Quadratic genetic modifications: a streamlined route to cosmological simulations with controlled merger history

Martin P. Rey★ and Andrew Pontzen
Department of Physics and Astronomy, University College London, London WC1E 6BT, UK

ABSTRACT

Recent work has studied the interplay between a galaxy's history and its observable properties using 'genetically modified' cosmological zoom simulations. The approach systematically generates alternative histories for a halo, while keeping its cosmological environment fixed. Applications to date altered linear properties of the initial conditions, such as the mean overdensity of specified regions; we extend the formulation to include quadratic features, such as local variance, that determines the overall importance of smooth accretion relative to mergers in a galaxy's history. We introduce an efficient algorithm for this new class of modification and demonstrate its ability to control the variance of a region in a one-dimensional toy model. Outcomes of this work are twofold: (i) a clarification of the formulation of genetic modifications and (ii) a proof of concept for quadratic modifications leading the way to a forthcoming implementation in cosmological simulations.

Key words: methods: numerical – galaxies: evolution – galaxies: formation – dark matter.

1 INTRODUCTION

Mergers and accretion are thought to play a key role in shaping the observed galaxy population; in the prevailing cosmological paradigm merger histories are in turn seeded from random inflationary perturbations. Numerical studies must make inferences about the galaxy population from a finite sample of such histories. Due to the limited computer time available, this generates a tension between resolution (for resolving the interstellar medium) and volume (for adequately sampling histories).

One attempt to sidestep this problem is to create and study a small number of carefully controlled tests of the relationship between a galaxy's history and its observable properties. This has long been attempted in idealized, non-cosmological settings (e.g. Hernquist 1993; Di Matteo, Springel & Hernquist 2005; Hopkins et al. 2012). More recently, Roth, Pontzen & Peiris (2016) proposed performing such tests within a fully cosmological environment by constructing a series of closely related initial conditions with targeted 'genetic modifications' (hereafter GMs). The formalism resembles that of constrained realizations (Bardeen et al. 1986; Bertschinger 1987; Hoffman & Ribak 1991) which generates realizations of Gaussian random fields satisfying user-defined constraints on initial densities, velocities or potentials (e.g. Bertschinger 2001). Simulations based on constrained realizations have been extensively applied to recreating the local universe using observed galaxy distributions as constraints (for recent examples, see Heß, Kitaura & Gottlöber 2013; Sorce et al. 2016; Wang et al. 2016; Hoffman et al. 2017).

Despite a resemblance, genetically modified simulations are markedly different from constrained simulations. The process of GM involves creating multiple versions of the initial conditions, each with carefully selected small changes. By re-simulating each scenario it becomes possible to study how the changes affect the non-linear evolution of structure. For example, modifications can be chosen such that they enhance or suppress merger ratios in incremental steps and so vary a galaxy's history in a systematic and controlled way. The first application of this technique in a hydrodynamic simulation was made by Pontzen et al. (2017); that work focuses on the response of a galaxy's central black hole and its ability to quench star formation as the merger history is changed gradually. Unlike studies based on fully idealized merger simulations, the GM-based approach is able to capture the effects of gradual gas accretion from filaments which is essential when probing the balance between star formation and black hole feedback.

On a technical level, Pontzen et al. (2017) used multiple linear modifications to alter the merger history. Such a method requires human effort on two fronts: (i) to identify and track particles forming the merging substructures; and (ii) to tune the modifications and understand their effects on one another. For instance, GMs suppressing a merger tend to increase the mass of other nearby substructures, which complicates interpretation of the final results (see section 2.3 and fig. 2 of Pontzen et al. 2017). Bypassing this behaviour would be possible by individually identifying all substructures and demanding the algorithm fix each one. However, the spiralling complexity of the set-up makes this option unattractive.

★ E-mail: martin.rey.16@ucl.ac.uk

Another possibility, which is the primary aim of this paper, is to find a new type of modification which automatically suppresses the merger ratios of *all* large substructures in a target galaxy's history. Such a modification would smooth the expected history while keeping its final mass and overall environment fixed. These modifications must be applicable to cosmological simulations, so our objective is an algorithm that remains tractable even working with fields on multidimensional grids. To achieve this goal, we start by clarifying the formulation of GMs (Section 2). We then expand the framework to quadratic modifications (Section 3), allowing control over the variance at different scales to tackle the problem of multiple mergers. We demonstrate the feasibility of our method on a one-dimensional model (Section 4); in forthcoming work, we will demonstrate the implementation for a full 3D zoom simulation. Results are discussed in Section 5 and we conclude in Section 6.

2 LINEAR CONSTRAINTS AND MODIFIED FIELDS

In this section, we contrast the method of constraints (Hoffman & Ribak 1991) against that of linear GMs. The aim is to clarify the status of the latter as a building block for non-linear GMs, which are introduced in Section 3.

2.1 Constrained ensemble

We start by reviewing the construction of constrained ensembles (see bottom panel of Fig. 1). In this case, constraints must be known a priori, i.e. independently of any specific realization. Constrained ensembles are therefore particularly useful when using observations as external inputs to constrain numerical simulations.

Consider a Gaussian random field sampled at n points to create a vector $\boldsymbol{\delta}$ with covariance matrix $\mathbf{C}_0 = \langle \boldsymbol{\delta} \boldsymbol{\delta}^\dagger \rangle$. The Hoffman & Ribak (1991) algorithm allows for an arbitrary number (denoted p) of linear constraints to be placed on $\boldsymbol{\delta}$; these can be expressed as $\mathbf{A}\boldsymbol{\delta} = \boldsymbol{b}$, where $\mathbf{A}$ is a $p \times n$ matrix and $\boldsymbol{b}$ is a length-p vector.

We start by constructing the ensemble of all fields $\boldsymbol{\delta}$ satisfying the constraint for a chosen $\boldsymbol{b}$, i.e. $P(\boldsymbol{\delta}|\boldsymbol{b})$. Applying Bayes' theorem, the probability reads

$$P(\boldsymbol{\delta}|\boldsymbol{b}) = \frac{P(\boldsymbol{b}|\boldsymbol{\delta})\, P(\boldsymbol{\delta})}{P(\boldsymbol{b})}. \tag{1}$$

Using the fact that $P(\boldsymbol{\delta})$ is Gaussian and disregarding normalization, this relation becomes

$$P(\boldsymbol{\delta}|\boldsymbol{b}) \propto \delta_{\mathrm{D}}(\mathbf{A}\,\boldsymbol{\delta} - \boldsymbol{b}) \exp\left(-\frac{1}{2}\boldsymbol{\delta}^\dagger \mathbf{C}_0^{-1}\boldsymbol{\delta}\right), \tag{2}$$

where δ_{D} is the (p-dimensional) Dirac delta function.

This expression suggests a brute force sampling solution: We could draw many trial $\boldsymbol{\delta}$s from the original ensemble and keep only the ones satisfying the constraints (within some tolerance). This solution is, however, computationally inefficient. Making use of the fact that the Dirac delta function can be represented as the zero-variance limit of a Gaussian, we can instead derive the following results (Bertschinger 1987):

$$P(\boldsymbol{\delta}|\boldsymbol{b}) \propto e^{-\frac{1}{2}(\boldsymbol{\delta}-\bar{\boldsymbol{\delta}})^\dagger \mathbf{C}^{-1}(\boldsymbol{\delta}-\bar{\boldsymbol{\delta}})}, \text{ with}$$
$$\bar{\boldsymbol{\delta}}(\boldsymbol{b}) = \mathbf{C}_0\,\mathbf{A}^\dagger\,(\mathbf{A}\,\mathbf{C}_0\,\mathbf{A}^\dagger)^{-1}\,\boldsymbol{b}, \text{ and}$$
$$\mathbf{C} = \mathbf{C}_0 - \mathbf{C}_0\,\mathbf{A}^\dagger(\mathbf{A}\,\mathbf{C}_0\,\mathbf{A}^\dagger)^{-1}\,\mathbf{A}\,\mathbf{C}_0, \tag{3}$$

where $\bar{\boldsymbol{\delta}}$ and $\mathbf{C}$ are the expectation and the covariance of the Gaussian distribution $P(\boldsymbol{\delta}|\boldsymbol{b})$. By construction, all fields drawn from this distribution will satisfy the constraints ($\mathbf{A}\boldsymbol{\delta} = \boldsymbol{b}$).
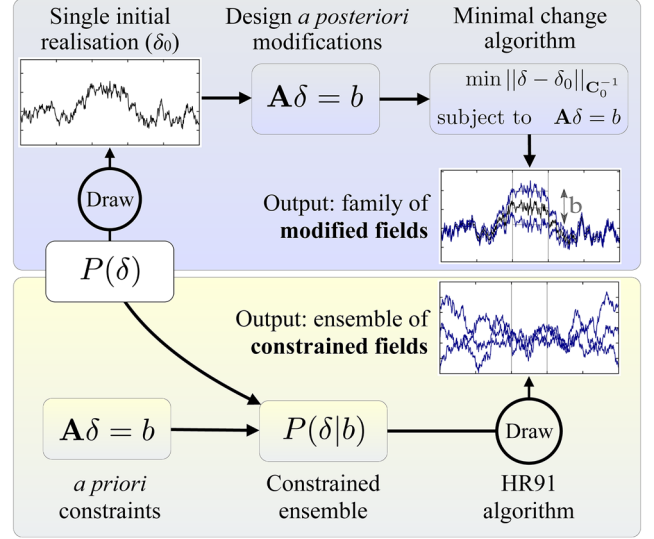

