Uncertainty quantification for large-scale nonlinear inverse problems

with some focus on seismic tomography

Andreas Fichtner

Andrea Zunino

Lars Gebraad

ETH Seismology & Wave Physics

faster supercomputers + improved numerical methods to solve the wave equation

faster supercomputers + improved numerical methods to solve the wave equation

properly account for

wave propagation through complex 3-D Earth models

finite-frequency sensitivity

nonlinear dependence of observations on Earth model parameters [velocities, attenuation, ...]

faster supercomputers + improved numerical methods to solve the wave equation

properly account for

wave propagation through complex 3-D Earth models

finite-frequency sensitivity

nonlinear dependence of observations on Earth model parameters [velocities, attenuation, ...]

reduced ability to

properly quantify uncertainties [machinery for linear inverse problems fails]

find alternative models that explain data equally well

interpret (tomographic) Earth models with high confidence

Goals of this Talk

Efficient method to produce alternative Earth models that explain the data equally well.

New Monte Carlo method for comprehensive uncertainty quantification in large nonlinear inverse problems.

Automatic tuning of Monte Carlo sampling on the fly using quasi-Newton methods.

Goals of this Talk		Outline
Efficient method to produce alternative Earth models that explain the data equally well.	→	PART I Nonlinear Nullspace Shuttles
New Monte Carlo method for comprehensive uncertainty quantification in large nonlinear inverse problems.	→	PART II Hamiltonian Monte Carlo Tomography
Automatic tuning of Monte Carlo sampling on the fly using quasi-Newton methods.	→	PART III Autotuning


PART I
Nonlinear Nullspace Shuttles

1.1. Preparations and Problem Statement

The effective nullspace:

- misfit functional: $\chi(\mathbf{m})$
- acceptable model: $\widehat{\mathbf{m}}$

The effective nullspace:

- misfit functional: $\chi(\mathbf{m})$
- acceptable model: $\widehat{\mathbf{m}}$
- effective nullspace = all models such that $\chi(\widehat{\mathbf{m}} + \Delta\mathbf{m}) \leq \chi(\widehat{\mathbf{m}}) + \varepsilon$

misfit tolerance

The effective nullspace:

- misfit functional: $\chi(\mathbf{m})$
- acceptable model: $\widehat{\mathbf{m}}$
- effective nullspace = all models such that $\chi(\widehat{\mathbf{m}} + \Delta\mathbf{m}) \leq \chi(\widehat{\mathbf{m}}) + \varepsilon$

Problem:

- How do we find models in the effective nullspace?

The effective nullspace:

- misfit functional: $\chi(\mathbf{m})$
- acceptable model: $\widehat{\mathbf{m}}$
- effective nullspace = all models such that $\chi(\widehat{\mathbf{m}} + \Delta\mathbf{m}) \leq \chi(\widehat{\mathbf{m}}) + \varepsilon$

Problem:

- How do we find models in the effective nullspace?

Why is this relevant?

- Uncertainty analysis. Are there alternative models that are very different?
- Construction of alternative models that contain some new feature.

1.2. Take-off

The trick – Part I:

- interpret any model as the position of a space shuttle in high dimension
- assign an artificial potential energy: $U(\mathbf{m}) = \chi(\mathbf{m})$

The trick – Part I:

- interpret any model as the position of a space shuttle in high dimension
- assign an artificial potential energy: $U(\mathbf{m}) = \chi(\mathbf{m})$
- assign an artificial momentum **p**.
- define an artificial kinetic energy: $K(\mathbf{p}) = \frac{1}{2}\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p}$

The trick – Part I:

- interpret any model as the position of a space shuttle in high dimension
- assign an artificial potential energy: $U(\mathbf{m}) = \chi(\mathbf{m})$
- assign an artificial momentum **p**.
- define an artificial kinetic energy: $K(\mathbf{p}) = \frac{1}{2}\mathbf{p}^T \mathbf{M}^{-1} \mathbf{p}$
- with some initial momentum such that: $K(\widehat{\mathbf{p}}) = \widehat{\mathbf{p}}^T \mathbf{M}^{-1} \widehat{\mathbf{p}}/2 = \varepsilon$
- total energy (Hamiltonian): $H = K + U$

The trick – Part II:

- let the shuttle fly along a trajectory determined by Hamilton's equations

$$\frac{dm_i}{dt} = \frac{\partial H}{\partial p_i}, \qquad \frac{dp_i}{dt} = -\frac{\partial H}{\partial m_i}$$

The trick – Part II:

- let the shuttle fly along a trajectory determined by Hamilton's equations

$$\frac{dm_i}{dt} = \frac{\partial H}{\partial p_i}, \qquad \frac{dp_i}{dt} = -\frac{\partial H}{\partial m_i}$$

- *H* is constant along a trajectory:

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

total energy after time *t* — *U(t)* — *K(t)*

The trick – Part II:

- let the shuttle fly along a trajectory determined by Hamilton's equations

$$\frac{dm_i}{dt} = \frac{\partial H}{\partial p_i}, \qquad \frac{dp_i}{dt} = -\frac{\partial H}{\partial m_i}$$

- *H* is constant along a trajectory:

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

total energy after time *t* | *U(t)* | *K(t)* | initial total energy

The trick – Part II:

- let the shuttle fly along a trajectory determined by Hamilton's equations

$$\frac{dm_i}{dt} = \frac{\partial H}{\partial p_i}, \qquad \frac{dp_i}{dt} = -\frac{\partial H}{\partial m_i}$$

- *H* is constant along a trajectory:

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

total energy after time *t* | *U(t)* | *K(t)* | initial total energy | *U(0)* | *K(0)*

The trick – Part II:

- let the shuttle fly along a trajectory determined by Hamilton's equations

$$\frac{dm_i}{dt} = \frac{\partial H}{\partial p_i}, \qquad \frac{dp_i}{dt} = -\frac{\partial H}{\partial m_i}$$

- *H* is constant along a trajectory:

$$\chi[\mathbf{m}(t)] + \frac{1}{2}\mathbf{p}(t)^T\mathbf{M}^{-1}\mathbf{p}(t) = \qquad \chi(\widehat{\mathbf{m}}) + \frac{1}{2}\widehat{\mathbf{p}}^T\mathbf{M}^{-1}\widehat{\mathbf{p}}$$

U(t) *K(t)>0* *U(0)* *tolerance ε*

The trick – Part II:

- let the shuttle fly along a trajectory determined by Hamilton's equations

$$\frac{dm_i}{dt} = \frac{\partial H}{\partial p_i}, \qquad \frac{dp_i}{dt} = -\frac{\partial H}{\partial m_i}$$

- H is constant along a trajectory:

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

total energy after time t | $U(t)$ | $K(t)>0$ | $U(0)$ | tolerance ε

➢ It follows that

$$\chi[\mathbf{m}(t)] \leq \chi(\widehat{\mathbf{m}}) + \varepsilon$$

➢ **All models along the trajectory are indeed in the effective nullspace!**

1.3. Navigating the Nullspace

<u>Depending on various choices the shuttle probes different parts of the nullspace</u>

- Zero-tolerance case:
  - Nullspace shuttle = **gradient descent**
  - Type of descent method depends on choice of the mass matrix

Depending on various choices the shuttle probes different parts of the nullspace

- Zero-tolerance case:
  - Nullspace shuttle = **gradient descent**
  - Type of descent method depends on choice of the mass matrix

- Prescribed take-off direction
  - Adding specific features to the model *a posteriori*

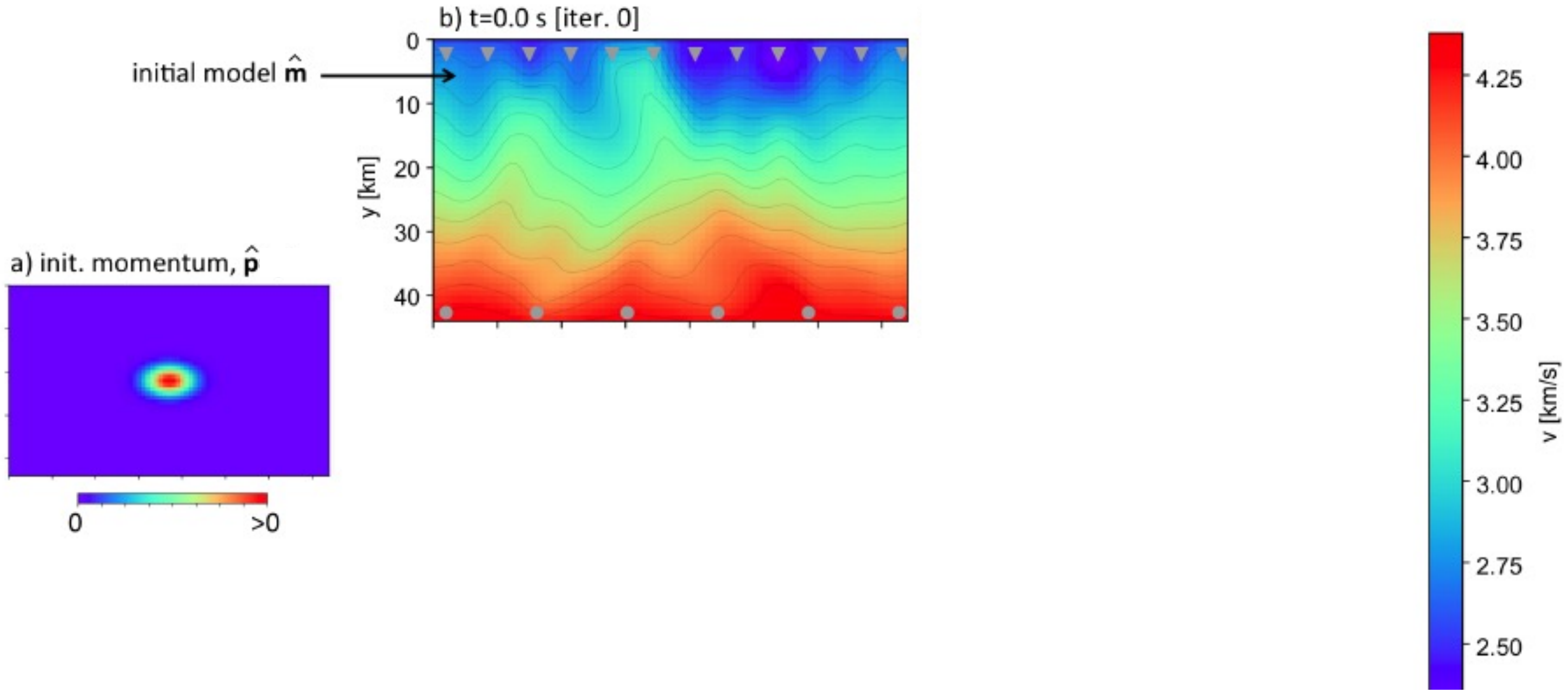
Depending on various choices the shuttle probes different parts of the nullspace

- Zero-tolerance case:
  - Nullspace shuttle = **gradient descent**
  - Type of descent method depends on choice of the mass matrix

