Adiabatic Monte Carlo

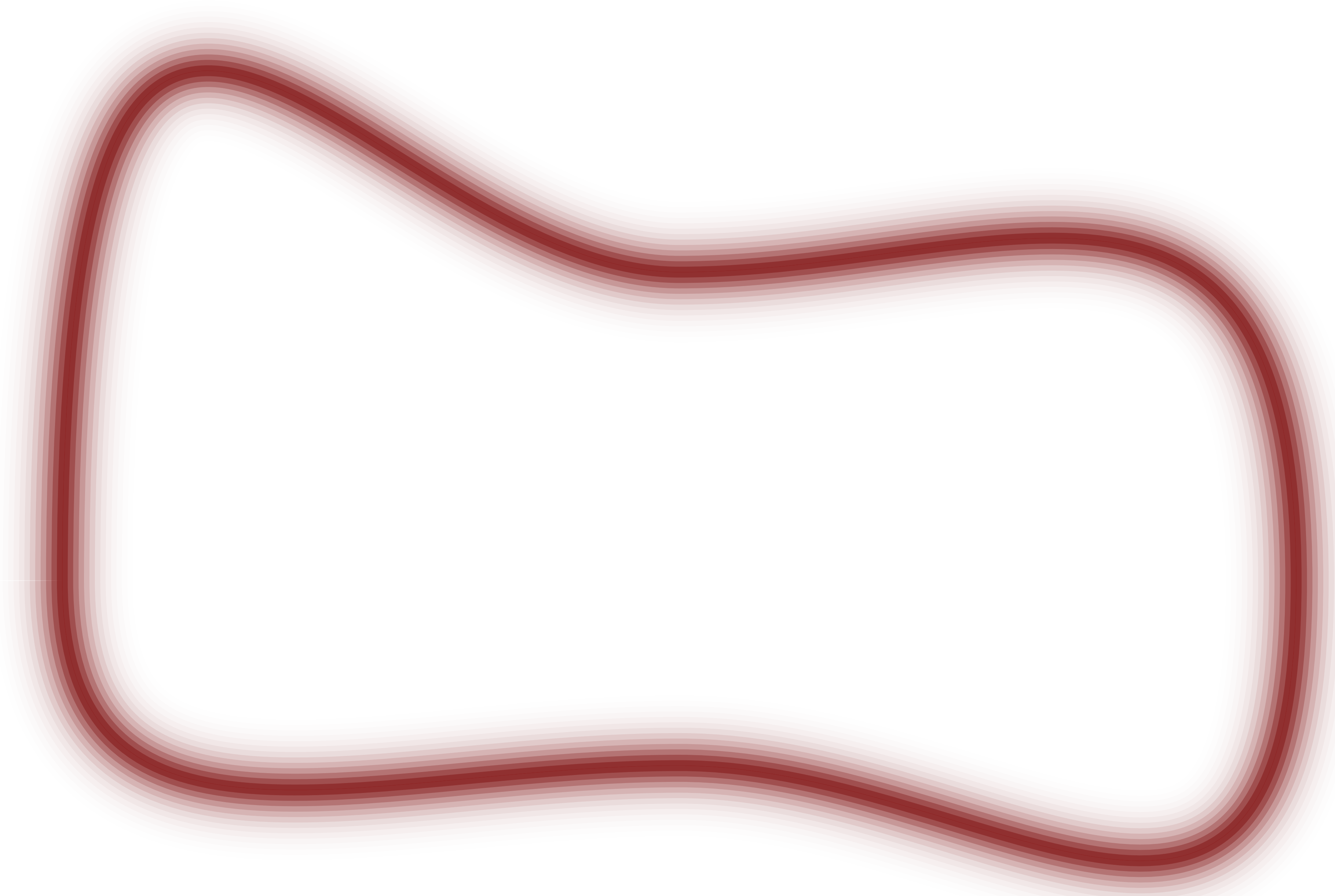
Michael Betancourt @betanalpha
University of Warwick

CRiSM Workshop:
Estimating Constants,
University of Warwick
April 21, 2016

Computational statistics is all about computing expectations with respect to a given target distribution.