Figure 1. GMs and constrained ensembles are two techniques to generate targeted initial conditions for numerical simulations. They have markedly different motivations and properties despite sharing similar mathematics. We illustrate the differences by showing the flow of information in the two cases. Upper Panel: In the GM case, a single initial realization (black) is first drawn from the underlying ensemble. Next, modifications are designed to alter chosen properties of this realization; each modification therefore depends on the specific $\boldsymbol{\delta}_0$. The modified fields are computed by demanding minimal changes while satisfying the requested modifications. In the illustrated example, we create two modified fields with enhanced and reduced mean values, corresponding to two different values of $\boldsymbol{b}$ inside the target region. Lower panel: In the constrained ensemble case, the constraints are independent of any particular realization and are used to define the ensemble $P(\boldsymbol{\delta}|\boldsymbol{b})$. This ensemble is efficiently sampled using the Hoffman & Ribak (1991) technique. In this example, three fields are drawn and by construction satisfy the same mean value inside the target region.

Hoffman & Ribak (1991) pointed out a convenient shortcut for efficiently sampling from the distribution specified by equation (3). Starting from a draw of the unconstrained ensemble, $\boldsymbol{\delta}_0$, we calculate $\boldsymbol{b}_0 = \mathbf{A}\,\boldsymbol{\delta}_0$. One can then rewrite $\boldsymbol{\delta}_0$ as the sum of the mean field $\bar{\boldsymbol{\delta}}(\boldsymbol{b}_0)$ from equation (3) and a residual term $\boldsymbol{\delta}_{\mathrm{residual}}$, defined by

$$\boldsymbol{\delta}_{\mathrm{residual}} \equiv \boldsymbol{\delta}_0 - \bar{\boldsymbol{\delta}}(\boldsymbol{b}_0)$$
$$= \boldsymbol{\delta}_0 - \mathbf{C}_0\,\mathbf{A}^\dagger\,(\mathbf{A}\,\mathbf{C}_0\,\mathbf{A}^\dagger)^{-1}\,\boldsymbol{b}_0. \tag{4}$$

From here, a draw from the constrained ensemble $\boldsymbol{\delta}_1$ can be generated by recombining the residuals with the corrected mean $\bar{\boldsymbol{\delta}}(\boldsymbol{b})$:

$$\boldsymbol{\delta}_1 = \mathbf{C}_0\,\mathbf{A}^\dagger\,(\mathbf{A}\,\mathbf{C}_0\,\mathbf{A}^\dagger)^{-1}\,\boldsymbol{b} + \boldsymbol{\delta}_{\mathrm{residual}}. \tag{5}$$

To verify that this procedure draws samples $\boldsymbol{\delta}_1$ from the constrained distribution, one first writes the mapping from $\boldsymbol{\delta}_0$ to $\boldsymbol{\delta}_1$ in a single step:

$$\boldsymbol{\delta}_1 = \boldsymbol{\delta}_0 - \mathbf{C}_0\,\mathbf{A}^\dagger\,(\mathbf{A}\,\mathbf{C}_0\,\mathbf{A}^\dagger)^{-1}\,(\mathbf{A}\,\boldsymbol{\delta}_0 - \boldsymbol{b}). \tag{6}$$

Then, by calculating $\langle \boldsymbol{\delta}_1 \rangle$ and $\langle \boldsymbol{\delta}_1\boldsymbol{\delta}_1^\dagger \rangle$, it is possible to check that the ensemble has the correct mean and covariance from equation (3). The fact that $\boldsymbol{\delta}_1$ is Gaussian follows from its construction as a linear transformation of $\boldsymbol{\delta}_0$. The underlying efficiency of this method is that the covariance matrix in equation (3) does not depend on the value of $\boldsymbol{b}$, allowing the $\boldsymbol{\delta}_{\mathrm{residual}}$ term to be the same for both expressions.

In summary, the Hoffman & Ribak (1991) algorithm creates a draw from the constrained ensemble in two steps, using the realiza-

tion δ_0 as an intermediate construction tool. It provides a computationally efficient way of generating Gaussian constrained fields.

2.2 Genetic modifications

We now turn to GMs (see the upper panel of Fig. 1) to contrast their formulation with that of constrained fields. The GM procedure can be summarized as follows.

(i) Draw the unmodified realization δ_0.

(ii) Define the modifications by choosing which properties of δ_0 are to be modified. Unlike in the constrained field case, this is accomplished with reference to specific features of the δ_0 realization (e.g. the location and properties of particular haloes). This reflects how GMs are intended for constructing numerical experiments rather than for recreating observationally motivated scenarios. We focus first on linear modifications, i.e. of the form $\mathbf{A}\,\delta = \boldsymbol{b}$.

(iii) Create the modified field (or multiple modified fields with different values of $\boldsymbol{b}$). We require changes between fields to be as small as possible, which relies on the definition of a distance in field space. In the context of Gaussian fields, the only available metric is defined by the χ^2 distance,

$$\chi^2 \equiv || \, \delta \, ||^2_{\mathbf{C}_0^{-1}} = \delta^{\dagger}\, \mathbf{C}_0^{-1}\, \delta. \tag{7}$$

Consequently, GMs can be formulated as finding the modified field solution of the following optimization problem:

$$\min_{\delta} \quad || \, \delta - \delta_0 \, ||^2_{\mathbf{C}_0^{-1}} \, ,$$
$$\text{subject to} \quad \mathbf{A}\,\delta = \boldsymbol{b}. \tag{8}$$

The problem is solved by minimizing the Lagrangian:

$$\mathcal{L} \equiv (\delta - \delta_0)^{\dagger}\, \mathbf{C}_0^{-1}\, (\delta - \delta_0) + \lambda^{\dagger}\, (\mathbf{A}\,\delta - \boldsymbol{b}), \tag{9}$$

where λ is a vector of size p containing the Lagrange multipliers for each modification.

By differentiating to find critical points with respect to δ and λ, we obtain a system of two vector equations with the solution

$$\delta_1 = \delta_0 - \mathbf{C}_0\,\mathbf{A}^{\dagger}\,(\mathbf{A}\,\mathbf{C}_0\,\mathbf{A}^{\dagger})^{-1}\,(\mathbf{A}\,\delta_0 - \boldsymbol{b}), \tag{10}$$

where δ_1 is the modified field.

Equation (10) has regenerated equation (6) using a different motivation and derivation. To summarize:

(i) In the case of (6), δ_0 is an intermediate construct that is never used in a simulation; it only exists to aid finding δ_1, which is a sample from the distribution (3).

(ii) In the case of (10), δ_0 and δ_1 are put on equal footing. They are both initial condition fields drawn from the original, underlying ensemble $P(\delta)$. The fact that the modifications (choice of $\mathbf{A}$ and $\boldsymbol{b}$) depend on δ_0, as emphasized by Porciani (2016), does not impact this interpretation.

(iii) We show in Appendix A that in the case of non-linear constraints, there is no joint expression for GMs and Hoffman & Ribak (1991), formalizing their intrinsic difference.

GMs should therefore be seen as a mapping between fields of the same ensemble. A family of modified fields is generated by choosing multiple values for $\boldsymbol{b}$; the resulting mapping between members of the family is continuous and invertible. These properties are highly valuable for providing controlled tests, allowing for systematic exploration of the effects of formation history on a galaxy.

