Geophysical Research Letters

RESEARCH LETTER
10.1029/2018GL080931

Hamiltonian Nullspace Shuttles

Andreas Fichtner[1] **and Andrea Zunino**[2]

[1]Department of Earth Sciences, ETH Zurich, Zurich, Switzerland, [2]Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

Key Points:
- We propose a method to explore the nullspace of nonlinear inverse problems without Monte Carlo sampling
- It has diverse tuning options for systematic hypothesis testing and for the discovery of alternative models with special predefined properties
- It represents a hybrid between Hamiltonian Monte Carlo and gradient-based descent methods

Supporting Information:
- Supporting Information S1

Correspondence to:
A. Fichtner,
andreas.fichtner@erdw.ethz.ch

Citation:
Fichtner, A., & Zunino, A. (2019). Hamiltonian nullspace shuttles. *Geophysical Research Letters*, *46*. https://doi.org/10.1029/2018GL080931

Received 15 OCT 2018
Accepted 19 DEC 2018
Accepted article online 27 DEC 2018

Abstract We present a method to explore the effective nullspace of nonlinear inverse problems without Monte Carlo sampling. This is based on the construction of an artificial Hamiltonian system where a model is treated as a high-dimensional particle. Depending on its initial momentum and mass matrix, the particle evolves along a trajectory that traverses the effective nullspace, thereby producing a series of alternative models that are consistent with observations and their uncertainties. Variants of the nullspace shuttle enable hypothesis testing, for example, by adding features or by producing smoother or rougher models. Furthermore, the Hamiltonian nullspace shuttle can serve as a tunable hybrid between deterministic and probabilistic inversion methods: Choosing random initial momenta, it resembles Hamiltonian Monte Carlo; requiring misfits to decrease along a trajectory, it transforms into gradient descent. We illustrate the concept with a low-dimensional toy example and with high-dimensional nonlinear inversions of seismic traveltimes and magnetic data, respectively.

Plain Language Summary All geophysical data are inherently limited in their ability to constrain the properties and dynamics of the Earth, such as material parameters deep inside our planet, future atmospheric circulation, or fluid flow in a geothermal reservoir. As a consequence, there are generally many representations (also called models) of the Earth that explain our data to within their uncertainties. The ensemble of these admissible representations is the "nullspace." Here we present an efficient tool to explore the nullspace. This is based on treating an Earth model as an artificial particle under the influence of gravity. Much like a real planet, the particle travels in an orbit. Each position in the orbit corresponds to an admissible model of the Earth. In addition to developing the theory, we demonstrate the efficiency of our method using examples from traveltime tomography and magnetic surveying.

1. Introduction

Geophysical research is largely driven by the need to solve inverse problems where the model parameters $\mathbf{m}$ of a system are to be estimated based on observed data that are typically sparse and noisy. An admissible set of model parameters $\widehat{\mathbf{m}}$ that explains the observations to within their uncertainties may be found through the approximate minimization of a suitable misfit functional $\chi(\mathbf{m})$ that quantifies differences between observed and simulated (synthetic) data (e.g., Menke, 2012; Parker, 1994; Tarantola, 2005).

Owing to the presence of observational uncertainties, deviations $\widehat{\mathbf{m}} + \Delta\mathbf{m}$ from the approximate minimizer $\widehat{\mathbf{m}}$ may still be plausible, provided that their misfit remains within a prescribed tolerance $\varepsilon \geq 0$, that is

$$\chi(\widehat{\mathbf{m}} + \Delta\mathbf{m}) \leq \chi(\widehat{\mathbf{m}}) + \varepsilon\,. \tag{1}$$

The ensemble of models $\widehat{\mathbf{m}} + \Delta\mathbf{m}$ that satisfy equation (1) defines the *effective nullspace*, the full exploration of which constitutes the complete solution of the inverse problem (Backus & Gilbert, 1970).

In special cases where the forward problem is linear and χ is the least squares misfit functional, the effective nullspace can be characterized exactly and independently of $\widehat{\mathbf{m}}$ using the linear nullspace shuttle (Deal & Nolet, 1996; de Wit et al., 2012). For nonlinear problems or more complicated misfit functionals, the effective nullspace may be explored with the help of Monte Carlo methods (e.g., Mosegaard & Tarantola, 1995; Sambridge & Mosegaard, 2002). However, even when the model space dimension n is only moderately high ($n \gtrsim$ 10–100, as orders of magnitude) the required number of forward problem solves is often prohibitively large, and so the problem of characterizing the effective nullspace remains open.

©2018. The Authors.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

The goal of this work is to develop an efficient method to explore the effective nullspace for inverse problems where the forward modeling equations are nonlinear and/or where the misfit functional is not quadratic. For this, we work under the assumption that an approximation $\hat{\mathbf{m}}$ of some optimal model has been found, using, for instance, iterative optimization methods (e.g., Nocedal & Wright, 1999).

This article is organized as follows: In section 2 we introduce the concept of the Hamiltonian nullspace shuttle, which treats a model as a high-dimensional particle that traverses the effective nullspace. Section 3 explains properties of the shuttle's orbit, that may be tuned, for instance, to add specific features to a model or to explore particularly rough or smooth parts of the effective nullspace. Depending on the choice of tuning parameters, the nullspace shuttle may be further specialized into either a Monte Carlo sampler or a gradient descent method. Following numerical considerations in section 4, we provide examples from nonlinear traveltime tomography and the inversion of magnetic data in section 5.