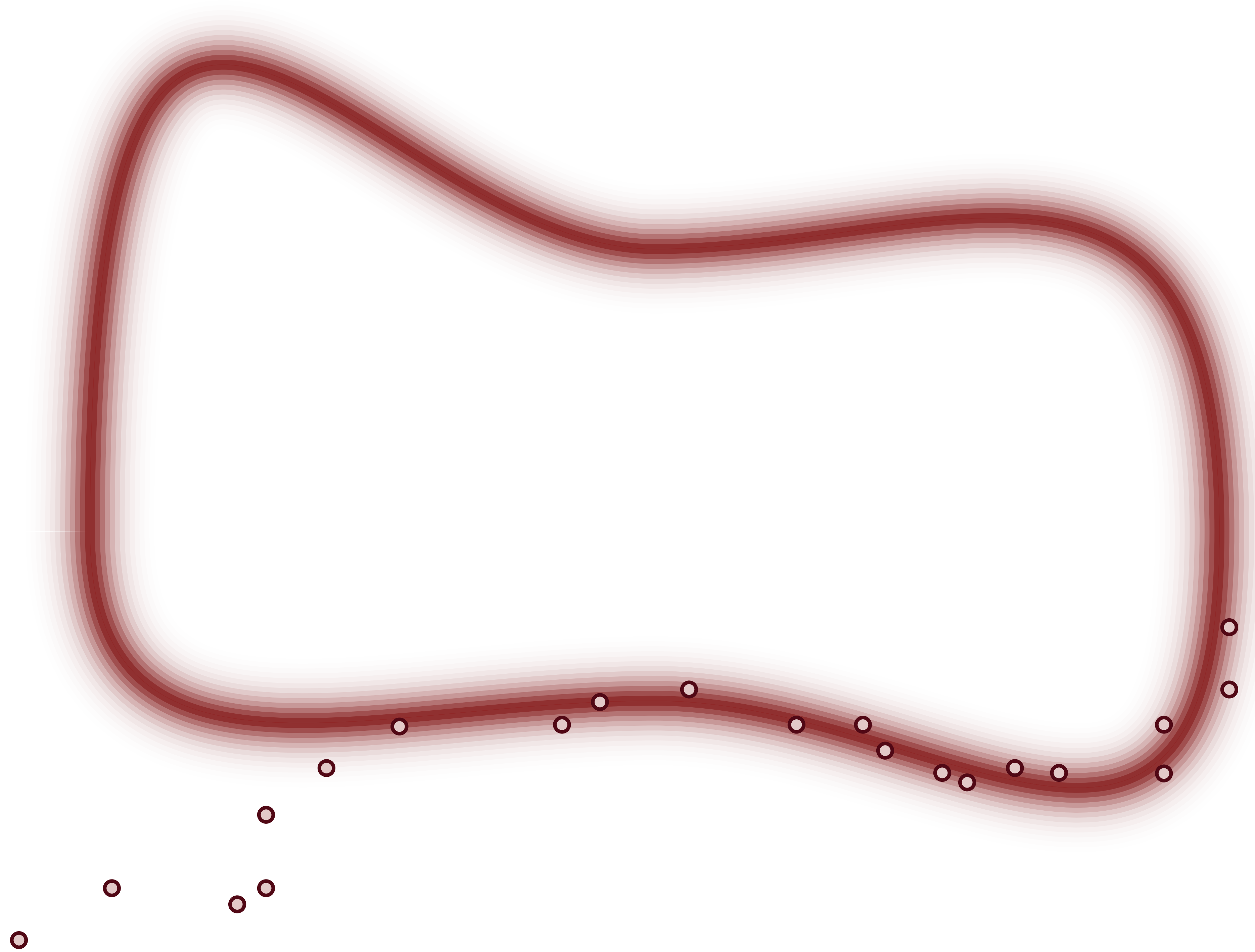
$$\mathbb{E}_{\pi}[f] = \int f(q) \pi(q) \, \mathrm{d}q$$

High-dimensional target distributions exhibit *concentration of measure*, which frustrates these computations.

