide sufficient gain to offset their increased number of evaluations.

B Hamiltonian formulation of the isokinetic Langevin dynamics

Here we show that we can rederive from an Hamiltonian the MCLMC equations of motion [49], without resorting to a momentum-dependent time rescaling used in [47, 48]. We follow the idea of [46] and consider a particle of position \mathbf{q} , momentum \mathbf{p} , and mass matrix M , evolving in a d -variate potential U , with the Gaussian IsoKinetic (GIK) Hamiltonian

$$\mathcal{H}(\mathbf{q}, \mathbf{p}) = \frac{\mathbf{p}^\top M^{-1} \mathbf{p}}{2m(\mathbf{q})} - Km(\mathbf{q}) \quad (\text{B.1})$$

where $m := e^{-U}$ and K is a constant. Note that the adjective *Gaussian* in GIK refers to Gauss's principle of least constraint, rather than to the Gaussian distribution.

Setting the auxiliary momentum $\mathbf{u} := m^{-1} M^{-1/2} \mathbf{p}$ and initial energy $E := \mathcal{H}(\mathbf{q}_0, \mathbf{p}_0)$, we obtain from energy conservation that the kinetic energy is $\dot{\mathbf{q}}^\top M \dot{\mathbf{q}}/2 = |\mathbf{u}|^2/2 = K + E/m(\mathbf{q})$. For the canonical phase state (\mathbf{q}, \mathbf{p}) , Hamilton equations are then

$$\begin{cases} \dot{\mathbf{q}} = \partial_{\mathbf{p}} \mathcal{H} = (m(\mathbf{q})M)^{-1} \mathbf{p} \\ \dot{\mathbf{p}} = -\partial_{\mathbf{q}} \mathcal{H} = -\nabla U(\mathbf{q}) m(\mathbf{q}) (2K + E/m(\mathbf{q})) \end{cases} \quad (\text{B.2})$$

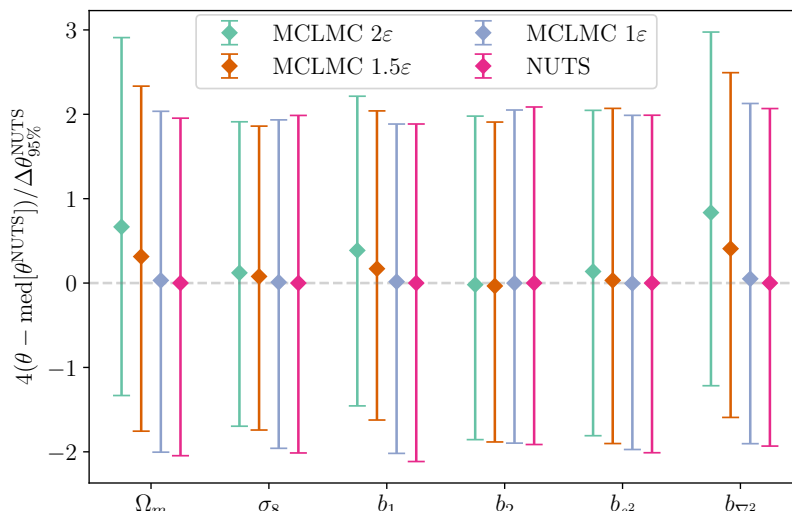


Figure 10. Comparison of posterior median and 95%-credible interval with respect to NUTS for the cosmological and bias parameters $\Omega, b \mid \delta_g$. We consider different step sizes: $1\varepsilon, 1.5\varepsilon, 2\varepsilon$ given $\varepsilon = 33$ the nominal step size. The corresponding EEVPD are $10^{-6}, 4 \times 10^{-6}, 10^{-5}$.

which translates for state (\mathbf{q}, \mathbf{u}) into

$$\begin{cases} \dot{\mathbf{q}} = M^{-1/2} \mathbf{u} \\ \dot{\mathbf{u}} = -((2K + E/m(\mathbf{q}))I - \mathbf{u}\mathbf{u}^\top)M^{-1/2}\nabla U(\mathbf{q}) \end{cases} \quad (\text{B.3})$$

This dynamics trivially admits the microcanonical ensemble $p_{\text{MC}}(\mathbf{q}, \mathbf{p}) \propto \delta(\mathcal{H}(\mathbf{q}, \mathbf{p}) - E)$ as a stationary density. Moreover, when starting at kinetic energy $|\mathbf{u}_0|^2/2 = K$, i.e. $E = 0$, this ensemble coincides with the isokinetic ensemble [46]

$$p_{\text{MC}}(\mathbf{q}, \mathbf{p}) \propto \delta(\mathcal{H}(\mathbf{q}, \mathbf{p})) \propto p_{\text{IK}}(\mathbf{q}, \mathbf{u}) \propto e^{-(d-1)U(\mathbf{q})} \delta(|\mathbf{u}|^2/2 - K) \quad (\text{B.4})$$