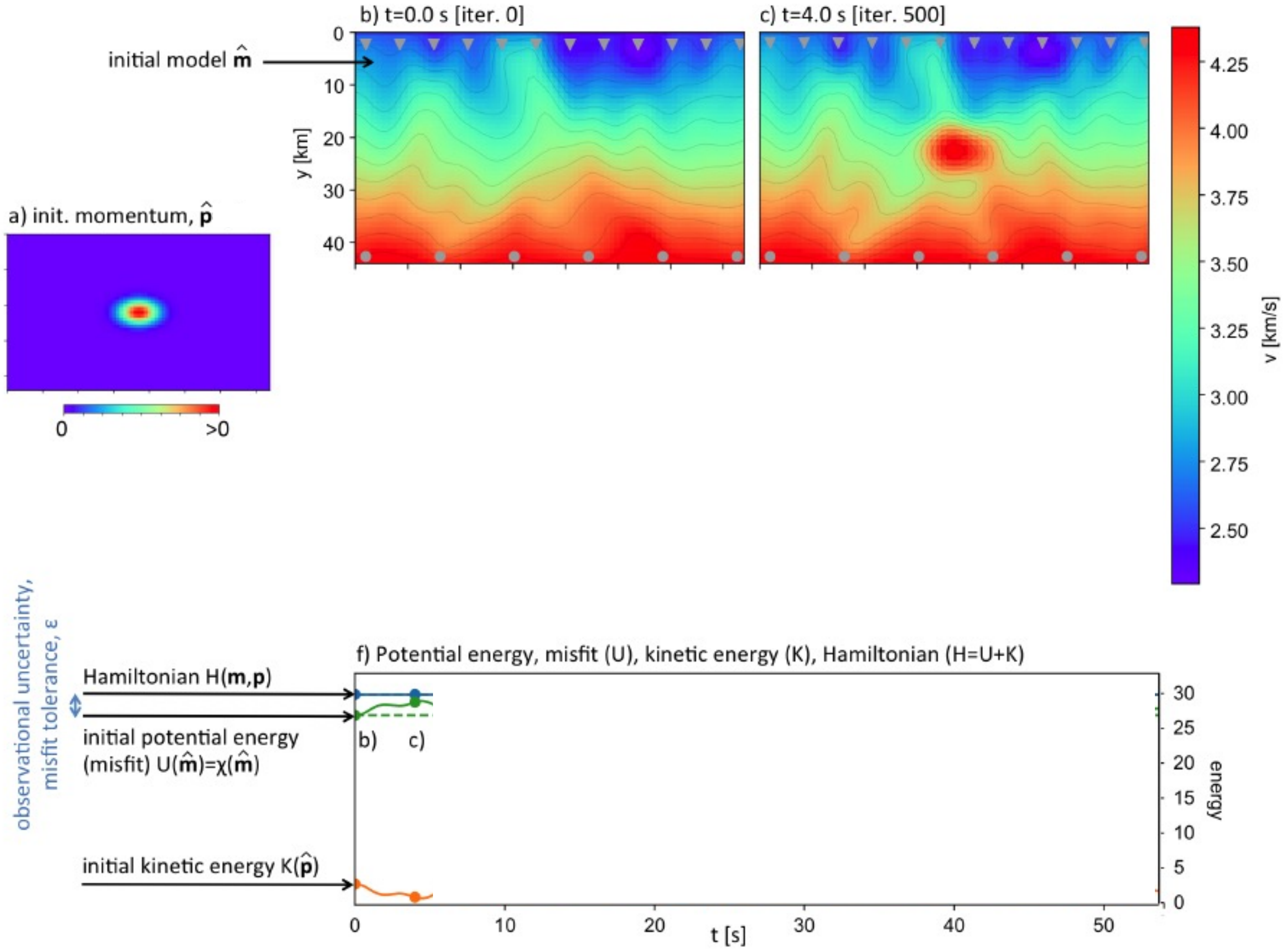
- Prescribed take-off direction
  - Adding specific features to the model *a posteriori*

- Rough and smooth parts of nullspace

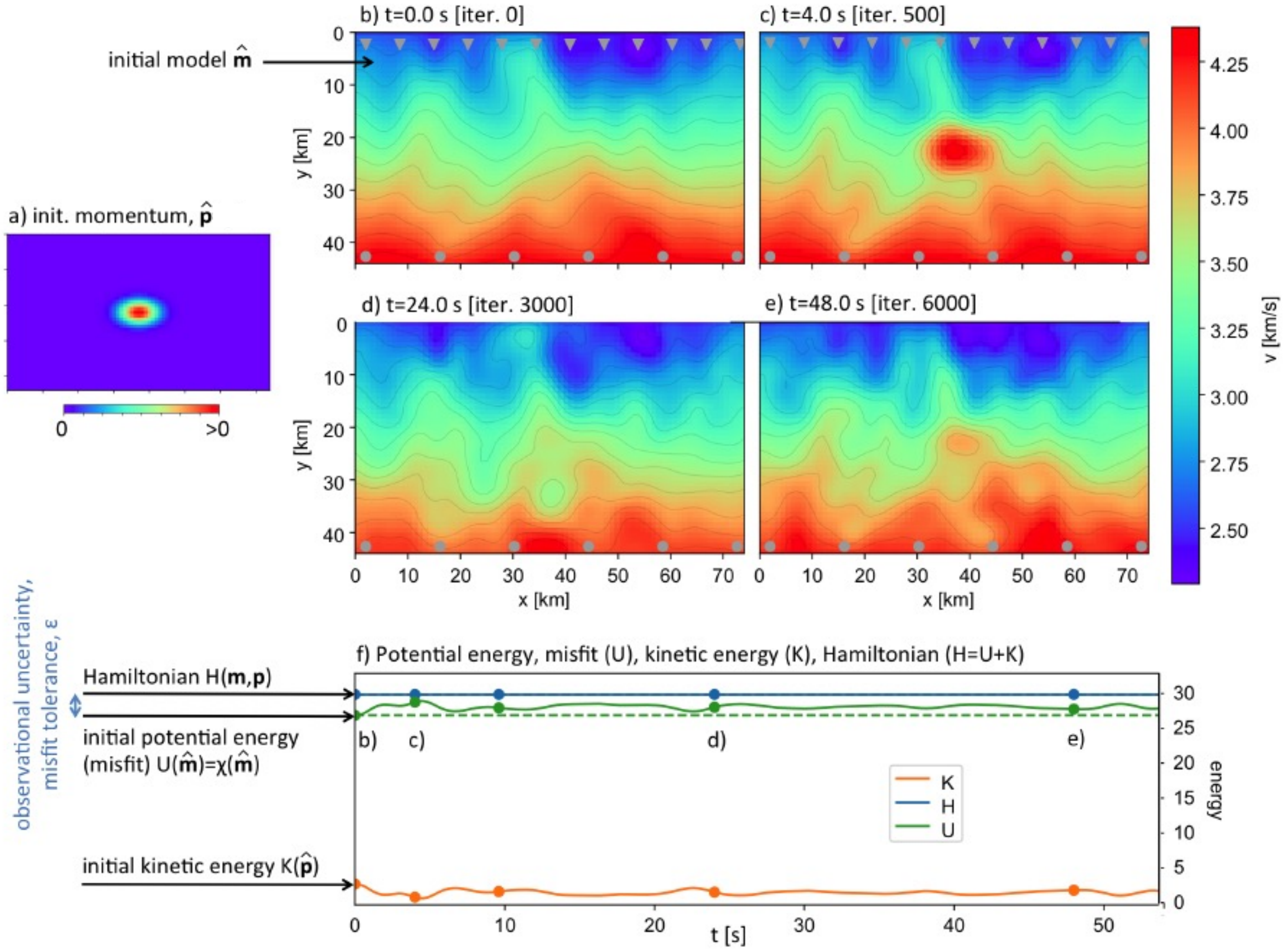
Example: Nonlinear traveltime tomography



Example: Nonlinear traveltime tomography



Example: Nonlinear traveltime tomography



Example: Nonlinear traveltime tomography

Example: Nonlinear traveltime tomography (random perturbations)

ETH Seismology & Wave Physics

PART II
Hamiltonian Monte Carlo Tomography

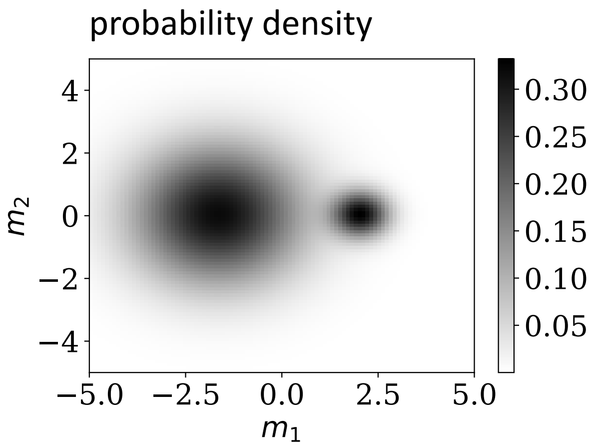
2.1. Motivation

Sampling model space efficiently

Hamiltonian Monte Carlo

introduced as *hybrid* Monte Carlo in quantum mechanics [Duane et al. 1987]

- **Random walk** method to sample posterior probability distribution of an inverse problem.
- Motivated by well-known **deficiency of Metropolis-Hastings** algorithm:

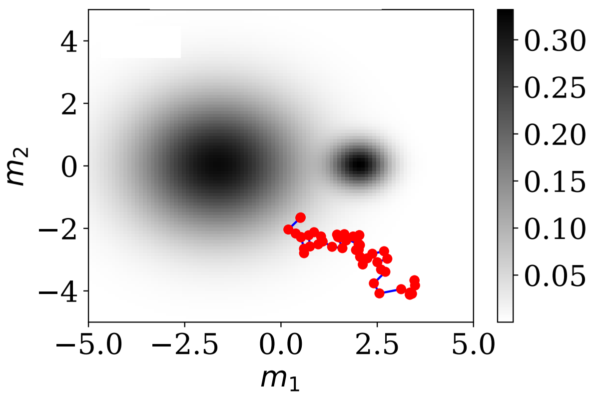


Hamiltonian Monte Carlo

introduced as *hybrid* Monte Carlo in quantum mechanics [Duane et al. 1987]

- **Random walk** method to sample posterior probability distribution of an inverse problem.
- Motivated by well-known **deficiency of Metropolis-Hastings** algorithm:

small steps
high acceptance rate
slow model space exploration

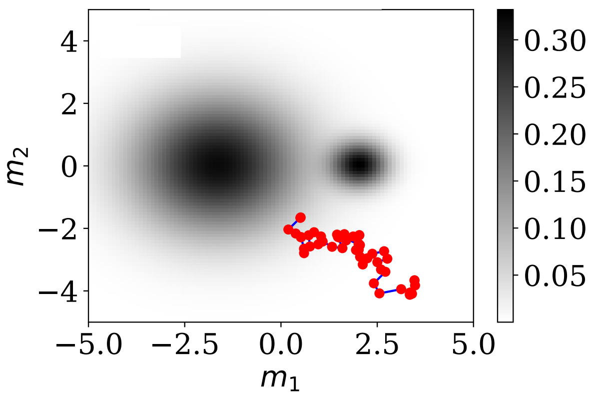


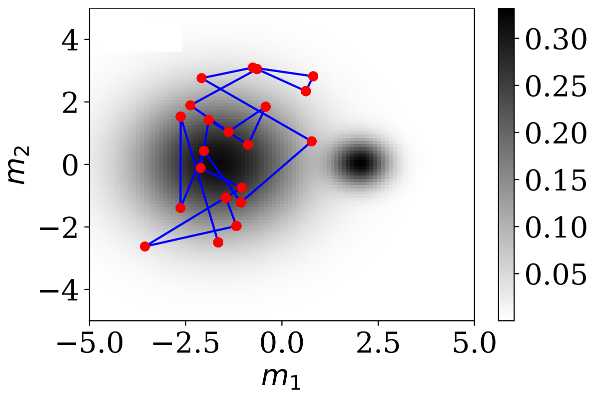
Hamiltonian Monte Carlo

introduced as *hybrid* Monte Carlo in quantum mechanics [Duane et al. 1987]

- **Random walk** method to sample posterior probability distribution of an inverse problem.
- Motivated by well-known **deficiency of Metropolis-Hastings** algorithm:

small steps
high acceptance rate
slow model space exploration





big steps
fast model space exploration
low acceptance rate

Hamiltonian Monte Carlo

introduced as *hybrid* Monte Carlo in quantum mechanics [Duane et al. 1987]

- **Random walk** method to sample posterior probability distribution of an inverse problem.
- Motivated by well-known **deficiency of Metropolis-Hastings** algorithm:

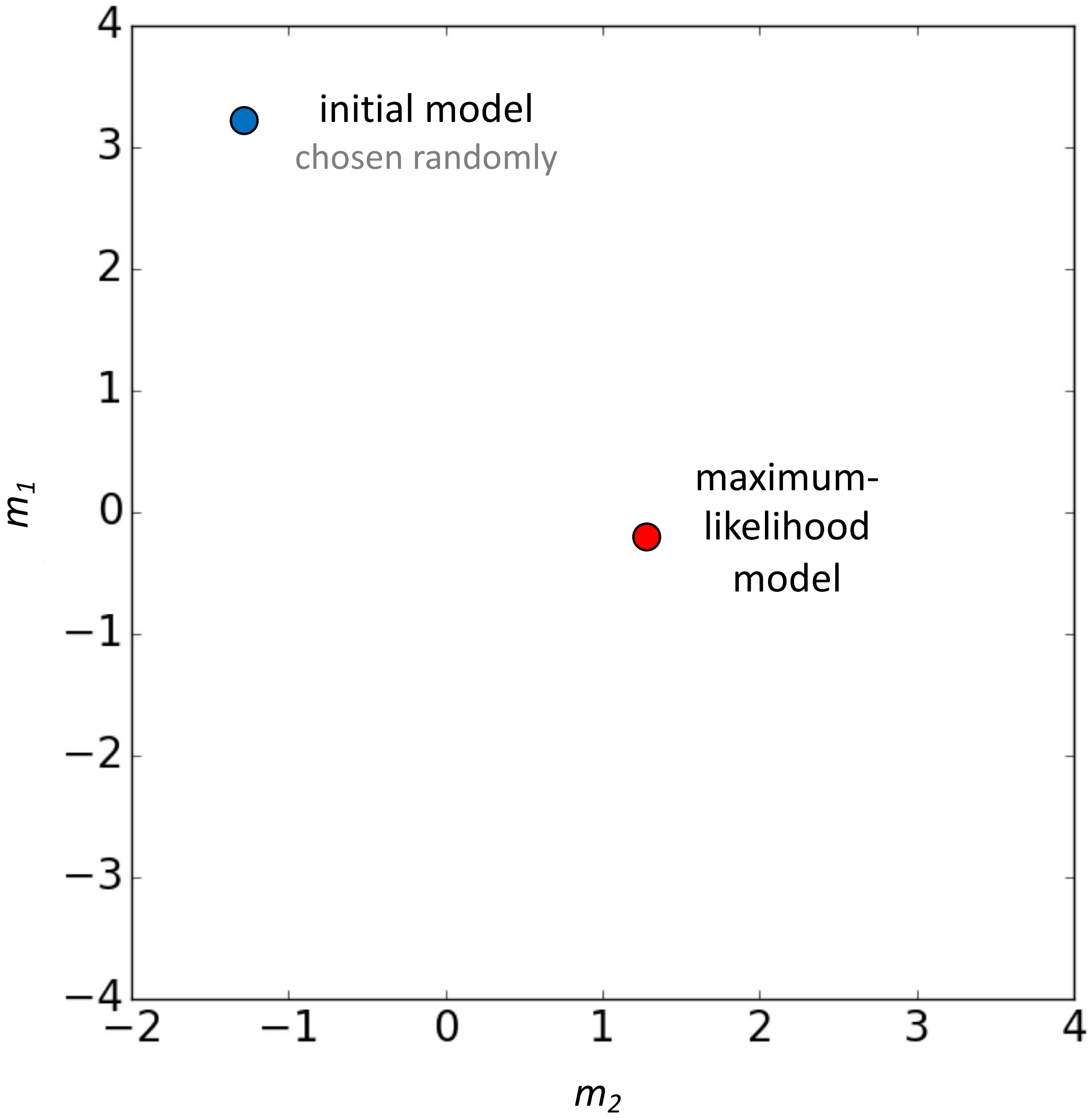
small steps
high acceptance rate
slow model space exploration