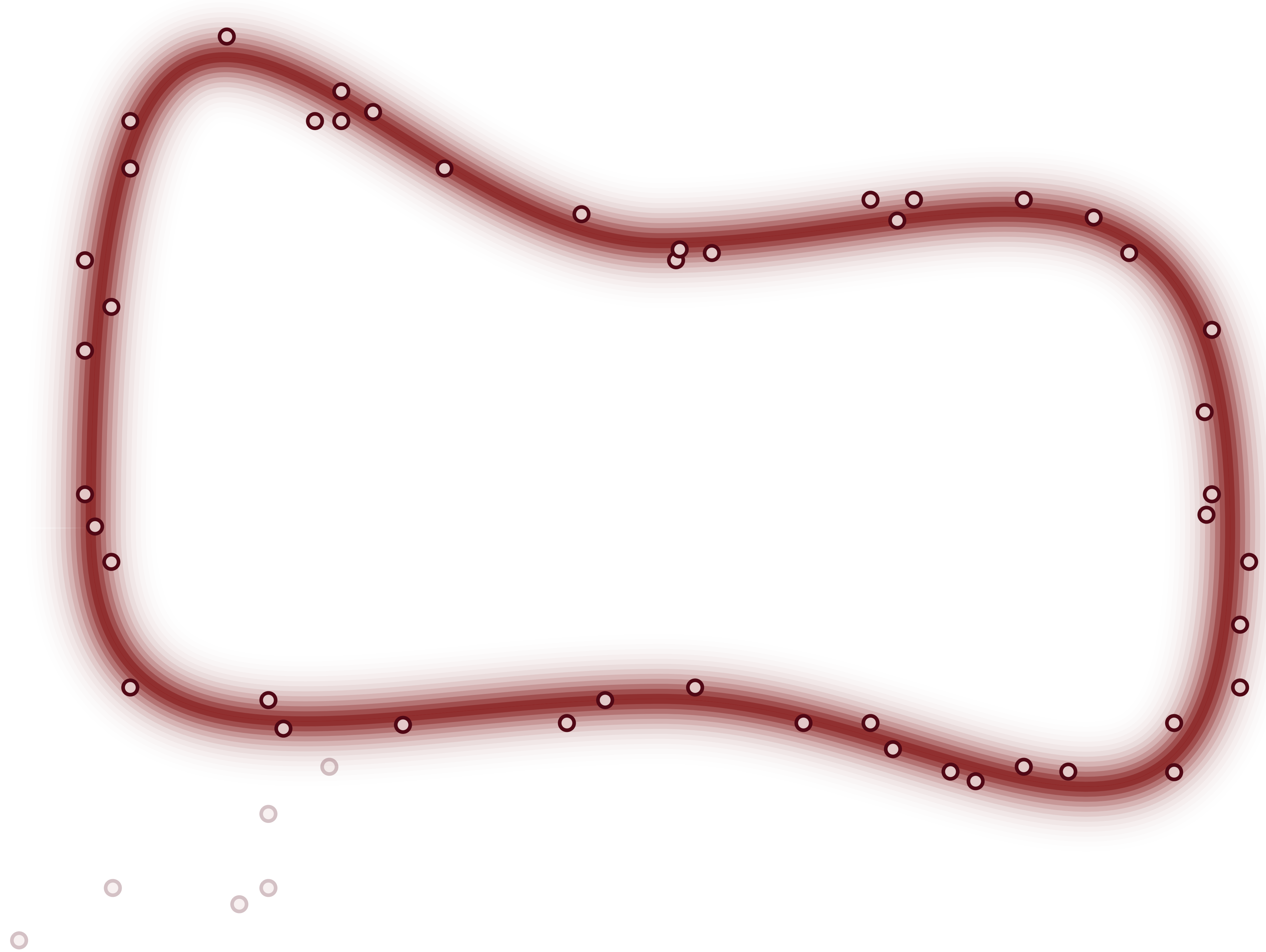


Markov chains provide a generic scheme for finding and then exploring the resulting typical set.