2. Navigating the Effective Nullspace

Our method rests on the interpretation of a model $\mathbf{m}$ in terms of an n-dimensional particle. The particle has a time-dependent position, with each position corresponding to a new realization of a model. To describe the movement of a model along a trajectory, we construct artificial equations of motion, known as Hamilton's equations in classical mechanics. This requires us to define a potential energy U, which we choose to be equal to the misfit, that is $U(\mathbf{m}) = \chi(\mathbf{m})$. The potential energy results in a force proportional to $-\nabla U$, which gives the model momentum, described by an (artificial) n-dimensional momentum vector $\mathbf{p}$. The mass of the particle is described by an $n \times n$ symmetric, positive definite mass matrix $\mathbf{M}$. This allows us to define a kinetic energy as

$$K(\mathbf{p}) = \frac{1}{2}\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p}. \tag{2}$$

Since the inverse of a symmetric, positive definite matrix is also symmetric and positive definite, the kinetic energy is generally greater than or equal to 0. The sum of potential and kinetic energies defines the artificial Hamiltonian, $H(\mathbf{m}, \mathbf{p}) = U(\mathbf{m}) + K(\mathbf{p})$ of the system. Propagating the particle in (artificial) time t according to Hamilton's equations (e.g., Landau & Lifshitz, 1976; Symon, 1971),

$$\frac{\mathrm{d}m_i}{\mathrm{d}t} = \frac{\partial H}{\partial p_i}, \qquad \frac{\mathrm{d}p_i}{\mathrm{d}t} = -\frac{\partial H}{\partial m_i}, \qquad i = 1, \ldots, n, \tag{3}$$

produces a continuous trajectory of models, along which the Hamiltonian is preserved. Choosing the approximate optimum $\hat{\mathbf{m}}$ as starting point of the trajectory therefore results in a sequence $\mathbf{m}(t)$ that satisfies

$$H[\mathbf{m}(t), \mathbf{p}(t)] = \chi[\mathbf{m}(t)] + \frac{1}{2}\mathbf{p}(t)^T\mathbf{M}^{-1}\mathbf{p}(t) = H(\hat{\mathbf{m}}, \hat{\mathbf{p}}) = \chi(\hat{\mathbf{m}}) + \frac{1}{2}\hat{\mathbf{p}}^T\mathbf{M}^{-1}\hat{\mathbf{p}}. \tag{4}$$

When the initial momentum $\hat{\mathbf{p}}$ is set such that $K(\hat{\mathbf{p}}) = \hat{\mathbf{p}}^T\mathbf{M}^{-1}\hat{\mathbf{p}}/2 = \varepsilon$, equation (4) implies

$$\chi[\mathbf{m}(t)] \leq \chi(\hat{\mathbf{m}}) + \varepsilon. \tag{5}$$

It follows that all models $\mathbf{m}(t)$ along the trajectory are within the effective nullspace, in the sense of equation (1), because $\mathbf{p}(t)^T\mathbf{M}^{-1}\mathbf{p}(t) \geq 0$. Thus, to explore the effective nullspace, it is sufficient to propagate some near optimum $\hat{\mathbf{m}}$ along a Hamiltonian trajectory where the initial kinetic energy is equal to the misfit tolerance ε. Freedom of choice, and potential to tune the method, lies in the design of the mass matrix $\mathbf{M}$ and the direction of the initial momentum $\hat{\mathbf{p}}$. Examples will be given in sections 3 and 5.

With the exception of few special cases, Hamilton's equations in (3) need to be solved numerically, using, for instance, the leapfrog method with time step Δt:

$$p_i(t + \Delta t/2) = p_i(t) - \frac{1}{2}\Delta t \left.\frac{\partial U}{\partial m_i}\right|_t, \tag{6}$$

$$m_i(t + \Delta t) = m_i(t) + \Delta t \left.\frac{\partial K}{\partial p_i}\right|_{t+\Delta t/2}, \tag{7}$$

$$p_i(t + \Delta t) = p_i(t + \Delta t/2) - \frac{1}{2}\Delta t \left.\frac{\partial U}{\partial m_i}\right|_{t+\Delta t}. \tag{8}$$

The choice of suitable numerical integrators will be further discussed in section 4.

3. Orbital Properties

Before considering concrete inverse problems, we study some special cases in order to understand basic properties of the Hamiltonian nullspace shuttle, as well as the role of the mass matrix $\mathbf{M}$ and the initial momentum $\hat{\mathbf{p}}$, which we are free to choose. Though our arguments will be based on the leapfrog method, they equally hold for other numerical integrators such as Newmark or Euler schemes.