While the algorithm makes the minimal changes to the field, δ_1 may still not be a particularly likely draw from $P(\delta)$ if the modifications are too extreme. To quantify the level of alteration, the relative likelihood of the two fields is given by $\exp(-\,\Delta\chi^2/2)$ with

$$\Delta\chi^2 = \delta_1^{\dagger}\, \mathbf{C}_0^{-1}\, \delta_1 - \delta_0^{\dagger}\, \mathbf{C}_0^{-1}\, \delta_0. \tag{11}$$

As long as $\Delta\chi^2$ stays small, we can regard the modified and unmodified fields as similarly likely draws from Λcolddarkmatter initial conditions.

Turning $\Delta\chi^2$ into a precise quantitative statement about the relative abundance of a particular galactic history remains a topic for future research. It relies on knowing the detailed Jacobian relating the initial conditions to properties of the final galaxy. This can so far only be estimated, and only in simple scenarios such as small modifications to the halo mass (Roth et al. 2016). There are multiple possible modifications (i.e. choices of $\mathbf{A}$ and $\boldsymbol{b}$) leading to a given effect in the target galaxy history (Porciani 2016); some will carry a smaller $\Delta\chi^2$ cost than others. Finding the minimum-cost route to a given change in the non-linear universe is not the aim of GMs; to perform galaxy formation experiments, we only need to find one choice of modification with an *acceptably* small $\Delta\chi^2$ penalty.

3 EXTENSION TO QUADRATIC MODIFICATIONS

The main aim of this paper is to formulate modifications that control the variance of a field. The variance on scales smaller than the parent halo scale relates to the number of substructures in haloes (Press & Schechter 1974; Bond et al. 1991), and is therefore a proxy for the overall importance of mergers. It is important to distinguish variance modifications of a region from alterations to the power spectrum. The power spectrum defines only the average variance over the entire box, and over all possible realizations. We propose on the other hand to modify the local variance, targeting *one* region of interest and making minimal changes to the remaining structures. Another way to picture this goal is as follows. In any one stochastic ensemble, two realizations might by chance have enhanced or reduced variance in an area. Our procedure aims to map between such realizations rather than to modify the underlying power spectrum.

Variance is quadratic in the field value and therefore the approach in Section 2.2 cannot be applied directly. One natural formulation of the problem is through a new minimization problem (analogous to the original linear case):

$$\min_{\delta} \quad || \, \delta - \delta_0 \, ||^2_{\mathbf{C}_0^{-1}} \, ,$$
$$\text{subject to} \quad \delta^{\dagger}\, \mathbf{Q}\,\delta = q, \tag{12}$$

where $\mathbf{Q}$ is an $n \times n$ matrix and q is a scalar. We can assume without loss of generality that $\mathbf{Q}$ is Hermitian. For a suitable choice of $\mathbf{Q}$ (see Section 4), q specifies the variance of a chosen region.

Following a similar approach to the linear modifications, we introduce the Lagrangian

$$\mathcal{L} = (\delta - \delta_0)^{\dagger}\, \mathbf{C}_0^{-1}\, (\delta - \delta_0) + \mu\,(\delta^{\dagger}\, \mathbf{Q}\,\delta - q), \tag{13}$$

where μ is a scalar Lagrange multiplier associated with the quadratic modification. Searching for critical points, we obtain two equations relating the modified field δ_1 and the multiplier:

$$\delta_1 = (\mathbf{I} + \mu\,\mathbf{C}_0\,\mathbf{Q})^{-1}\,\delta_0 \text{ and} \tag{14}$$

$$\delta_0^{\dagger}\,(\mathbf{I} + \mu\,\mathbf{C}_0\,\mathbf{Q})^{-1}\,\mathbf{Q}\,(\mathbf{I} + \mu\,\mathbf{C}_0\,\mathbf{Q})^{-1}\,\delta_0 = q. \tag{15}$$

Equations (14) and (15) provide a closed system for μ and δ_1 given a target q. Unlike the linear case, the system cannot be solved analytically. A possibility would be to solve equation (15) numerically for μ but direct matrix inversions are prohibited due to their computational cost. One would therefore need to perform approximate matrix inversion at each step of a root-finding scheme for μ, making the worst-case complexity of such method infeasible.

There are moreover deeper reasons why such procedures cannot be straightforwardly adapted to GMs. In the linear case discussed above, we defined GMs as a continuous and invertible mapping. Both of these properties are lost when looking at equations (14) and (15). First, it is not clear that equation (15) has a real solution for μ. Consequently, a real-valued δ_1 may not exist[1] for any chosen value of q.

Secondly, the relationship between δ_0 and δ_1 is asymmetric: if a new field δ' is constructed by taking q back to its original value q_0, we will have

$$\delta' = \left(\mathbf{I} + (\mu + \mu')\mathbf{C}_0\mathbf{Q} + \mu\mu'\,(\mathbf{C}_0\mathbf{Q})^2\right)^{-1}\delta_0, \tag{16}$$

for suitable choices of μ and μ'. To obtain a solution μ' allowing recovery of the initial field ($\delta_0 = \delta'$), it must hold that $\mathbf{C}_0\mathbf{Q} \propto (\mathbf{C}_0\mathbf{Q})^2$. This will not generally be the case for our applications, and so we conclude that in general $\delta' \neq \delta_0$. Such asymmetry would be problematic for GM; the sense of a unique 'family' of fields is lost.

The combination of computational intractability and loss of key properties for GMs lead us to focus on an alternate method. We describe next a Newton-like method which efficiently approximates a solution to the optimization problem, equation (12), while reinstating the desired properties of the GM mapping.

3.1 Linearized solution

In this section, we restate the quadratic problem in a way that has a guaranteed solution and that generates a single family as a function of q. The trick is to make only infinitesimally small changes to the value of q, building up finite changes by following a path through field space that is *locally* minimal. This leads to an iterative procedure for quadratic GMs, which we will demonstrate is both unique and computationally tractable.

3.1.1 One infinitesimal step

We start by defining the displacement ϵ from the unmodified field $\delta = \delta_0 + \epsilon$; for sufficiently small changes, we may then neglect $\mathcal{O}(\epsilon^2)$ terms. We will discuss in Section 3.1.3 how to practically decompose a macroscopic change into a series of such minor modifications.

At first order, the updated variance (or other quadratic property) is given by

$$\delta^\dagger\,\mathbf{Q}\,\delta = \delta_0^\dagger\,\mathbf{Q}\,\delta_0 + 2\,\delta_0^\dagger\,\mathbf{Q}\,\epsilon + \mathcal{O}(\epsilon^2) \tag{17}$$

where we have assumed δ is real and made use of the previously stated Hermitian assumption, $\mathbf{Q}^\dagger = \mathbf{Q}$. Having linearized the modification, we can now find an analytic solution for the displacement and the multiplier μ:

$$\epsilon = \mu\,\mathbf{C}_0\,\mathbf{Q}\,\delta_0\,,\ \ \text{with} \tag{18}$$

---

[1] We note in passing that, since variance is a positive quantity, $\mathbf{Q}$ is a positive semi-definite matrix. By definition, $\mathbf{C}_0$ is positive definite. These conditions ensure that δ_1 is unique if it exists – but they do not guarantee existence.

$$\mu = \frac{1}{2}\,\frac{q - \delta_0^\dagger\,\mathbf{Q}\,\delta_0}{\delta_0^\dagger\,\mathbf{Q}\,\mathbf{C}_0\,\mathbf{Q}\,\delta_0}. \tag{19}$$

Equation (19) does not involve matrix inversions and can therefore be efficiently evaluated, even in a 3D cosmological simulation context.