big steps
fast model space exploration
low acceptance rate

- **Hamiltonian Monte Carlo**
  - taking advantage of derivative information
  - **long-distance moves + high acceptance rate**
  - ➢ solve **high-dimensional** problems

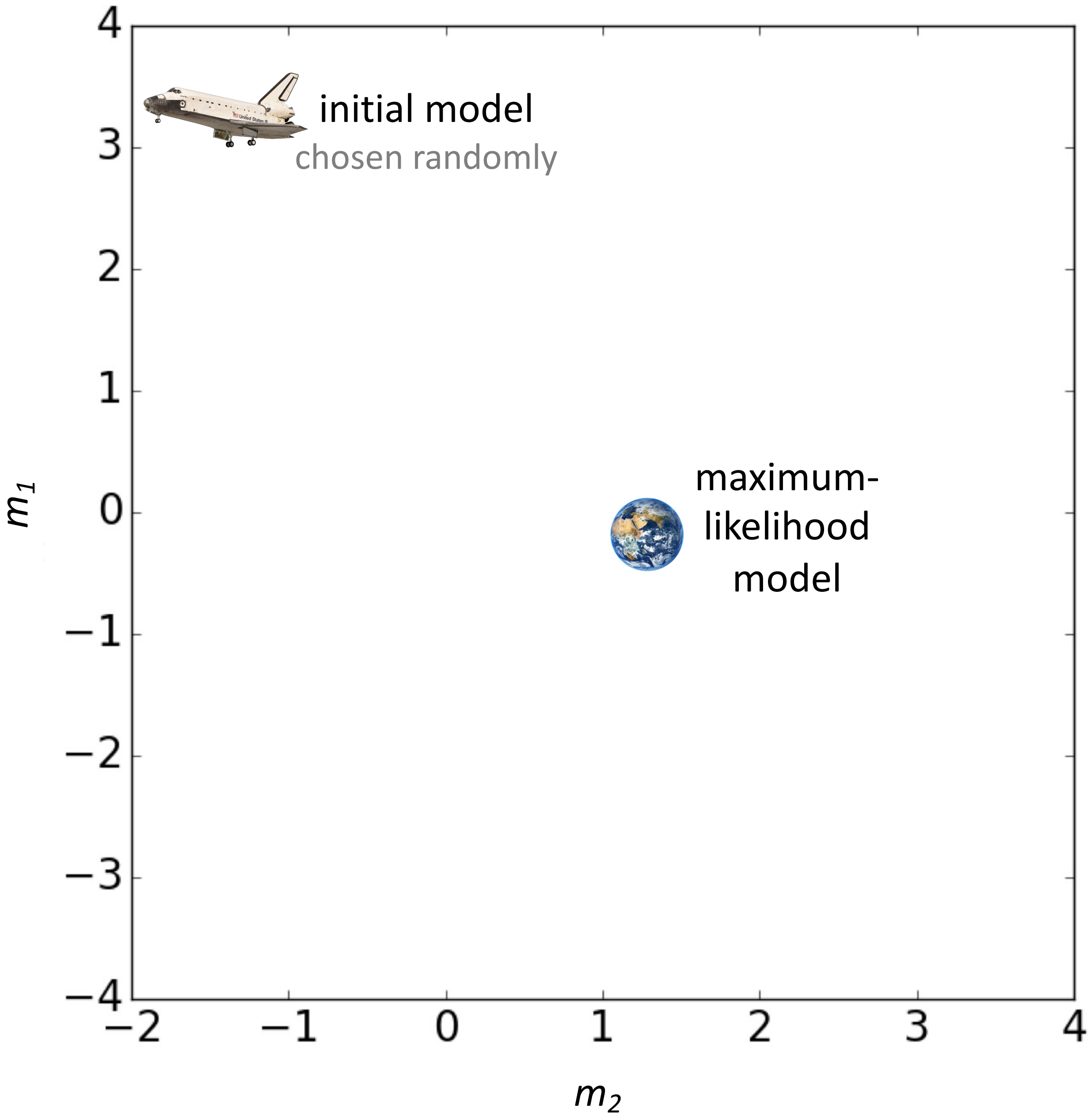
2.2. Conceptual Introduction

Hamiltonian Monte Carlo in pictures

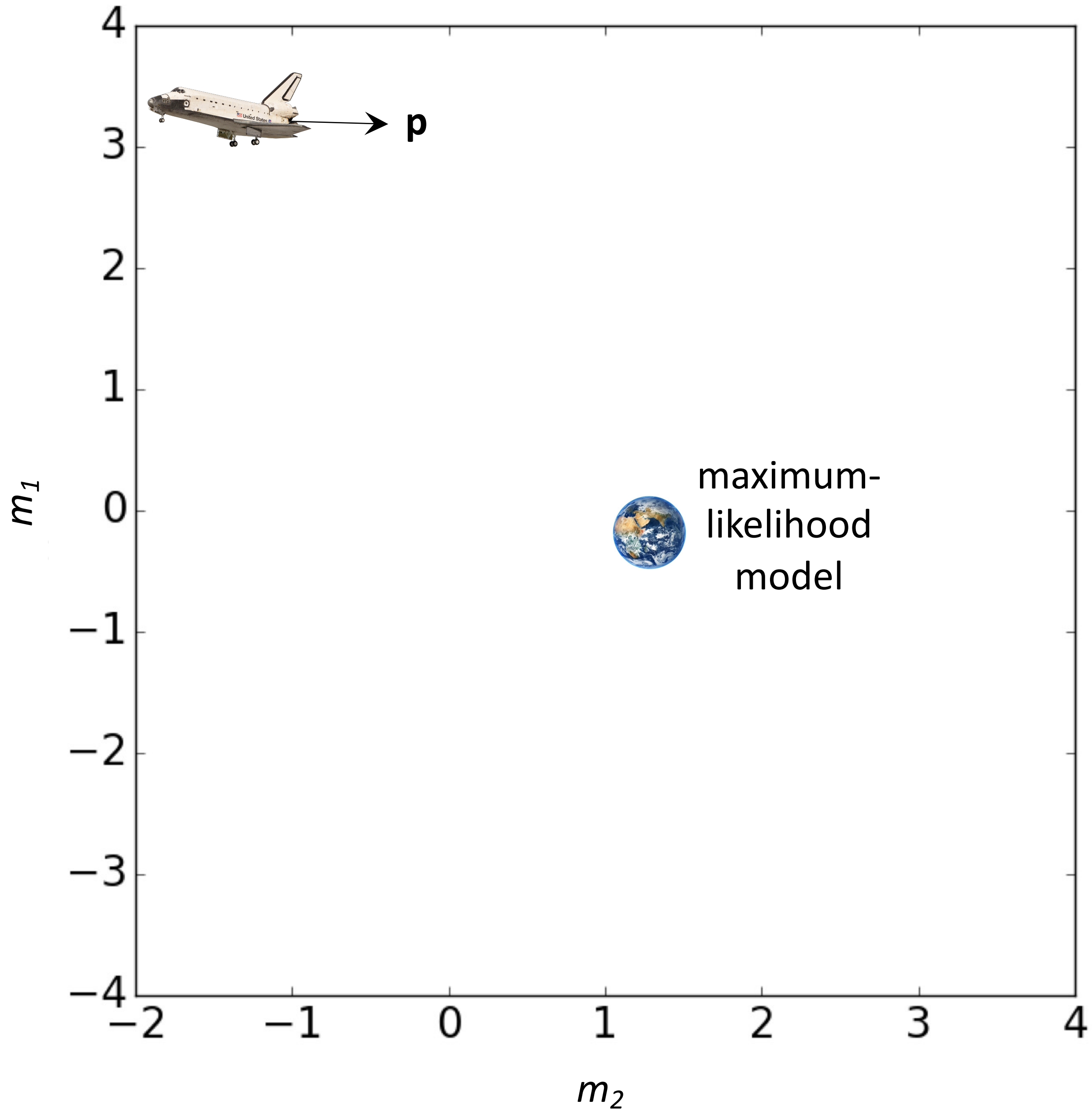


1. Initial model **m**, chosen randomly.

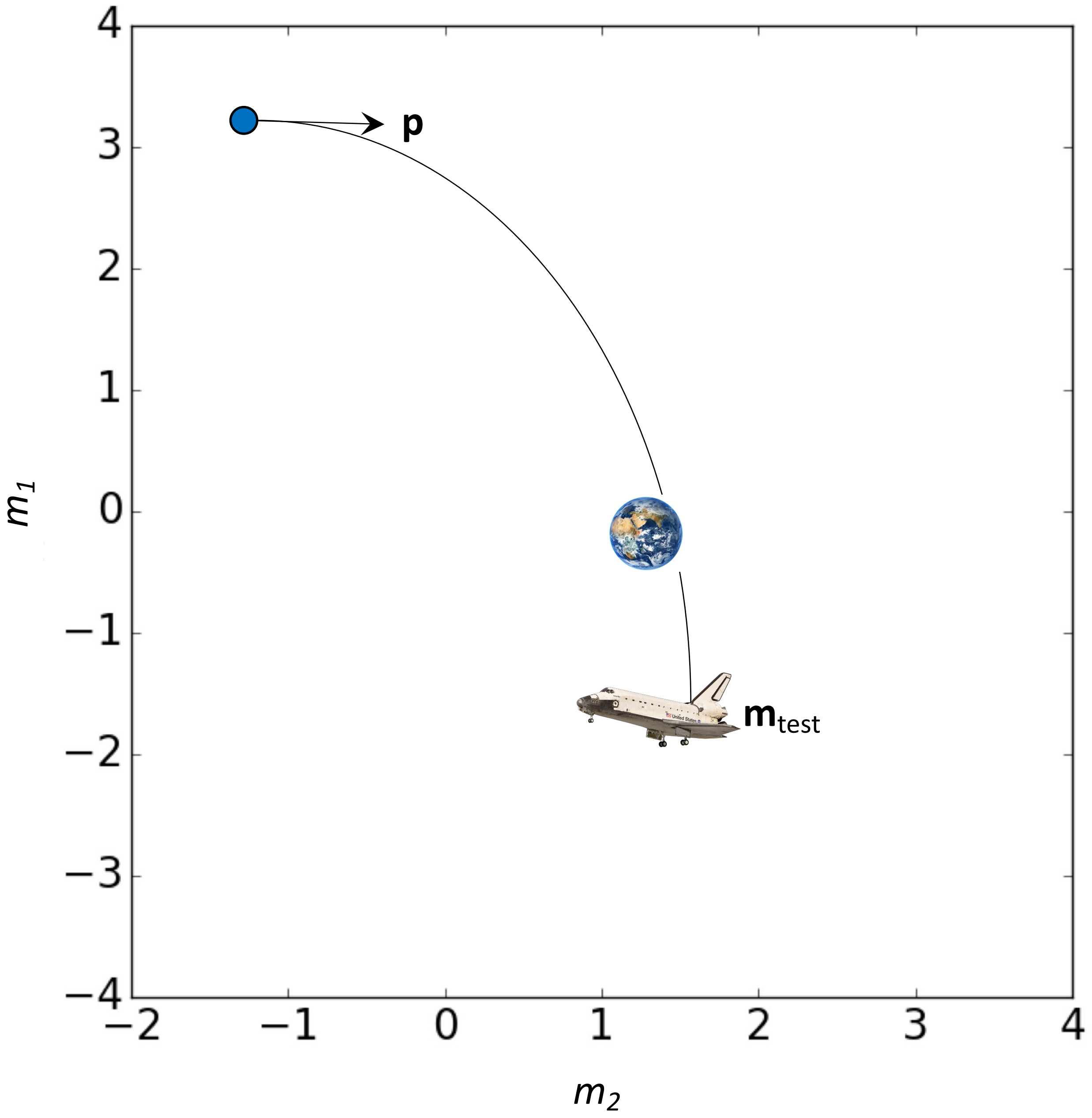
2. Misfit U(**m**) defines <u>potential energy</u>.