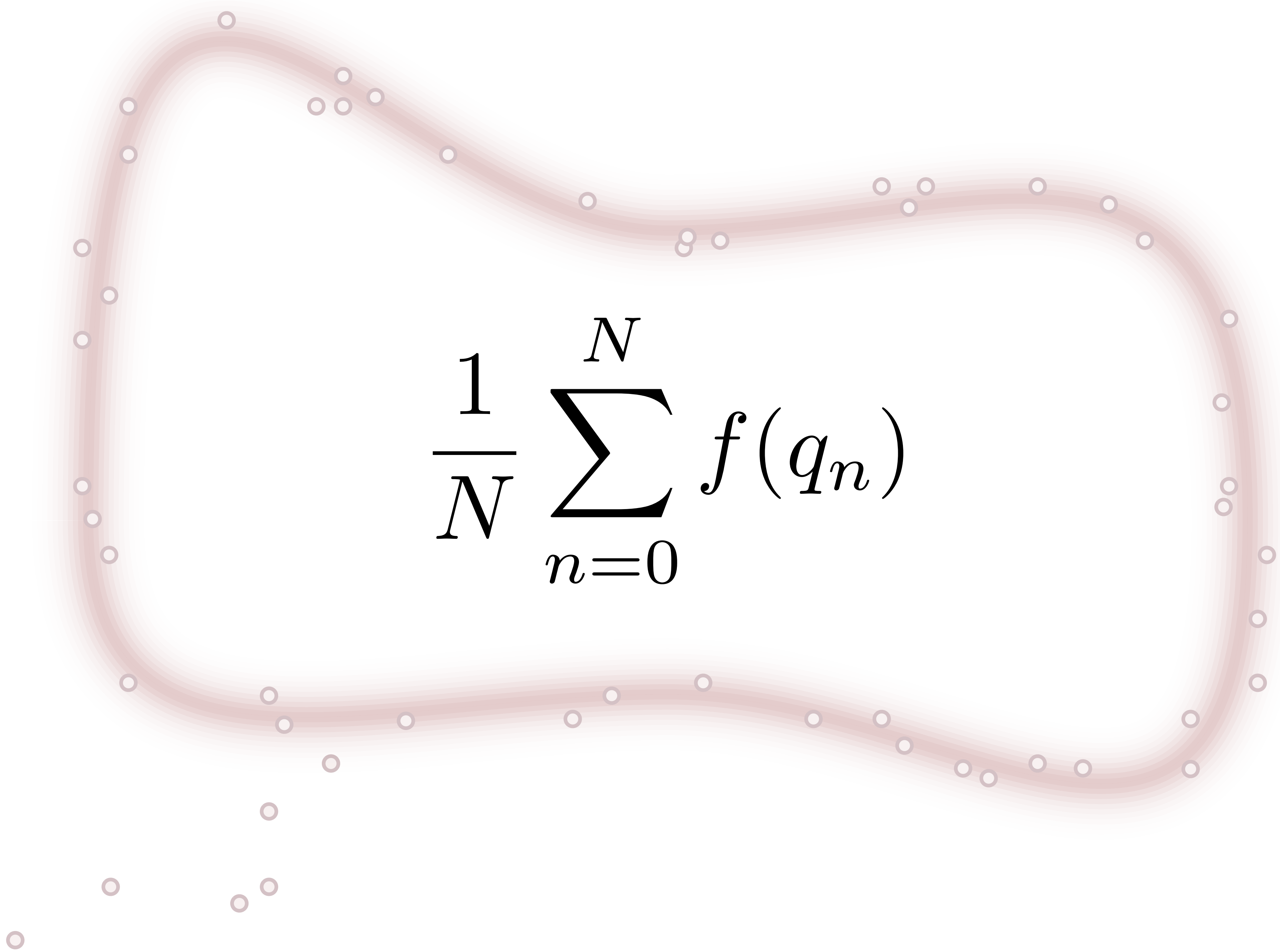
Markov chains provide a generic scheme for finding and then exploring the resulting typical set.



In particular, Markov chain define consistent *Markov Chain Monte Carlo estimators* of the desired expectations.