3.1.2 Building finite changes by successive infinitesimal updates

We now want to construct a macroscopic change in the field by iterating the infinitesimal steps of equation (18). Performing a finite number of steps N, the modified field reads

$$\delta_1 = \prod_{j=0}^{N}\left(\mathbf{I} + \mu_j\,\mathbf{C}_0\,\mathbf{Q}\right)\delta_0, \tag{20}$$

where μ_j is the Lagrange multiplier at step j. The value of each μ_j depends on how the fixed interval is divided, i.e. implicitly on N. In the limit of increasing number of steps, each individual μ_j becomes infinitesimally small and the final solution is

$$\begin{aligned}\delta_1 &= \lim_{\substack{N\to\infty\\ \mu_j\to 0}}\prod_{j=0}^{N}\left(\mathbf{I} + \mu_j\,\mathbf{C}_0\,\mathbf{Q}\right)\delta_0\\ &= \prod_{j=0}^{\infty}\exp\left(\mu_j\mathbf{C}_0\,\mathbf{Q}\right) \equiv \exp\left(\alpha\mathbf{C}_0\,\mathbf{Q}\right)\delta_0,\end{aligned} \tag{21}$$

where $\alpha = \sum_{j=0}^{\infty}\mu_j$ is the overall displacement and is finite. The right-hand side of equation (21) defines the matrix exponential operator, which is guaranteed to exist and is invertible.

The matrix exponential is a useful formal expression to show that there is a unique result, but does not help computationally since the required value of α to reach the objective $\delta_1^\dagger\mathbf{Q}\,\delta_1 = q$ is unknown. In practice, we use the finite approximation, equation (20). The μ_j at each step are chosen by targeting N intermediate modifications linearly spaced between the starting value $q_0 \equiv \delta_0^\dagger\mathbf{Q}\,\delta_0$ and the target q. At each step, μ_j is calculated using equation (19); ϵ_j is deduced with equation (18); and the field is updated, $\delta \to \delta + \epsilon_j$.

3.1.3 Step choice for a practical algorithm

When calculating equation (20) as an approximation to equation (21), the accuracy will increase with the number of steps N. One must choose a minimal N (for computational efficiency) while ensuring that linearly approximating the modification at each step is sufficiently accurate.

We first perform the calculation with a fixed number of steps N_{initial}. This gives rise to an initial estimate for the modified field that we denote $\delta_{1,\text{initial}}$. The error on the resulting modification can be characterized by the magnitude of η_{initial}, where

$$\eta_{\text{initial}} \equiv \delta_{1,\text{initial}}^\dagger\mathbf{Q}\,\delta_{1,\text{initial}} - q. \tag{22}$$

Because second-order terms are neglected in the modification, the error term η_{initial} should scale inverse-quadratically with the number of steps N_{initial}. We verified this behaviour numerically for a variety of fields and modifications. If η_{initial} is smaller than a desired precision, η_{target}, we retain the initial estimate as our final output field. Otherwise, the calculation must be repeated; the required number of steps to achieve the target precision is inferred from the quadratic scaling as

$$N = N_{\text{initial}}\sqrt{\frac{\eta_{\text{initial}}}{\eta_{\text{target}}}}. \tag{23}$$

Note that N_{initial} should be kept small to avoid unnecessary iterations; $N_{\text{initial}} = 10$ has been chosen for our test scenarios below.

The final algorithm has a worst-case complexity of $\mathcal{O}(\eta_{\text{target}}^{-1/2} n^3)$, where n is the number of elements in the field δ. The n^3 arises from matrix multiplications required to compute each step; in practice, the matrices will be sparse either in Fourier space (for the covariance matrix) or in real space (for the variance $\mathbf{Q}$ matrix). Therefore, one can speed up the matrix multiplications by transforming back and forth from real to Fourier space, improving the complexity to $\mathcal{O}(\eta_{\text{target}}^{-1/2} n \log n)$.

The final procedure shares numerous similarities with Newton methods, used in large-scale optimization (see Nocedal & Wright 2006 for a comprehensive review). It retains quadratic information in the objective and linear information in the modification at each step and has a quadratic rate of convergence to the solution.

3.2 Joint quadratic and linear modifications

The algorithm above can be generalized to the case where we have both a quadratic modification and p linear modifications of the form $\mathbf{A}\,\delta = \boldsymbol{b}$. We first apply the linear modifications using equation (8), then turn to the iterative quadratic modifications. However, if equation (20) is applied directly, the linear objective will no longer be satisfied; in other words, we need to enforce $\mathbf{A}\,\epsilon = \boldsymbol{0}$ at each step. Constructing and solving the appropriate minimization, expression (18) is replaced by

$$\epsilon = -\mu\,\mathbf{C}_0\,\mathbf{Q}\,\delta + \mu\,\mathbf{C}_0\,\mathbf{A}^\dagger\,(\mathbf{A}\,\mathbf{C}_0\,\mathbf{A}^\dagger)^{-1}\,\mathbf{A}\,\mathbf{C}_0\,\mathbf{Q}\,\delta, \tag{24}$$

where

$$\mu = \frac{1}{2}\,\frac{q - \delta^\dagger\mathbf{Q}\delta}{\delta^\dagger\mathbf{Q}\mathbf{C}_0\mathbf{A}^\dagger(\mathbf{A}\mathbf{C}_0\mathbf{A}^\dagger)^{-1}\mathbf{A}\mathbf{C}_0\mathbf{Q}\delta - \delta^\dagger\mathbf{Q}\mathbf{C}_0\mathbf{Q}\delta}. \tag{25}$$

These results can be iterated to achieve the final modified field in exactly the same way as for the pure-quadratic modification.

Despite the complexity of these expressions, the evaluation will remain $\mathcal{O}(\eta_{\text{target}}^{-1/2} n \log n)$ for reasons discussed previously. To help interpret the method, there is a clear geometric meaning for each term, which we present in Appendix B.

4 DEMONSTRATION

In this section, we demonstrate our algorithm in an n-pixel, one-dimensional setting as a proof of concept and as a reference for future implementation on cosmological simulations. We choose an example of red power spectrum, as typically encountered on the scales from which galaxies collapse. Specifically, we adopt $P(k) = P_0\,(k_0 + k)^{-2}$, where P_0 is an arbitrary normalization and $k_0 = 2\pi/n$, an offset that prevents divergence of $P(k)$ at $k = 0$.

4.1 Defining an example modification

The framework developed in Section 3 can alter any property that is quadratic in the field by suitable choice of $\mathbf{Q}$. We now specialize to the case that $\mathbf{Q}$ corresponds to the variance of a length-R region of the field. We start by defining the windowing operator $\mathbf{W}$ as a rectangular matrix picking out the desired R entries from the n pixels in δ. To calculate the variance of the region, one then calculates $\delta^\dagger\mathbf{Q}_{\sigma^2}\,\delta$ where $\mathbf{Q}_{\sigma^2}$ can be written

$$\mathbf{Q}_{\sigma^2} = \frac{1}{R^2}\mathbf{W}^\dagger\,(R\,\mathbf{I} - \boldsymbol{1}\otimes\boldsymbol{1})\,\mathbf{W}. \tag{26}$$

Here, $\mathbf{I}$ is the $R \times R$ identity matrix and $\boldsymbol{1}$ is a length-R vector of ones. Expression (26) is readily verified by constructing $\delta^\dagger\mathbf{Q}_{\sigma^2}\,\delta$ and seeing that it does boil down to the variance of the chosen region.

We wish to consider the field variance only on scales smaller than the region size (corresponding to substructures with mass lower than that of the parent halo). To achieve this, $\mathbf{Q}_{\sigma^2}$ can be high-pass filtered; we use a standard Gaussian high-pass filter $\mathbf{F}$ where in Fourier space the elements of $\tilde{\mathbf{F}}$ are given by

$$\tilde{F}_{lm} = \delta_{lm}\left(1 - \exp\left[-\frac{1}{2}\left(\frac{k_l}{k_{\text{f}}}\right)^2\right]\right). \tag{27}$$

Here, $k_l = 2\pi l/n$ is the wavenumber of the lth Fourier series element and k_{f}, the filtering scale, is defined in our case by $k_{\text{f}} = 2\pi/R$. The most appropriate choice of filtering scales and shapes in the context of cosmological simulations will be discussed in a forthcoming paper.