In particular, if ergodicity holds, the particle position \mathbf{q} samples the potential U at the effective temperature $(d-1)^{-1}$. Therefore, by replacing U by $U/(d-1)$, setting $K = 1/2$, and adding Gaussian white noise as in [49], we recover the desired isokinetic Langevin equations (2.9) that sample from $p_{\text{IK}}(\mathbf{q}, \mathbf{u}) \propto e^{-U(\mathbf{q})} \delta(|\mathbf{u}|^2 - 1)$.

C Handling bias in unadjusted Langevin dynamics

In section 3.2, we showed that unadjusted MCLMC can efficiently provide posterior samples without noticeable bias. Here, we illustrate the biased behavior of MCLMC with a poor choice of step size, or equivalently an inappropriate Energy Error Variance Per Dimension (EEVPD) threshold.

Figure 10 compares the posterior distributions of cosmological and bias parameters obtained from MCLMC with different step sizes to those obtained from NUTS, which is adjusted and therefore asymptotically unbiased. The model and sampler configurations are identical to those in section 3.2. We observe that our chosen EEVPD threshold of $\langle \Delta E^2 \rangle / d = 10^{-6}$ produces median and credible interval estimates nearly identical to those

from NUTS, with differences likely dominated by Monte Carlo error. However, when using step sizes that are $1.5\times$ or $2\times$ larger, a clear bias appears in the estimation of some parameters, particularly Ω_m , b_1 , and b_{∇^2} .

We showed in section 3.2 that for high-dimensional applications, NUTS is a magnitude less efficient than MCLMC, making it unsuitable for testing potential MCLMC bias. As a best practice, we recommend performing MCLMC sampling with at least two different EEVPD values, e.g. $\langle \Delta E^2 \rangle / d$ and $\langle \Delta E^2 \rangle / (10d)$, and checking for convergence of the posterior for the parameters of interest. Based on our experiments, once we identified an appropriate EEVPD threshold for our field-level inference problem (including preconditioning), we did not need to tune it further across the different settings considered in section 3.4.

Acknowledgments

The authors thank Marco Bonici and Natalia Porqueres for thoughtful discussions on field-level inference, Jakob Robnik for his feedback on the benchmark, and Wassim Kabalan for valuable insights on numerical implementation with JAX. They are especially grateful to Adrian Bayer for his constructive feedback and helpful comments on the manuscript. This work was granted access to the HPC resources of IDRIS under the allocations 2024-AD010414933 and 2025-475 AD010414933R1 made by GENCI.

References

- [1] DESI collaboration, *Validation of the Scientific Program for the Dark Energy Spectroscopic Instrument*, *Astron. J.* **167** (2024) 62 [[arXiv:2306.06307](#)] [[INSPIRE](#)].
- [2] EUCLID collaboration, *Euclid. I. Overview of the Euclid mission*, *Astron. Astrophys.* **697** (2025) A1 [[arXiv:2405.13491](#)] [[INSPIRE](#)].
- [3] LSST SCIENCE and LSST PROJECT collaborations, *LSST Science Book, Version 2.0*, [arXiv:0912.0201](#) [[DOI:10.2172/1156415](#)] [[INSPIRE](#)].
- [4] DESI collaboration, *DESI 2024 V: Full-Shape galaxy clustering from galaxies and quasars*, *JCAP* **09** (2025) 008 [[arXiv:2411.12021](#)] [[INSPIRE](#)].
- [5] C.H. Hahn et al., *Cosmological constraints from non-Gaussian and nonlinear galaxy clustering using the SimBIG inference framework*, *Nature Astron.* **8** (2024) 1457 [[arXiv:2310.15246](#)] [[INSPIRE](#)].
- [6] BEYOND-2PT collaboration, *A Parameter-Masked Mock Data Challenge for Beyond-Two-Point Galaxy Clustering Statistics*, *Astrophys. J.* **990** (2025) 99 [[arXiv:2405.02252](#)] [[INSPIRE](#)].
- [7] N.S. Sugiyama et al., *New constraints on cosmological modified gravity theories from anisotropic three-point correlation functions of BOSS DR12 galaxies*, *Mon. Not. Roy. Astron. Soc.* **523** (2023) 3133 [[arXiv:2302.06808](#)] [[INSPIRE](#)].
- [8] O.H.E. Philcox and M.M. Ivanov, *BOSS DR12 full-shape cosmology: Λ CDM constraints from the large-scale galaxy power spectrum and bispectrum monopole*, *Phys. Rev. D* **105** (2022) 043517 [[arXiv:2112.04515](#)] [[INSPIRE](#)].
- [9] DES collaboration, *Dark energy survey year 3 results: Cosmology with peaks using an emulator approach*, *Mon. Not. Roy. Astron. Soc.* **511** (2022) 2075 [[arXiv:2110.10135](#)] [[INSPIRE](#)].