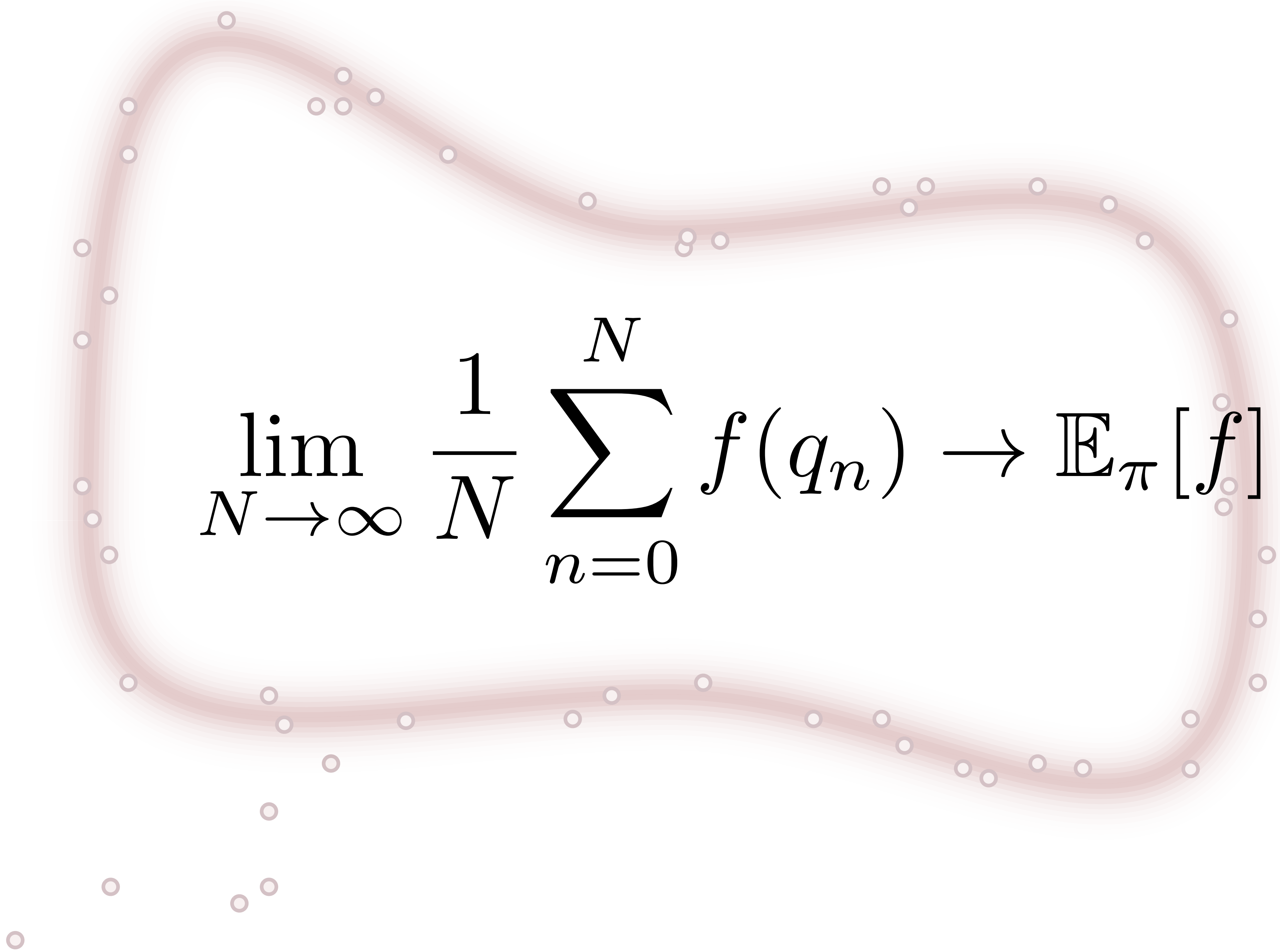


In particular, Markov chain define consistent *Markov Chain Monte Carlo estimators* of the desired expectations.