In real space, the matrix $\mathbf{F}$ is defined by $\mathbf{F} = \mathbf{U}^\dagger\tilde{\mathbf{F}}\mathbf{U}$, where $\mathbf{U}$ is the unitary Fourier transform matrix. Finally, to localize the target modification fully, we can re-window the matrix after smoothing. The operator $\mathbf{W}^\dagger\mathbf{W}$ achieves this by setting pixels outside the target window to zero. With this set of choices, the final quadratic objective is set by

$$\begin{aligned}\mathbf{Q} &\equiv \mathbf{W}^\dagger\mathbf{W}\mathbf{F}^\dagger\mathbf{Q}_{\sigma^2}\mathbf{F}\mathbf{W}^\dagger\mathbf{W} \\ &= \frac{1}{R^2}\mathbf{W}^\dagger\mathbf{W}\mathbf{F}^\dagger\mathbf{W}^\dagger\,(R\,\mathbf{I} - \boldsymbol{1}\otimes\boldsymbol{1})\,\mathbf{W}\mathbf{F}\mathbf{W}^\dagger\mathbf{W}.\end{aligned} \tag{28}$$

In practice, we never calculate the matrix $\mathbf{Q}$ explicitly but rather implement a routine to efficiently calculate $\mathbf{Q}\,\delta$ for any field δ, which is then used by the algorithm described in Section 3. The ability to bypass storing or manipulating $\mathbf{Q}$ is essential to permit the computation to operate on a 3D cosmological simulation.

4.2 Results

Fig. 2 shows examples of modified fields obtained with our algorithm. We alter the variance of a region of width $R = 100$ pixels enclosed by vertical lines, showing two quadratic modifications with the variance reduced by a factor of 3 (light grey) and 10 (dark blue). In both cases, the mean of the field is held fixed at the unmodified value (horizontal line). In the setting of a cosmological simulation, we expect to be able to fix the parent halo mass (through the mean value) while modifying the smoothness of accretion (through the variance).

We verified that these fields achieve the linear modification $\mathbf{A}\,\delta_1 - \boldsymbol{b}$ to within numerical accuracy and the quadratic modification $\delta_1^\dagger\mathbf{Q}\delta_1 - q$ to $\eta_{\text{target}} = 10^{-6}$ accuracy. The heights of small-scale peaks inside the enclosed region are successfully reduced and brought closer to the mean value for the modified fields. Visually, it can be seen that the changes to the field are minimal, maintaining as much as possible of the structure of the unmodified field in the modified versions. This underlines how the analytic minimization (equation 12), and its refinement to a linearized procedure (Section 3.1) agrees with the intuitive sense of making minimal changes. The different versions of the field form a continuous family as illustrated by the smooth deformation when reducing the variance by different factors.

Despite the modification objective $\mathbf{Q}$ being strictly confined to the target region, modifications can be seen to 'leak' outside (beyond the vertical lines). This effect, which is also seen in linear GMs, is an intentional aspect of the minimization construction – any sharp
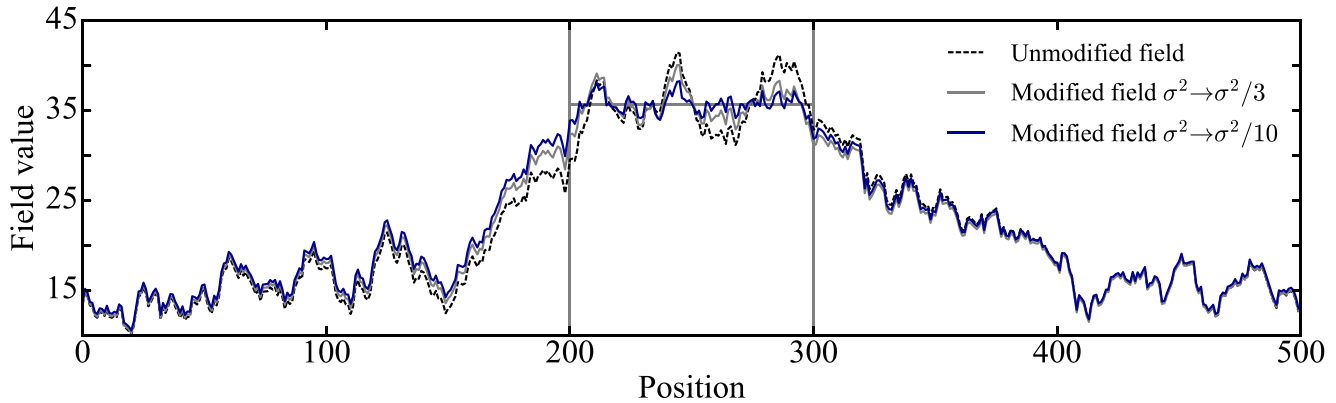
Figure 2. Example GM of a Gaussian random field with power spectrum $P(k) \propto (k_0 + k)^{-2}$. The unmodified and modified fields are shown, respectively, by dashed and solid curves. The region targeted for alteration is enclosed by vertical lines. We use simultaneous linear and quadratic modifications to conserve the mean value of this region (horizontal line) while reducing the small-scale variance by a factor of 3 (grey) and 10 (blue). In the context of galaxy formation, this would maintain the mass of a galaxy and make its formation history smoother, while making minimal alterations to the large-scale environment.

discontinuities in the field value or its gradients would give rise to a power spectrum inconsistent with the ensemble. In this specific example, the leakage appears more significant to the left than to the right of the target region. In general, the algorithm is spatially symmetric but its effect in any given case is not.

5 DISCUSSION

5.1 The advantage of quadratic over linear modifications

Pontzen et al. (2017) showed that using multiple linear modifications was sufficient to change the merger ratios in the history of a galaxy; substructures can be diminished or enhanced by manually modifying individual peak heights.

None the less, we expect the new quadratic approach to bring considerable benefits when making such manipulations; the advantages are illustrated in Fig. 3. The top panel shows a field representing the density in initial conditions expected for a halo. The field has a broad overdensity enclosed by vertical lines and two narrower peaks labelled (1) and (2). According to the excursion-set formalism (Bond et al. 1991), (1) and (2) will collapse to form two separate haloes that later merge. This, together with smooth accretion, will form the final halo.

Suppose we wish to generate a smoother accretion history by reducing the heights of peaks (1) and (2) while maintaining the large-scale overdensity. In the original approach, we use linear GMs to set the mean values of the peaks to the mean value of the broad overdensity (horizontal line). The middle panel of Fig. 3 presents the resulting field. However, a number of problems arise when performing the alteration using this approach.

(i) We had to identify (1) and (2) as the most interesting substructures and define specific modifications for each. In the context of *N*-body simulations, this requires manually identifying which particles of the initial conditions constitute each individual subhalo.

(ii) More importantly, spatially neighbouring modifications interact and create new substructures (peak labelled (3) in our example). One solution to prevent the appearance of new substructures could be to add a new linear objective forcing problematic regions such as (3) to remain unmodified. Identifying and mitigating side effects in this way adds a layer of complexity to the linear GM procedure.

Depending on the specific problem and the number of modifications at play, time spent at this tuning phase can rise steeply.

On the other hand, a single quadratic modification can avoid these problems by defining a variance target across the region. The third panel of Fig. 3 shows the same field with variance reduced by a factor of 10 (using the method from Section 4). The two local peaks are successfully reduced in amplitude while conserving the remaining small-scale structure of the parent halo. By construction, the variance modification naturally avoids compensation problems inherent to linear GMs. For this reason, quadratic GMs provide a cleaner, streamlined way to control merger histories.

5.2 Multiple quadratic modifications

The formalism discussed so far applies a single quadratic modification to a field (possibly in combination with linear objectives). Simultaneously applying *multiple* quadratic modifications would allow one to act concurrently on two separate haloes, or to further fine-tune the merger history of a single object. For instance, decreasing the variance on intermediate scales while increasing on small scales should increase the frequency of minor mergers.

To study this generalization, we introduce $i = 1, \ldots, P$ quadratic modifications, each with matrix $\mathbf{Q}_i$. For an infinitesimal update, the change in the field $\boldsymbol{\epsilon}$ is then given by

$$\boldsymbol{\epsilon} = \sum_i \mu_i \mathbf{C}_0 \mathbf{Q}_i \boldsymbol{\delta}, \qquad (29)$$

with $\quad \boldsymbol{\delta}^\dagger \mathbf{Q}_i \boldsymbol{\delta} + 2\boldsymbol{\delta}^\dagger \mathbf{Q}_i \boldsymbol{\epsilon} = q_i \quad$ for all i, $\qquad (30)$

where μ_i are the Lagrange multipliers associated with each modification. Equation (30) defines a system of P equations to be solved. The resulting value of a specific μ_i depends on the whole set of q_i and $\mathbf{Q}_i$, i.e. modifications are interdependent.

