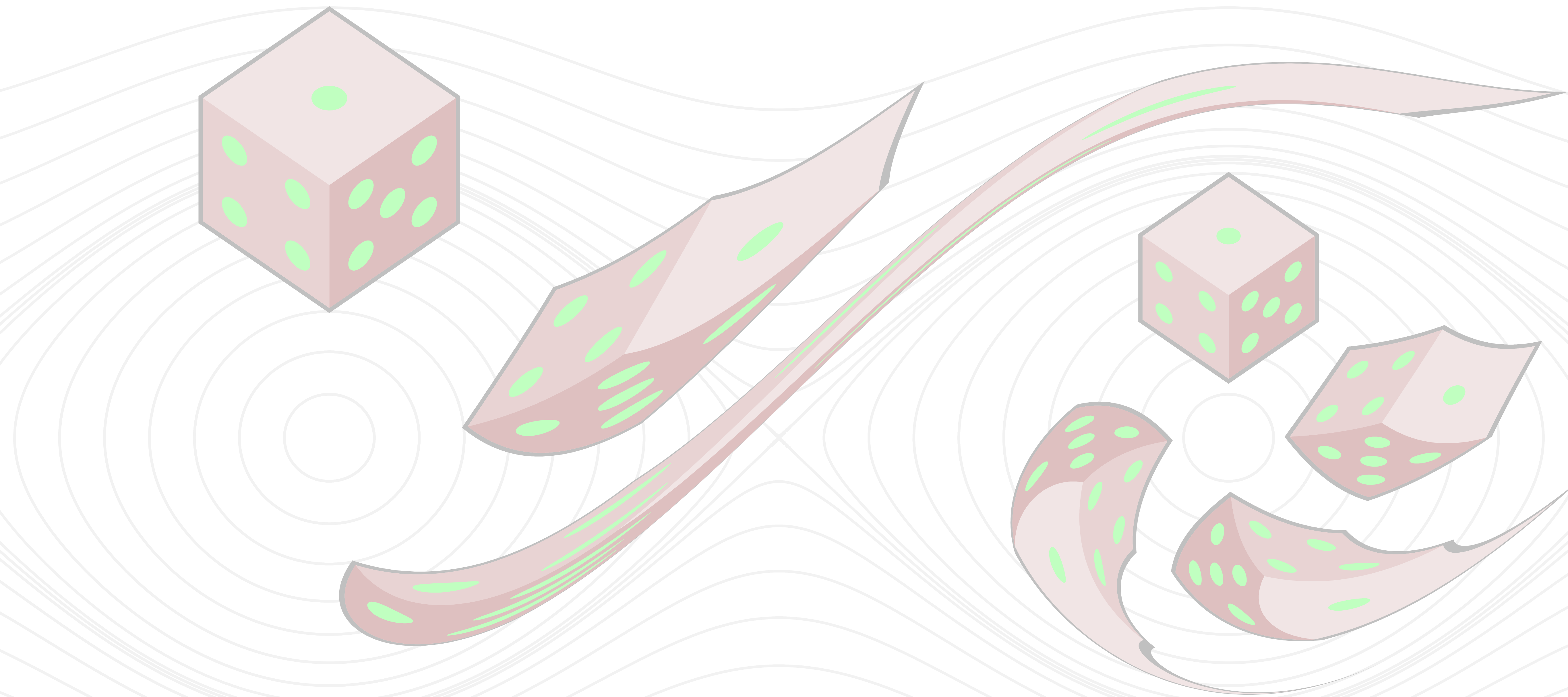
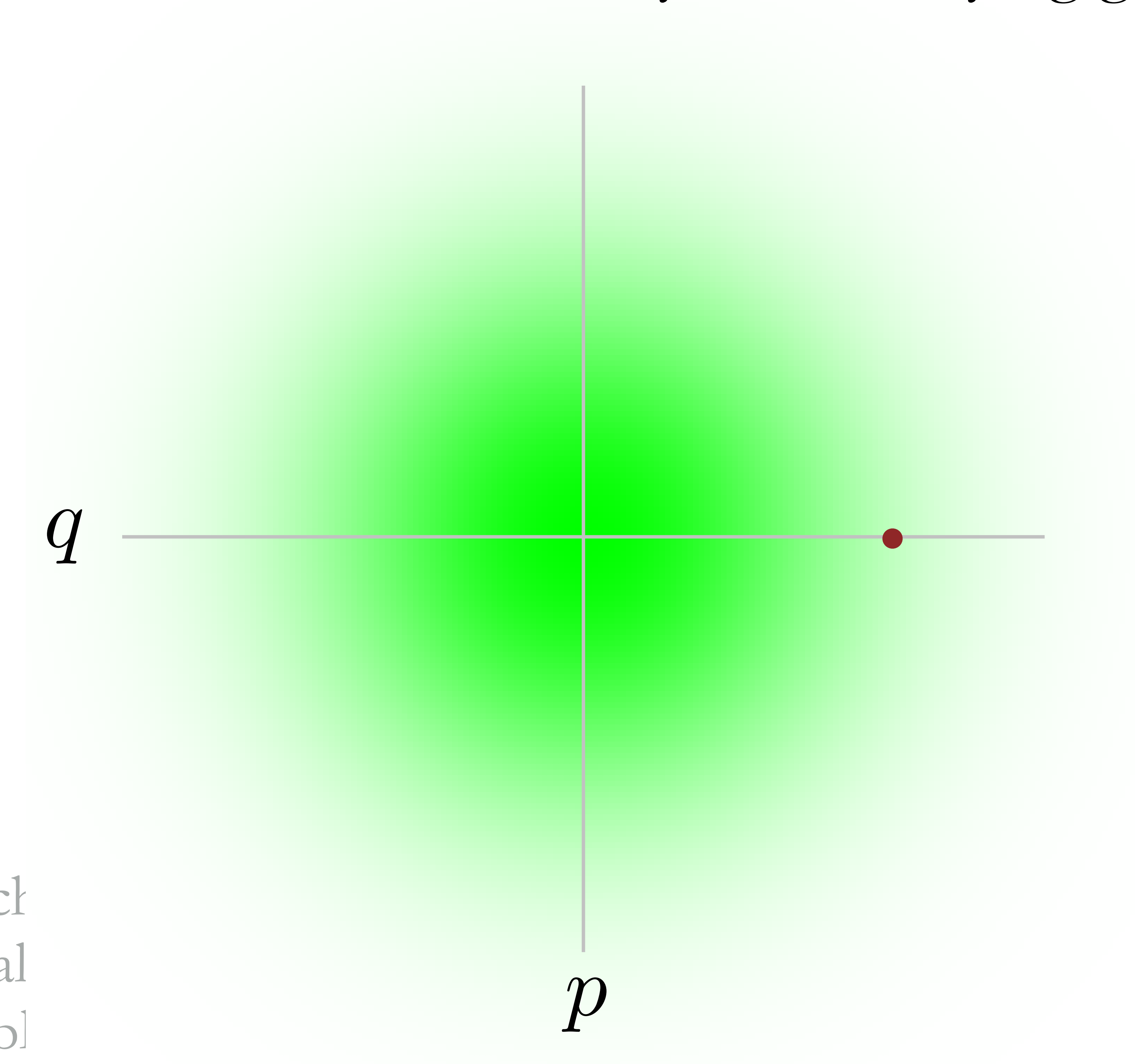
The Optimal Implementation of Hamiltonian Monte Carlo

Michael Betancourt
@betanalpha
Symplectomorphic, LLC

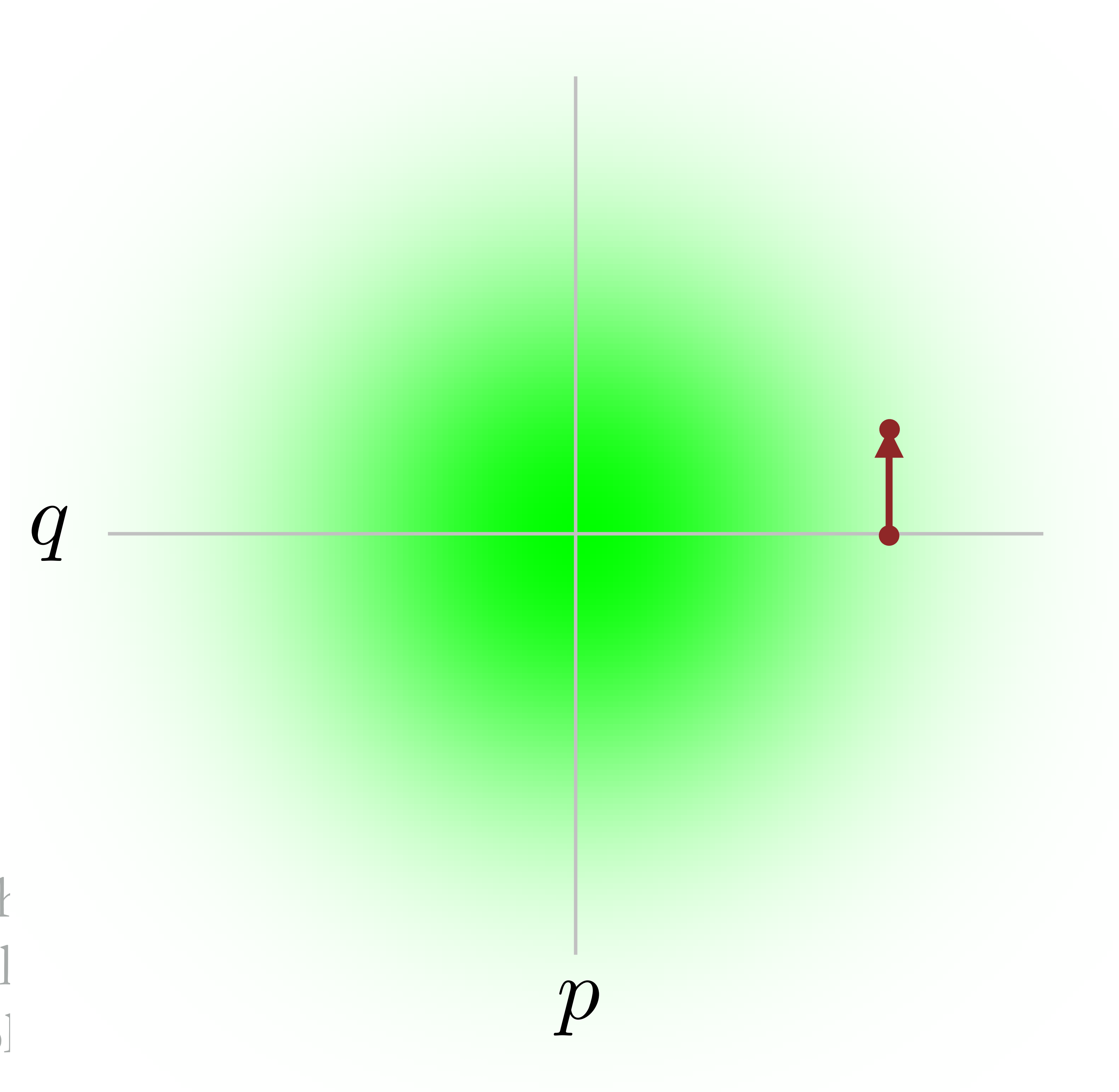


Machine Learning Summer School
London, United Kingdom
July 23, 2019

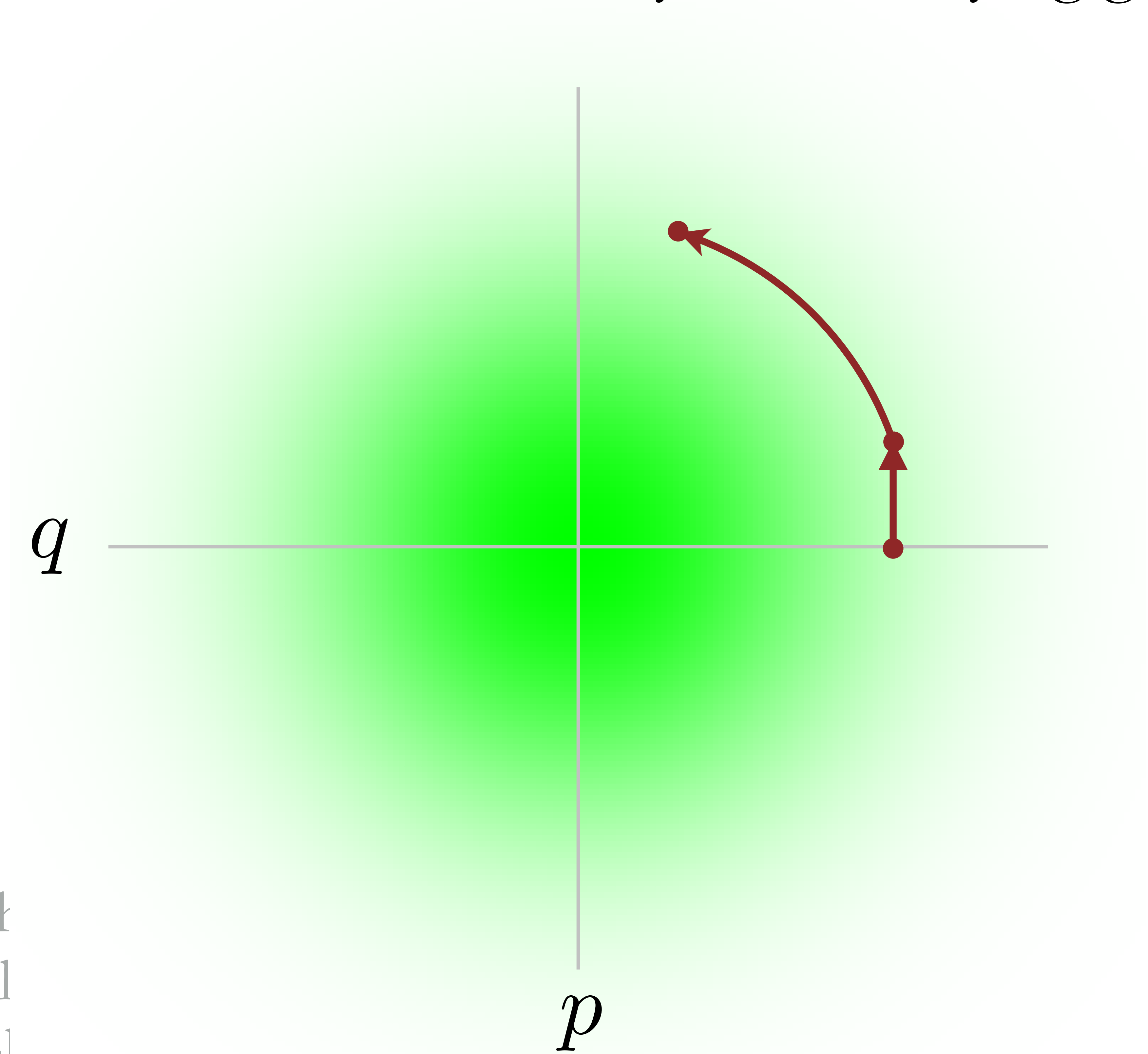
The performance of an implementation of Hamiltonian Monte Carlo is determined by its underlying geometry.