3.1. Zero Tolerance: Relation to Iterative Gradient Descent

When a misfit increase from the approximate optimum $\hat{\mathbf{m}}$ to some $\hat{\mathbf{m}} + \Delta\mathbf{m}$ cannot be tolerated, we set $\varepsilon = 0$, which implies vanishing initial momentum, $\hat{\mathbf{p}} = \mathbf{0}$. Using equations (6) and (7), it follows that the first leapfrog step from $t = 0$ to $t = \Delta t$ is

$$m_i(\Delta t) = \hat{m}_i - \frac{1}{2}\Delta t^2 M^{-1}_{ik}\frac{\partial\chi(\hat{\mathbf{m}})}{\partial m_k}, \tag{9}$$

where the summation convention for repeated indices is implied. Equation (9) establishes a relation between the first step of effective nullspace exploration under zero tolerance and gradient descent methods. Choosing the mass matrix $\mathbf{M}$ to be the identity matrix $\mathbf{I}$, transforms (9) into a steepest-descent step with steplength $\Delta t^2/2$. Alternatively, when $\mathbf{M}$ equals the Hessian $\mathbf{H}$ of the misfit function χ, we retrieve Newton's method with modified steplength. In both cases, the trajectory descends on the misfit surface, as required by setting $\varepsilon = 0$.

3.2. Prescribed Takeoff Direction: Adding Special Features

For approximately optimal models $\hat{\mathbf{m}}$ with nearly vanishing misfit derivative, we may use the nullspace shuttle to add specifically designed features $\Delta\mathbf{m}$ to $\hat{\mathbf{m}}$ while remaining within the tolerance. Indeed, inserting $\partial\chi(\hat{\mathbf{m}})/\partial m_i = \partial U(\hat{\mathbf{m}})/\partial m_i \approx 0$ into equation (6), specializes equation (7) to

$$m_i(\Delta t) = \hat{m}_i + \Delta t M^{-1}_{ik}\hat{p}_k. \tag{10}$$

Thus, by choosing the initial momentum

$$\hat{\mathbf{p}} = \gamma\,\mathbf{M}\,\Delta\mathbf{m}, \tag{11}$$

with scaling γ set to satisfy $K(\hat{\mathbf{p}}) = \varepsilon$, the first leapfrog iteration adds $\Delta\mathbf{m}$ to $\hat{\mathbf{m}}$. Further iterations will continue in a similar direction until the growing misfit derivative away from the optimum starts to dominate equation (6).

3.3. Smooth and Bumpy Rides

In both previous cases, the inverse mass matrix, $\mathbf{M}^{-1}$, modifies the initial direction of the trajectory, as shown in equations (9) and (10). This property can be used to enforce model variations $\Delta\mathbf{m}$ with a prescribed smoothness or roughness. For instance, choosing $\mathbf{M}^{-1}$ to be a smoothing operator (e.g., a discrete averaging or integration), will produce a trajectory of smooth models, thereby exploring the smooth part of the effective nullspace. Conversely, a roughening operator (e.g., a discrete differentiation) will generate a continuous sequence of rough models. This is schematically illustrated in Figure 1.

4. Numerical Considerations

The nullspace shuttle concept critically relies on the conservation of total energy, H, expressed in equation (4). However, all explicit numerical integrators, including leapfrog, are not energy preserving. For linear problems, where the nullspace of sufficiently small problems can already be fully characterized by singular-value analysis, the implicit Crank-Nicolson scheme (Crank & Nicolson, 1947) preserves energy exactly. For the more interesting nonlinear case, energy-preserving schemes have been developed more recently (e.g., Quispel & McLaren, 2008; Simo et al., 1992). These schemes are generally implicit and therefore prohibitively expensive for modern large-scale inverse problems.

We can nevertheless obtain an efficient algorithm by realizing that our principal goal is not to solve Hamilton's equations but to have some recipe that produces a sequence of models in the effective nullspace. This can be achieved most easily by rescaling the momentum after the time step Δt from $\mathbf{p}(t + \Delta t)$ to $s\mathbf{p}(t + \Delta t)$.

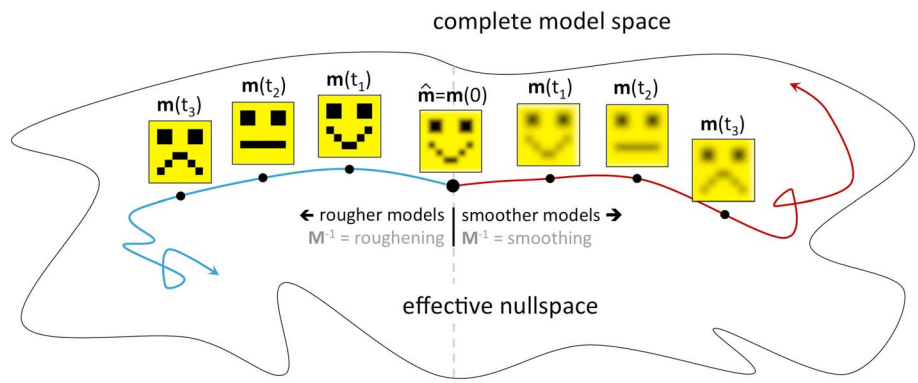

Figure 1. Schematic illustration of nullspace trajectories departing from the current model $\hat{\mathbf{m}} = \mathbf{m}(0)$ into rougher (left) or smoother (right) regions of the effective nullspace. Smileys represent rough or smooth realizations of models along the trajectories.

The positive scaling factor s is determined such that the total energy at time t equals the total energy at time $t = 0$, that is,

$$s^2 K(t + \Delta t) + U(t + \Delta t) = H(0). \tag{12}$$

The rescaling corresponds to a deceleration or acceleration of the nullspace shuttle so as to preserve its total energy, and it may be applied after each time step or only when $K(t + \Delta t) + U(t + \Delta t)$ exceeds an upper threshold value. The modified scheme may not generally solve Hamilton's equations, but it serves the purpose of traversing the effective nullspace. Numerical examples are provided in section 5.2. The time step Δt is usually chosen to be as large as possible while still ensuring numerical stability of the integration scheme.