1. Initial model **m**, chosen randomly.
2. Misfit U(**m**) defines potential energy.



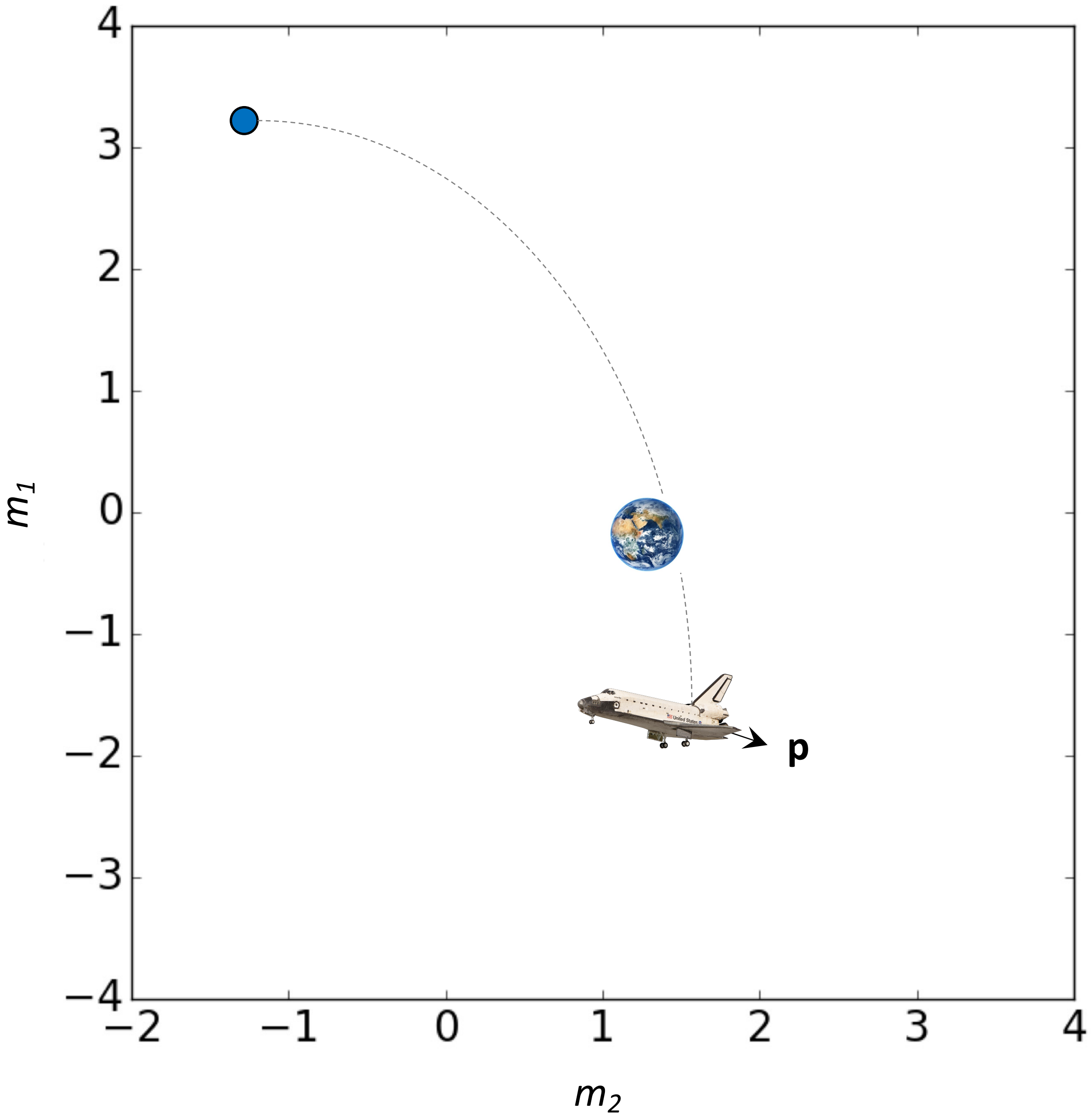
1. Initial model **m**, chosen randomly.
2. Misfit U(**m**) defines potential energy.
3. Random momentum **p** [auxiliary quantity].
4. Defines kinetic energy K(**p**).



1. Initial model **m**, chosen randomly.
2. Misfit U(**m**) defines potential energy.
3. Random momentum **p** [auxiliary quantity].
4. Defines kinetic energy K(**p**).
5. Solve Hamilton's equations with Hamiltonian

   H(**m**,**p**) = U(**m**) + K(**p**) .

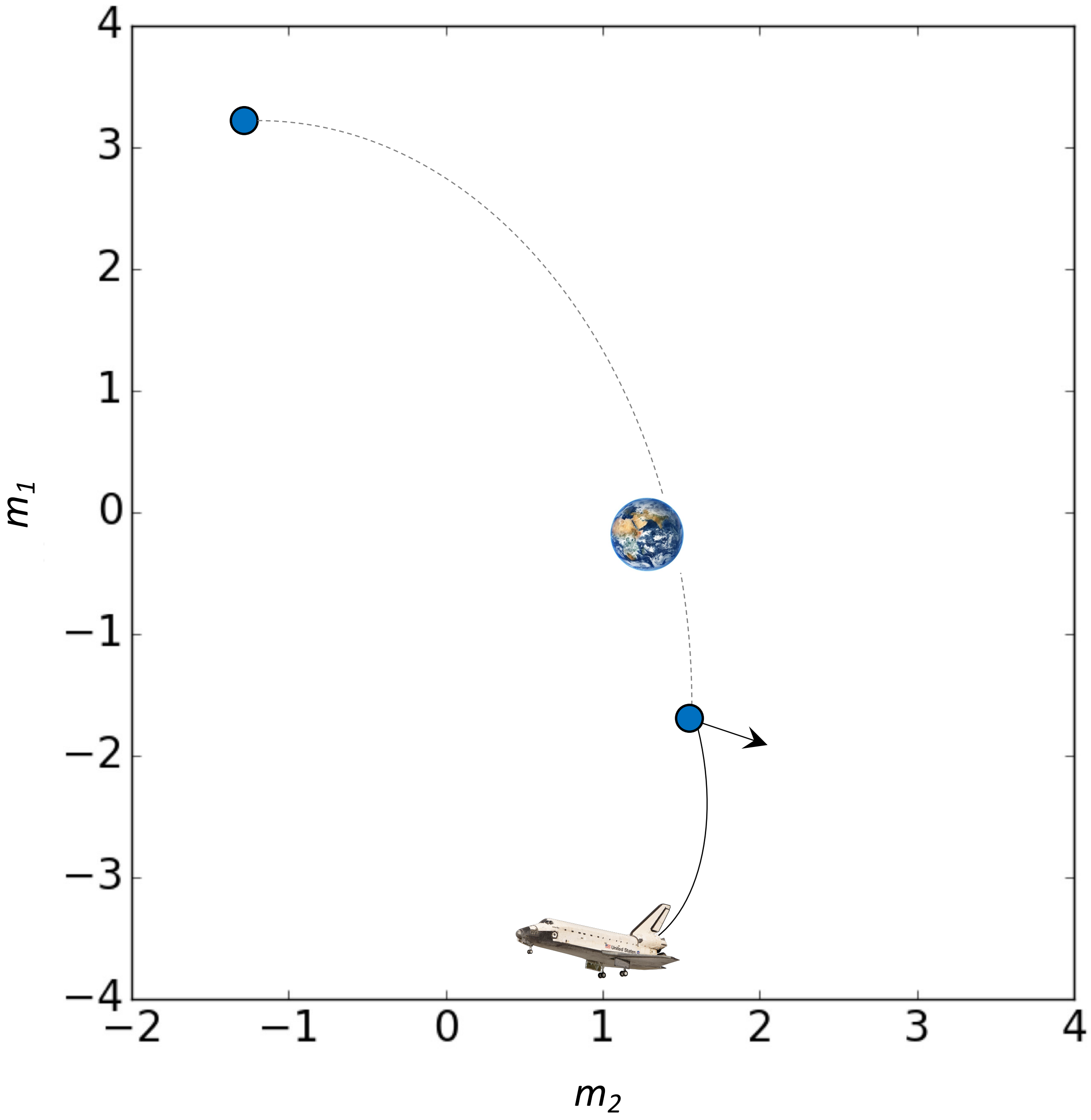
6. Move towards a new test model, $\mathbf{m}_{\text{test}}$.



1. Initial model **m**, chosen randomly.
2. Misfit U(**m**) defines potential energy.
3. Random momentum **p** [auxiliary quantity].
4. Defines kinetic energy K(**p**).
5. Solve Hamilton's equations with Hamiltonian

   H(**m**,**p**) = U(**m**) + K(**p**) .

6. Move towards a new test model, $\mathbf{m}_{test}$.
7. Evaluate Metropolis rule:
   - If rejected: go back.
   - If accepted: move on.



1. Initial model **m**, chosen randomly.

2. Misfit U(**m**) defines potential energy.

3. Random momentum **p** [auxiliary quantity].

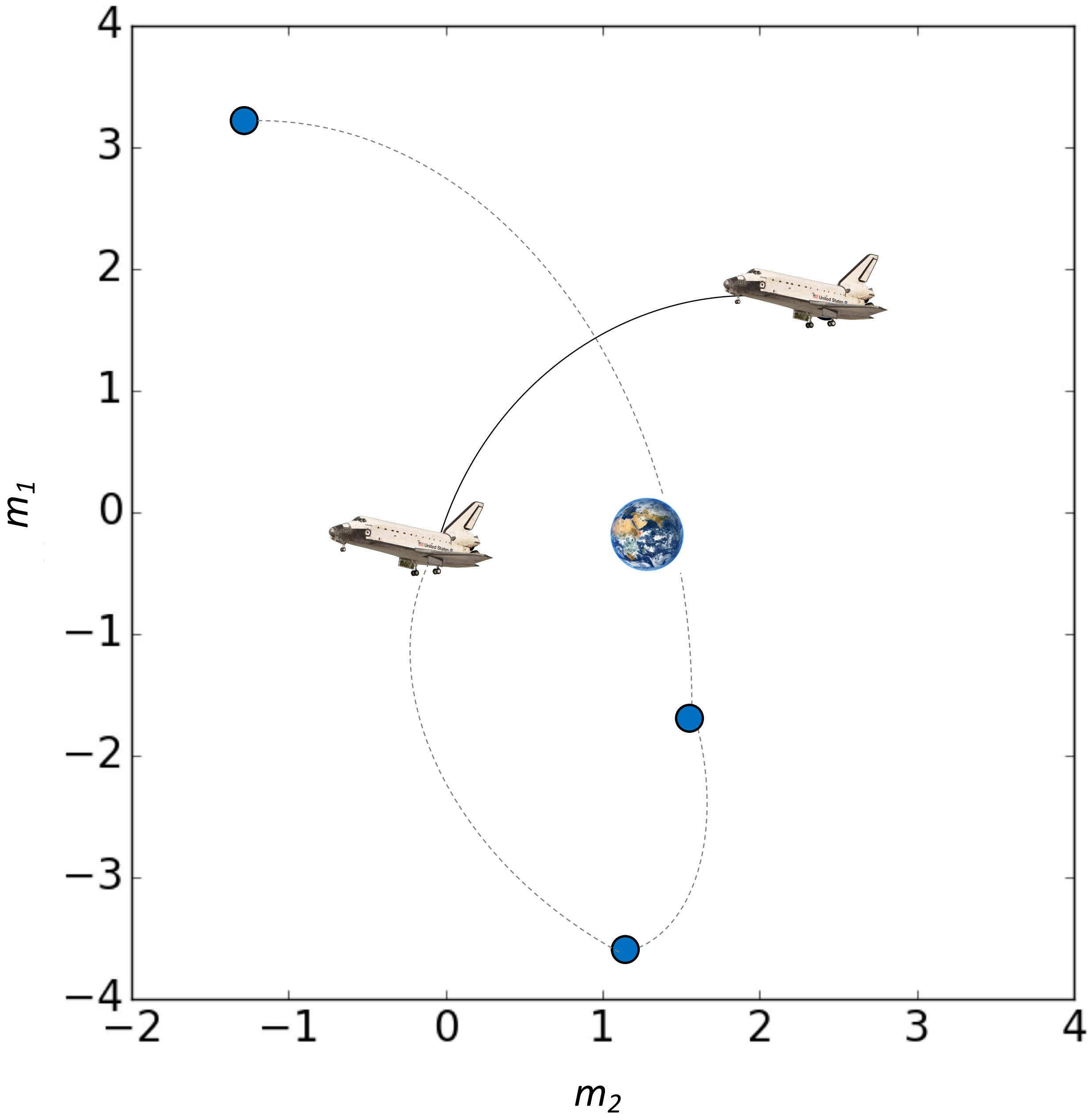
4. Defines kinetic energy K(**p**).