In the same way as Section 3.1, the update (29) can be iterated to create finite changes. Performing N steps, the modified field reads

$$\boldsymbol{\delta}_1 = \prod_{j=0}^{N} \left(\mathbf{I} + \sum_{i=0}^{P} \mu_{ij} \, \mathbf{C}_0 \, \mathbf{Q}_i \right) \boldsymbol{\delta}_0, \qquad (31)$$
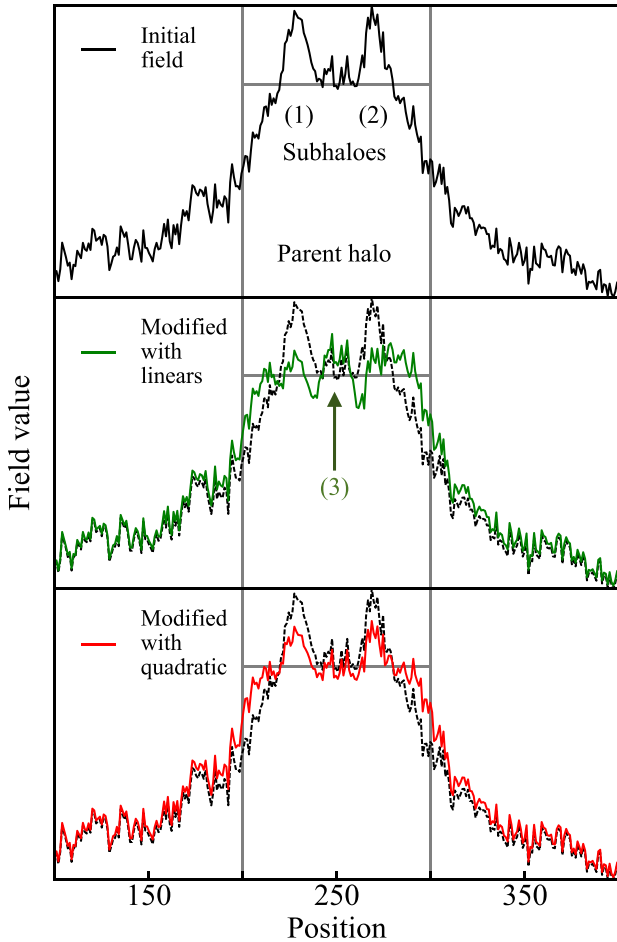
Figure 3. Comparison of pure linear against combined linear-quadratic GMs. Top panel: The unmodified field contains three distinct features: a broad overdensity that would generate a parent halo (enclosed by vertical lines) as well as two localized substructures labelled (1) and (2) that would lead to a merger during the formation history. The objective is to reduce the peak heights of these substructures while conserving the mean height of the parent (horizontal line). Middle panel: a GM field with linear modifications designed to bring peaks (1) and (2) to the mean value of the broad region. This approach has successfully smoothed the peak structure. However, as explained in the text, it suffers from the creation of an artificial substructure (3). Bottom panel: The same objective has been achieved through a variance modification. This quadratic modification does not require identifying individual subhaloes and by construction prevents unhelpful compensations such as (3).

where μ_{ij} is the multiplier μ_i at step j. However, in the limit that the number of steps $N \rightarrow \infty$, convergence to the matrix exponential,

$$\boldsymbol{\delta}_1 = \exp\left(\sum_i \alpha_i \mathbf{C}_0 \mathbf{Q}_i\right)\boldsymbol{\delta}_0, \tag{32}$$

is only guaranteed if either the $\mathbf{Q}_i$ commute with respect to $\mathbf{C}_0$ (i.e. $\mathbf{Q}_i \mathbf{C}_0 \mathbf{Q}_j = \mathbf{Q}_j \mathbf{C}_0 \mathbf{Q}_i$) or each μ_{ij} is directly proportional to α_i. Because α_is are not known in advance, the latter option is hard to arrange; the previously noted interdependence of the μ_is on all q_i and $\mathbf{Q}_i$ exacerbates the difficulty.

With our current algorithms, convergence to the matrix exponential is therefore only assured when the $\mathbf{Q}_i$ matrices commute. The easiest way to arrange for the commutation is to use orthogonal

modifications, i.e.

$$\mathbf{Q}_i \mathbf{C}_0 \mathbf{Q}_j \approx 0. \tag{33}$$

Physically, this requirement can be achieved by imposing modifications that are spatially separated by a sufficient number of correlation lengths or address distinct Fourier modes. This condition even allows one to apply the formalism of Section 3.1 to each modification one-by-one and still converge to the correct overall matrix exponential of equation (32). We leave the case of non-orthogonal multiple quadratic modifications to further work.

6 CONCLUSIONS

We have presented an efficient algorithm to modify the variance in a particular region of a Gaussian random field realization, with the aim of manipulating initial conditions for cosmological simulations. The modification produces a field that is as close as possible to the original realization. In this way, it provides a route for controlled tests of galaxy formation where multiple versions of the same galaxy are simulated within a fixed cosmological environment, but with altered accretion history.

We argued that quadratic controls, as developed here, offer a useful complement to the existing linear technique (Roth et al. 2016). In particular, variance on different filtering scales relates to dark matter halo substructure and merging history (Press & Schechter 1974; Bond et al. 1991). The new algorithm can construct GM fields with simultaneous control on the mean value and filtered variance of a region (Fig. 2). This provides a route to altering merger history and accretion over the lifetime of a given halo in a way that is more streamlined than modifying individual substructures (see Fig. 3).

In both linear and quadratic GM, the algorithm searches for fields which are nearby in the sense of the χ^2 distance measure. In the quadratic case, this definition is further refined: For large shifts in the control parameter q (which represents the variance in our test cases), the path through field space is defined by following a series of small shifts. Each of these individually minimize the χ^2 distance travelled. We demonstrated a formal convergence property for this series and argued that the approach is desirable for (a) returning a continuously deforming field $\boldsymbol{\delta}$ as a function of the changing target variance q; (b) being reversible, so that returning the variance to its initial value also returns the field to its initial state; and (c) being numerically tractable even for 3D zoom simulations.

In the process, we clarified the mathematical formulation of GM, carefully distinguishing it from the constrained ensemble of Hoffman & Ribak (1991); see Fig. 1 for an overview. The status of fields constructed in the two approaches is distinct – unlike constrained realizations, GMs should be seen as a mapping between two fields from the same ensemble. In the case of quadratic objectives such as variance, even the cosmetic similarities between constraints and modifications are lost (Appendix A).

The next step is an implementation of the new algorithm in a full N-body initial conditions generator, including on varying-resolution grids appropriate to zoom simulations. This will be presented in a forthcoming paper where we will evaluate the effectiveness of quadratic GM (alongside the existing linear technique) for constructing controlled tests of galaxy formation.

REFERENCES

Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15
Bertschinger E., 1987, ApJ, 323, L103
Bertschinger E., 2001, ApJS, 137, 1
Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440
Di Matteo T., Springel V., Hernquist L., 2005, Nature, 433, 604
Hernquist L., 1993, ApJS, 86, 389
Heß S., Kitaura F.-S., Gottlöber S., 2013, MNRAS, 435, 2065
Hoffman Y., Ribak E., 1991, ApJ, 380, L5
Hoffman Y., Pomarède D., Tully R. B., Courtois H. M., 2017, Nat. Astron., 1, 0036
Hopkins P. F., Kereš D., Murray N., Quataert E., Hernquist L., 2012, MN-RAS, 427, 968
Nocedal J., Wright S. J., 2006, Numerical Optimization, 2nd edn. Springer-Verlag, Berlin
Pontzen A., Tremmel M., Roth N., Peiris H. V., Saintonge A., Volonteri M., Quinn T., Governato F., 2017, MNRAS, 465, 547
Porciani C., 2016, MNRAS, 463, 4068
Press W. H., Schechter P., 1974, ApJ, 187, 425
Roth N., Pontzen A., Peiris H. V., 2016, MNRAS, 455, 974
Sorce J. G. et al., 2016, MNRAS, 455, 2078
Wang H. et al., 2016, ApJ, 831, 164

APPENDIX A: CONSTRUCTING CONSTRAINED ENSEMBLES FOR QUADRATIC CONSTRAINTS

In Section 2, we contrasted the notion of a linearly constrained ensemble (Section 2.1) against that of GMs (Section 2.2). While conceptually different, constrained ensembles can be sampled using the Hoffman & Ribak (1991) procedure which can in turn be seen as applying suitable modifications to realizations from the unconstrained pdf.