5. Examples

5.1. Toy Problem: Cycle Skipping in 1-D

To gain intuition, we first consider the simple example of a 1-D misfit functional, defined for the purpose of illustration as

$$\chi(m) = U(m) = 1 - \cos m + 0.025 m^2, \tag{13}$$

with a scalar model parameter m. The global minimum is at $m = 0$, but local minima exist roughly at multiples of 2π, when m is small. Misfit functionals of this kind are characteristic of cycle skipping, where time shifts of bandlimited time series by few periods may be plausible, whereas large shifts become increasingly unlikely.

The Hamiltonian and selected trajectories corresponding to the potential energy defined in (13) are presented in Figure 2a for a 1×1 mass matrix equal to 1. Starting near the optimum at $\hat{m} = 0.1$ with a small tolerance of $\varepsilon = 0.12$, produces an elliptical trajectory in phase space that orbits around the global minimum. The effective nullspace explored in this case approximately ranges from -0.5 to 0.5, as shown in the top panel of Figure 2b. Successively increasing the tolerance to $\varepsilon = 2.4$ and $\varepsilon = 5.1$ allows the trajectory to cross the high-energy barriers and to explore more distant energy and misfit valleys.

While being simplistic, this example illustrates that the Hamiltonian nullspace shuttle explores locally unless an increased tolerance allows it to travel further. A large tolerance may either result from genuinely large observational errors or be chosen deliberately for the purpose of wider exploration.

5.2. Nonlinear Traveltime Tomography

Equipped with the intuition from the previous example, we consider the more relevant problem of nonlinear traveltime tomography that accounts for the deflection of raypaths due to lateral heterogeneities (e.g., Cerveny, 2001; Nolet, 2008). The model is parameterized using 70×40 grid points, leading to a 2,800-dimensional problem in a 70 km $\times$ 40 km-wide domain, shown in Figure 3. Six sources at the bottom and 12 receivers at the surface produce artificial traveltimes, treated as observations T_i^{obs}, with $i = 1, \ldots, 72$.

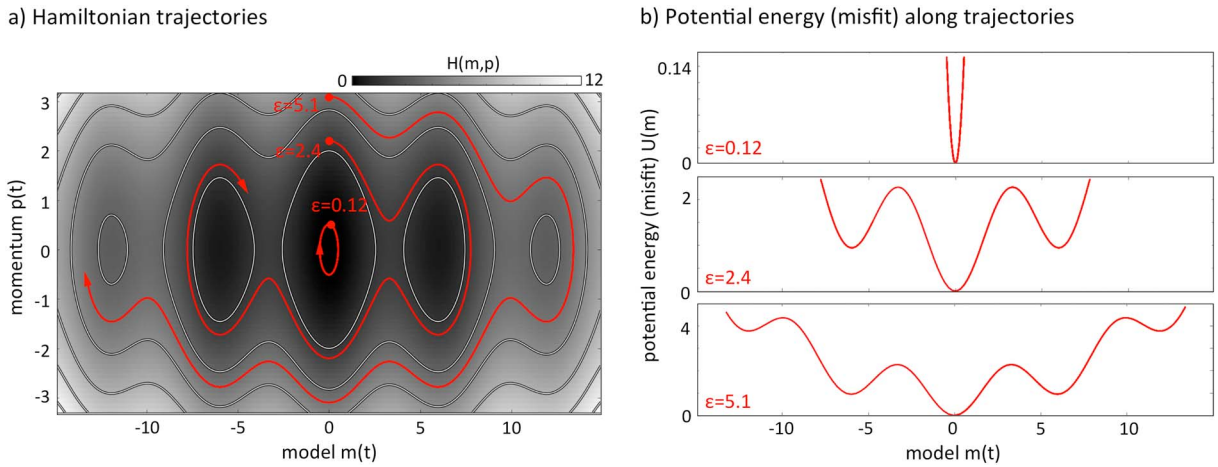

Figure 2. Nullspace exploration for the 1-D cycle skipping example. (a) The Hamiltonian for the potential energy from equation (13) and a 1×1 mass matrix $M = 1$ is shown in gray scale with contours. Trajectories for various tolerances ε at a total integration time of 23 artificial seconds are superimposed in red. (b) Potential energy (misfit) as a function of the models $m(t)$ explored along the trajectories.

The forward problem is solved by finite-differencing the eikonal equation (Podvin & Lecomte, 1991; Rawlinson & Sambridge, 2004; Sethian, 1996), and the misfit (potential energy) is

$$\chi(\mathbf{m}) = U(\mathbf{m}) = \frac{1}{2\sigma^2} \sum_i \left[T_i^{\text{obs}} - T_i(\mathbf{m})\right]^2 + \frac{1}{2}(\mathbf{m} - \mathbf{m}_{\text{prior}})^T \mathbf{C}_M^{-1}(\mathbf{m} - \mathbf{m}_{\text{prior}}), \tag{14}$$

with traveltimes T_i for a model $\mathbf{m}$, composed of velocity values at the 2,800 grid points. The prior model $\mathbf{m}_{\text{prior}}$ contains a purely depth-dependent (finely layered) velocity model. Since we use artificial data, we arbitrarily add random noise with a prior standard deviation of $\sigma = 0.6$ s. With the help of adjoint techniques, the derivatives of U, needed to solve Hamilton's equations, are computed numerically (Leung & Qian, 2006; Taillandier et al., 2009).