5. Solve Hamilton's equations with Hamiltonian

   H(**m**,**p**) = U(**m**) + K(**p**) .

6. Move towards a new test model, $\mathbf{m}_{test}$.

7. Evaluate Metropolis rule:
   - If rejected: go back.
   - If accepted: move on.



1. Initial model **m**, chosen randomly.
2. Misfit U(**m**) defines potential energy.
3. Random momentum **p** [auxiliary quantity].
4. Defines kinetic energy K(**p**).
5. Solve Hamilton's equations with Hamiltonian

   H(**m**,**p**) = U(**m**) + K(**p**) .

6. Move towards a new test model, $\mathbf{m}_{test}$.
7. Evaluate Metropolis rule:
   - If rejected: go back.
   - If accepted: move on.

Key features

Pros:

- Trajectories orbit around plausible models. [Earth stays near the Sun.]
- Long-distance moves still plausible.
- Fast model space exploration.

Cons:

- Requires derivatives of the forward problem.

Key features

Pros:

- Trajectories orbit around plausible models. [Earth stays near the Sun.]
- Long-distance moves still plausible.
- Fast model space exploration.

~~Cons~~:

- Requires derivatives of the forward problem.
- Easy thanks to adjoint techniques.

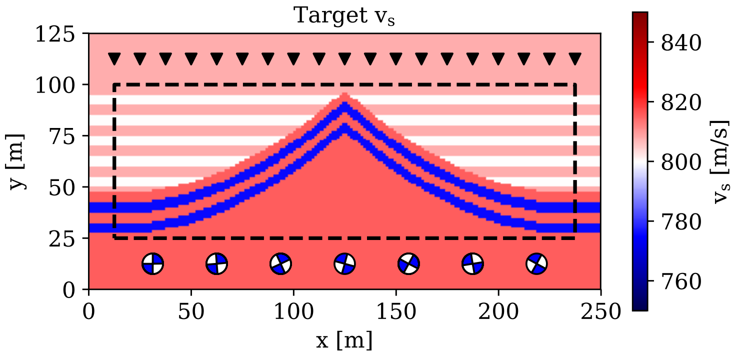
2.3. Towards Applications

Probabilistic full-waveform inversion

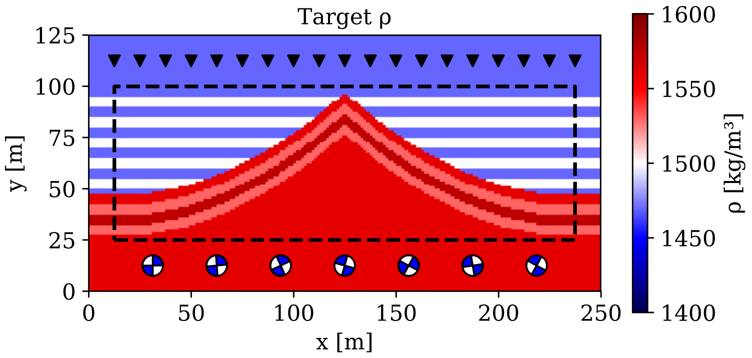
- **2D elastic wave propagation** [staggered-grid FD, f_{max}=50 Hz].
- Model parameters: v_p, v_s, ρ.
- Grid points: 10’800.
- **Model space dimension: 32’400**

- **2D elastic wave propagation** [staggered-grid FD, f_{max}=50 Hz].
- Model parameters: v_p, v_s, ρ.
- Grid points: 10’800.
- **Model space dimension: 32’400**

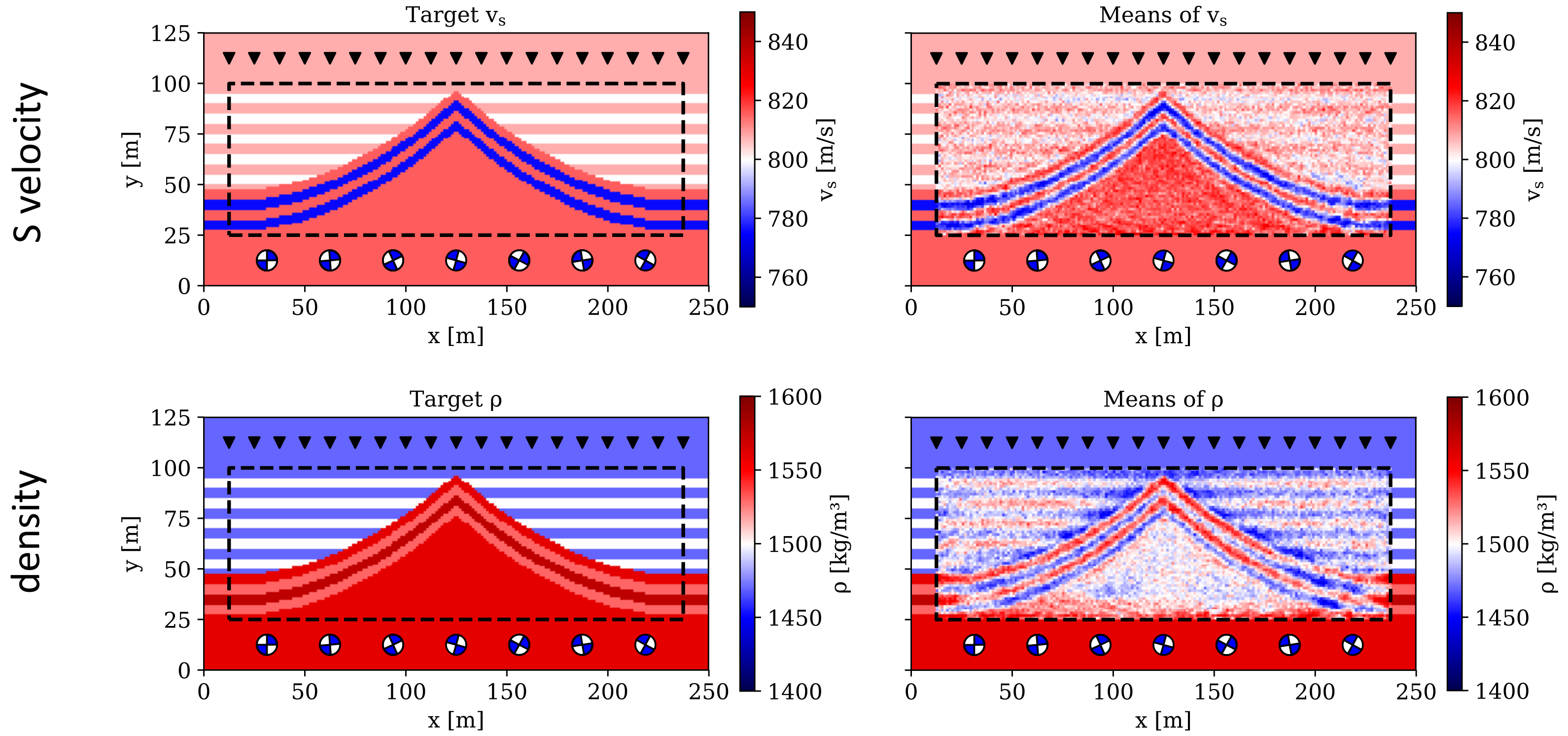
S velocity



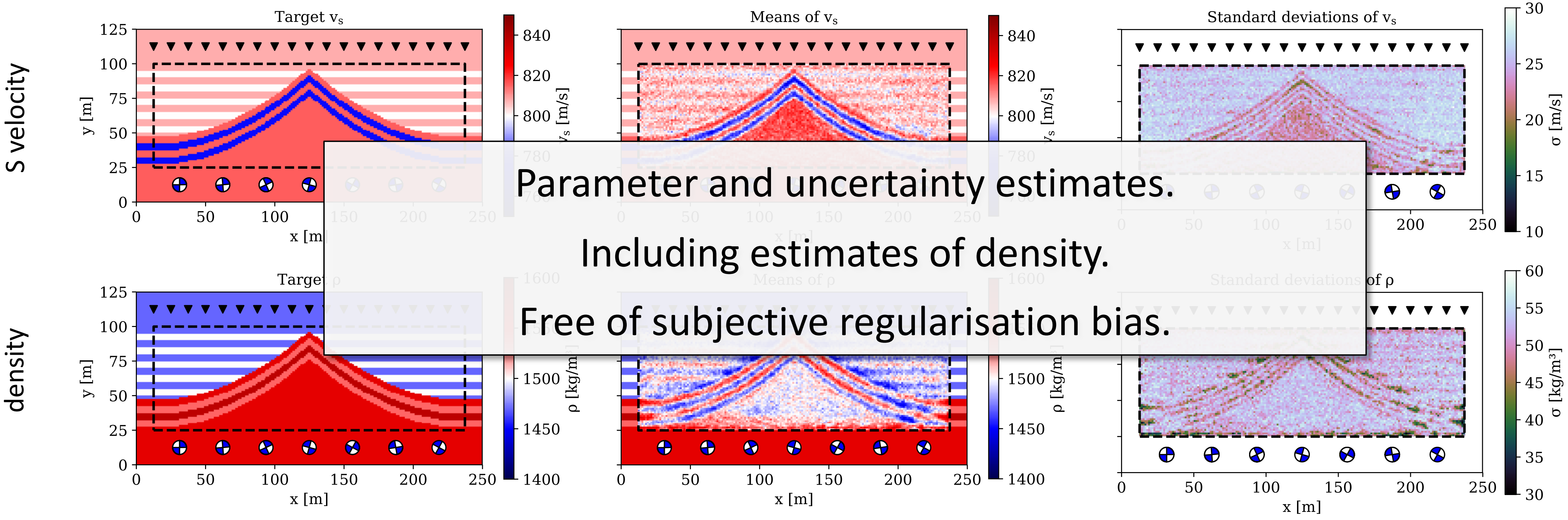
density



- **2D elastic wave propagation** [staggered-grid FD, f_{max}=50 Hz].
- Model parameters: v_p, v_s, ρ.
- Grid points: 10’800.
- **Model space dimension: 32’400**



- **2D elastic wave propagation** [staggered-grid FD, f_{max}=50 Hz].
- Model parameters: v_p, v_s, ρ.
- Grid points: 10’800.
- **Model space dimension: 32’400**



PART III
Autotuning

The actual challenge:

Drawing independent models

The actual challenge:

Drawing independent models

important for

Efficient model- or null-space exploration

Convergence error of Monte Carlo integrals $\propto 1/\sqrt{\mathrm{N}_{\mathrm{independent}}}$