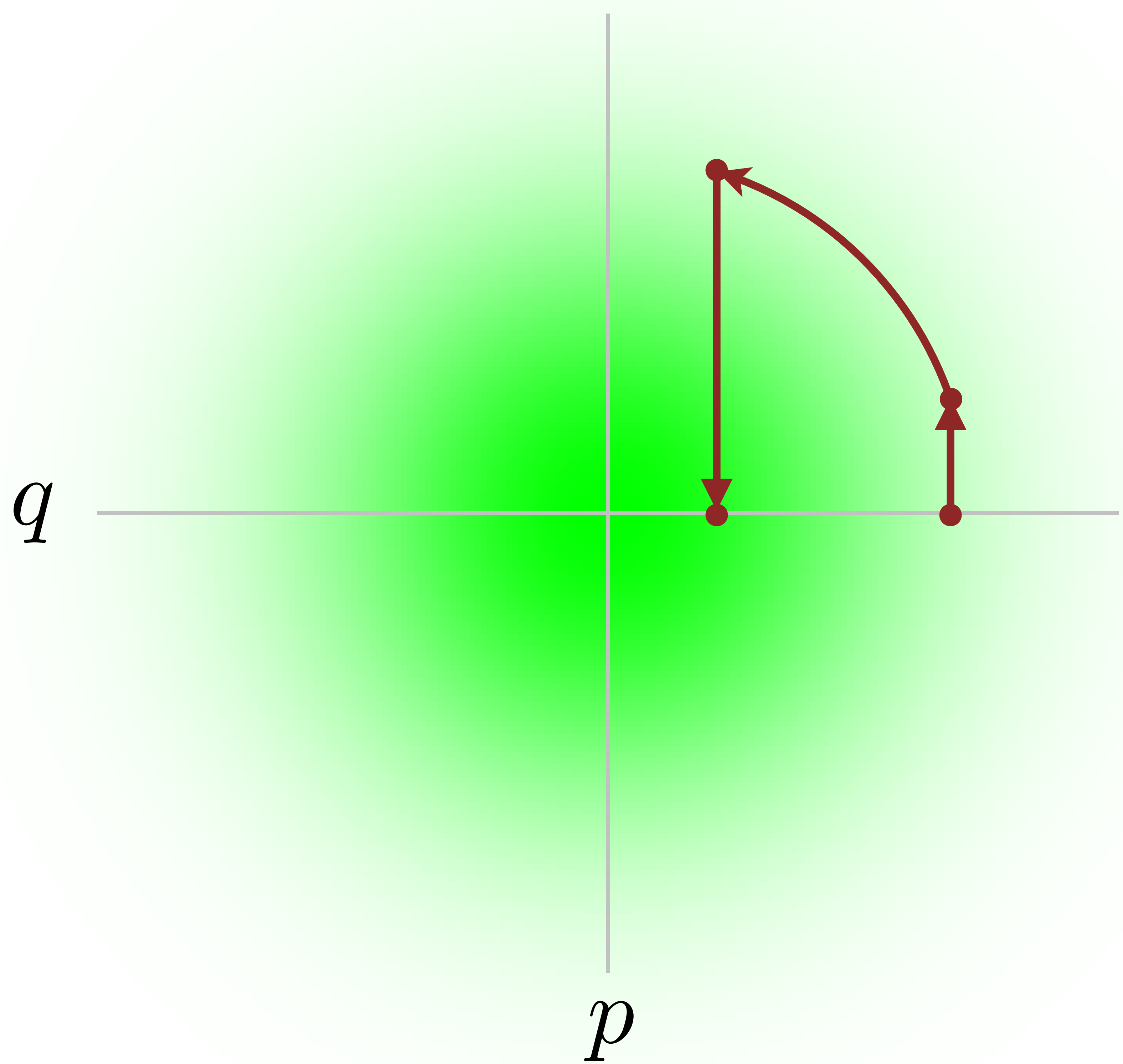
The performance of an implementation of Hamiltonian Monte Carlo is determined by its underlying geometry.



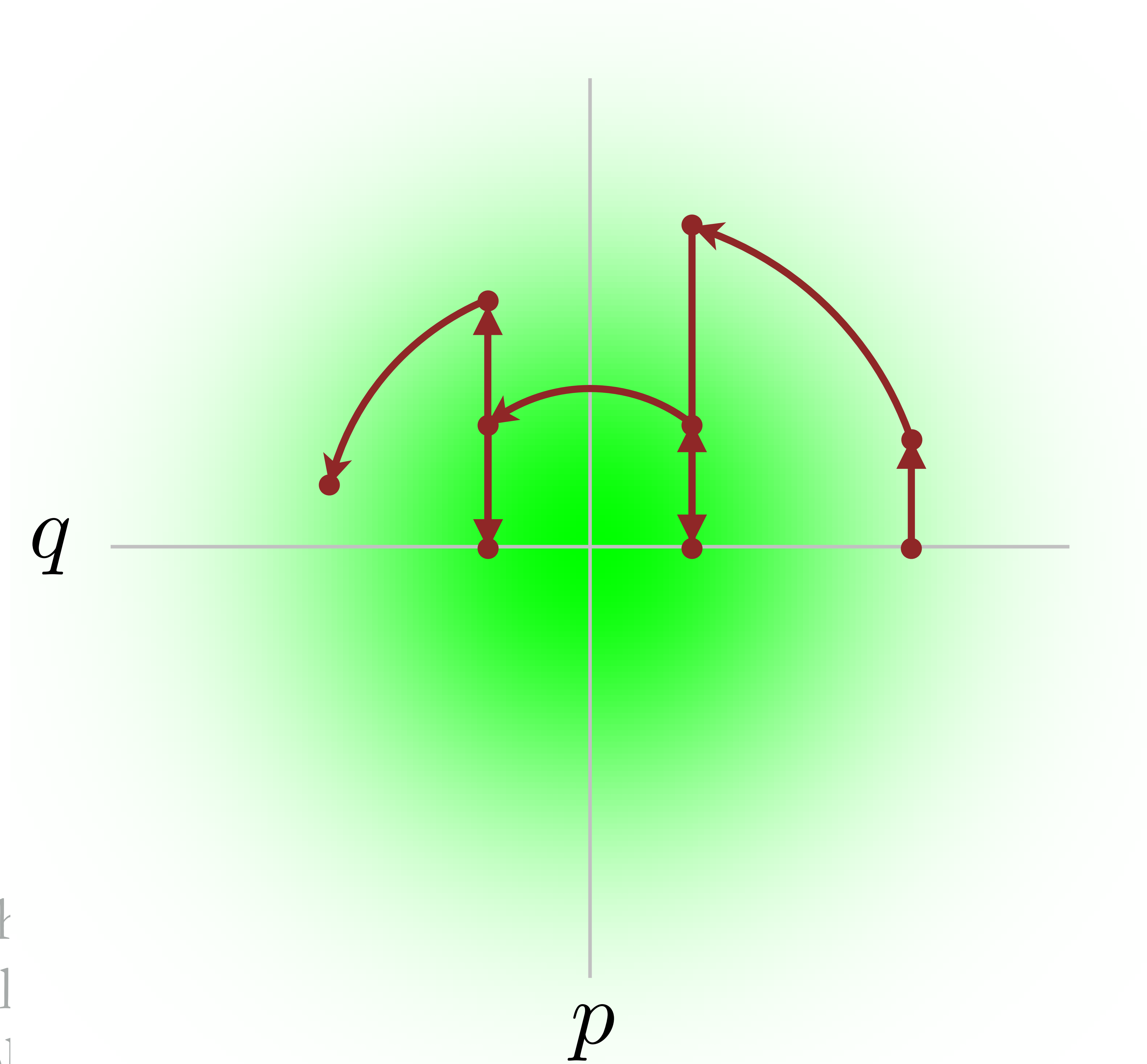
The performance of an implementation of Hamiltonian Monte Carlo is determined by its underlying geometry.



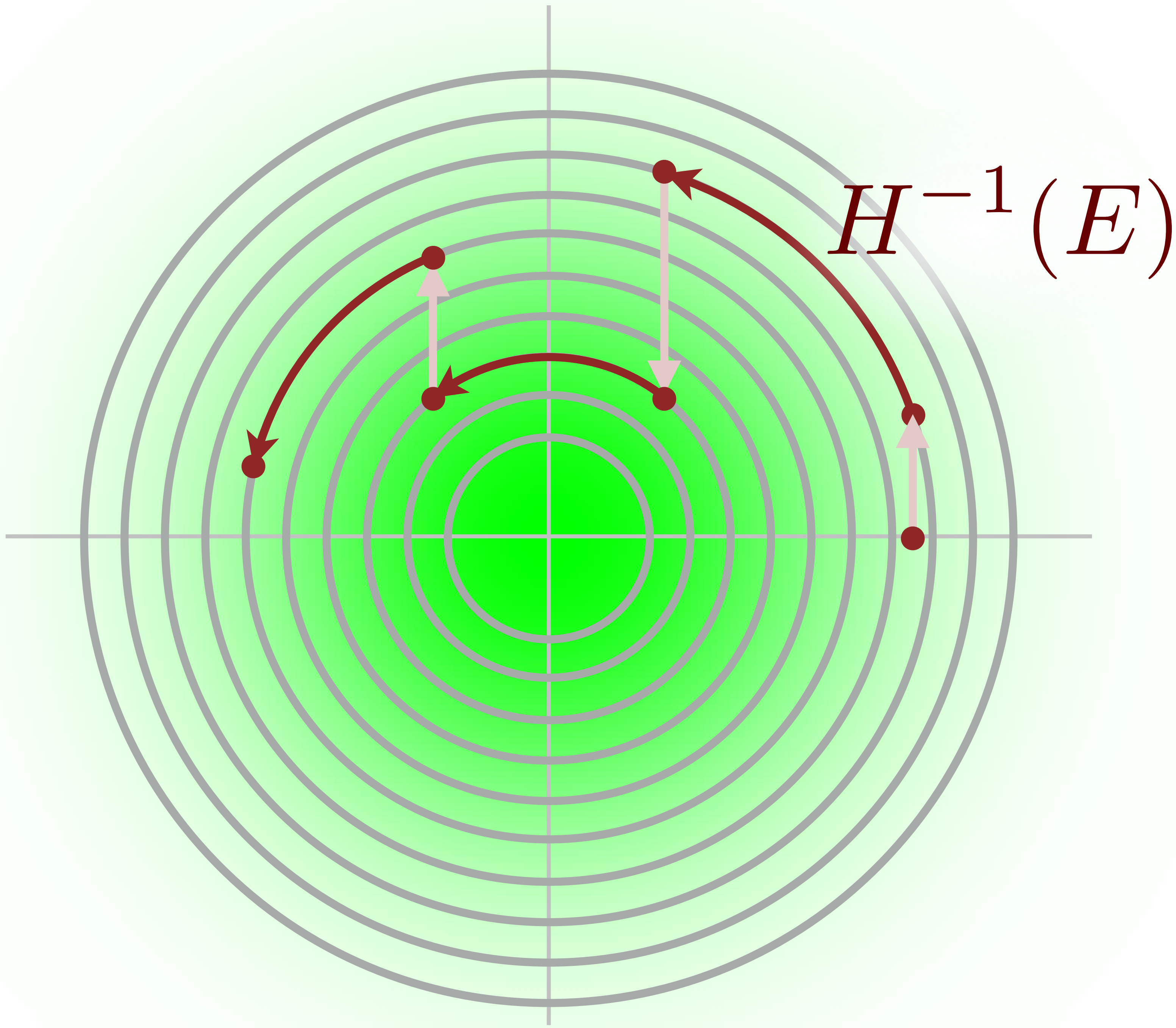
The performance of an implementation of Hamiltonian Monte Carlo is determined by its underlying geometry.



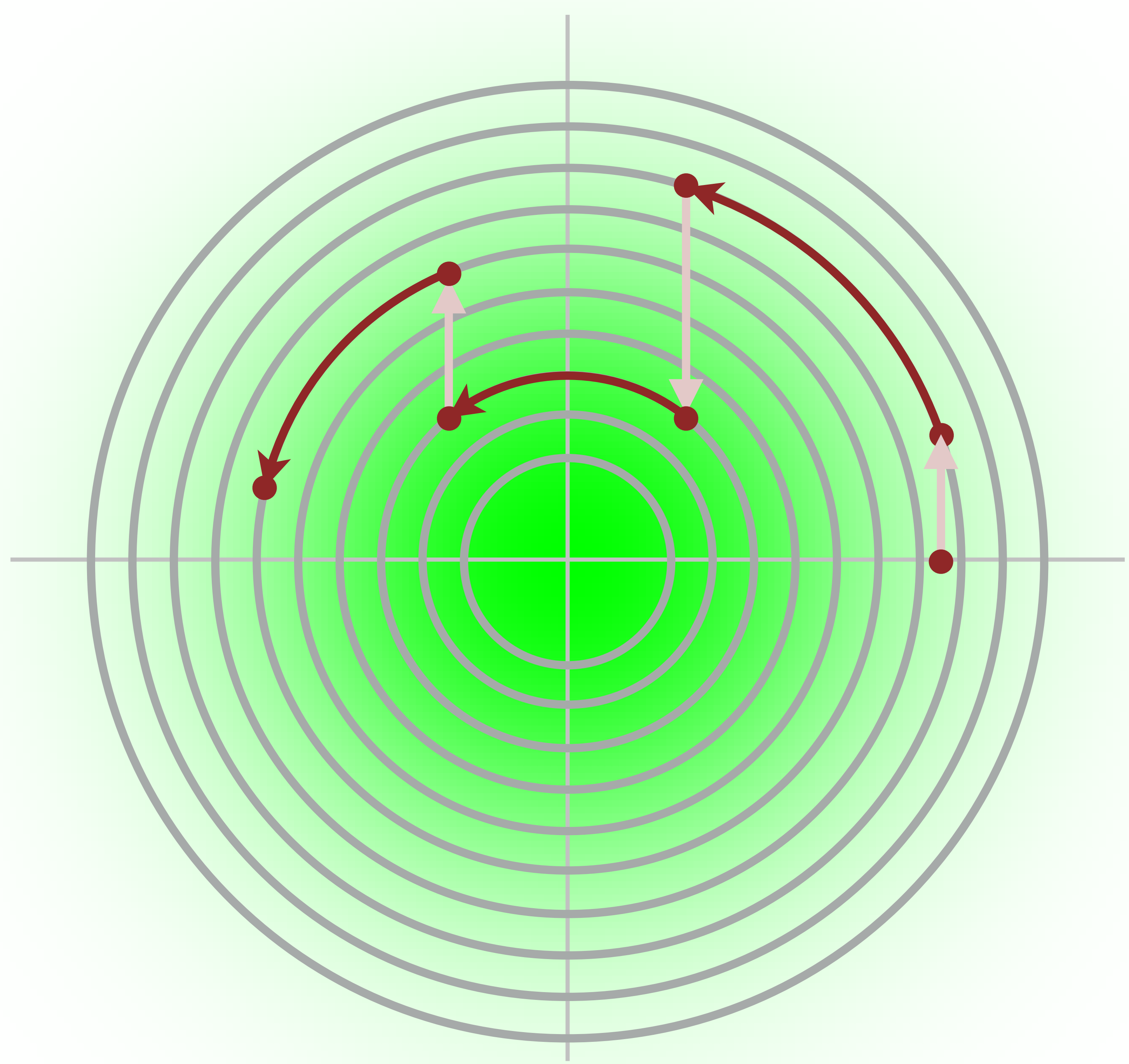
The performance of an implementation of Hamiltonian Monte Carlo is determined by its underlying geometry.



The performance of an implementation of Hamiltonian Monte Carlo is determined by its underlying geometry.

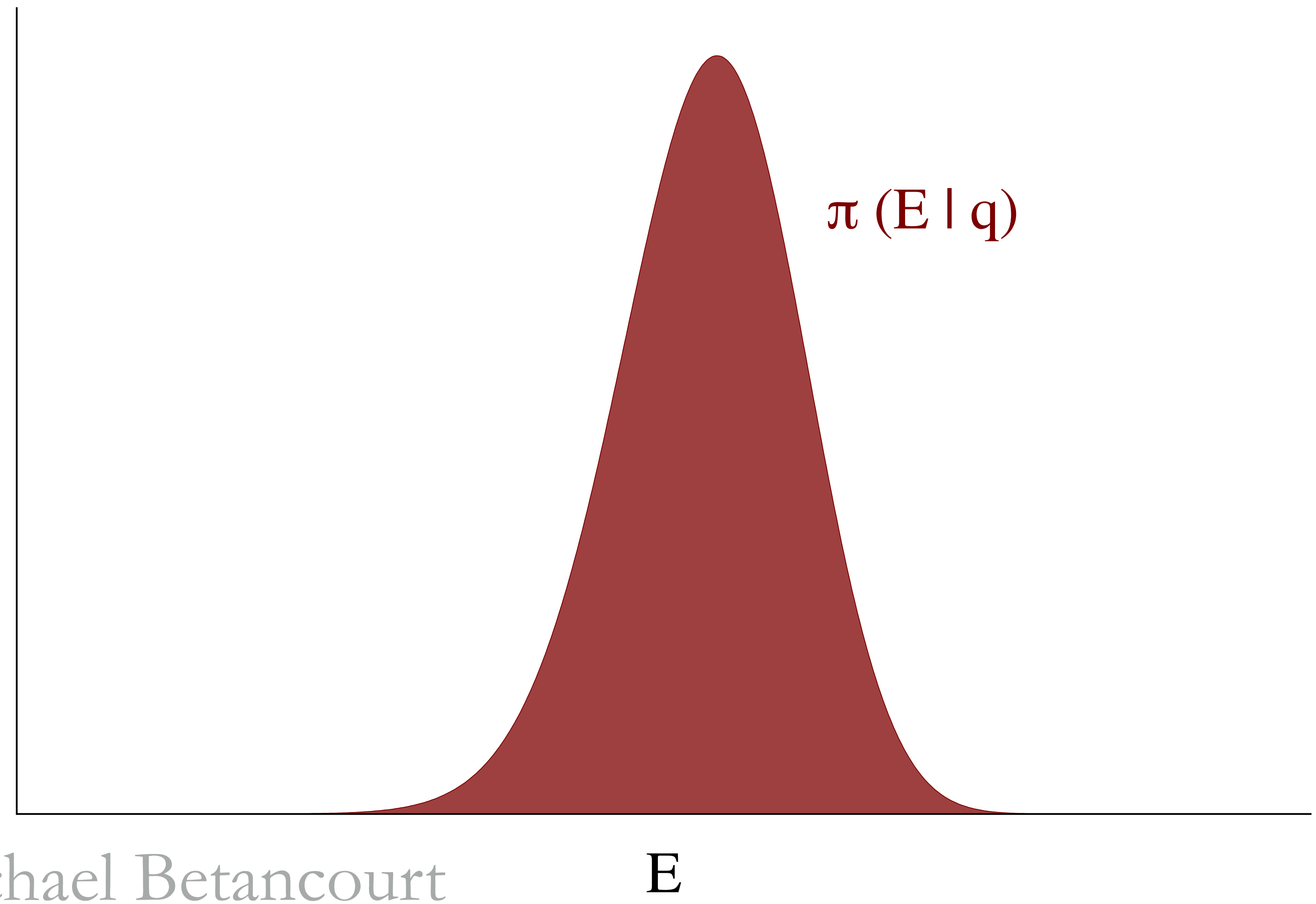


The performance of an implementation of Hamiltonian Monte Carlo is determined by its underlying geometry.