The nullspace trajectory starts at some acceptable model $\hat{\mathbf{m}}$ (Figure 3b), found via gradient-based, iterative optimization. The notion of acceptable is to some extent arbitrary because the prior standard deviation observational errors σ, in a synthetic inversion are subjectively chosen and not determined by actual data. To $\hat{\mathbf{m}}$ we try to add an elliptically shaped anomaly (Figure 3a), which determines the initial momentum $\hat{\mathbf{p}}$ according to equation (11). The mass matrix is a scaled unit matrix, $\mathbf{M} = \sigma^{-2}\mathbf{I}$. After a short time on the trajectory, the nullspace shuttle produces, by construction, alternative acceptable models that approximately contain scaled versions of the anomaly. See, for instance, Figure 3b for $t = 4.0$ s. As the elliptical anomaly increases in strength, the potential energy increases, at the expense of decreasing kinetic energy. Thus, the trajectory reaches a turning point, and the elliptical anomaly is again removed as the shuttle continues its journey. Subsequent alternative models often show more complex deviations from the initial model (Figures 3d and 3e). Note that the potential energy always stays within the tolerance ε and below the constant Hamiltonian H.

Figure 3, in the form of a movie can be found in the supporting information (Movie S1). This also contains an example with random model perturbations and a nondiagonal mass matrix that promotes smooth variations from the initial model (Movie S2). As demonstrated in Fichtner et al. (2018) and Zunino and Mosegaard (2018) and discussed in section 6, this tomographic version of the nullspace shuttle can easily be modified into a Hamiltonian Monte Carlo (HMC) sample of the posterior probability density.

5.3. Nonlinear Inversion of Magnetic Data

The nullspace shuttle concept can be straightforwardly transferred from traveltime tomography to other nonlinear inverse problems, including the inversion of magnetic anomaly data for subsurface magnetization (e.g., Blakely, 1996). In most magnetic prospecting surveys, magnetometers only measure the modulus of the total magnetic field at positions $\mathbf{x}_i$, denoted $|\mathbf{B}|_i$. Knowing the inducing geomagnetic field $\mathbf{g}_i$,