Example 1: 1000-D Gaussian

Covariance matrix: $C_{1,1}$=0.100, $C_{2,2}$=0.101, ..., $C_{1000,1000}$=1.100

Mass matrix: **M**=**I**

auto-correlation of sample chain

measure of the independence of successive samples

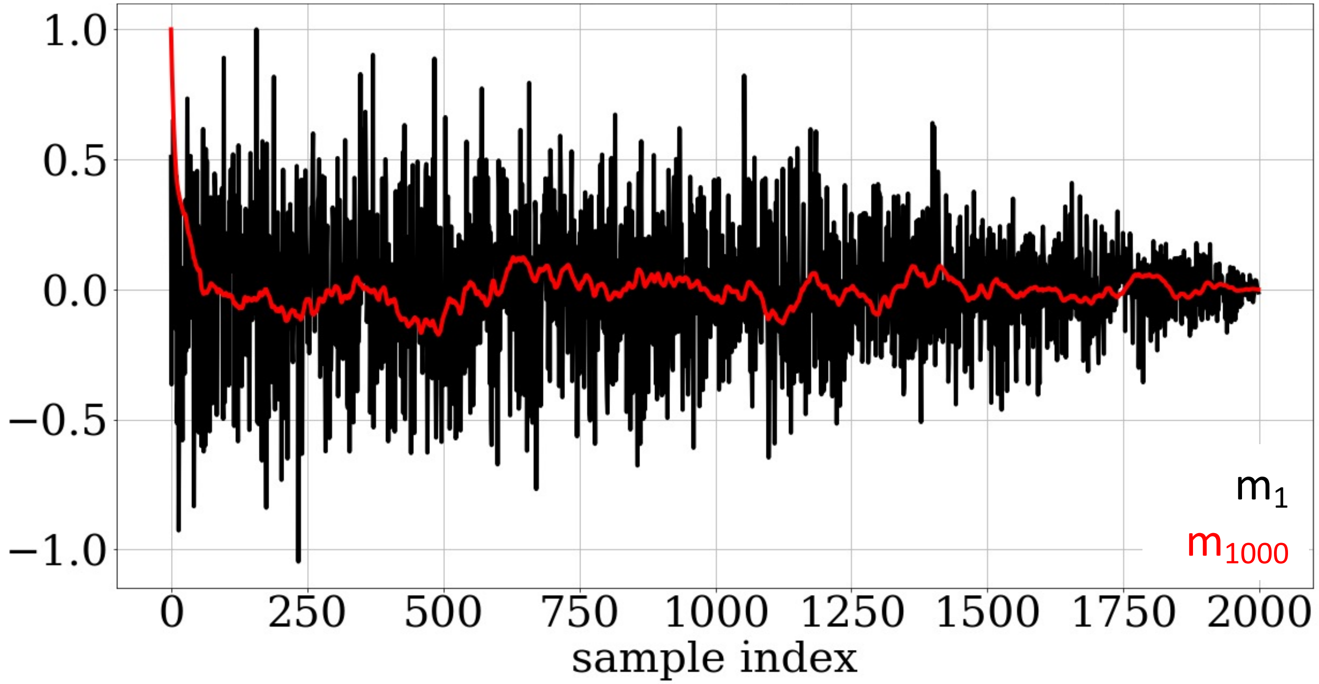
Example 1: 1000-D Gaussian

Covariance matrix: $C_{1,1}$=0.100, $C_{2,2}$=0.101, ..., $C_{1000,1000}$=1.100

Mass matrix: **M=I**

auto-correlation of sample chain

measure of the independence of successive samples



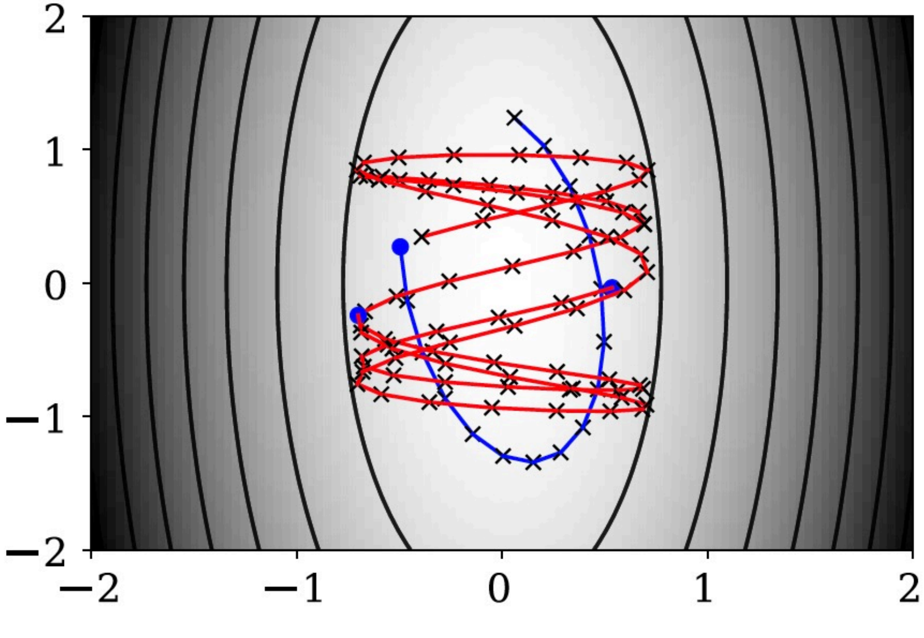
Example 1: 1000-D Gaussian

Covariance matrix: $C_{1,1}$=0.100, $C_{2,2}$=0.101, ..., $C_{1000,1000}$=1.100

Mass matrix: **M=I**

Mass matrix: **M=C^{-1}**

Hamiltonian trajectory [2-D projection]



Ideal: $\mathbf{M}=\mathbf{C}^{-1}=\mathbf{H}$ [mass matrix = Hessian]

Problems: Hessian cannot be computed or stored explicitly

Autotuning: Approximate the Hessian on the fly

- Use last couple of samples to approximate **H·vector**.
- Closely related to L-BFGS method from nonlinear optimisation [Nocedal, 1980].
- Use approximate **H** as **M** in computation of Hamiltonian trajectories.

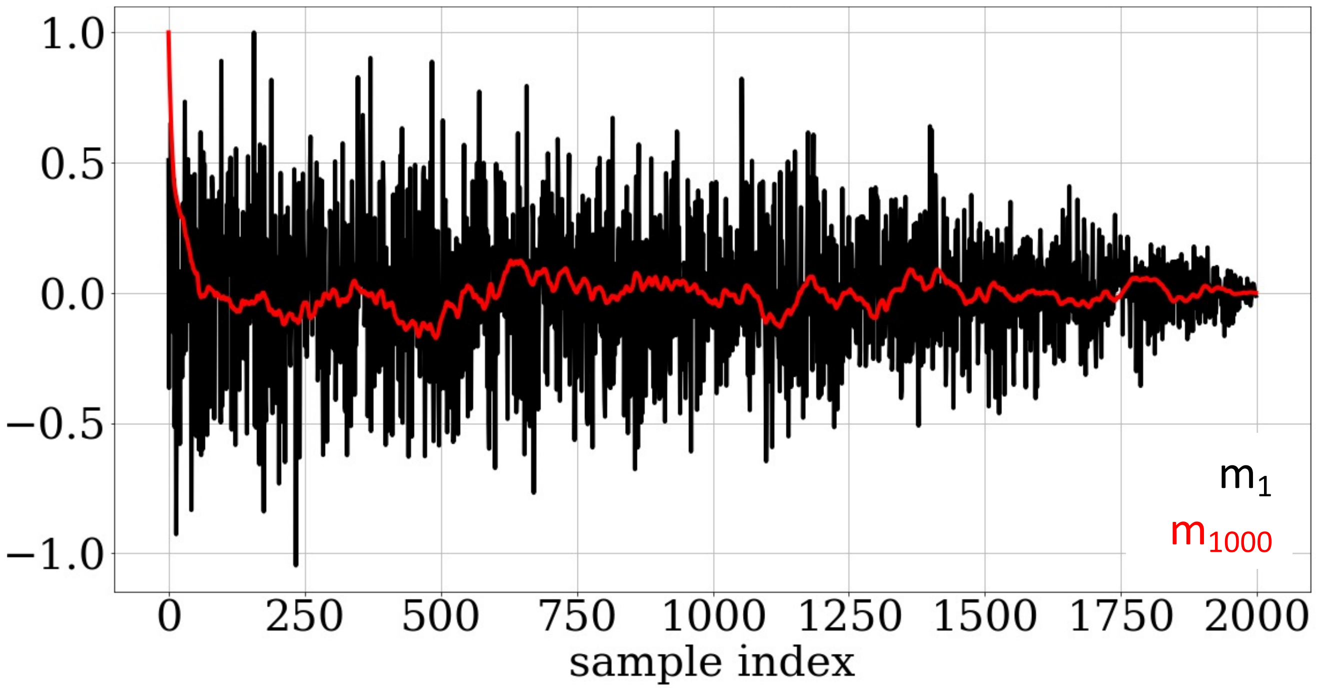
Return to Example 1:

Autotuning: Approximate the Hessian on the fly

auto-tuning off

[mass matrix: **M=I**]

auto-correlation of sample chain

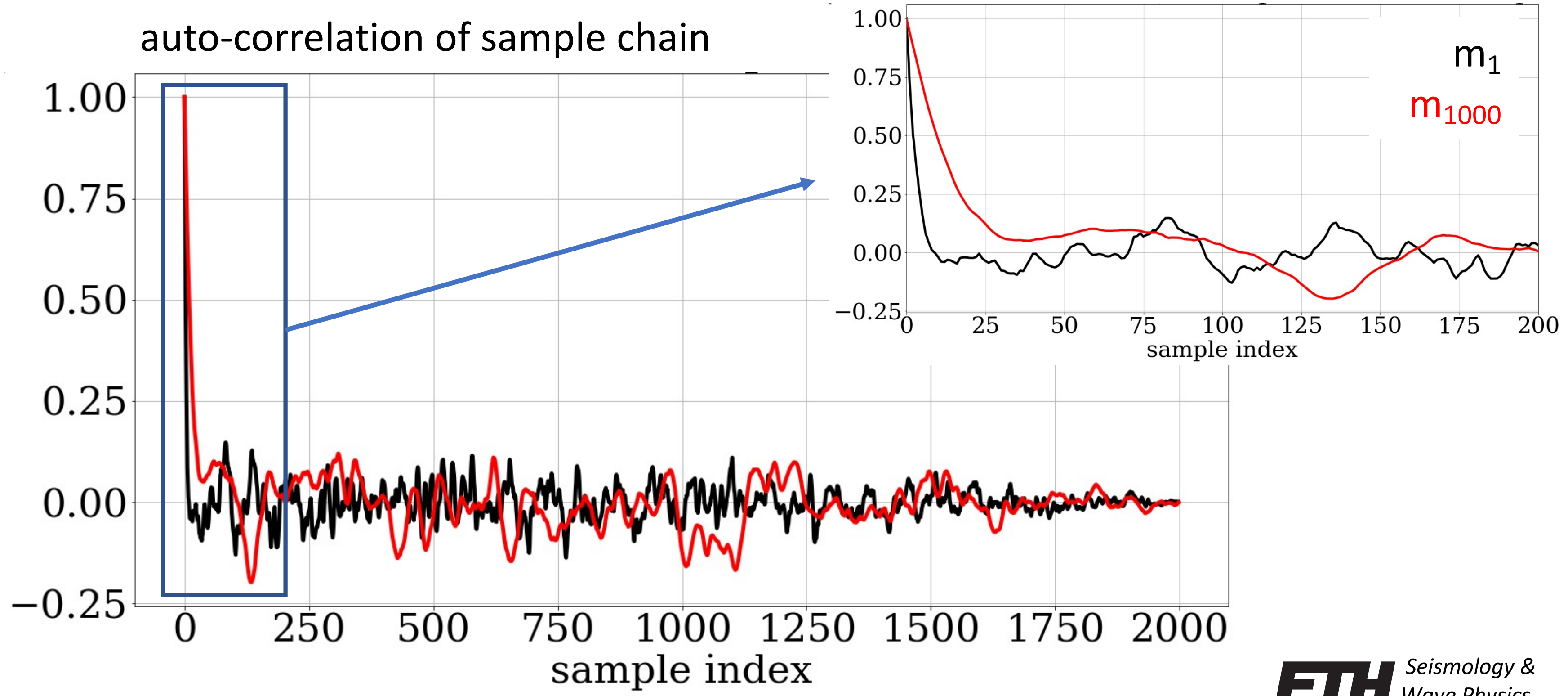


Return to Example 1:

Autotuning: Approximate the Hessian on the fly

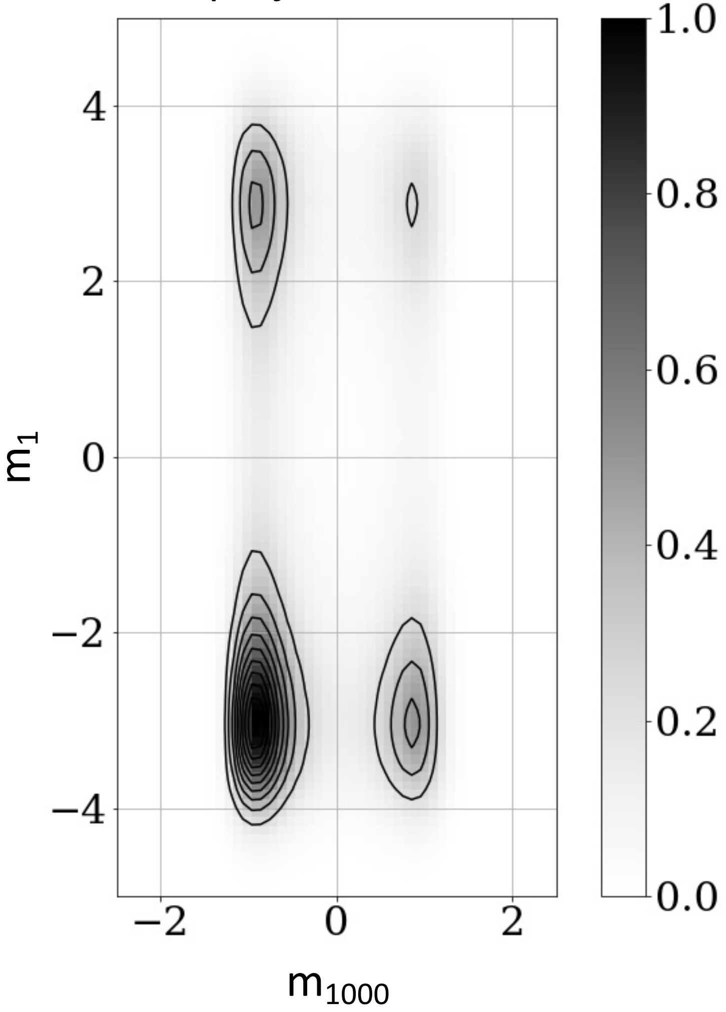
auto-tuning on:

[mass matrix: **M≈H**]

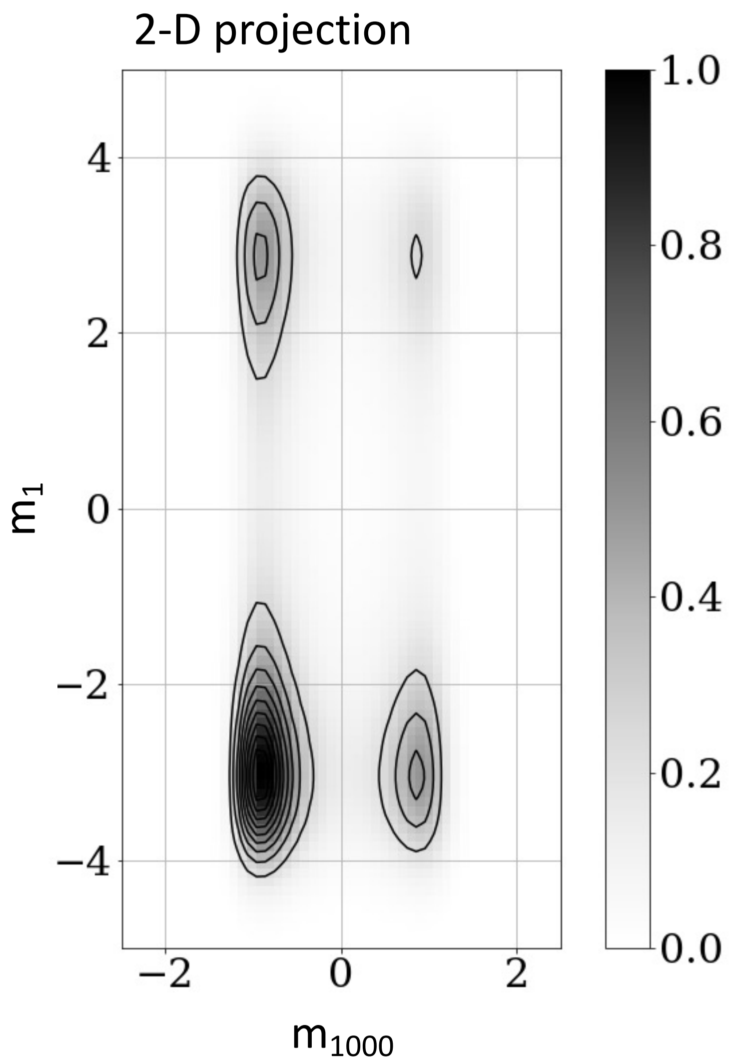


Example 2: 1000-D Styblinski-Tang function [4^{500} local minima]

2-D projection

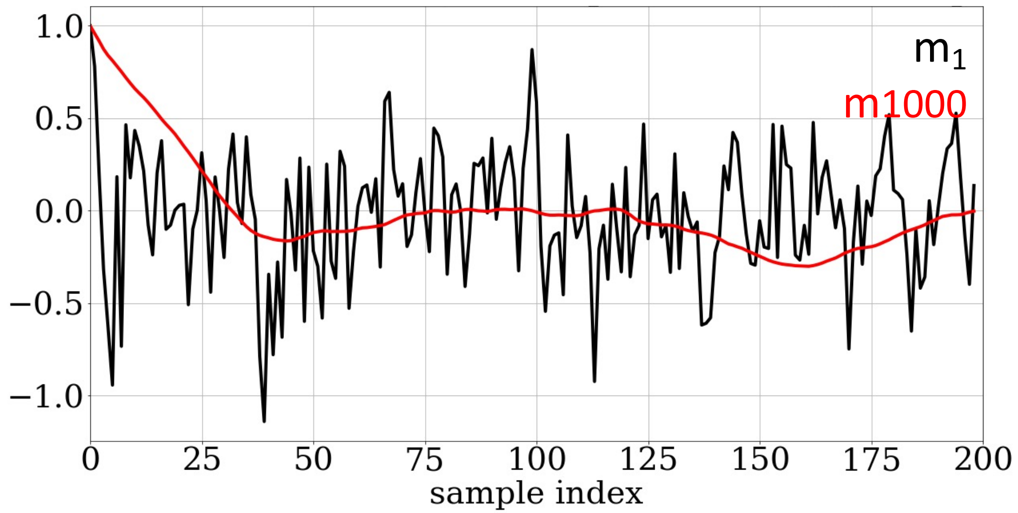


Example 2: 1000-D Styblinski-Tang function [4^{500} local minima]

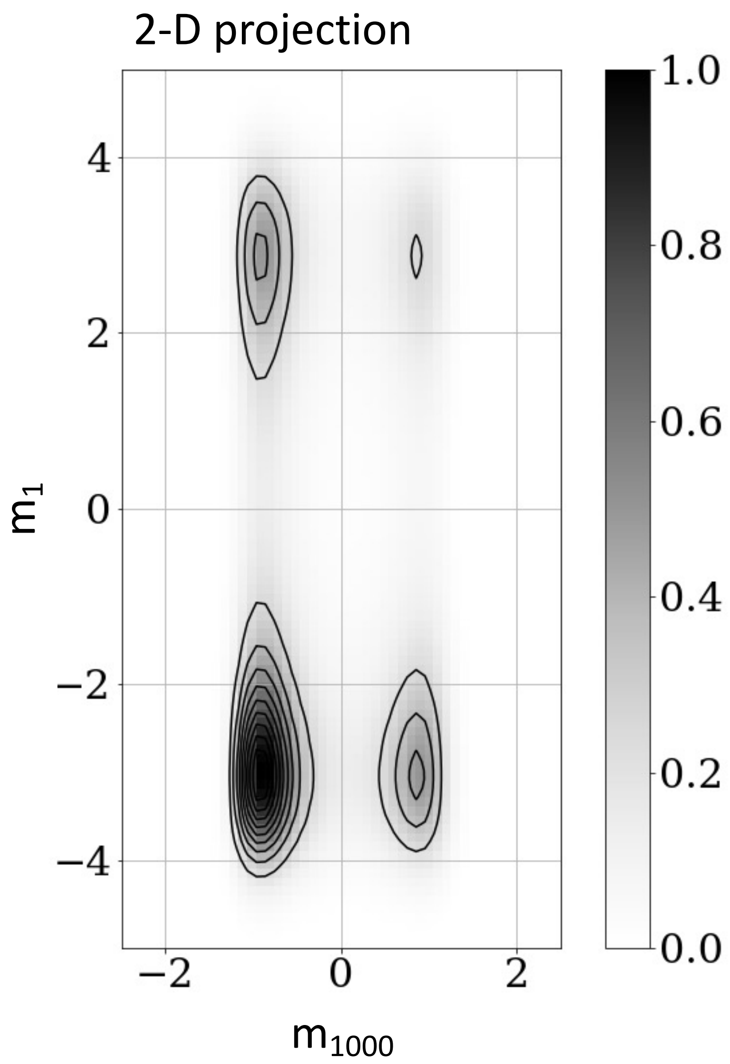


auto-tuning off

highly correlated m_1

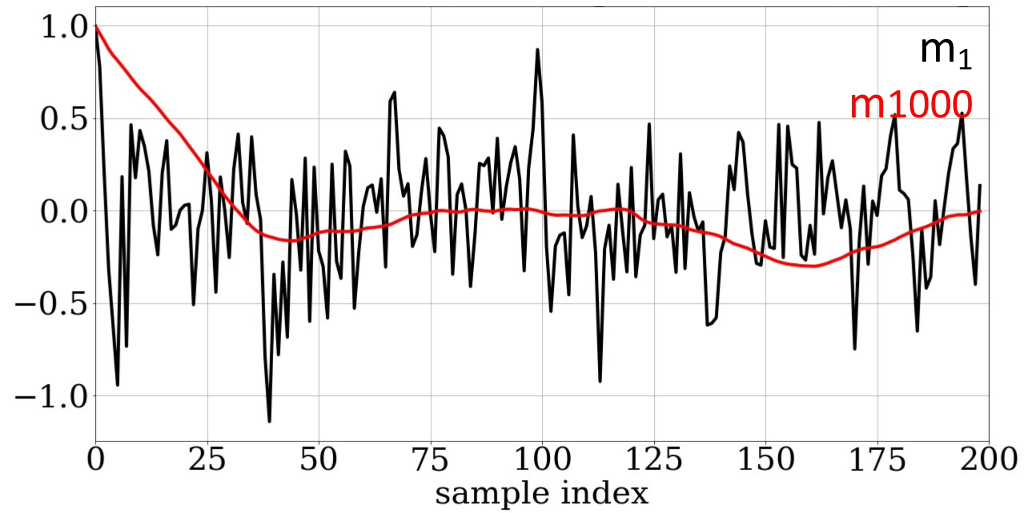


Example 2: 1000-D Styblinski-Tang function [4^{500} local minima]