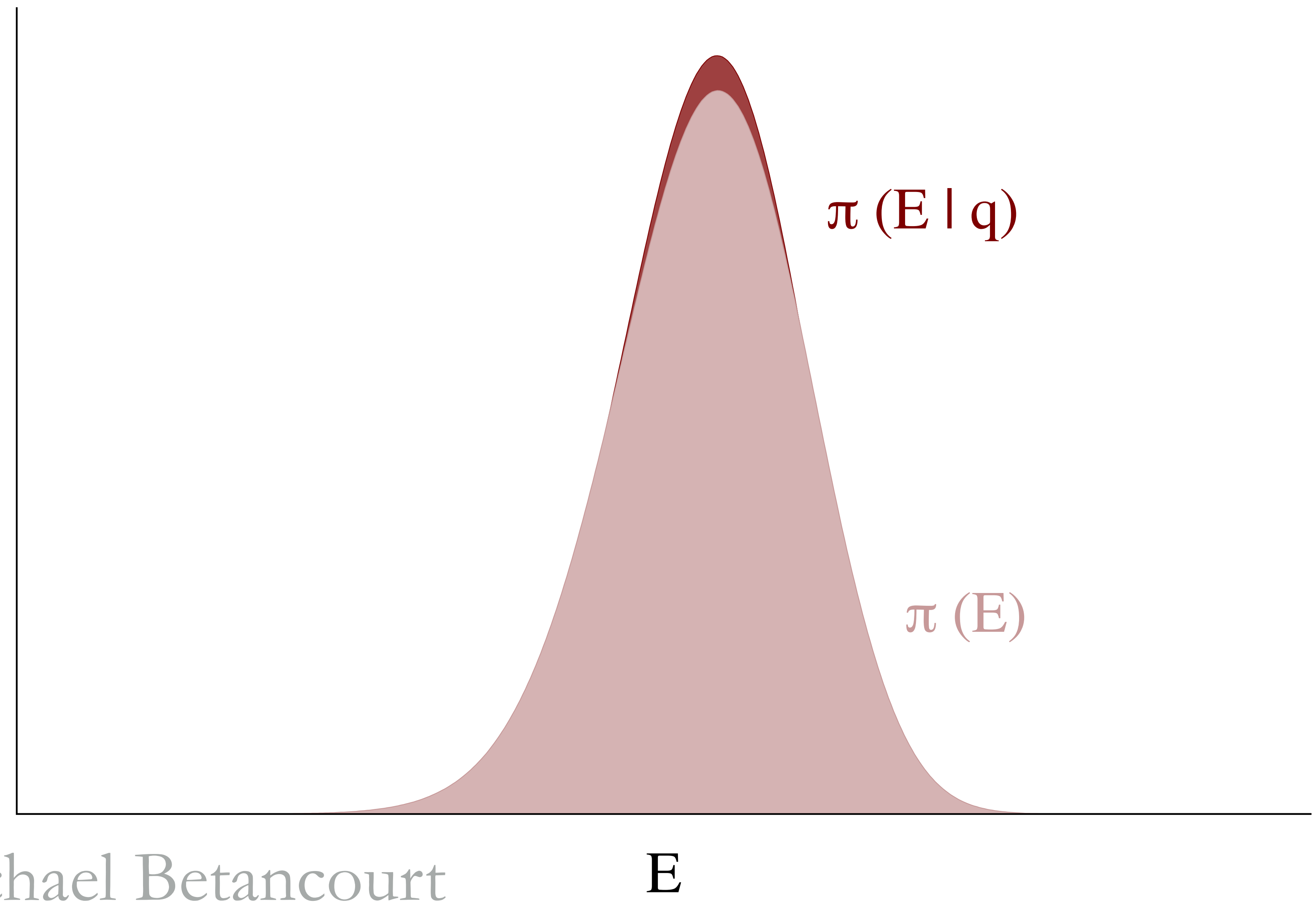


$$\pi_H = \pi_{H^{-1}(E)} \wedge \pi_E$$

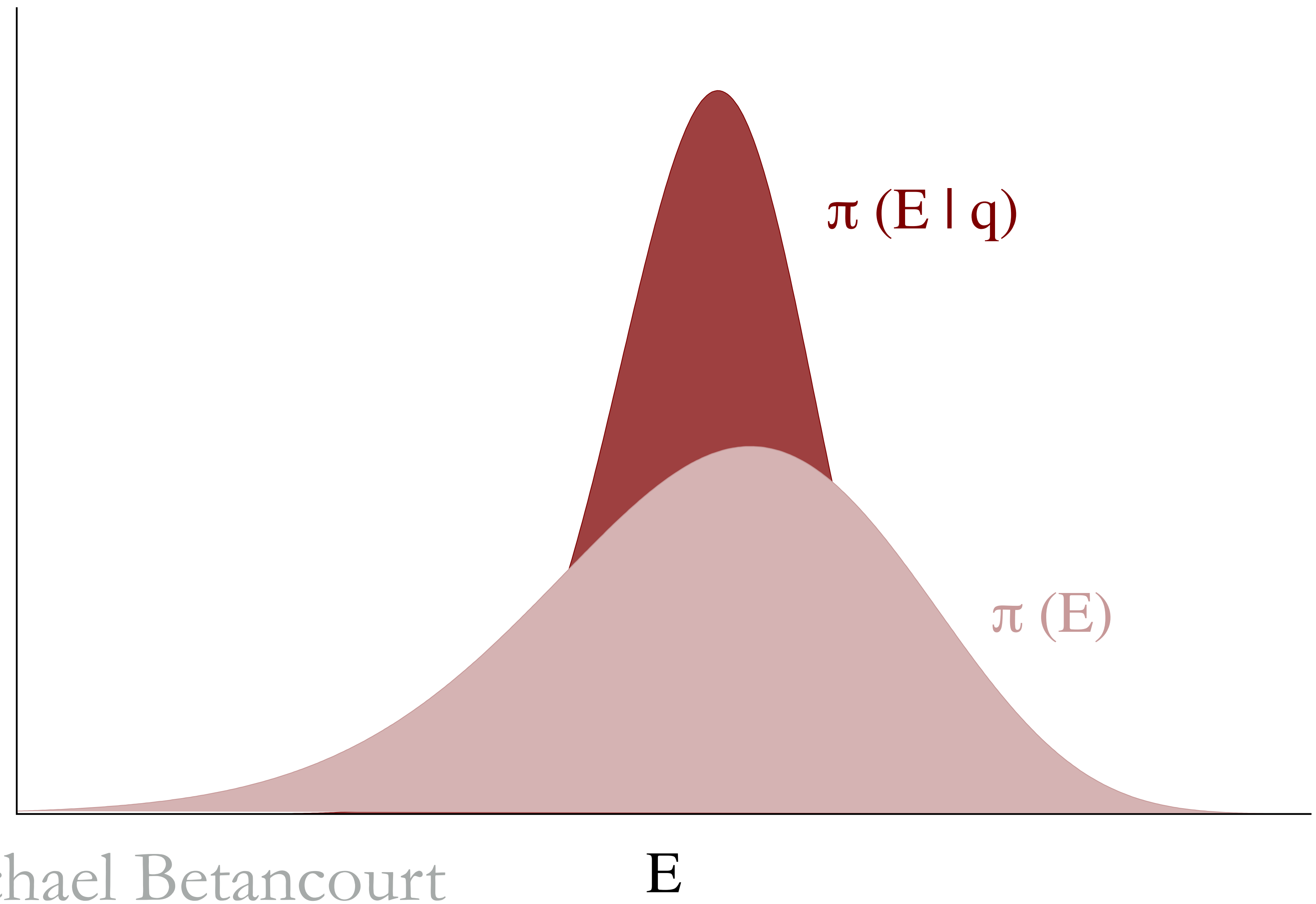
The momenta resampling, for example, determines how effectively we explore between energy level sets.



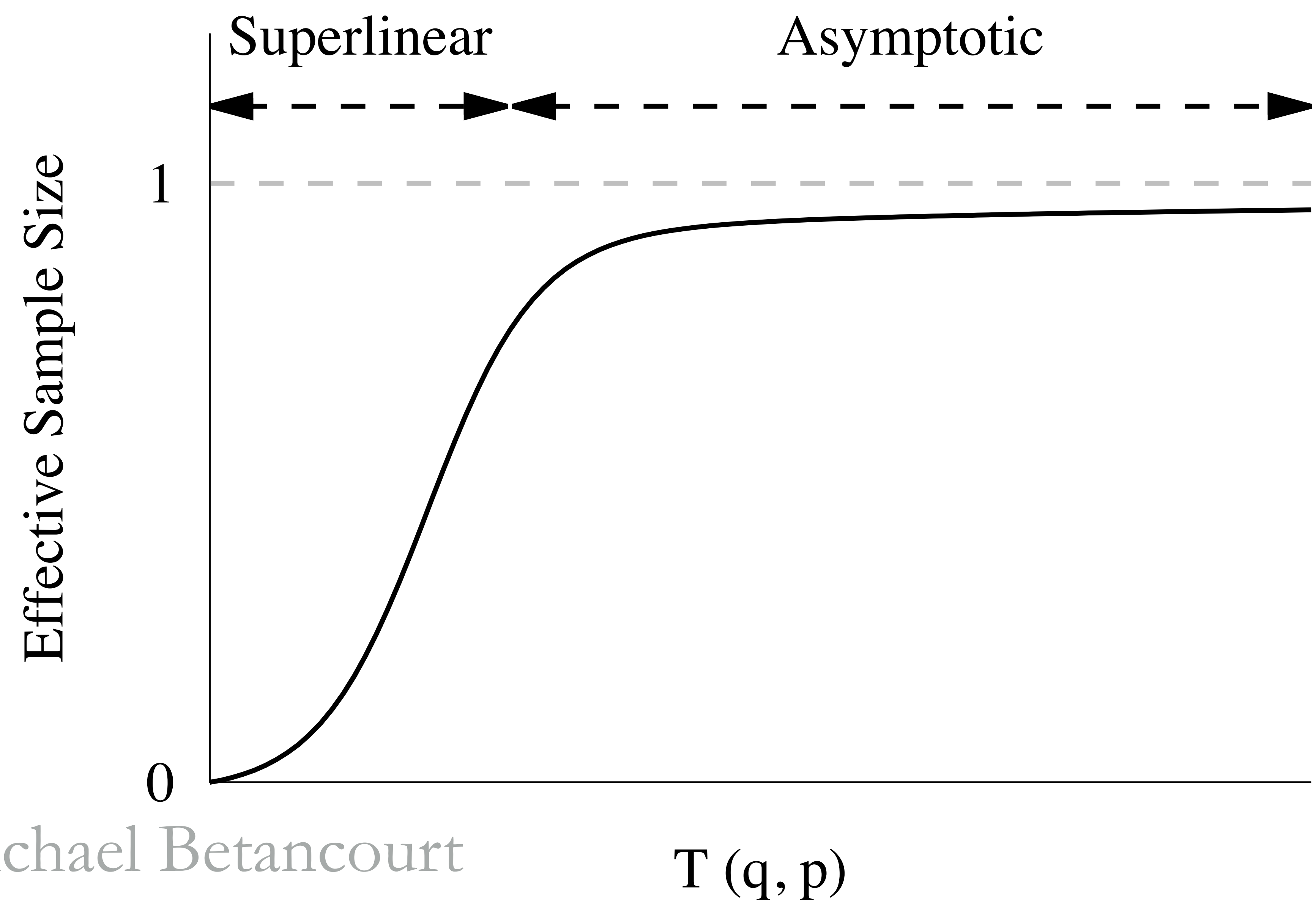
The momenta resampling, for example, determines how effectively we explore between energy level sets.



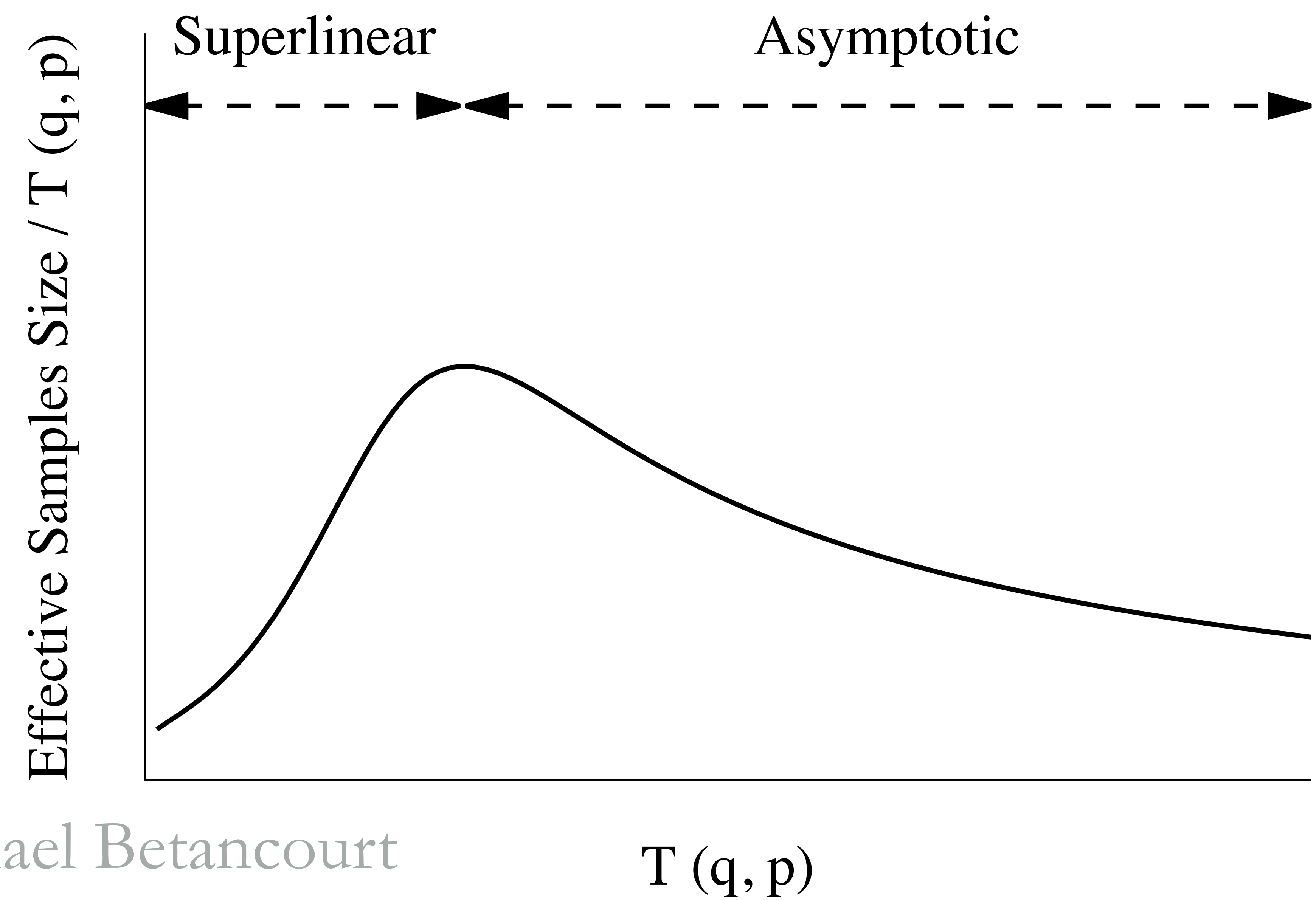
The momenta resampling, for example, determines how effectively we explore between energy level sets.