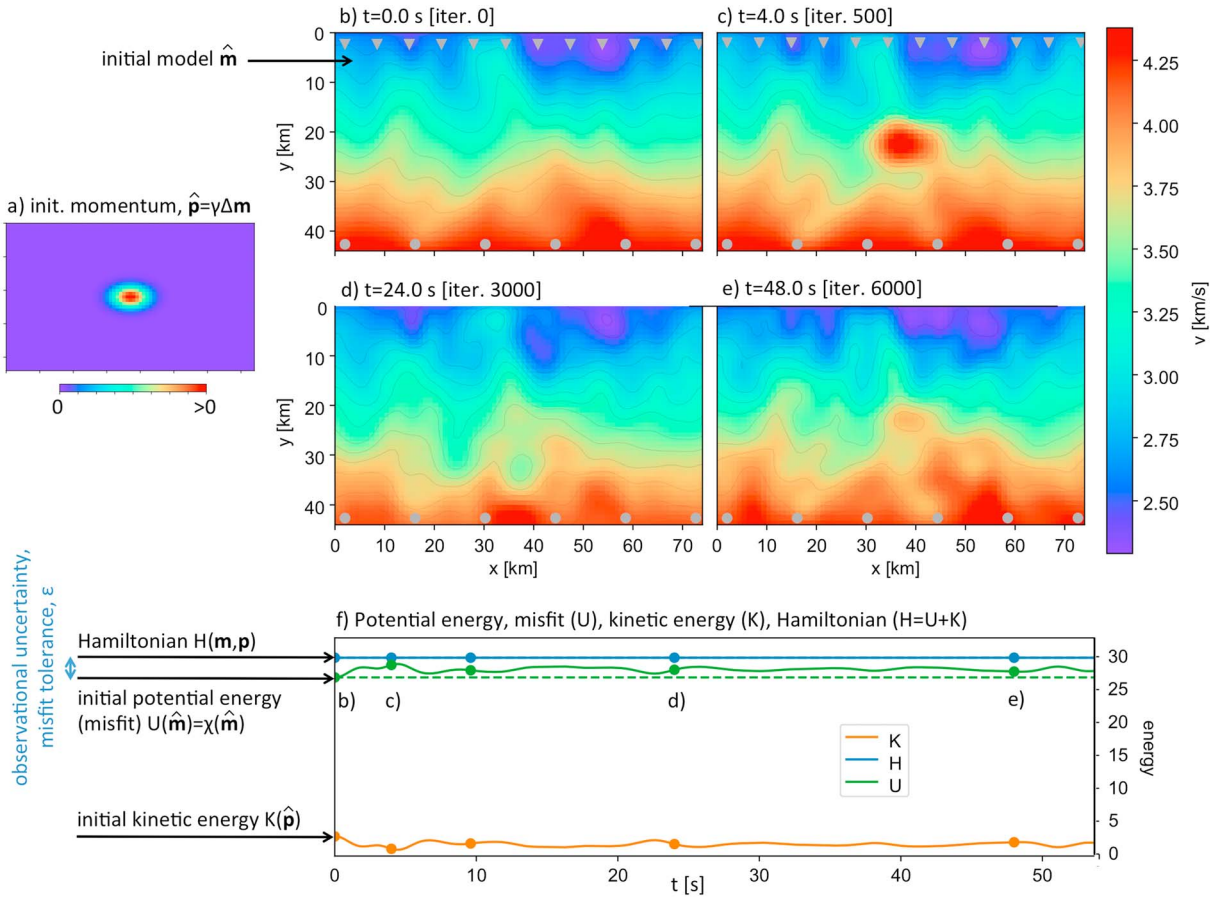

Figure 3. Nullspace exploration for nonlinear traveltime tomography. (a) Chosen initial momentum $\hat{\mathbf{p}}$, which is proportional to the initial model update $\Delta\mathbf{m}$. (b) Initial (near-optimal) velocity (v) model $\hat{\mathbf{m}}$, corresponding to time $t = 0$ s along the trajectory. (c) Snapshot of velocity models along the trajectory at time $t = 4$ s (500 leapfrog iterations), showing an alternative velocity model within the effective nullspace. The initial model perturbation $\Delta\mathbf{m}$ from panel (a) is clearly visible in the center of the model. (d, e) Further alternative models taken along the trajectory at $t = 24$ s and $t = 48$ s. (f) Evolution of potential energy, kinetic energy, and the Hamiltonian as a function of time along the trajectory. Note that the potential energy U (the misfit) always remains within the tolerance ε and below the constant Hamiltonian H.

the forward model,

$$|\mathbf{B}|_i = \sqrt{\sum_{j=1}^{3}\left(\sum_{k=1}^{n} A_{i,jk}\, m_k + g_{i,j}\right)^2}, \tag{15}$$

relates $|\mathbf{B}|_i$ to a discretized model of intensity of magnetization $\mathbf{m}$ via the kernels $\mathbf{A}_i$ (Li & Oldenburg, 2000; Zunino et al., 2009). We assume absence of remanent magnetization (Blakely, 1996) and consider the case of apparent susceptibility mapping (Grant, 1973) using a 2-D set of observations and unknowns, simulating an airborne survey. Our example consists of a grid with 1,760 cells and 440 observations at an altitude of 200 m (Figure 4). As mistfit, we again use the sum of squared differences from equation (14), adapted to measurements of $|\mathbf{B}|_i$ instead of T_i. The derivatives can be computed analytically based on equation (15).

The nullspace shuttle trajectory again starts at an acceptable model found by gradient-based optimization (Figure 4b). To this we initially add a scaled version of the magnetization perturbation $\Delta\mathbf{m}$ (Figure 4a) by choosing the appropriate initial momentum $\hat{\mathbf{p}}$. The mass matrix is the unit matrix. Unlike the velocity perturbation in Figure 3, the elliptical magnetization perturbation is hardly visible during the nullspace shuttle journey. This indicates that perturbations of this specific shape cannot be added to the initial model $\hat{\mathbf{m}}$ without increasing the misfit beyond the tolerance, $U(\hat{\mathbf{m}}) + \varepsilon$. Instead, more complex alternative models with smaller-scale features are produced along the trajectory. A movie version of Figure 4 can again be found in the supporting information (Movie S3).