- [10] S. Nadathur et al., *Beyond BAO: Improving cosmological constraints from BOSS data with measurement of the void-galaxy cross-correlation*, *Phys. Rev. D* **100** (2019) 023504 [[arXiv:1904.01030](#)] [[INSPIRE](#)].
- [11] EBOSS collaboration, *Constraints on the growth of structure around cosmic voids in eBOSS DR14*, *JCAP* **06** (2020) 012 [[arXiv:1909.04394](#)] [[INSPIRE](#)].
- [12] E. Paillas et al., *Cosmological constraints from density-split clustering in the BOSS CMASS galaxy sample*, *Mon. Not. Roy. Astron. Soc.* **531** (2024) 898 [[arXiv:2309.16541](#)] [[INSPIRE](#)].
- [13] M. Pinon et al., *A theoretical approach to density-split clustering*, *JCAP* **05** (2025) 090 [[arXiv:2501.14638](#)] [[INSPIRE](#)].
- [14] G. Valogiannis, S. Yuan and C. Dvorkin, *Precise cosmological constraints from BOSS galaxy clustering with a simulation-based emulator of the wavelet scattering transform*, *Phys. Rev. D* **109** (2024) 103503 [[arXiv:2310.16116](#)] [[INSPIRE](#)].
- [15] S. Codis et al., *Encircling the dark: constraining dark energy via cosmic density in spheres*, *Mon. Not. Roy. Astron. Soc.* **460** (2016) 1549 [[arXiv:1603.03347](#)] [[INSPIRE](#)].
- [16] W. Liu, A. Jiang and W. Fang, *Probing massive neutrinos with the Minkowski functionals of the galaxy distribution*, *JCAP* **09** (2023) 037 [[arXiv:2302.08162](#)] [[INSPIRE](#)].
- [17] A. Jiang et al., *Minkowski functionals of large-scale structure as a probe of modified gravity*, *Phys. Rev. D* **109** (2024) 083537 [[arXiv:2305.04520](#)] [[INSPIRE](#)].
- [18] A. Banerjee and T. Abel, *Nearest neighbour distributions: New statistical measures for cosmological clustering*, *Mon. Not. Roy. Astron. Soc.* **500** (2020) 5479 [[arXiv:2007.13342](#)] [[INSPIRE](#)].
- [19] A. Ouellette and G. Holder, *Cosmological information content of Betti curves and k-nearest neighbor distributions*, [arXiv:2502.09709](#) [[INSPIRE](#)].
- [20] P. Villanueva-Domingo and F. Villaescusa-Navarro, *Learning Cosmology and Clustering with Cosmic Graphs*, *Astrophys. J.* **937** (2022) 115 [[arXiv:2204.13713](#)] [[INSPIRE](#)].
- [21] G. Lavaux, J. Jasche and F. Leclercq, *Systematic-free inference of the cosmic matter density field from SDSS3-BOSS data*, [arXiv:1909.06396](#) [[INSPIRE](#)].
- [22] S. Duane, A.D. Kennedy, B.J. Pendleton and D. Roweth, *Hybrid Monte Carlo*, *Phys. Lett. B* **195** (1987) 216 [[INSPIRE](#)].
- [23] R.M. Neal, *Handbook of Markov Chain Monte Carlo*, [arXiv:1206.1901](#) [[DOI:10.1201/b10905](#)] [[INSPIRE](#)].
- [24] J. Bradbury, R. Frostig, P. Hawkins, M.J. Johnson, C. Leary, D. Maclaurin et al., *JAX: composable transformations of Python+NumPy programs*, (2018).
- [25] J. Jasche and B.D. Wandelt, *Bayesian physical reconstruction of initial conditions from large scale structure surveys*, *Mon. Not. Roy. Astron. Soc.* **432** (2013) 894 [[arXiv:1203.3639](#)] [[INSPIRE](#)].
- [26] J. Jasche and G. Lavaux, *Physical Bayesian modelling of the non-linear matter distribution: new insights into the Nearby Universe*, *Astron. Astrophys.* **625** (2019) A64 [[arXiv:1806.11117](#)] [[INSPIRE](#)].
- [27] N. Porqueres, O. Hahn, J. Jasche and G. Lavaux, *A hierarchical field-level inference approach to reconstruction from sparse Lyman- α forest data*, *Astron. Astrophys.* **642** (2020) A139 [[arXiv:2005.12928](#)] [[INSPIRE](#)].