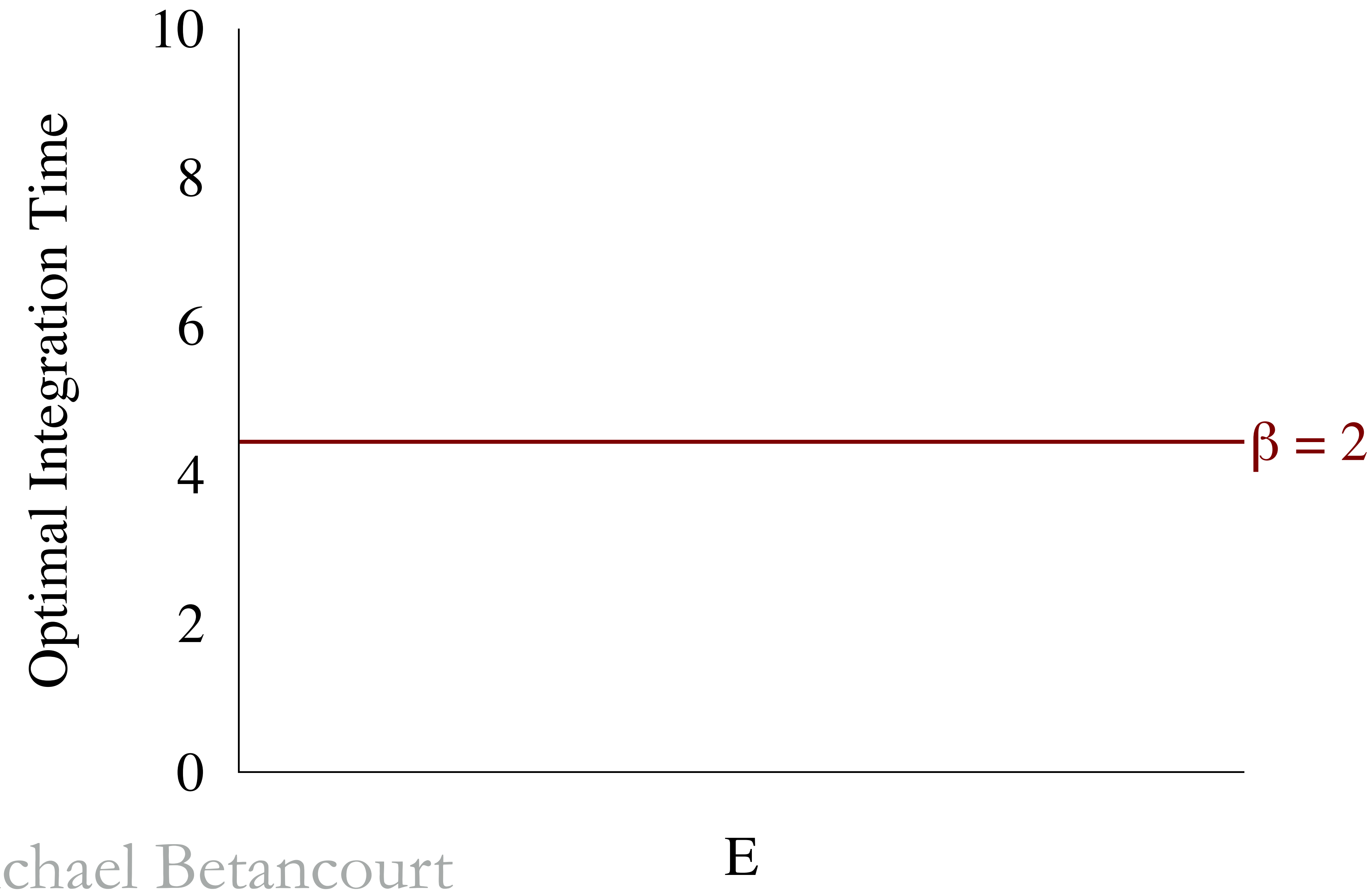
The efficacy in which we can explore the level sets themselves is determined by the integration time.



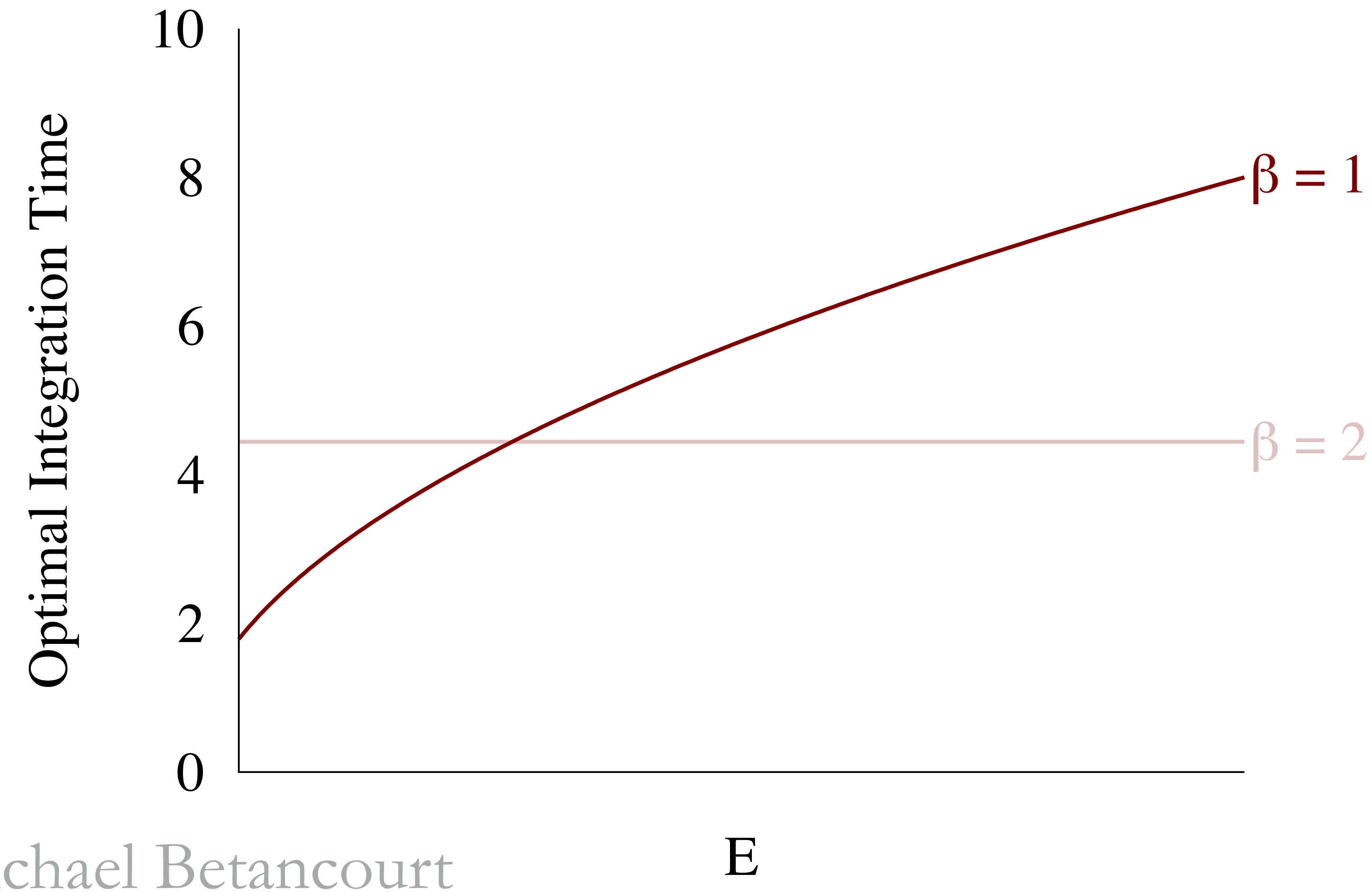
The efficacy in which we can explore the level sets themselves is determined by the integration time.



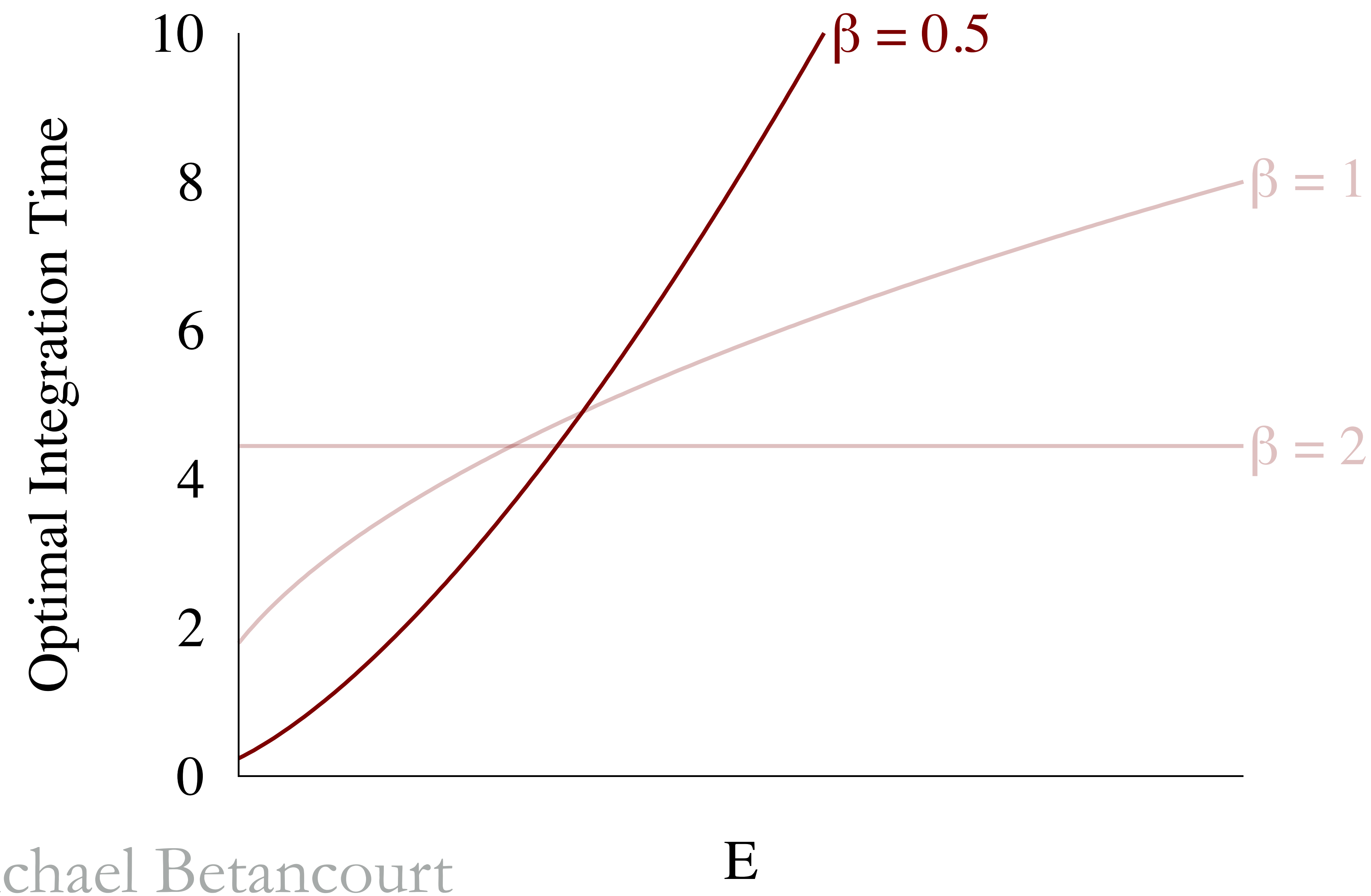
The optimal integration time, however, will in general *vary* from level set to level set.



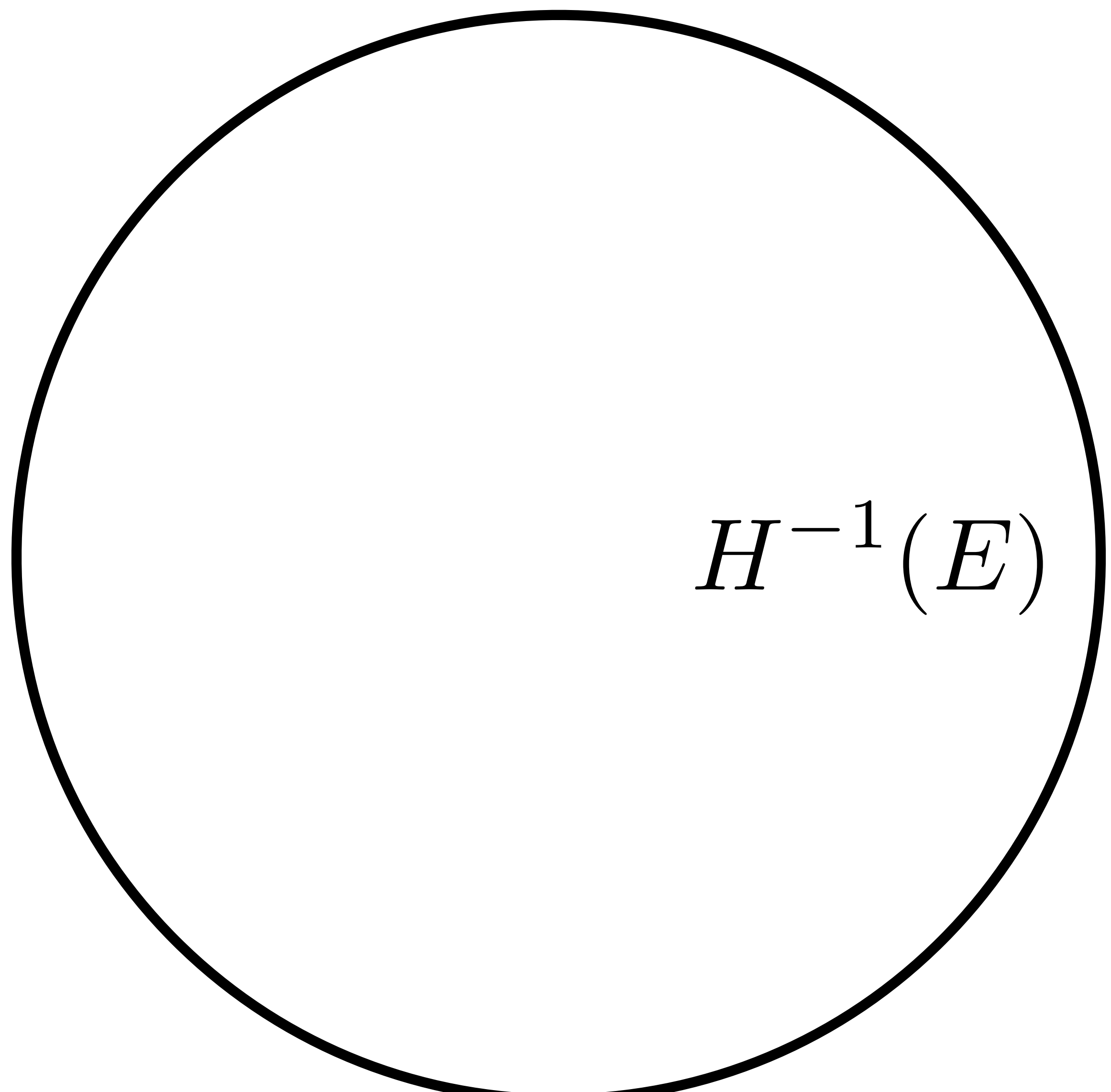
The optimal integration time, however, will in general *vary* from level set to level set.