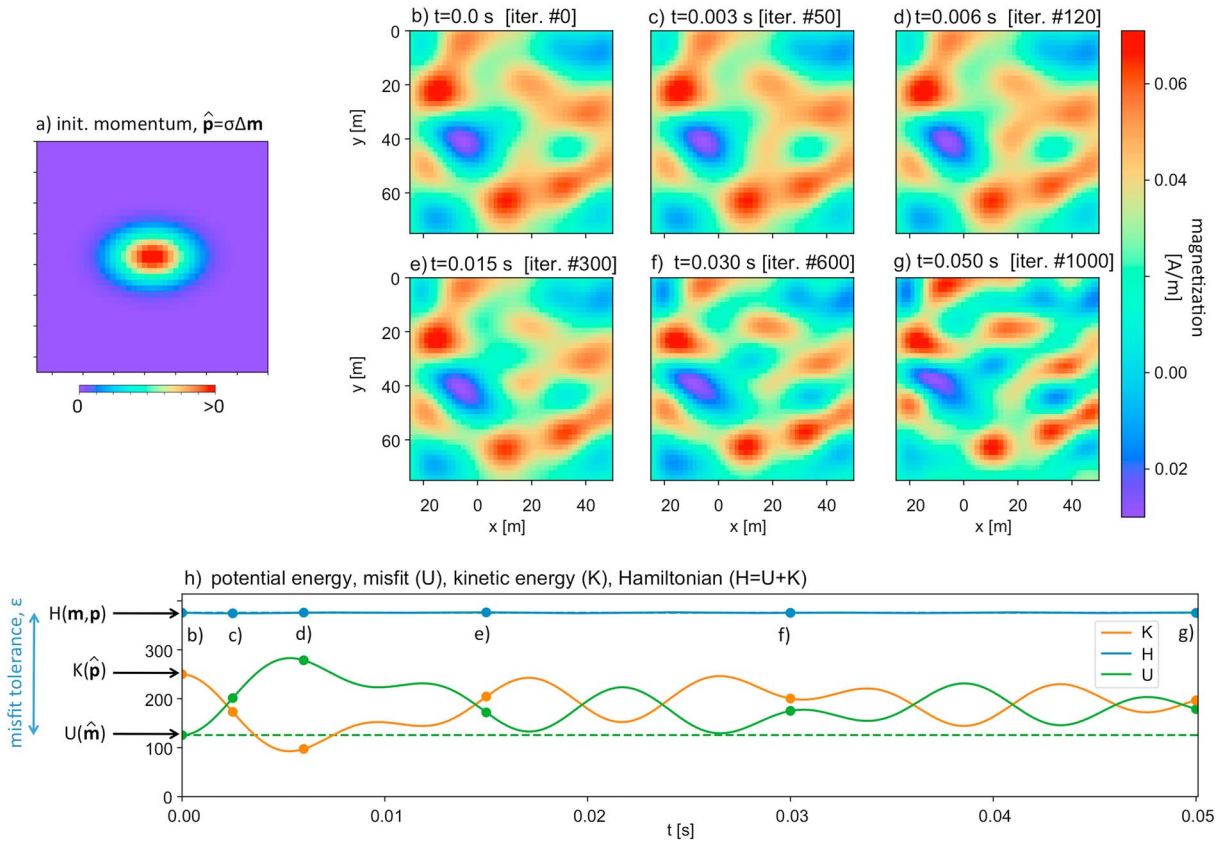

Figure 4. Nullspace exploration for nonlinear magnetic data inversion, similar to Figure 3 for traveltime inversion. (a) Chosen initial momentum $\hat{\mathbf{p}}$, which is proportional to the initial model update $\Delta\mathbf{m}$. (b) Initial (near-optimal) magnetization model $\hat{\mathbf{m}}$. (c–g) Snapshots of magnetization (μ) models at different times along the trajectory. (h) Evolution of potential energy, kinetic energy, and the Hamiltonian along the trajectory.

6. Discussion

We proposed a method to explore the effective nullspace of general nonlinear inverse problems. It rests on the construction of an artificial Hamiltonian system where models are treated as high-dimensional particles. The particles evolve along trajectories, thereby producing new models that are consistent with the observations and their uncertainties. In the following paragraphs, we place the method in the context of other methods, and we elaborate on its most significant advantages and limitations.

The most significant advantage of the nullspace shuttle is its ability to produce alternative, acceptable models without Monte Carlo sampling or inversion restarts from different initial models. In the examples of section 5, a large variety of different models could be produced within a few hundred to a few thousand iterations, despite the high dimension of the model space. An important aspect of the nullspace shuttle is that alternative models can be produced in a guided fashion. For instance, specific features may be added to the extent permitted by the data, and rougher or smoother alternatives may be found. In this sense, the Hamiltonian nullspace shuttle is a tool for efficient hypothesis testing.

Section 3 shows that the Hamiltonian nullspace shuttle can be interpreted as a hybrid between Monte Carlo samplers and gradient-based descent methods. The category into which it falls depends on the tuning, that is, the choice of the mass matrix, the initial momentum, and the tolerance. For instance, repeated runs of the nullspace shuttle with randomly chosen initial momenta generates the test models of an HMC scheme. Selecting or rejecting samples based on a Metropolis rule produces samples of the posterior distribution $e^{-\chi}$ (Betancourt, 2017; Fichtner & Simute, 2018; Fichtner et al., 2018; Neal, 2011). In contrast, vanishing initial momentum produces a family of gradient descent methods, as shown in equation (9). In this sense, the nullspace shuttle may be considered as a superclass of methods that contains Monte Carlo samplers and gradient descent as special cases. (An algorithmic detail that makes the nullspace shuttle different from

HMC is the enforcement of energy conservation at the price of breaking symplecticity of the numerical integrator.)

A limitation of the Hamiltonian nullspace shuttle is its inability to explore the effective nullspace entirely within a finite amount of (computing) time. Alternative, admissible models are found along trajectories that may not fill the effective nullspace. Also, as illustrated in section 5.1, the nullspace shuttle may not be able to cross a high-energy barrier between its current and a neighboring misfit valley. The nullspace shuttle does serve the important purpose of finding other admissible models, but it may not find all of them.

Acknowledgments
This work was supported by the CSCS computing time grant s741 and the European Research Council (ERC) under the EU's Horizon 2020 program (grant 714069). Part of this work has been funded by the DHRTC centre (Denmark) through the project "Outcrop analog studies of chalk—integrated geology, geophysics and geostatistics." The authors gratefully acknowledge constructive comments from Guust Nolet and another anonymous reviewer, as well as fruitful discussions with David Al-Attar, Christian Boehm, Klaus Mosegaard, and Malcolm Sambridge. No data have been used for this work. The code for the computation of traveltimes follows the procedure of Podvin and Lecomte (1991).

References

Backus, G. E., & Gilbert, F. (1970). Uniqueness in the inversion of inaccurate gross Earth data. *Philosophical Transactions of the Royal Society of London Series A*, *266*, 123–192.