- [28] S.S. Boruah, E. Rozo and P. Fiedorowicz, *Map-based cosmology inference with lognormal cosmic shear maps*, *Mon. Not. Roy. Astron. Soc.* **516** (2022) 4111 [[arXiv:2204.13216](#)] [[INSPIRE](#)].
- [29] N. Porqueres et al., *Field-level inference of cosmic shear with intrinsic alignments and baryons*, [arXiv:2304.04785](#) [[INSPIRE](#)].
- [30] N.-M. Nguyen et al., *How Much Information Can Be Extracted from Galaxy Clustering at the Field Level?*, *Phys. Rev. Lett.* **133** (2024) 221006 [[arXiv:2403.03220](#)] [[INSPIRE](#)].
- [31] M.D. Hoffman and A. Gelman, *The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo*, [arXiv:1111.4246](#) [[INSPIRE](#)].
- [32] LSST DARK ENERGY SCIENCE collaboration, *Simulation-based inference benchmark for weak lensing cosmology*, *Astron. Astrophys.* **699** (2025) A327 [[arXiv:2409.17975](#)] [[INSPIRE](#)].
- [33] B. Horowitz and Z. Lukic, *Differentiable Cosmological Hydrodynamics for Field-Level Inference and High Dimensional Parameter Constraints*, *Mon. Not. Roy. Astron. Soc.* **960** (2025) 974 [[arXiv:2502.02294](#)] [[INSPIRE](#)].
- [34] A.E. Bayer, U. Seljak and C. Modi, *Field-Level Inference with Microcanonical Langevin Monte Carlo*, in the proceedings of the *40th International Conference on Machine Learning*, Honolulu, U.S.A., July 23–29 (2023) [[arXiv:2307.09504](#)] [[INSPIRE](#)].
- [35] N.-M. Nguyen, F. Schmidt, G. Lavaux and J. Jasche, *Impacts of the physical data model on the forward inference of initial conditions from biased tracers*, *JCAP* **03** (2021) 058 [[arXiv:2011.06587](#)] [[INSPIRE](#)].
- [36] A. Kostić, N.-M. Nguyen, F. Schmidt and M. Reinecke, *Consistency tests of field level inference with the EFT likelihood*, *JCAP* **07** (2023) 063 [[arXiv:2212.07875](#)] [[INSPIRE](#)].
- [37] D. Phan, N. Pradhan and M. Jankowiak, *Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro*, [arXiv:1912.11554](#) [[INSPIRE](#)].
- [38] D.J. Eisenstein and W. Hu, *Baryonic features in the matter transfer function*, *Astrophys. J.* **496** (1998) 605 [[astro-ph/9709112](#)] [[INSPIRE](#)].
- [39] J.-E. Campagne et al., *JAX-COSMO: An End-to-End Differentiable and GPU Accelerated Cosmology Library*, *Open J. Astrophys.* **6** (2023) 1 [[arXiv:2302.05163](#)] [[INSPIRE](#)].
- [40] C. Modi, F. Lanusse and U. Seljak, *FlowPM: Distributed TensorFlow implementation of the FastPM cosmological N-body solver*, *Astron. Comput.* **37** (2021) 100505 [[arXiv:2010.11847](#)] [[INSPIRE](#)].
- [41] C. Rampf, F. List and O. Hahn, *BullFrog: multi-step perturbation theory as a time integrator for cosmological simulations*, *JCAP* **02** (2025) 020 [[arXiv:2409.19049](#)] [[INSPIRE](#)].
- [42] Y. Feng, M.-Y. Chu, U. Seljak and P. McDonald, *FastPM: a new scheme for fast simulations of dark matter and haloes*, *Mon. Not. Roy. Astron. Soc.* **463** (2016) 2273 [[arXiv:1603.00476](#)] [[INSPIRE](#)].
- [43] T. Matsubara, *Nonlinear perturbation theory with halo bias and redshift-space distortions via the Lagrangian picture*, *Phys. Rev. D* **78** (2008) 083519 [Erratum *ibid.* **78** (2008) 109901] [[arXiv:0807.1733](#)] [[INSPIRE](#)].
- [44] C. Modi, S.-F. Chen and M. White, *Simulations and symmetries*, *Mon. Not. Roy. Astron. Soc.* **492** (2020) 5754 [[arXiv:1910.07097](#)] [[INSPIRE](#)].
- [45] DESI collaboration, *Enhancing DESI DR1 full-shape analyses using HOD-informed priors*, *JCAP* **11** (2025) 049 [[arXiv:2504.10407](#)] [[INSPIRE](#)].