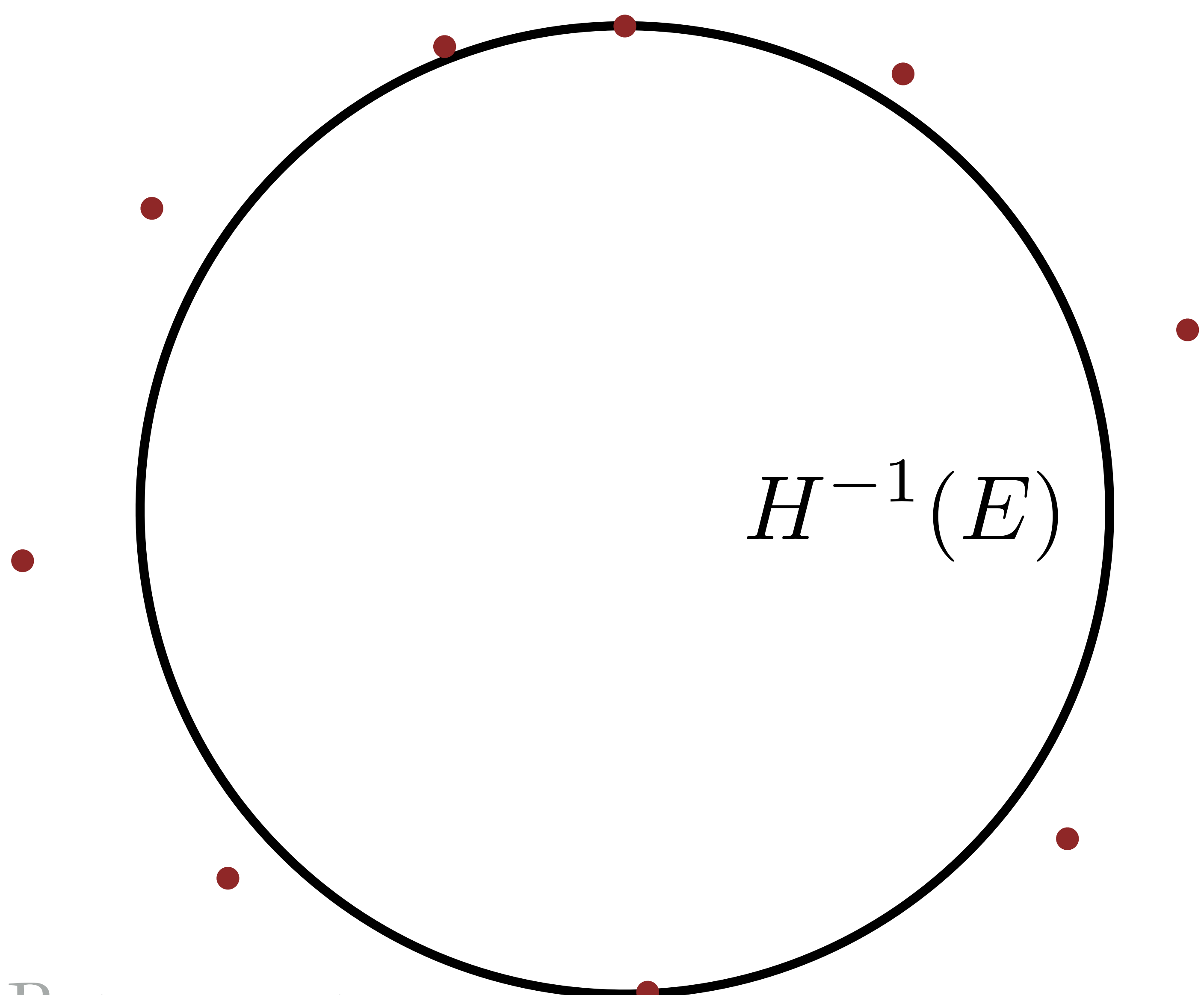
The optimal integration time, however, will in general *vary* from level set to level set.



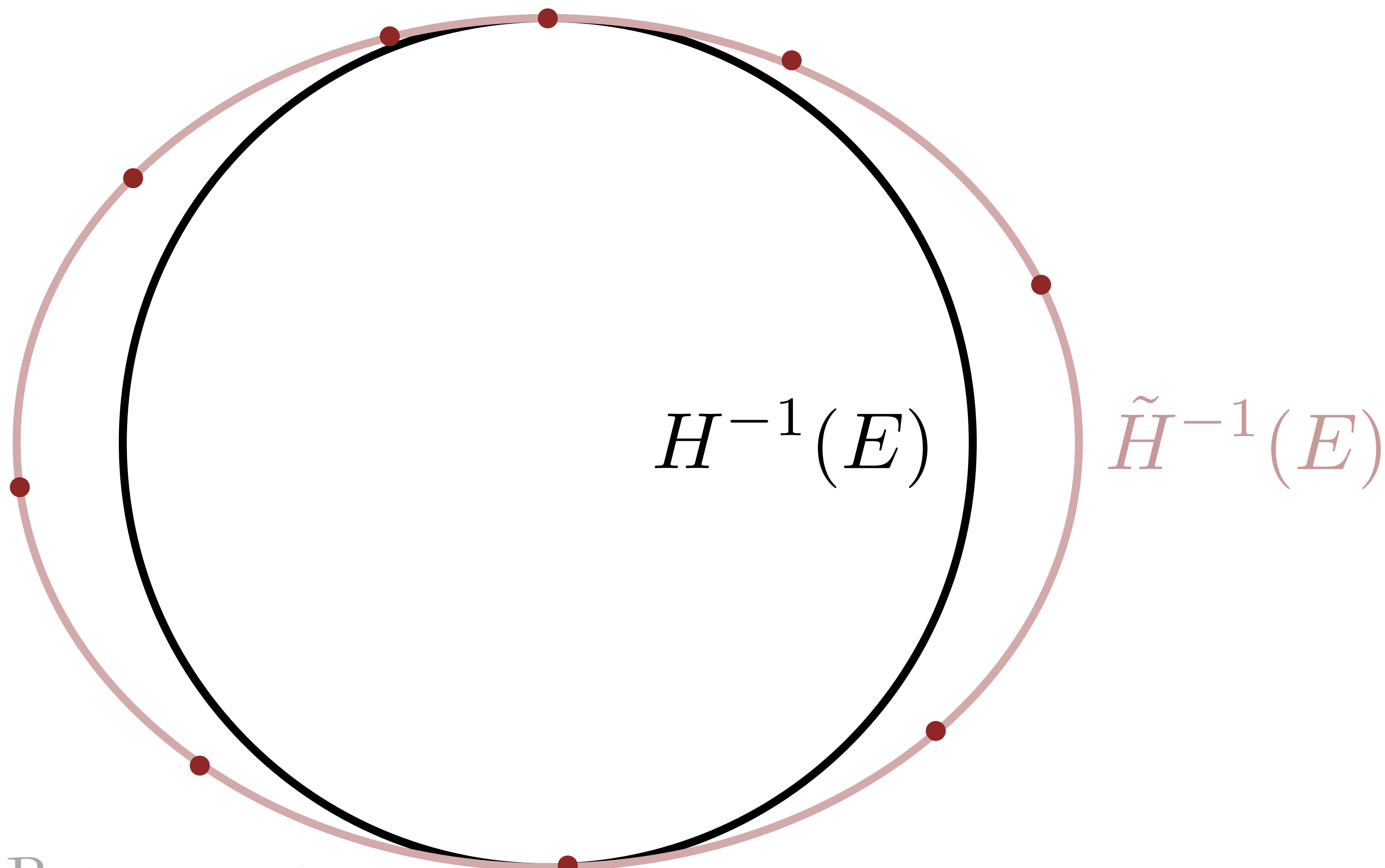
The particular geometry of the level sets also determines the accuracy of symplectic integration.



The particular geometry of the level sets also determines the accuracy of symplectic integration.



The particular geometry of the level sets also determines the accuracy of symplectic integration.



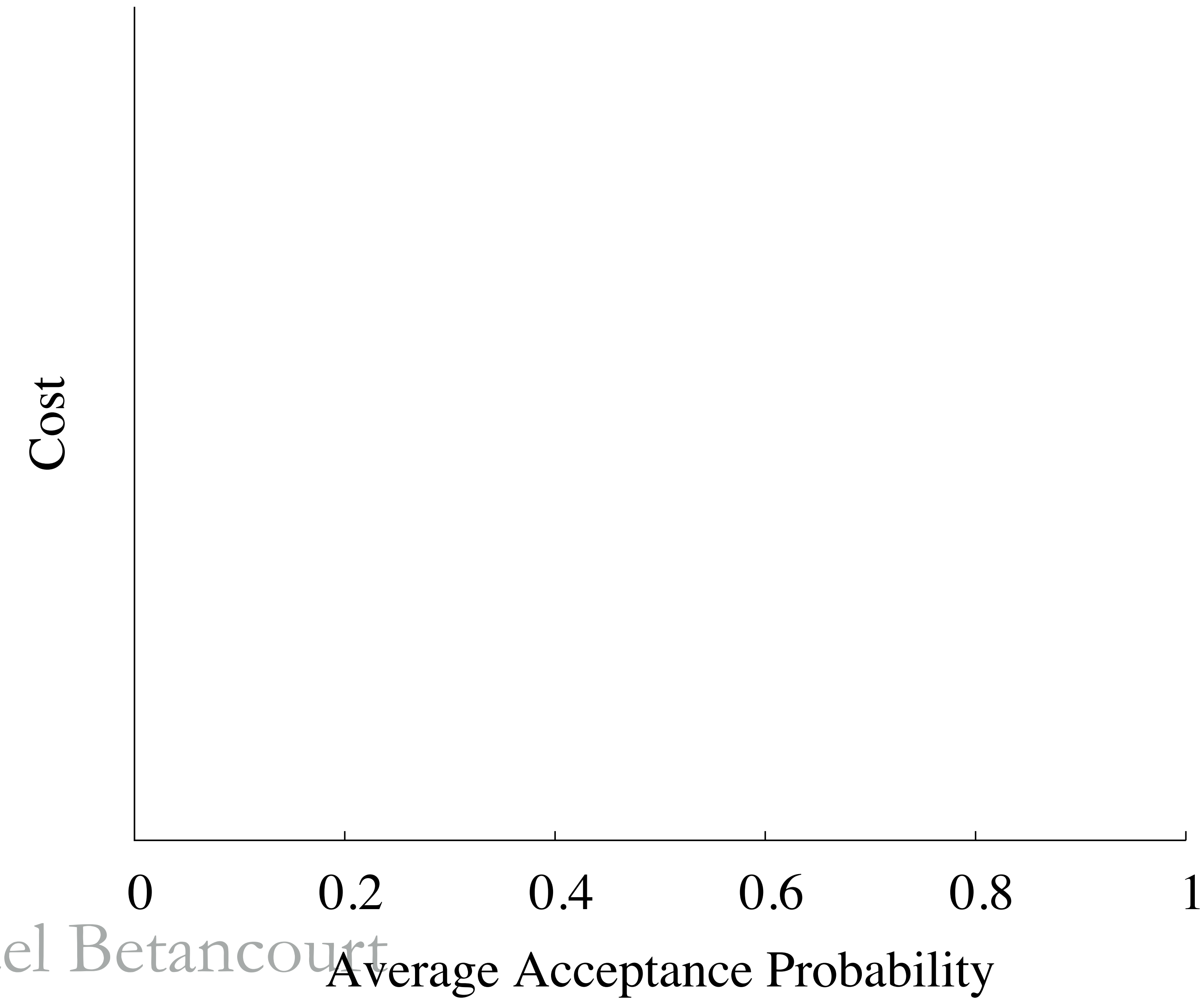
The particular geometry of the level sets also determines the accuracy of symplectic integration.

$$\tilde{H}(q,p) = H(q,p) + \epsilon^2 H_2(q,p) + \ldots$$

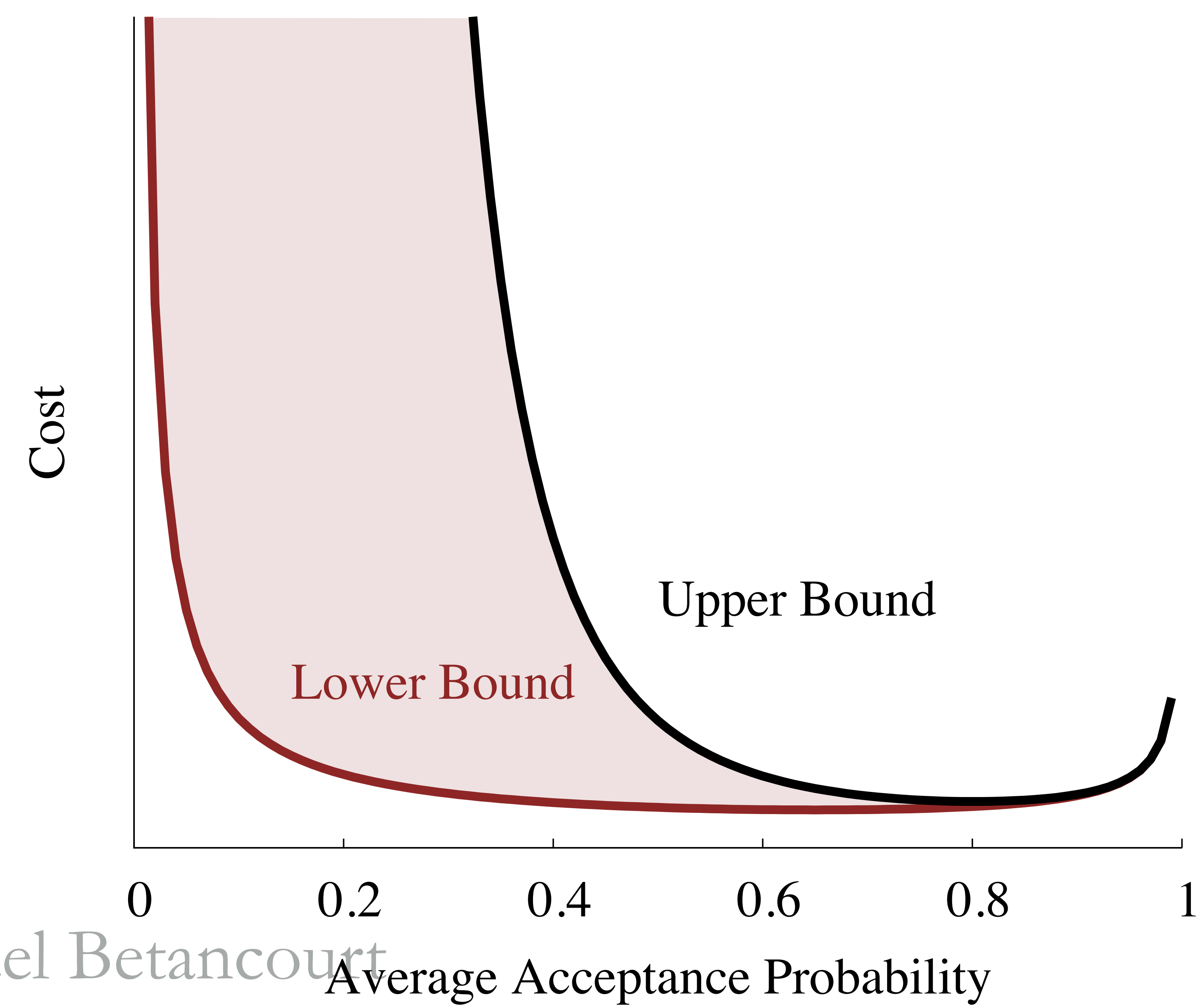
The particular geometry of the level sets also determines the accuracy of symplectic integration.

$$\tilde{H}(q,p) = H(q,p) + \epsilon^2 H_2(q,p) + \ldots$$

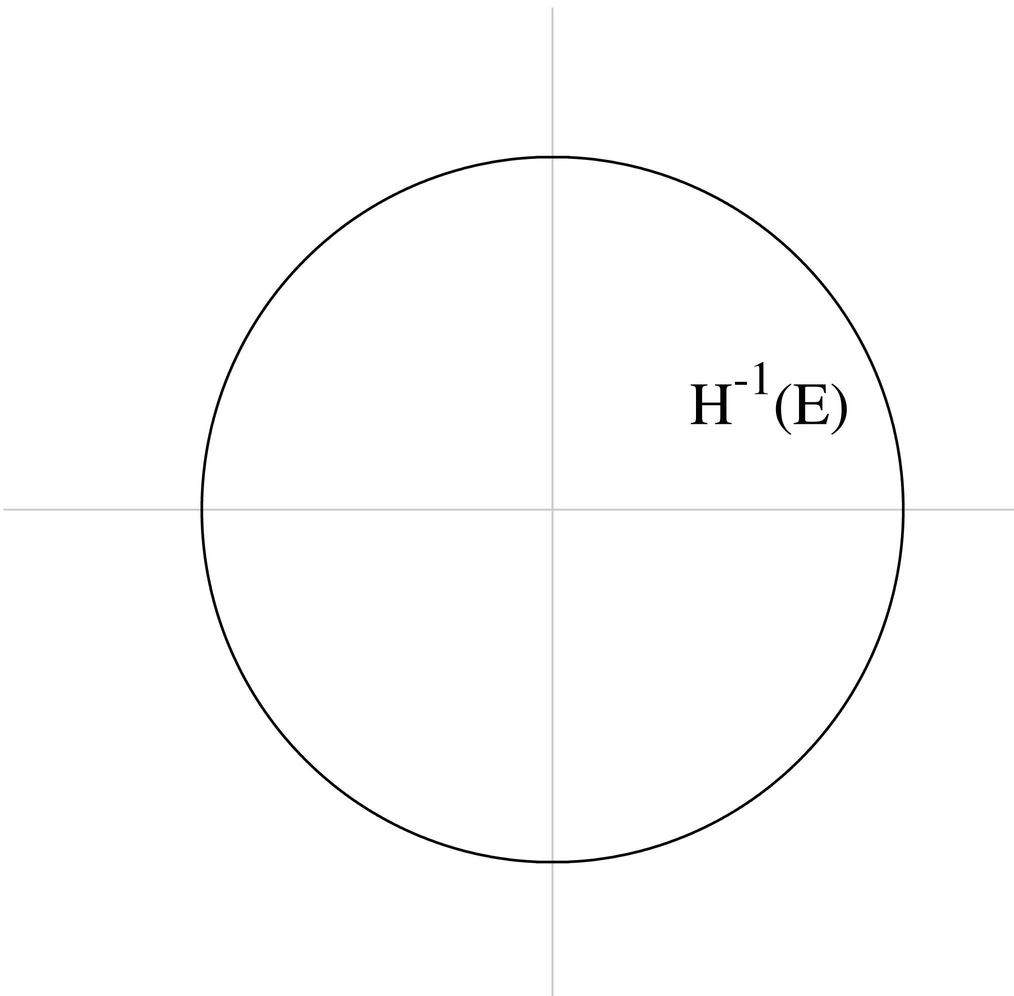
We can exploit this geometric relationship to develop criteria to optimize the choice of integrator step size.