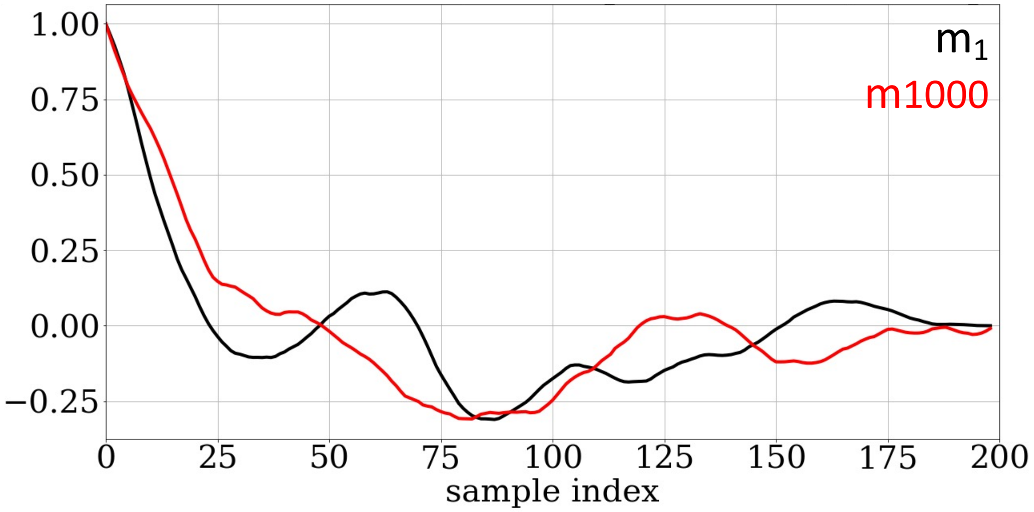
auto-tuning off

highly correlated m_1



auto-tuning on

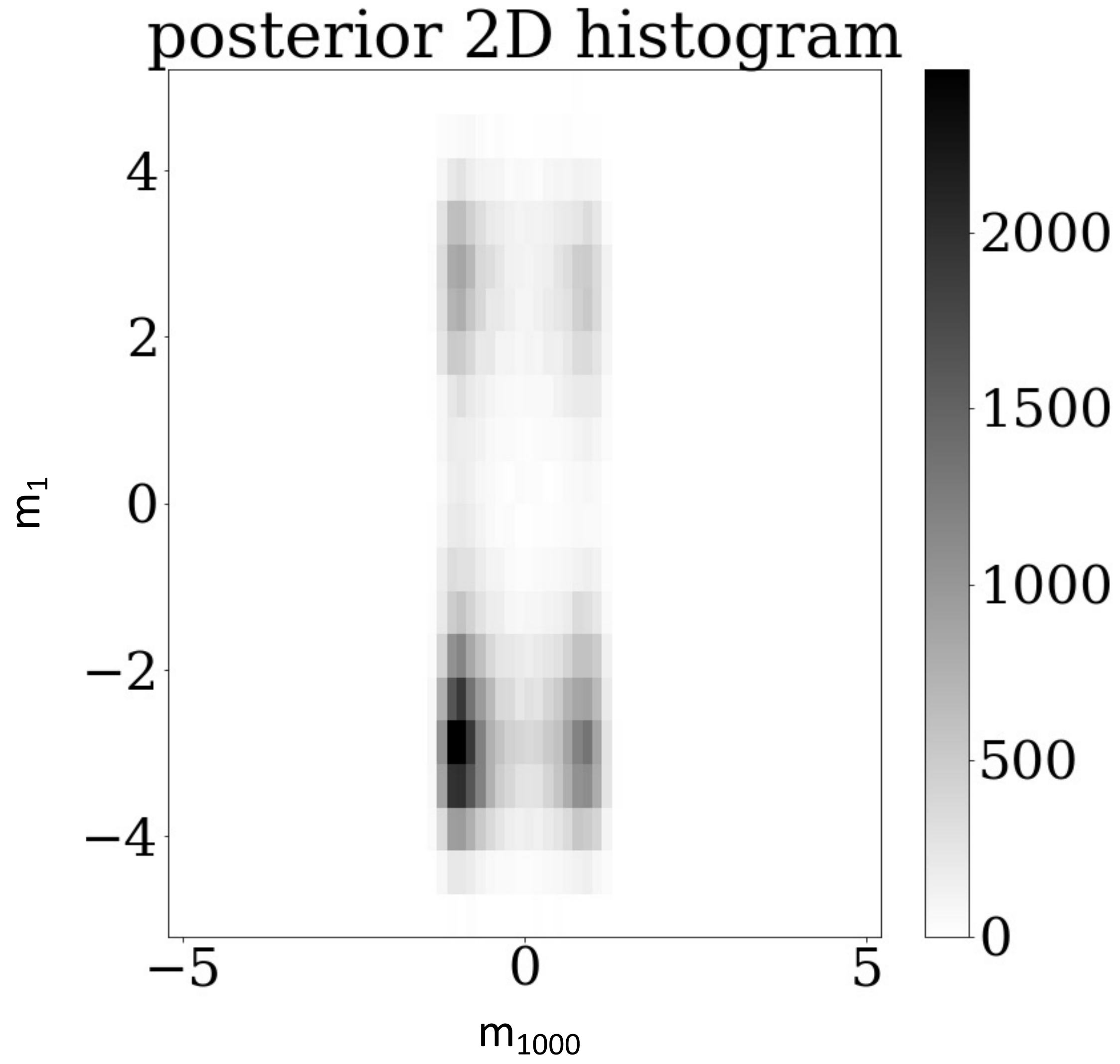
largely uncorrelated m_1



Example 2: 1000-D Styblinski-Tang function [10‘000 samples]

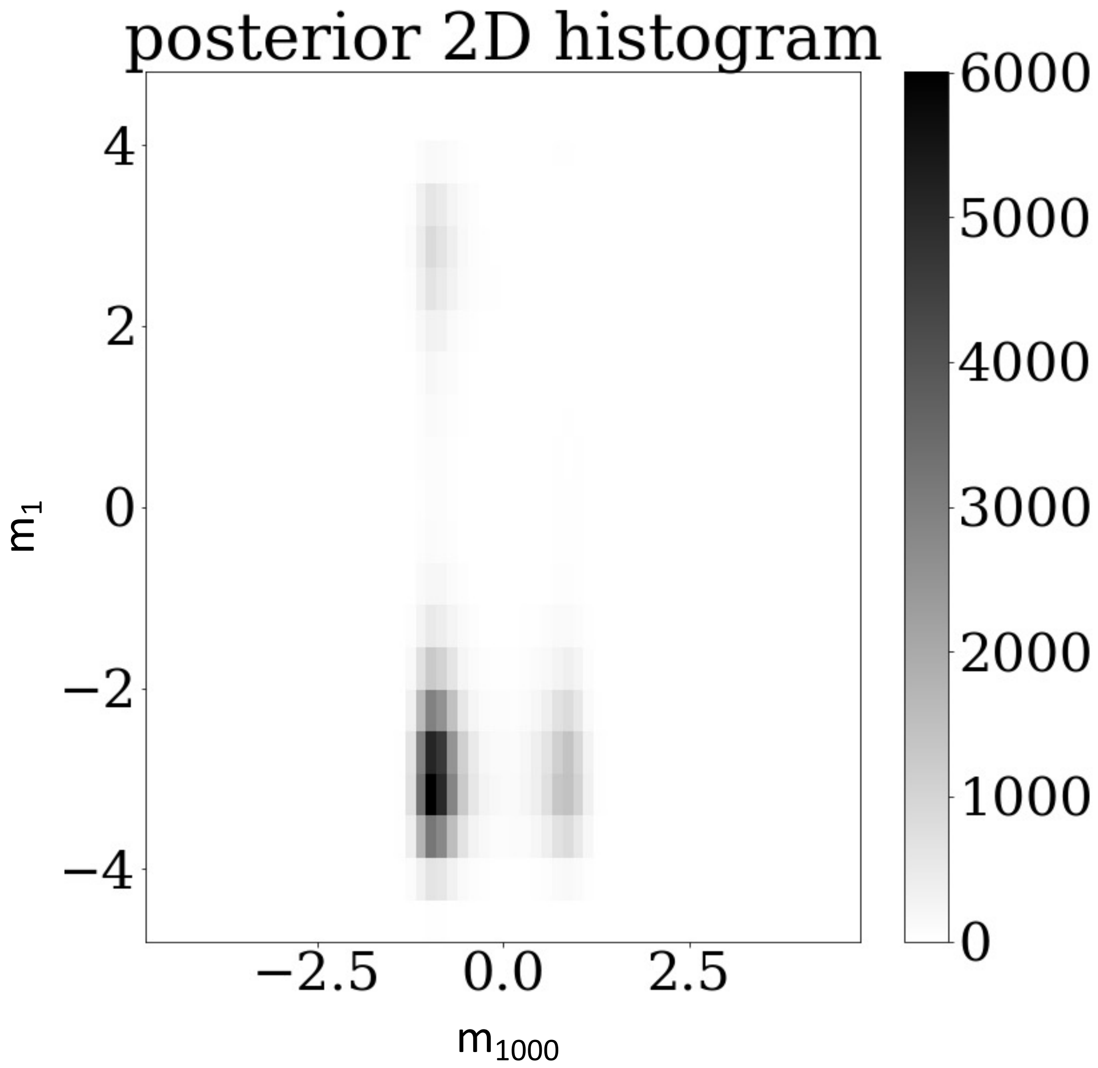
auto-tuning on

largely uncorrelated m_1



auto-tuning off

highly correlated m_1



Discussion & Conclusions

Discussion & Conclusions

Hamiltonian Nullspace Shuttles

- Efficient nullspace exploration of nonlinear inverse problems [without Monte Carlo sampling].
- Enables quantitative hypothesis testing.
- Targeted construction of alternative models.

Discussion & Conclusions

Hamiltonian Nullspace Shuttles

- Efficient nullspace exploration of nonlinear inverse problems [without Monte Carlo sampling].
- Enables quantitative hypothesis testing.
- Targeted construction of alternative models.

Hamiltonian Monte Carlo for tomographic inversion

- Radomised version of the nullspace shuttle.
- $\mathcal{O}$**(10'000)** model parameters in 2D FWI without any supercomputing.
- **Fully probabilistic tomography in 3D is within reach!**

Discussion & Conclusions

Hamiltonian Nullspace Shuttles

- Efficient nullspace exploration of nonlinear inverse problems [without Monte Carlo sampling].
- Enables quantitative hypothesis testing.
- Targeted construction of alternative models.

Hamiltonian Monte Carlo for tomographic inversion

- Radomised version of the nullspace shuttle.
- $\mathcal{O}$**(10’000)** model parameters in 2D FWI without any supercomputing.
- **Fully probabilistic tomography in 3D is within reach!**

Autotuning

- HMC + factorised version of L-BFGS.
- Rapid generation of independent models and improved convergence.
- Great promise for realistic applications.

Literature

Fichtner, A., Simute, S., 2018. *Hamiltonian Monte Carlo Inversion of seismic sources in complex media*. Journal of Geophysical Research 123, doi: 10.1002/2017JB015249

Fichtner, A., Zunino, A., Gebraad, L., 2019. *Hamiltonian Monte Carlo solution of tomographic inverse problems*. Geophysical Journal International, 216, 1344-1363, doi: 10.1093/gji/ggy496.

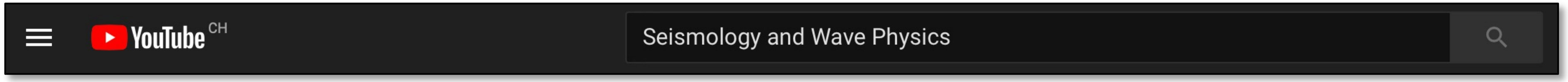
Fichtner, A., Zunino, A., 2019. *Hamiltonian Nullspace Shuttles*. Geophysical Research Letters, doi: 10.1029/2018GL08931.

Gebraad, L., Boehm, C., Fichtner, A., 2020. *Bayesian elastic full-waveform inversion using Hamiltonian Monte Carlo*. Journal of Geophysical Research, 125, doi:10.1029/2019JB018428.

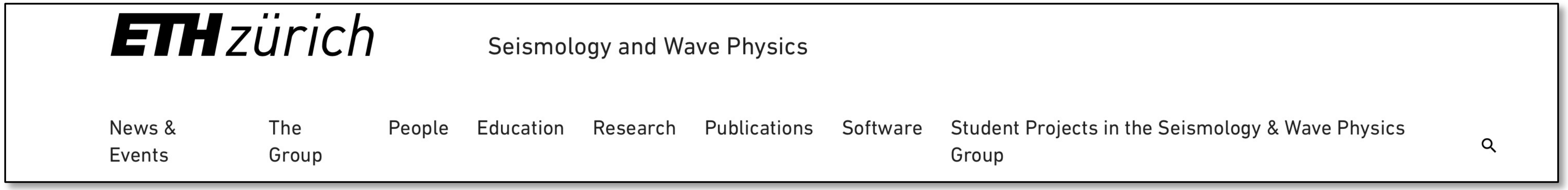
Fichtner, A., Zunino, A., Gebraad, L., Boehm, C., 2021. *Autotuning Hamiltonian Monte Carlo for efficient generalised nullspace exploration*. Geophysical Journal International, in press.

Other resources

YouTube channel with educational movies



Software Library @ www.swp.ethz.ch



HMC Tomography Package @ upon request

HMC Tomography

Literature

Fichtner, A., Simute, S., 2018. *Hamiltonian Monte Carlo Inversion of seismic sources in complex media*. Journal of Geophysical Research 123, doi: 10.1002/2017JB015249

Fichtner, A., Zunino, A., Gebraad, L., 2019. *Hamiltonian Monte Carlo solution of tomographic inverse problems*. Geophysical Journal International, 216, 1344-1363, doi: 10.1093/gji/ggy496.

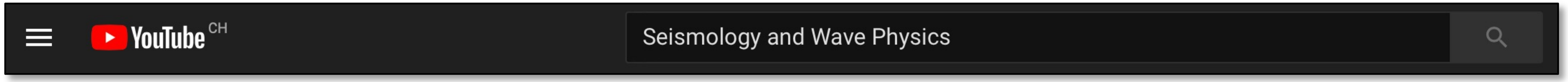
Fichtner, A., Zunino, A., 2019. *Hamiltonian Nullspace Shuttles*. Geophysical Research Letters, doi: 10.1029/2018GL08931.

Gebraad, L., Boehm, C., Fichtner, A., 2020. *Bayesian elastic full-waveform inversion using Hamiltonian Monte Carlo*. Journal of Geophysical Research, 125, doi:10.1029/2019JB018428.

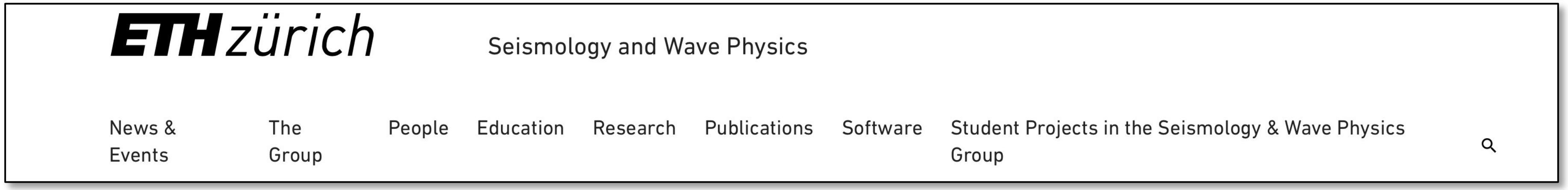
Fichtner, A., Zunino, A., Gebraad, L., Boehm, C., 2021. *Autotuning Hamiltonian Monte Carlo for efficient generalised nullspace exploration*. Geophysical Journal International, in press.

Other resources

YouTube channel with educational movies



Software Library @ www.swp.ethz.ch



HMC Tomography Package @ upon request

HMC Tomography

Thank you for your attention!