Betancourt, M. (2017). A conceptual introduction to Hamiltonian Monte Carlo. arXiv:1701.02434 [stat.ME].

Blakely, R. J. (1996). *Potential theory in gravity and magnetic applications*. Cambridge, UK: Cambridge University Press.

Cerveny, V. (2001). *Seismic ray theory*. Cambridge, UK: Cambridge University Press.

Crank, J., & Nicolson, P. (1947). A practical method for numerical evaluation of solutions of partial differential equations of the heat conduction type. *Proceedings of the Cambridge Philosophical Society*, *43*, 50–67.

de Wit, R. W. L., Trampert, J., & van der Hilst, R. D. (2012). Toward quantifying uncertainty in travel time tomography using the null-space shuttle. *Journal of Geophysical Research*, *117*, 2156–2202. https://doi.org/10.1029/2011JB008754

Deal, M. M., & Nolet, G. (1996). Nullspace shuttles. *Geophysical Journal International*, *124*, 372–380.

Fichtner, A., & Simute, S. (2018). Hamiltonian Monte Carlo inversion of seismic sources in complex media. *Journal of Geophysical Research: Solid Earth*, *123*, 2984–2999. https://doi.org/10.1002/2017JB015249

Fichtner, A., Zunino, A., & Gebraad, L. (2018). Hamiltonian Monte Carlo solution of tomographic inverse problems. *Geophysical Journal International*, *216*(2), 1344–1363. https://doi.org/10.1093/gji/ggy496

Grant, F. (1973). Magnetic susceptibility mapping: The first year's experience. In *43rd Annual International meeting* (pp. 1201–1204), Society of Exploration Geophysicists.

Landau, L. D., & Lifshitz, E. M. (1976). *Course of theoretical physics, volume 1, mechanics* (3rd ed.). Amsterdam: Elsevier Butterworth Heinemann.

Leung, S., & Qian, J. (2006). An adjoint state method for three-dimensional transmission traveltime tomography using first-arrivals. *Communications in Mathematical Sciences*, *4*(1), 249–266.

Li, Y., & Oldenburg, D. W. (2000). Joint inversion of surface and three-component borehole magnetic data. *Geophysics*, *65*(2), 540–552. https://doi.org/10.1190/1.1444749

Menke, W. (2012). *Geophysical data analysis: Discrete inverse theory*. London: Academic Press.

Mosegaard, K., & Tarantola, A. (1995). Monte Carlo sampling of solutions to inverse problems. *Journal of Geophysical Research*, *100*, 12,431–12,447.

Neal, R. M. (2011). MCMC using Hamiltonian dynamics, *Handbook of Markov chain Monte Carlo* (Chapter 5). Boca Raton, FA: Chapman & Hall / CRC Press.

Nocedal, J., & Wright, S. J. (1999). *Numerical optimization*. New York: Springer.

Nolet, G. (2008). *A breviary of seismic tomography*. Cambridge, UK: Cambridge University Press.

Parker, R. L. (1994). *Geophysical inverse theory*. Princeton, NJ: Princeton University Press.

Podvin, P., & Lecomte, I. (1991). Finite difference computation of traveltimes in very contrasted velocity models: A massively parallel approach and its associated tools. *Geophysical Journal International*, *105*, 271–284. https://doi.org/10.1111/j.1365-246X.1991.tb03461.x

Quispel, G. R. W., & McLaren, D. I. (2008). A new class of energy-preserving numerical integration methods. *Journal of Physics A*, *41*, 045206. https://doi.org/10.1088/1751–8113/41/4/045,206

Rawlinson, N., & Sambridge, M. (2004). Wave front evolution in strongly heterogeneous layered media using the fast marching method. *Geophysical Journal International*, *156*(3), 631–647. https://doi.org/10.1111/j.1365-246X.2004.02153.x

Sambridge, M. S., & Mosegaard, K. (2002). Monte Carlo methods in geophysical inverse problems. *Reviews of Geophysics*, *40*(3), 1009. https://doi.org/10.1029/2000RG000089

Sethian, J. A. (1996). A fast marching level set method for monotonically advancing fronts. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(4), 1591–1595.

Simo, J. C., Tarnow, N., & Wong, K. K. (1992). Exactly energy-momentum conserving algorithms and symplectic schemes for nonlinear dynamics. *Computer Methods in Applied Mechanics and Engineering*, *100*, 63–116.

Symon, K. R. (1971). *Mechanics*. Reading, MA: Addison-Wesley.

Taillandier, C., Noble, M., Chauris, H., & Calandra, H. (2009). First arrival travel time tomography based on the adjoint state methods. *Geophysics*, *74*(6), WCB57–WCB66.

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation* (2nd ed.). Philadephia: Society for Industrial and Applied Mathematics.

Zunino, A., Benvenuto, F., Armadillo, E., Bertero, M., & Bozzo, E. (2009). Iterative deconvolution and semiblind deconvolution methods in magnetic archaeological prospecting. *Geophysics*, *74*(4), L43—L51. https://doi.org/10.1190/1.3129263

Zunino, A., & Mosegaard, K. (2018). Integrating gradient information with probabilistic traveltime tomography using the Hamiltonian Monte Carlo algorithm. In *80th EAGE Conference & Exhibition*, Copenhagen. https://doi.org/10.3997/2214-4609.201801971