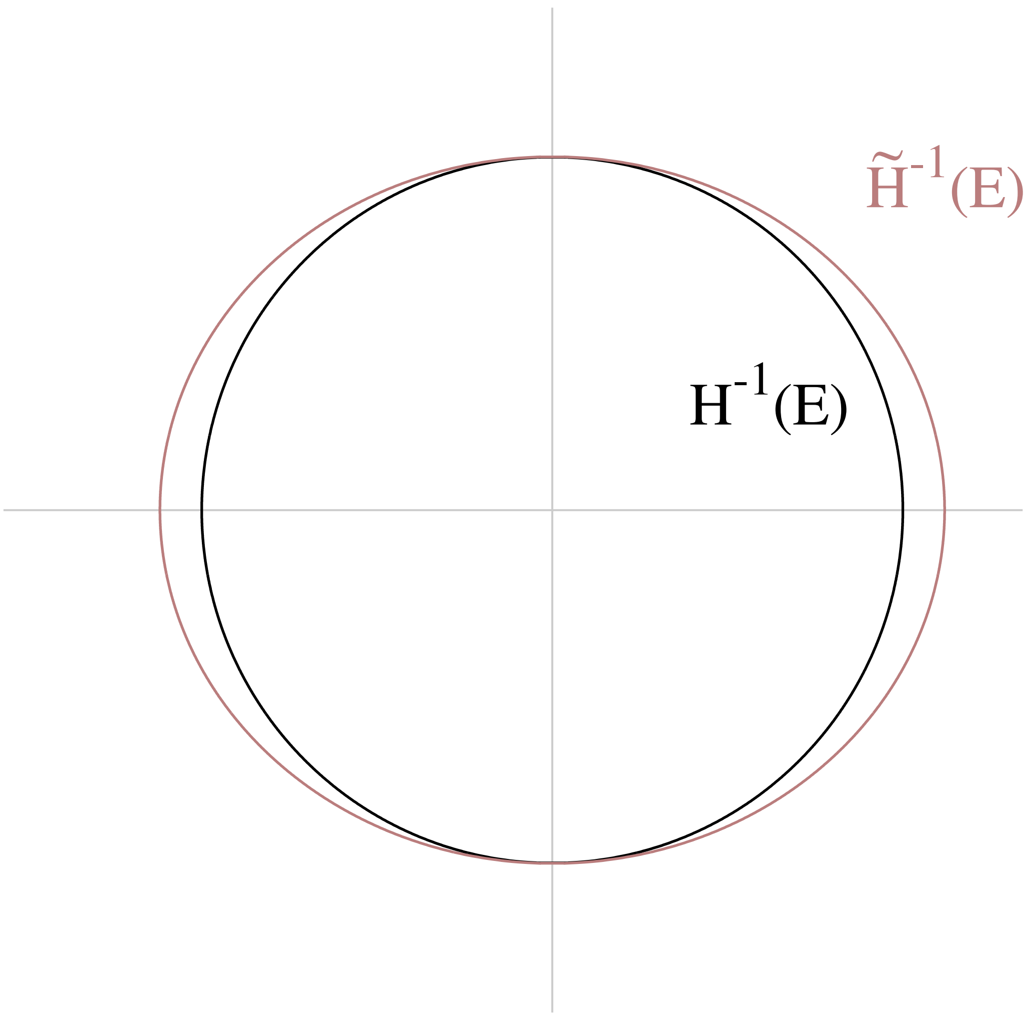
We can exploit this geometric relationship to develop criteria to optimize the choice of integrator step size.



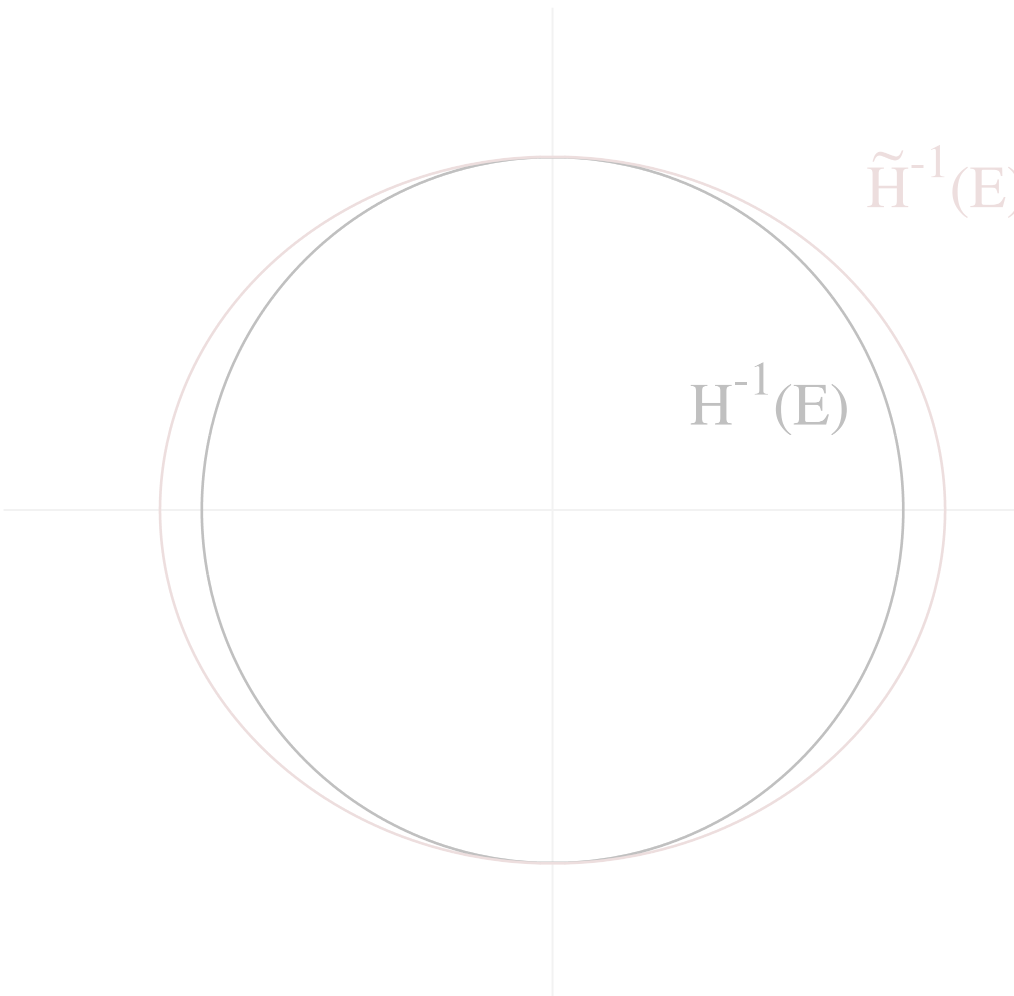
The geometry of the modified level sets also provides sensitive diagnostic information about convergence.

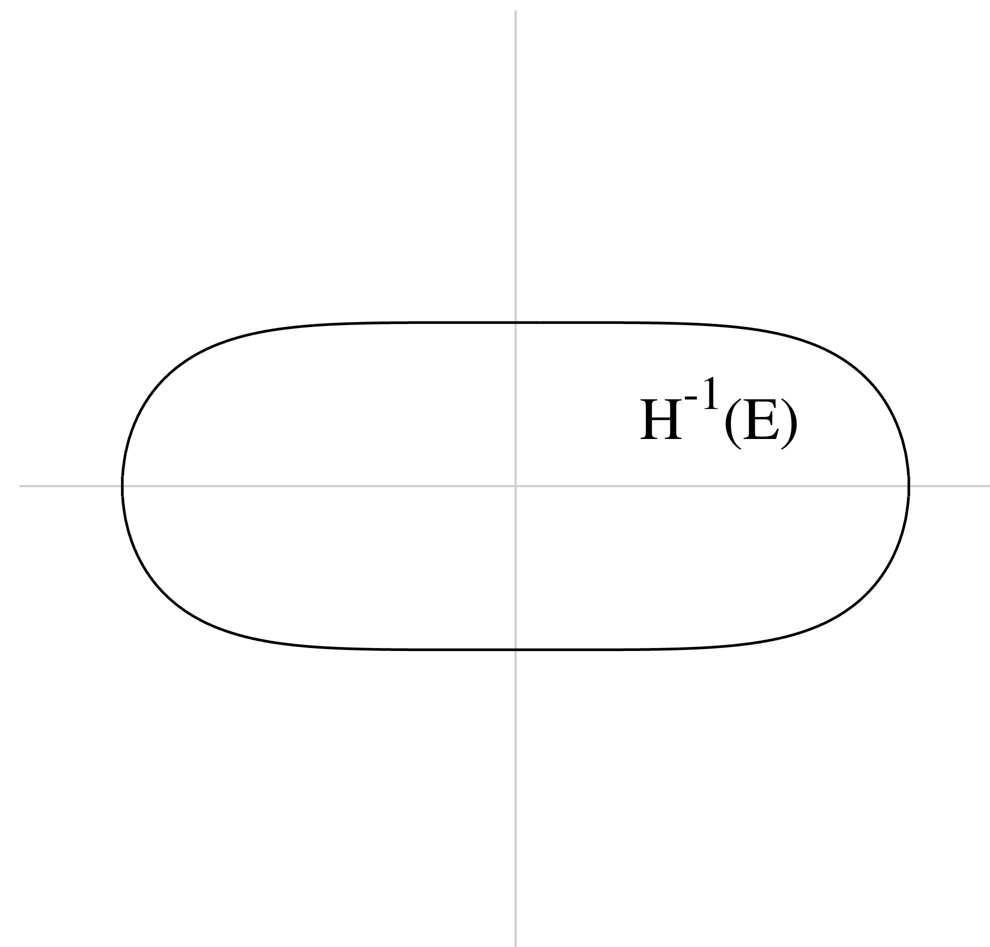


The geometry of the modified level sets also provides sensitive diagnostic information about convergence.

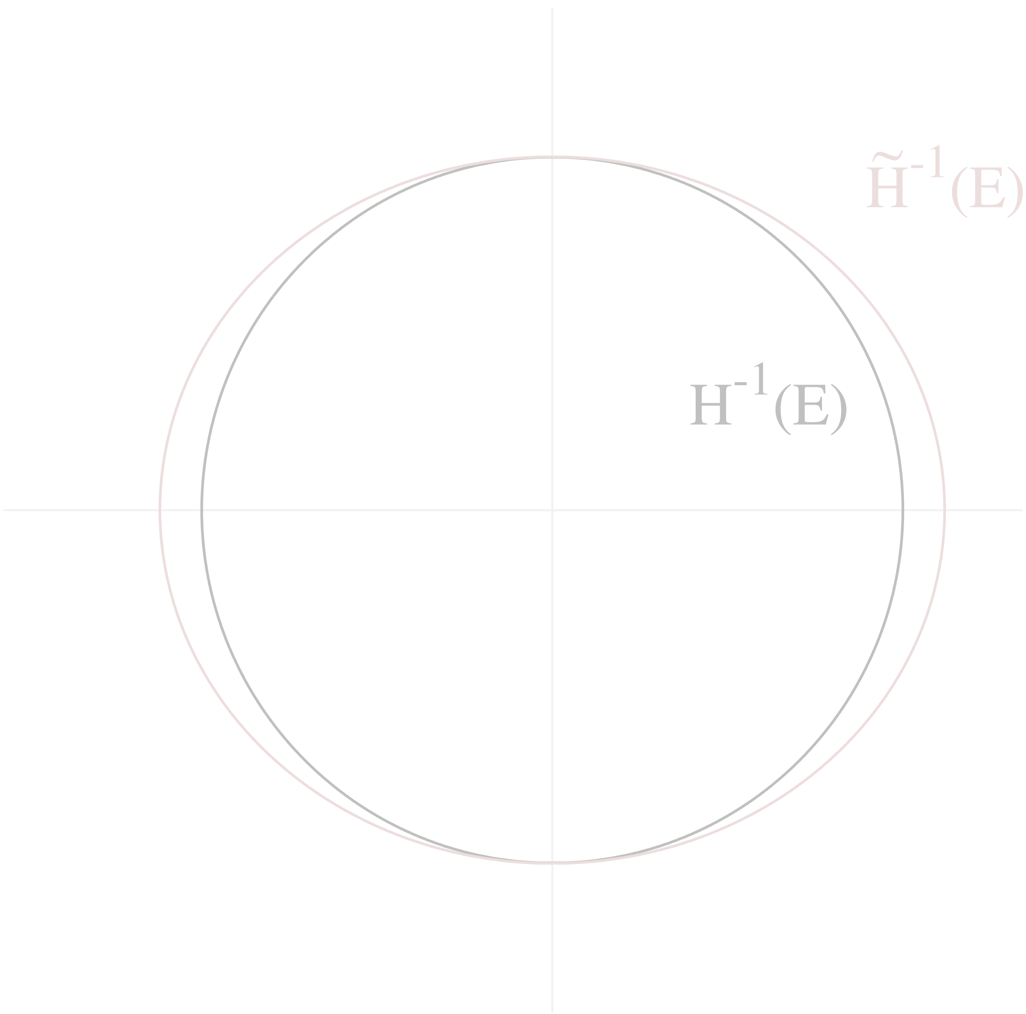


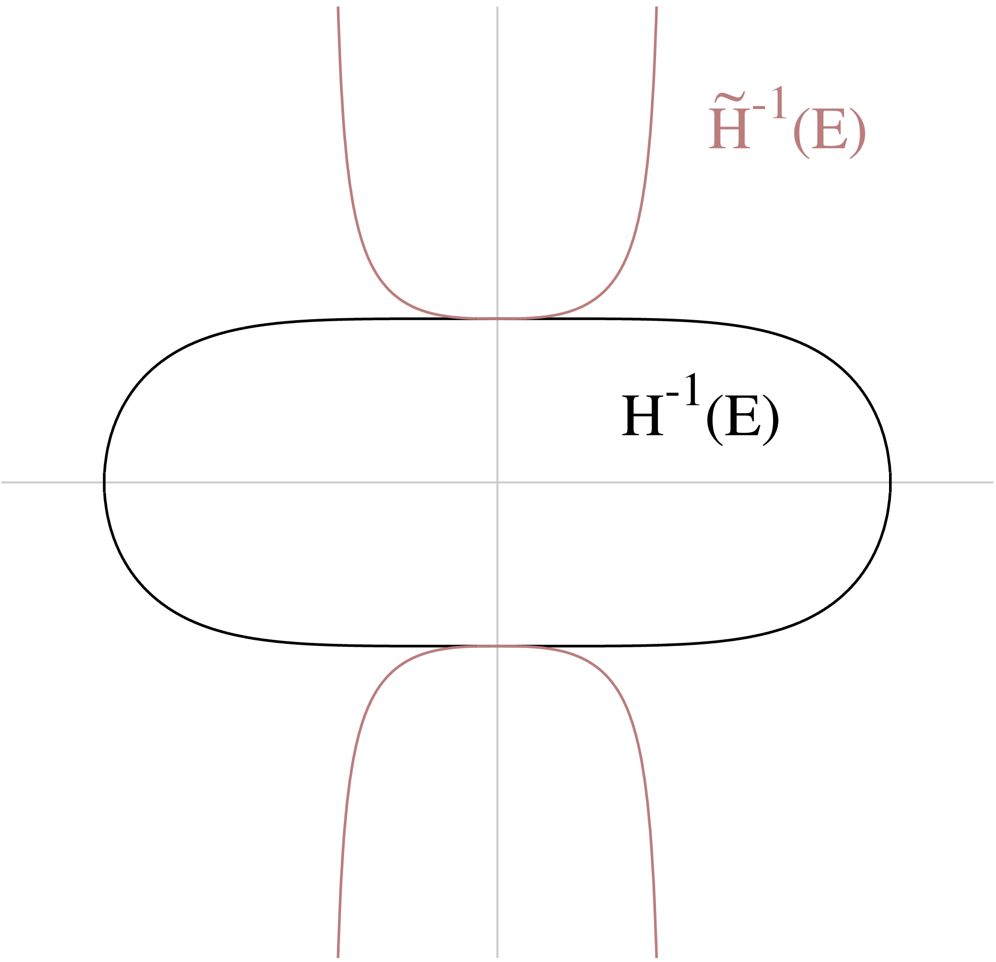
The geometry of the modified level sets also provides sensitive diagnostic information about convergence.