In this appendix, we show that there is no such similarity between quadratically constrained ensembles and quadratically modified fields. To put it another way, there is no Hoffman & Ribak (1991)-like method for generating samples from a quadratically constrained ensemble.

Following the same Bayesian argument as in Section 2.1, we define the quadratically constrained ensemble for a fixed $\mathbf{Q}$ and q by

$$P(\boldsymbol{\delta}|q) \propto \exp\left(-\frac{1}{2}\boldsymbol{\delta}^\dagger \mathbf{C}_0^{-1}\boldsymbol{\delta}\right) \, \delta_\mathrm{D}(\boldsymbol{\delta}^\dagger \mathbf{Q}\,\boldsymbol{\delta} - q). \tag{A1}$$

We will show that the modification procedure does not generate samples from the ensemble (A1), even when $\mathbf{Q}$ and q are known and fixed in advance.

We start by defining the alternative ensemble $P(\boldsymbol{\delta}_1|q)$ to be that sampled by drawing an unconstrained field from $P(\boldsymbol{\delta})$ and using the GM procedure to enforce the constraint $\boldsymbol{\delta}^\dagger \mathbf{Q}\boldsymbol{\delta} = q$. In Section 3.1, the mapping $\boldsymbol{\delta}_0 \rightarrow \boldsymbol{\delta}_1$ was given by

$$\boldsymbol{\delta}_1 = \exp\left(\alpha(\boldsymbol{\delta}_0, q)\,\mathbf{C}_0\,\mathbf{Q}\right)\boldsymbol{\delta}_0, \tag{A2}$$

where $\mathbf{C}_0$ is the covariance matrix of the Gaussian distribution $P(\boldsymbol{\delta})$ and the value of $\alpha(\boldsymbol{\delta}_0, q)$ is implicitly defined by the need to satisfy the quadratic constraint $\boldsymbol{\delta}_1^\dagger \, \mathbf{Q}\,\boldsymbol{\delta}_1 = q$.

To incorporate this implicit requirement to choose the correct value of α into an expression for the ensemble, we make use of Bayes' theorem:

$$\begin{aligned}
P(\boldsymbol{\delta}_1|q) &= \int P(\boldsymbol{\delta}_1|\alpha)P(\alpha|q)\,\mathrm{d}\alpha \\
&= \iint P(\boldsymbol{\delta}_1|\alpha, \boldsymbol{\delta}_0)P(\alpha|q, \boldsymbol{\delta}_0)P(\boldsymbol{\delta}_0)\,\mathrm{d}\alpha\,\mathrm{d}\boldsymbol{\delta}_0 \\
&= \iint P(\boldsymbol{\delta}_1|\alpha, \boldsymbol{\delta}_0)P(q|\alpha, \boldsymbol{\delta}_0)\frac{P(\alpha|\boldsymbol{\delta}_0)}{P(q)}P(\boldsymbol{\delta}_0)\,\mathrm{d}\alpha\,\mathrm{d}\boldsymbol{\delta}_0. \tag{A3}
\end{aligned}$$

Note that the constraint demands $P(q|\alpha, \boldsymbol{\delta}_0) = \delta_\mathrm{D}(\boldsymbol{\delta}_1^\dagger \mathbf{Q}\boldsymbol{\delta}_1 - q)$, where $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_0$ are related by the condition (A2). Writing out the normalization condition for $P(\alpha|q, \boldsymbol{\delta})$ then gives

$$1 = \int P(\alpha|q, \boldsymbol{\delta})\,\mathrm{d}\alpha = \int \delta_\mathrm{D}(\boldsymbol{\delta}_1^\dagger \mathbf{Q}\boldsymbol{\delta}_1 - q)\frac{P(\alpha|\boldsymbol{\delta}_0)}{P(q)}\,\mathrm{d}\alpha. \tag{A4}$$

Because $\mathbf{Q}$ and $\mathbf{C}_0\mathbf{Q}$ are positive semidefinite, q is a monotonically increasing function of α; there is only one value of α which satisfies the Dirac delta function on the right-hand side. Consequently, we can perform the integration by a change of variables to yield

$$\left.\frac{P(\alpha|\boldsymbol{\delta}_0)}{P(q)}\right|_{\boldsymbol{\delta}_1^\dagger \mathbf{Q}\boldsymbol{\delta}_1 = q} = \left.\frac{\partial}{\partial\alpha}\right|_{\boldsymbol{\delta}_0}\left(\boldsymbol{\delta}_1^\dagger \mathbf{Q}\boldsymbol{\delta}_1\right) = 2\boldsymbol{\delta}_1^\dagger \mathbf{Q}\mathbf{C}_0\mathbf{Q}\boldsymbol{\delta}_1. \tag{A5}$$

Substituting this result back into equation (A3) and performing the integral over $\boldsymbol{\delta}_0$ using $P(\boldsymbol{\delta}_1|\alpha, \boldsymbol{\delta}_0) = \delta_\mathrm{D}(\boldsymbol{\delta}_1 - \exp(\alpha\,\mathbf{C}_0\,\mathbf{Q})\boldsymbol{\delta}_0)$, one obtains

$$\begin{aligned}
P(\boldsymbol{\delta}_1|q) \propto {}& \delta_\mathrm{D}(q - \boldsymbol{\delta}_1^\dagger \mathbf{Q}\boldsymbol{\delta}_1)\,\boldsymbol{\delta}_1^\dagger \mathbf{Q}\mathbf{C}_0\mathbf{Q}\boldsymbol{\delta}_1 \\
&\times \int \mathrm{d}\alpha\,\left|\mathrm{e}^{-\alpha\mathbf{C}_0\mathbf{Q}}\right|\exp\left(-\frac{1}{2}\boldsymbol{\delta}_1^\dagger \mathrm{e}^{-\alpha\mathbf{Q}\mathbf{C}_0}\mathbf{C}_0^{-1}\mathrm{e}^{-\alpha\mathbf{C}_0\mathbf{Q}}\boldsymbol{\delta}_1\right), \tag{A6}
\end{aligned}$$

where normalization factors depending only on $\mathbf{C}_0$ have been dropped. This expression no longer has any explicit reference to $\boldsymbol{\delta}_0$, which was our primary aim. It can now be compared with the distribution for a true constrained ensemble (equation A1). The two distributions appear different (as expected, given our claim of inequivalence), but the comparison is complicated by the unsolved integral over α which obscures the content of the expression.

We can see that this integral will never regenerate the true quadratic constrained ensemble by taking an illustrative example. Consider a three-dimensional field $\boldsymbol{\delta}_1 = (x, y, z)$ with unit power spectrum ($\mathbf{C}_0 = \mathbf{I}$). Let us further choose an explicit form for $\mathbf{Q}$ such that

$$\mathbf{Q} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \Rightarrow \mathrm{e}^{\alpha\mathbf{C}_0\mathbf{Q}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{-\alpha} & 0 \\ 0 & 0 & e^{\alpha} \end{pmatrix}. \tag{A7}$$

Inserting these results into equation (A6) gives

$$\begin{aligned}
P(\boldsymbol{\delta}_1|q) \propto {}& \delta_\mathrm{D}(q + y^2 - z^2)(y^2 + z^2) \\
&\times \int_0^\infty \frac{\mathrm{d}\beta}{\beta}\exp\left(-\frac{1}{2}\left(x^2 + \beta^{-2}y^2 + \beta^2 z^2\right)\right), \tag{A8}
\end{aligned}$$

where we have made the substitution $\beta = \mathrm{e}^{-\alpha}$. The integral over β has an analytical solution using the further substitution $t = (\beta z)^2/2$ and introducing the modified Bessel function of the second kind

$$K_0(x) = \frac{1}{2}\int_0^\infty \frac{\mathrm{d}t}{t}\mathrm{e}^{-t-\frac{x^2}{4}}. \tag{A9}$$

Equation (A6) can then be evaluated explicitly to obtain

$$P(\boldsymbol{\delta}_1|q) \propto \mathrm{e}^{-\frac{x^2}{2}}\,\delta_\mathrm{D}(q + y^2 - z^2)(y^2 + z^2)\,K_0(|yz|). \tag{A10}$$

For comparison, the quadratic constrained ensemble in this example is given by

$$P(\boldsymbol{\delta}|q) \propto \mathrm{e}^{-\frac{x^2}{2}}\,\delta_\mathrm{D}(q + y^2 - z^2)\mathrm{e}^{-\frac{y^2+z^2}{2}}. \tag{A11}$$

The distributions defined by equations (A10) and (A11) have identical *x*-dependence. This is a general property: degrees of freedom for which **Q** has a null direction are unconstrained and, similarly, left unchanged by our GM transformation. The distribution generated by these degrees of freedom will therefore coincide at all times with the constrained ensemble. However, the *y* and *z* dependences differ between equations (A10) and (A11). In general, non-null directions in field space will behave differently between the GM and constrained ensemble cases.