- [46] G.P. Morriss and C.P. Dettmann, *Thermostats: Analysis and application*, *Chaos* **8** (1998) 321.
- [47] G.V. Steeg and A. Galstyan, *Hamiltonian Dynamics with Non-Newtonian Momentum for Rapid Sampling*, [arXiv:2111.02434](#).
- [48] J. Robnik, G.B. De Luca, E. Silverstein and U. Seljak, *Microcanonical Hamiltonian Monte Carlo*, [arXiv:2212.08549](#) [INSPIRE].
- [49] J. Robnik and U. Seljak, *Fluctuation without dissipation: Microcanonical Langevin Monte Carlo*, [arXiv:2303.18221](#) [INSPIRE].
- [50] R.I. McLachlan, *On the Numerical Integration of Ordinary Differential Equations by Symmetric Composition Methods*, *SIAM J. Sci. Comput.* **16** (1995) 151.
- [51] S. Geman and D. Geman, *Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images*, *IEEE Trans. Pattern Anal. Machine Intell. P AMI-6* (1984) 721 [INSPIRE].
- [52] J. Robnik, R. Cohn-Gordon and U.Å. Seljak, *Metropolis Adjusted Microcanonical Hamiltonian Monte Carlo*, [arXiv:2503.01707](#) [INSPIRE].
- [53] A. Mijatović, V. Mramor and G. Uribe Bravo, *A note on the exact simulation of spherical Brownian motion*, *Statist. Probab. Lett.* **165** (2020) 108836.
- [54] A. Durmus and A. Eberle, *Asymptotic bias of inexact Markov chain Monte Carlo methods in high dimension*, *Annals Appl. Probab.* **34** (2024).
- [55] J. Robnik and U. Seljak, *Controlling the asymptotic bias of the unadjusted (Microcanonical) Hamiltonian and Langevin Monte Carlo*, [arXiv:2412.08876](#) [INSPIRE].
- [56] O. Papaspiliopoulos, G.O. Roberts and M. Sköld, *A General Framework for the Parametrization of Hierarchical Models*, *Statist. Sci.* **22** (2007).
- [57] N. Kaiser, *Clustering in real space and in redshift space*, *Mon. Not. Roy. Astron. Soc.* **227** (1987) 1 [INSPIRE].
- [58] M. Girolami and B. Calderhead, *Riemann Manifold Langevin and Hamiltonian Monte Carlo Methods*, *J. Roy. Statist. Soc. B* **73** (2011) 123.
- [59] S. Livingstone, M.F. Faulkner and G.O. Roberts, *Kinetic energy choice in Hamiltonian/hybrid Monte Carlo*, *Biometrika* **106** (2019) 303.
- [60] A. Cabezas et al., *BlackJAX: Composable Bayesian inference in JAX*, [arXiv:2402.10797](#).
- [61] D. Vats and C. Knudson, *Revisiting the Gelman–Rubin Diagnostic*, *Statist. Sci.* **36** (2021) 518.
- [62] M.J. Betancourt, S. Byrne and M. Girolami, *Optimizing The Integrator Step Size for Hamiltonian Monte Carlo*, [arXiv:1411.6669](#).
- [63] Q. Jiang, *On the dissipation of ideal Hamiltonian Monte Carlo sampler*, *Stat* **12** (2023).
- [64] S. Blanes, F. Casas and J.M. Sanz-Serna, *Numerical Integrators for the Hybrid Monte Carlo Method*, *SIAM J. Sci. Comput.* **36** (2014) A1556 [[arXiv:1405.3153](#)].
- [65] I.P. Omelyan, I.M. Mryglod and R. Folk, *Symplectic analytically integrable decomposition algorithms: classification, derivation, and application to molecular dynamics, quantum and celestial mechanics simulations*, *Comput. Phys. Commun.* **151** (2003) 272 [INSPIRE].
- [66] T. Takaishi and P. de Forcrand, *Testing and tuning new symplectic integrators for hybrid Monte Carlo algorithm in lattice QCD*, *Phys. Rev. E* **73** (2006) 036706 [[hep-lat/0505020](#)] [INSPIRE].
- [67] A.E. Bayer, C. Modi and S. Ferraro, *Joint velocity and density reconstruction of the Universe with nonlinear differentiable forward modeling*, *JCAP* **06** (2023) 046 [[arXiv:2210.15649](#)] [INSPIRE].