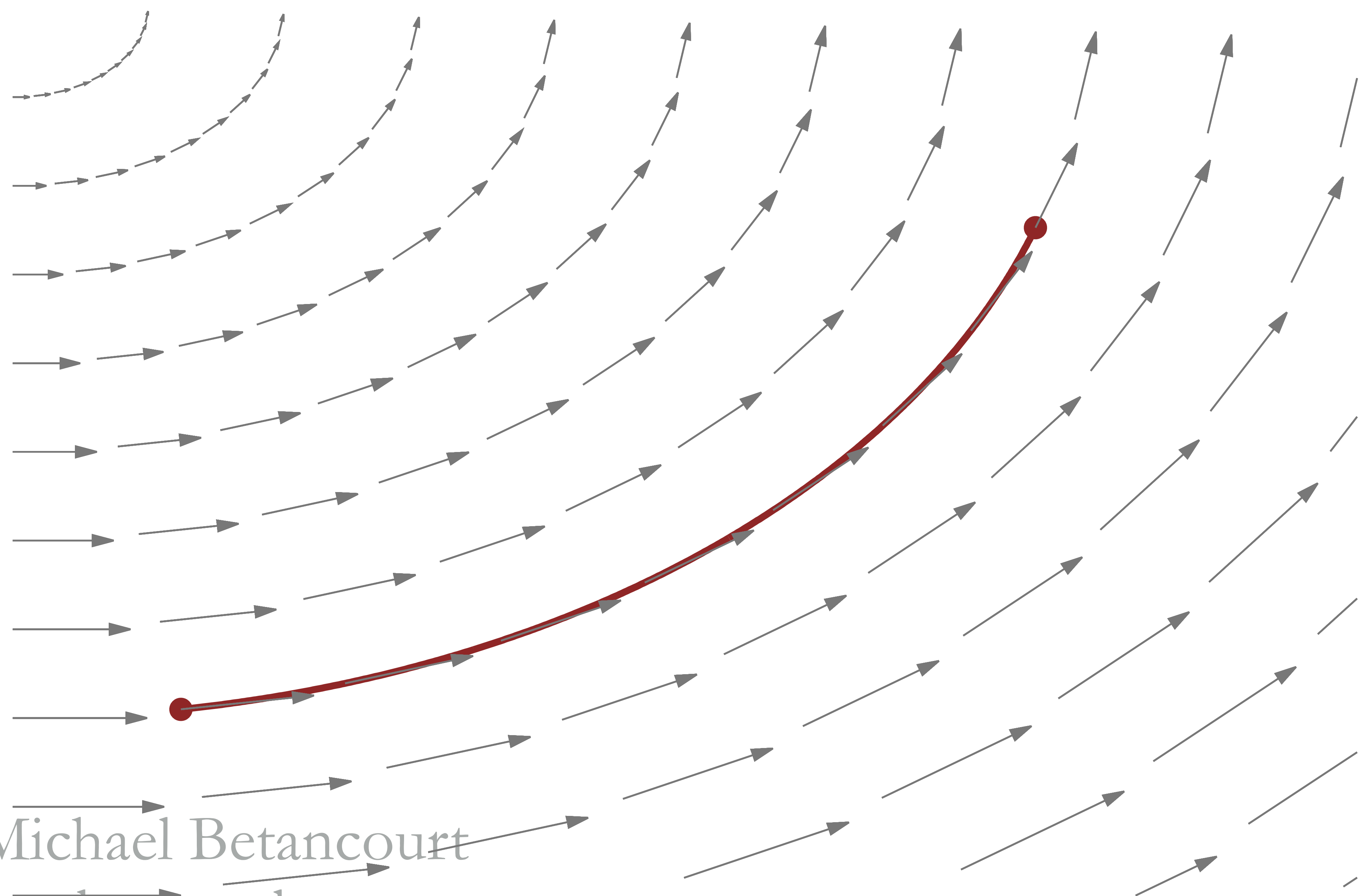


The geometry of the modified level sets also provides sensitive diagnostic information about convergence.

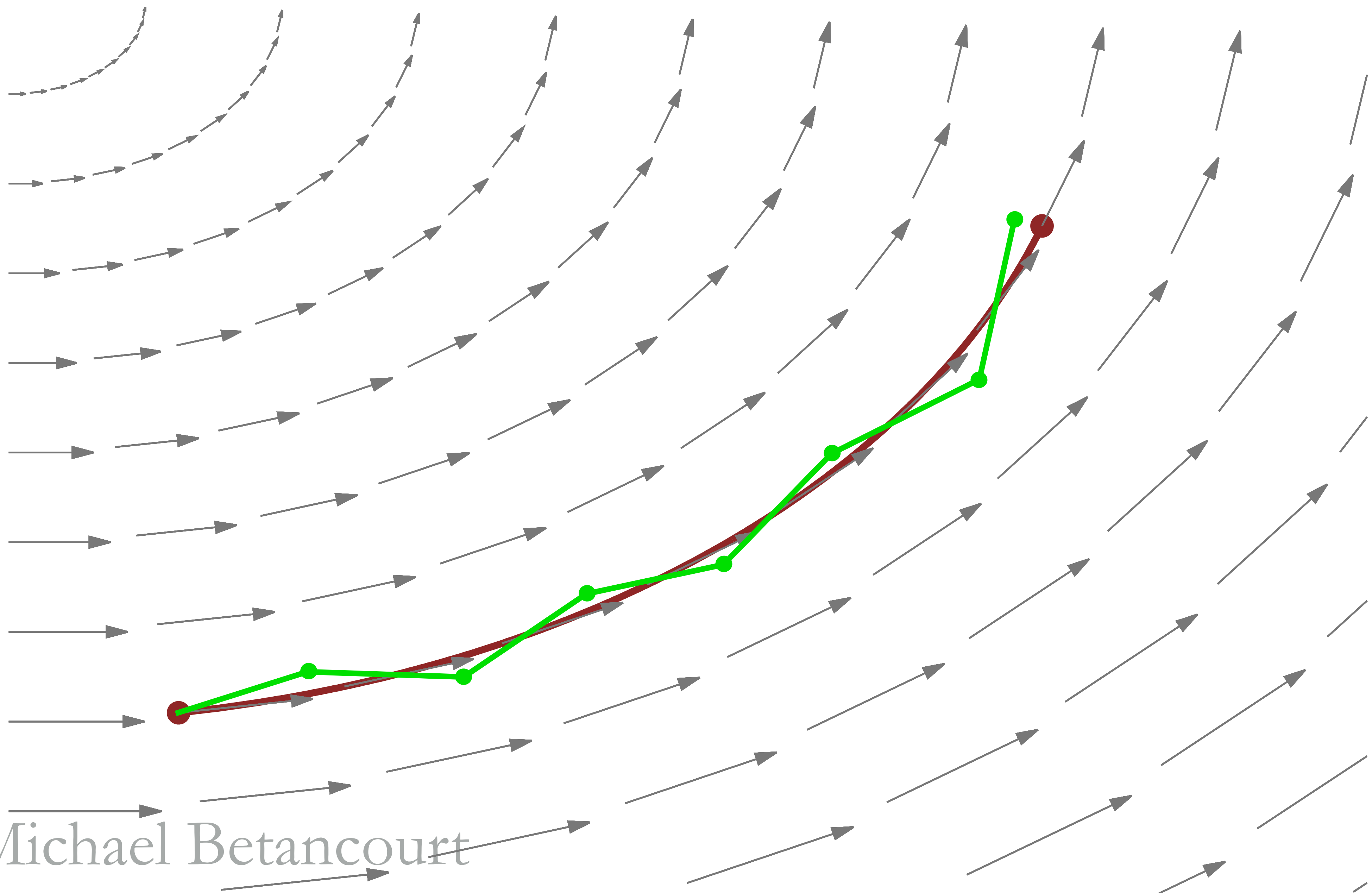




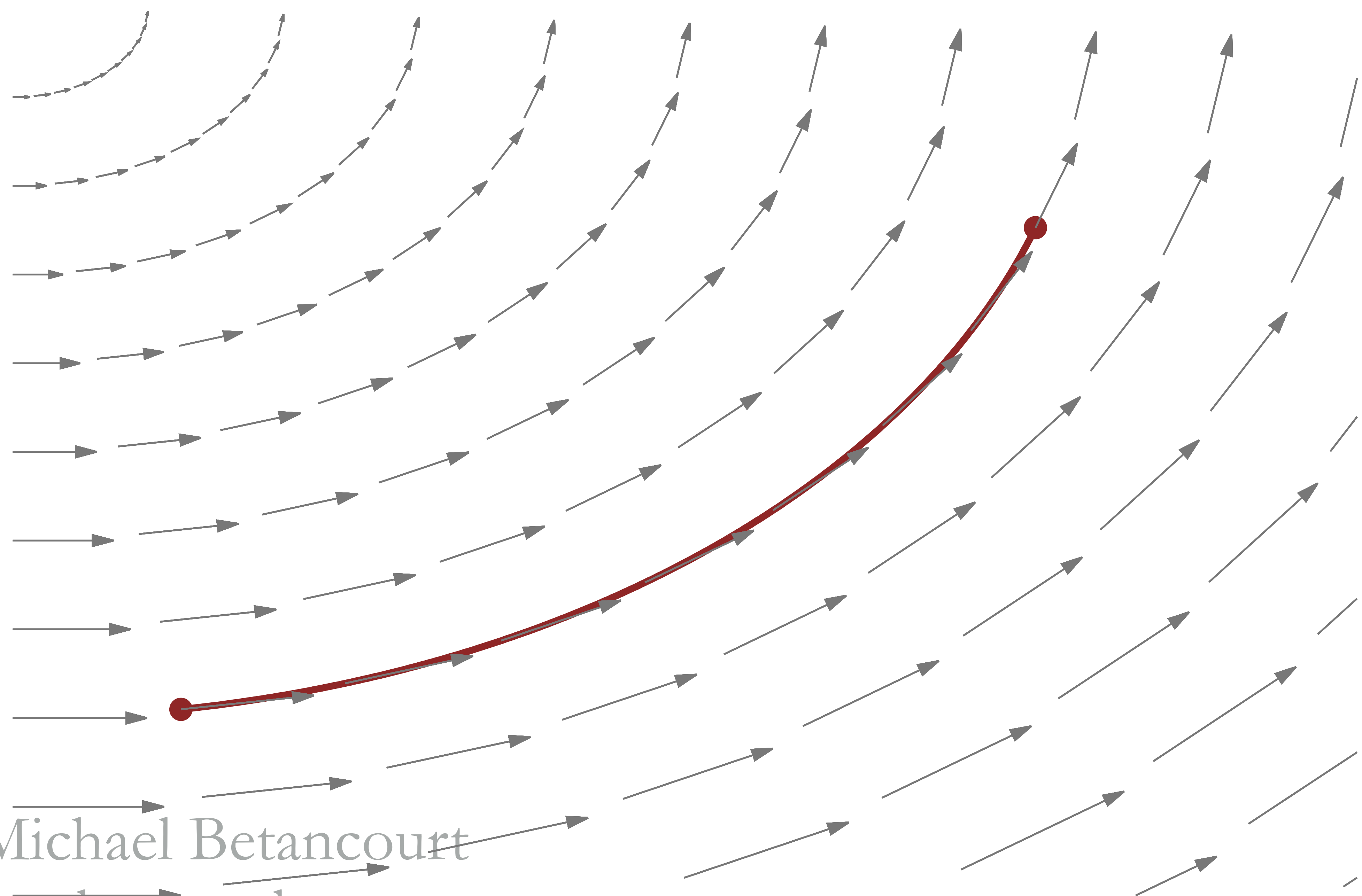
Stable numerical integration indicates accurate exploration of the given level set.



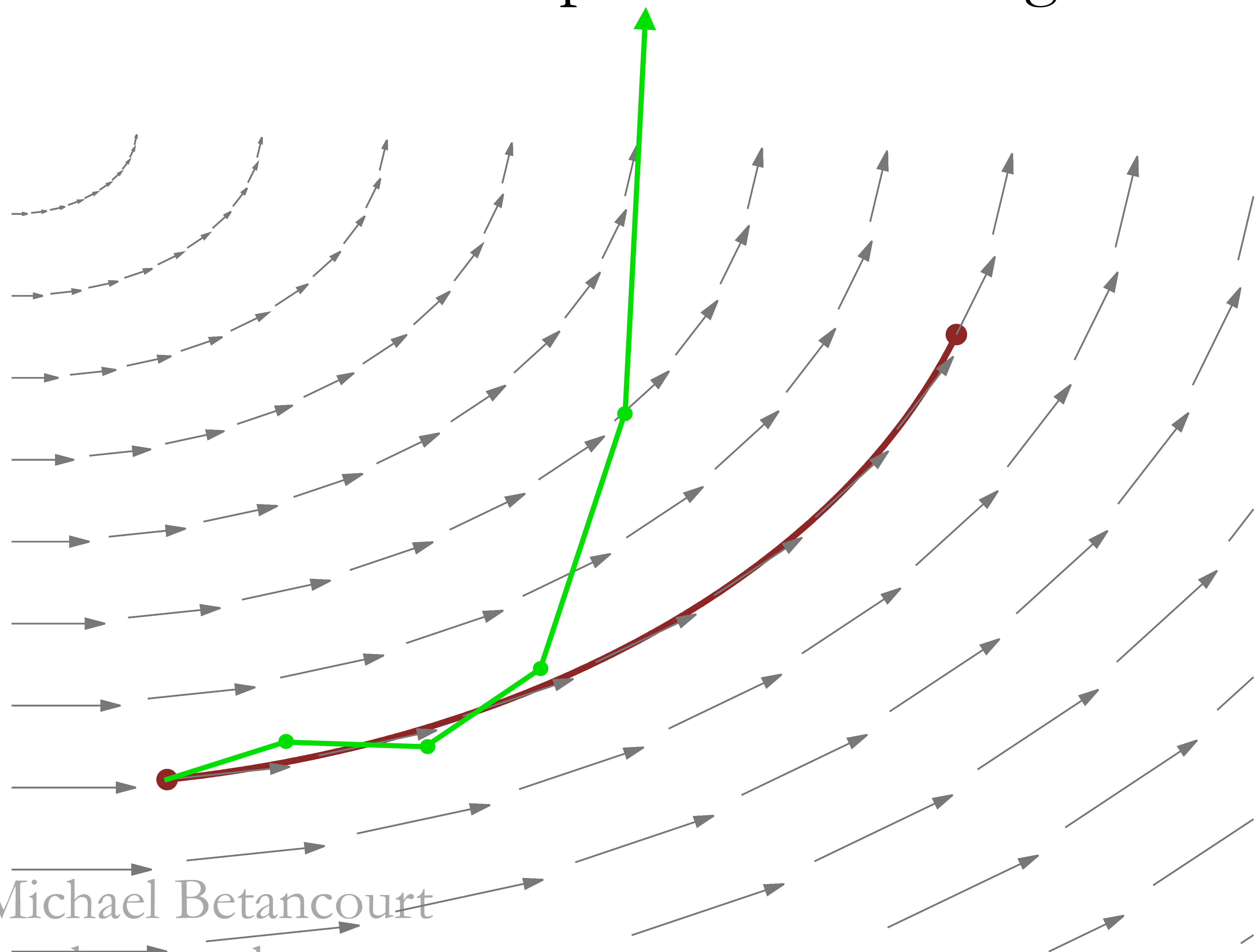
Stable numerical integration indicates accurate exploration of the given level set.