The result establishes that our formulation of quadratic GM as a matrix exponential mapping does not reproduce a quadratically constrained ensemble when used analogously to the Hoffman & Ribak (1991) algorithm. A similar argument allows one to verify that applying the alternative non-linear modification specified by equation (14) also fails to regenerate the constrained result. In fact, one can go even further and write a general power series expansion for the mapping between δ_0 and δ_1, writing

$$\delta_1 = \sum_{i=0}^{\infty} A_i (\mu \mathbf{C}_0 \mathbf{Q})^i \delta_0, \tag{A12}$$

without further specifying the power series coefficients A_i. Even in this case, which generalizes away from a specific mapping, it is not possible to generate a constrained ensemble from the modification procedure. This underlines that modifications and constraints need to be regarded as entirely separate procedures. Only in the linear case do they appear to be cosmetically related.

APPENDIX B: GEOMETRICAL INTERPRETATION

Throughout the main text, we used fields sampled at a finite number of points *n*; the resulting algorithms can therefore be interpreted geometrically as acting on vectors in an *n*-dimensional space. For instance, Roth et al. (2016) noted that the linear GM procedure is equivalent to an orthonormal projection of the unmodified field on to a subspace defining the modification objective (see their appendix A). In this appendix, we provide the geometric interpretation for our extended formulation of GM.

For the purposes of visualizing the connection, we use fields with two samples, $\delta = (x, y)$. The arguments of this appendix generalize to higher dimensions but are easiest to visualize with $n = 2$. Fig. A1 shows the resulting two-dimensional geometry in terms of the displacements Δx and Δy from the unmodified field. By construction, the unmodified field is at the origin.

The left-hand panel shows the elliptical geometry generated by the covariance matrix in the Δx–Δy plane; specifically, the ellipses

are of constant distance Δs^2 from the origin, where

$$\Delta s^2 \equiv (\, \Delta x \ \, \Delta y \,) \, \mathbf{C}_0^{-1} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}. \tag{B1}$$

The linear objective $\mathbf{A}\delta = b$ defines a line in two dimensions. The modification consists of finding the value of $(\Delta x, \Delta y)$ lying on the line, while minimizing Δs^2. Since Δs^2 is measured in terms of $\mathbf{C}_0^{-1}$, the solution does not correspond to the closest point on the page but to the point at which a covariance ellipse is tangent to the modification line.

Similarly, the quadratic modifications (right-hand panel of Fig. A1) are associated with ellipses of constant $q = (x, y) \, \mathbf{Q} \, (x, y)^\top$. These targets are shown as dotted lines; note that they are centred on $(x, y) = (0, 0)$ and therefore appear offset from the origin in the Δx–Δy plane.

The right-hand panel of Fig. A1 also illustrates the algorithm for finding the modified field with a simultaneous quadratic and linear objective. For visual clarity, an unrealistically small ($N = 3$) number of steps are taken. We start by defining three intermediate ellipses (red dotted) between the value of the modification at the unmodified field and the target. As explained in Section 3.2, we first apply the global linear modifications from equation (8)

$$\delta \rightarrow \delta - \underbrace{\mathbf{C}_0 \, \mathbf{A}^\dagger \, (\mathbf{A} \, \mathbf{C}_0 \, \mathbf{A}^\dagger)^{-1} \, (\mathbf{A} \, \delta - b)}_{\mathrm{a}}. \tag{B2}$$

The algorithm then iterates the step ϵ defined by equation (24)

$$\epsilon = - \underbrace{\mu \, \mathbf{C}_0 \, \mathbf{Q} \, \delta}_{\mathrm{b}} + \underbrace{\mu \, \mathbf{C}_0 \, \mathbf{A}^\dagger \, (\mathbf{A} \, \mathbf{C}_0 \, \mathbf{A}^\dagger)^{-1} \, \mathbf{A} \, \mathbf{C}_0 \, \mathbf{Q} \, \delta}_{\mathrm{c}}. \tag{B3}$$

These operations can be understood geometrically as follows.

(a) A projection of the current field on the linear modification. This term is similar to the case with linear modifications only.

(b) A displacement along the normal of the ellipse at the current field value. This term is towards the next intermediate ellipse.

(c) The projection of the previous term back on to the linear modification to ensure that both are always satisfied.

Term (c) ensures that the current field at the end of each step always lies on the linear constraint. Term (a) therefore vanishes after the first step; it is an overall offset that needs to be applied only once. Together (b) and (c) are locally orthogonalizing the quadratic modification with respect to the global linear modification. The orthogonalization must be repeated at each step since the local linearization changes as we progress towards the final value of *q*.
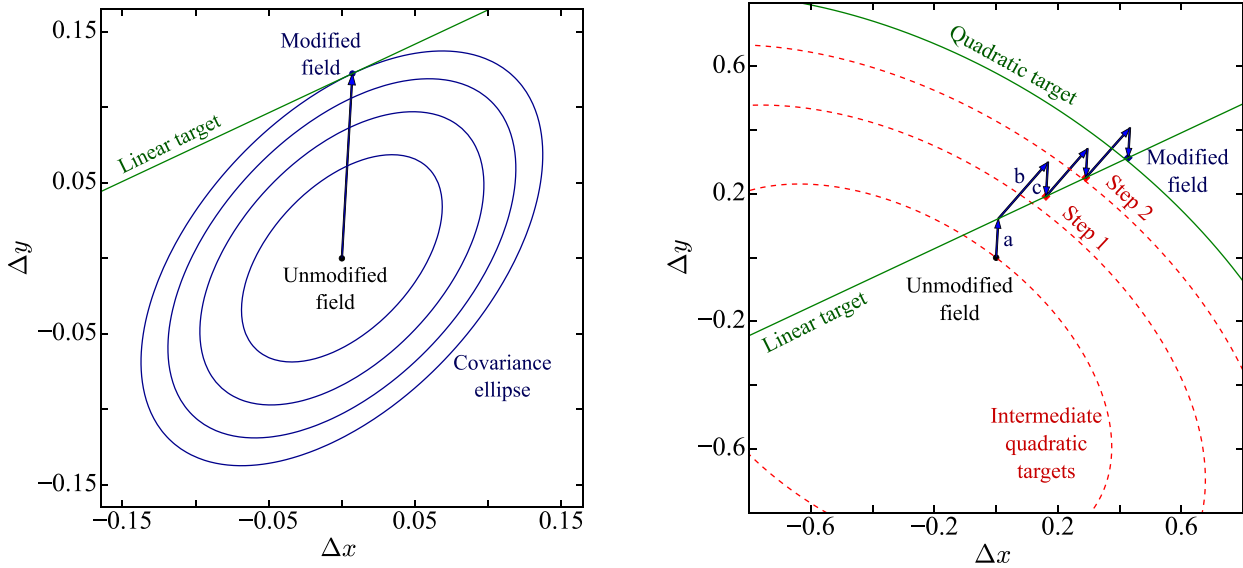
Figure A1. Left-hand panel: geometry of linear GM for a field with two components $\boldsymbol{\delta} = (x, y)^\top$. The axes represent displacements $\Delta\boldsymbol{\delta} = (\Delta x, \Delta y)^\top$ from the unmodified realization. The distance measure (equation B1) gives rise to elliptical surfaces of constant distance (blue). The linear target corresponds to a line (green). The GM algorithm (arrow) takes the unmodified realization (black dot) to the first intersection between this line and ellipses of increasing distance, defining the modified field. Right-hand panel: geometry of making simultaneous quadratic and linear modifications using the algorithm from Section 3.1. Two target modifications are shown, a linear (green line) and a quadratic (green ellipse). The algorithm defines intermediate quadratic modifications (red dotted ellipses) to step towards the final result. The first operation is the projection of the unmodified field on to the linear modification (a); each iterative step then displaces the field along the normal of the ellipse (b), and projects it again on to the linear modification (c).

This paper has been typeset from a TEX/LATEX file prepared by the author.