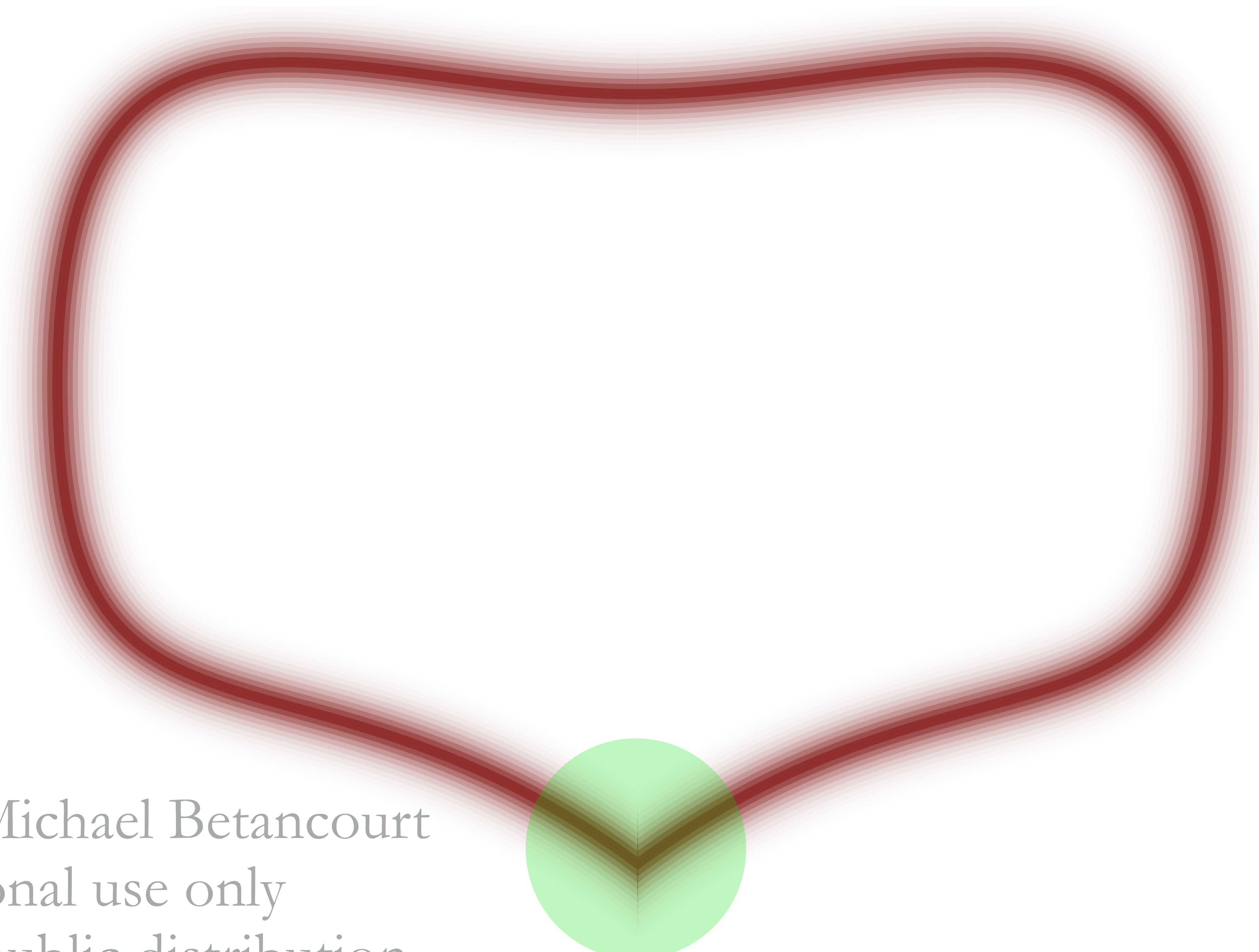
Unstable numerical integration, which yields a *divergence*, indicates inaccurate exploration of the given level set.



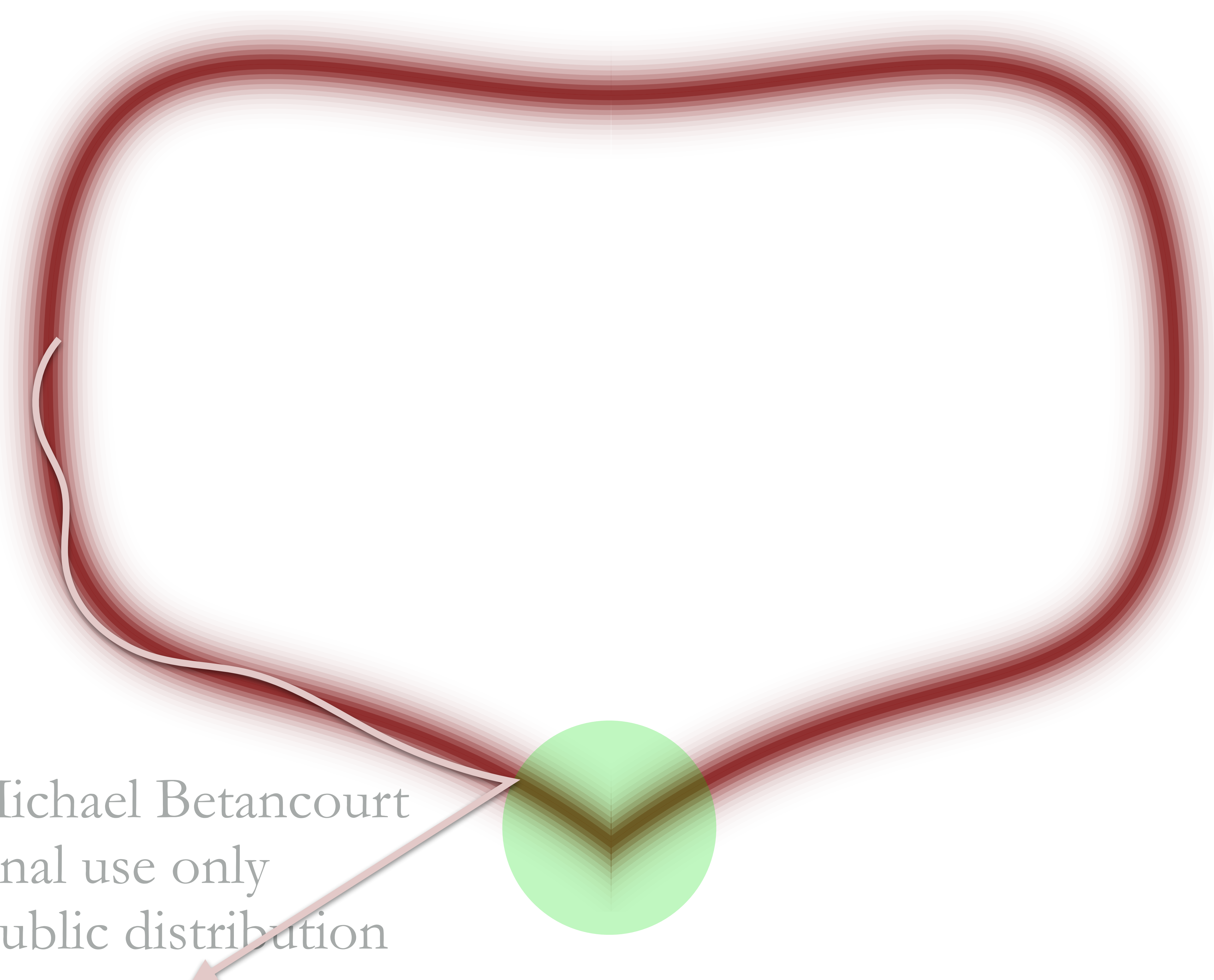
Unstable numerical integration, which yields a *divergence*, indicates inaccurate exploration of the given level set.



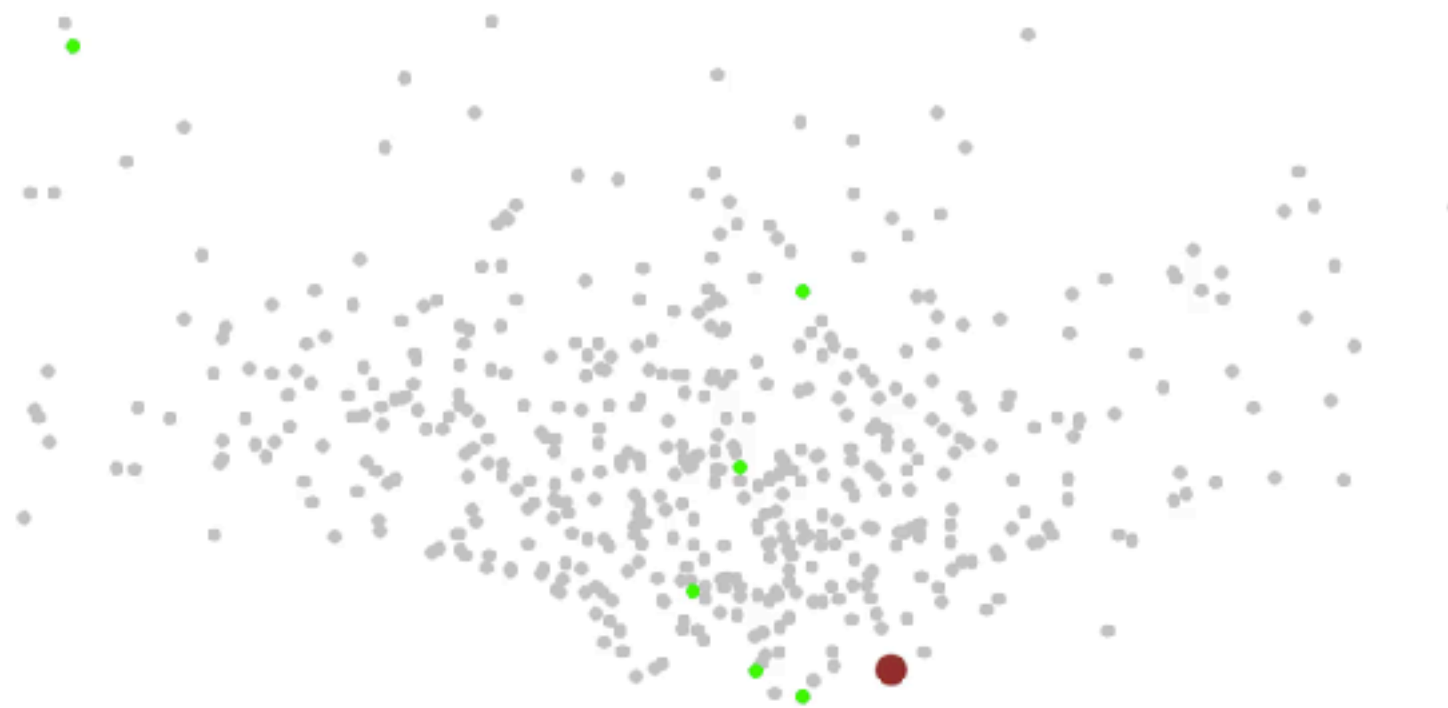
More formally these divergences identify some of the pathologies which obstruct geometric ergodicity.



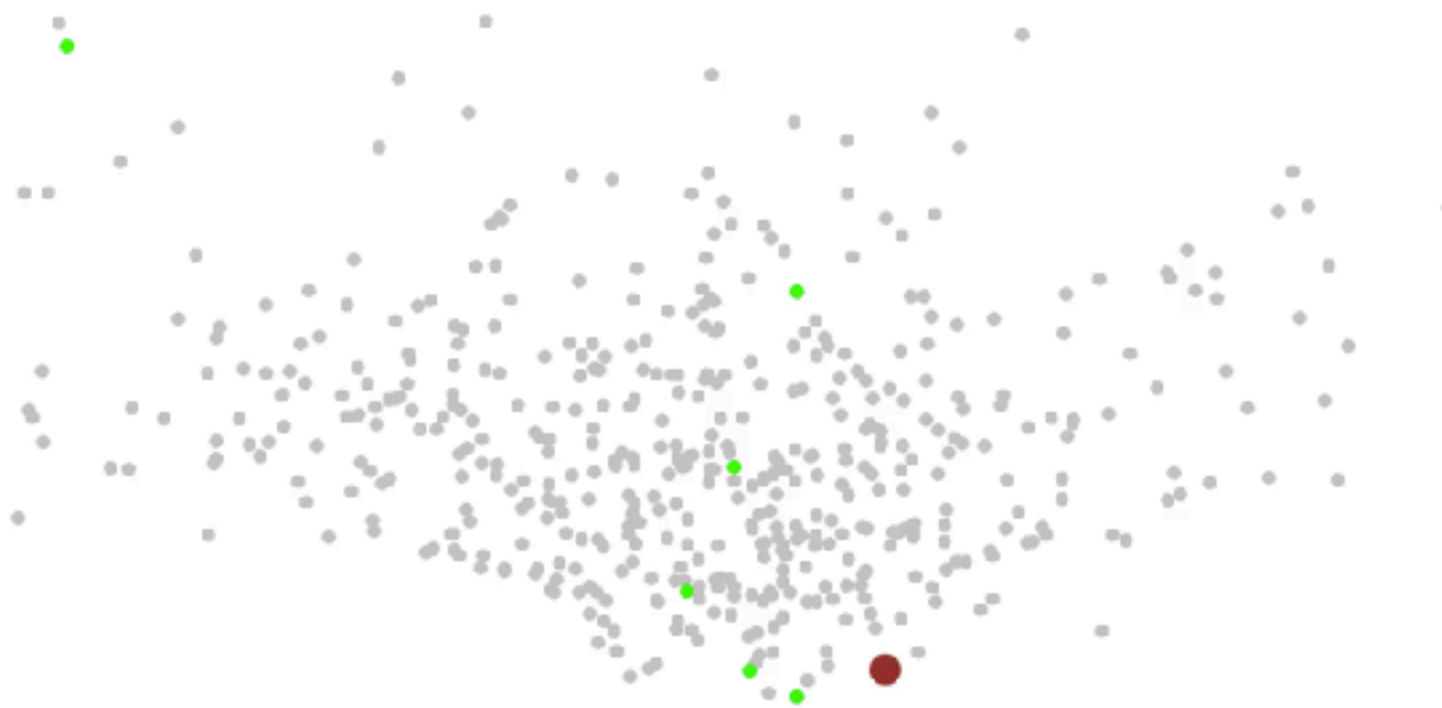
More formally these divergences identify some of the pathologies which obstruct geometric ergodicity.



More formally these divergences identify some of the pathologies which obstruct geometric ergodicity.



More formally these divergences identify some of the pathologies which obstruct geometric ergodicity.



These geometric properties defines novel diagnostics of geometric ergodicity for Hamiltonian Monte Carlo.

```
Warning messages:
1: There were 2 chains where the estimated Bayesian
Fraction of Missing Information was low. See
http://mc-stan.org/misc/warnings.html#bfmi-low
```

```
Warning messages:
1: There were 2 divergent transitions after warmup.
Increasing adapt_delta above 0.8 may help. See
http://mc-stan.org/misc/warnings.html#divergent-
transitions-after-warmup
```

Software tools like Stan implement MCMC, allowing users to focus on building models and analyzing chains.

Software tools like Stan implement MCMC, allowing users to focus on building models and analyzing chains.

Modeling Language

Software tools like Stan implement MCMC, allowing users to focus on building models and analyzing chains.

Modeling Language

Automatic Differentiation

Software tools like Stan implement MCMC, allowing users to focus on building models and analyzing chains.

Modeling Language

Automatic Differentiation

Hamiltonian Monte Carlo

https://betanalpha.github.io/writing

betanalpha.github.io

Consulting Speaking Writing Resources

Michael Betancourt

Once and future physicist masquerading as an applied statistician.

New York, NY

Email

Twitter

GitHub

Writing

Preprints of my work are posted on the arXiv as much as possible. Highlights include a long but comprehensive introduction to statistical computing and Hamiltonian Monte Carlo targeted at applied researches and a more theoretical treatment of the geometric foundations of Hamiltonian Monte Carlo.

I've also been experimenting with a nontraditional introduction to some of the more formal aspects of probability theory and statistical computation, although fair warning that this draft is long overdue for a reorganization. Still, comments are always welcome.

Recently I have been taking advantage of notebook environments such as knitr and Jupyter to develop case studies demonstrating important concepts in statistical workflow and modeling with Stan.

Bayes Sparse Regression

In this case study I'll review how sparsity arises in frequentist and Bayesian analyses and discuss the often subtle challenges in implementing sparsity in practical Bayesian analyses.

View (HTML)
betanalpha/knitr_case_studies/bayes_sparse_regression (GitHub)
Dependences: `R, knitr, RStan`
Code License: BSD (3 clause)