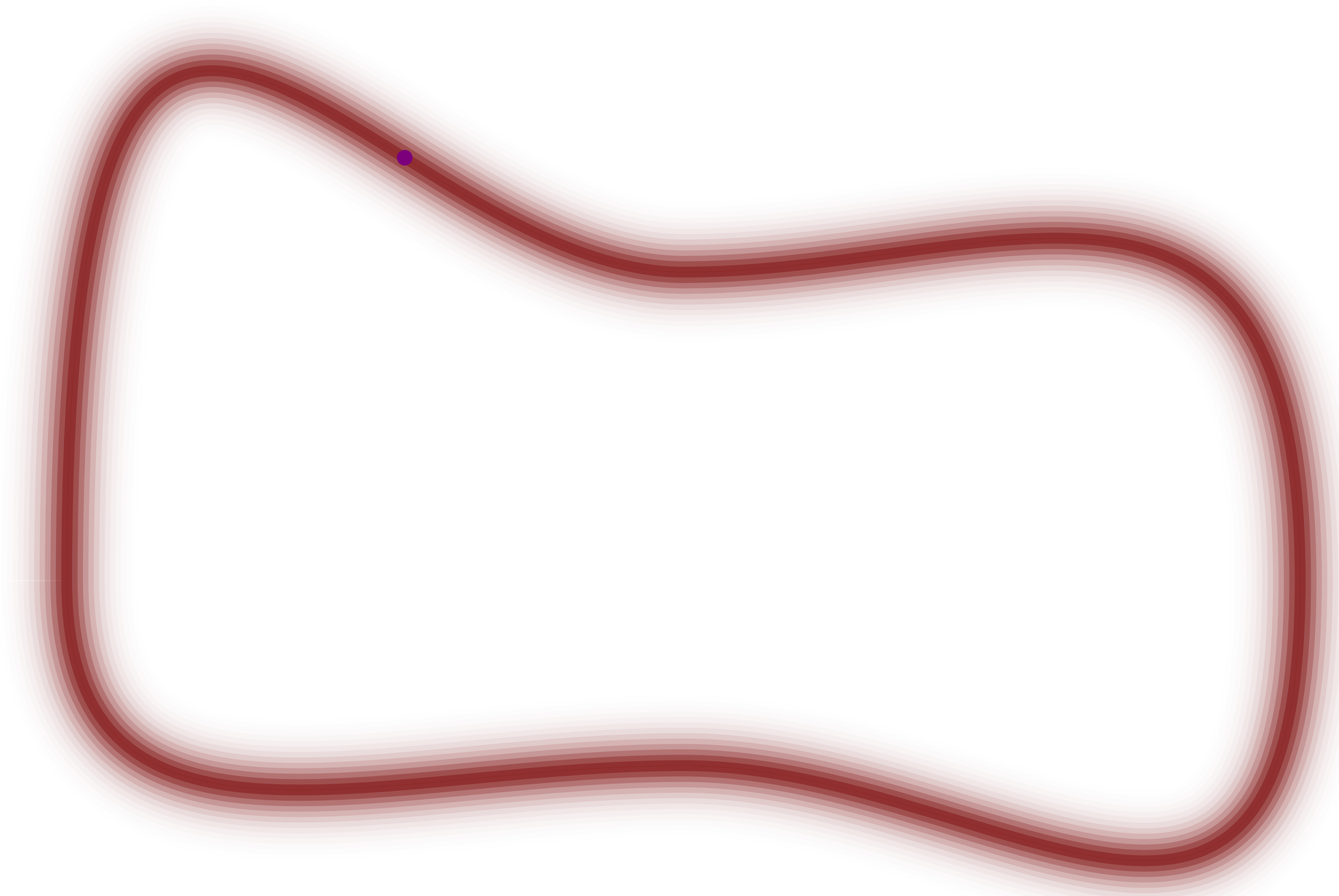
$$\frac{1}{N}\sum_{n=0}^{N} f(q_n)$$

In particular, Markov chain define consistent *Markov Chain Monte Carlo estimators* of the desired expectations.

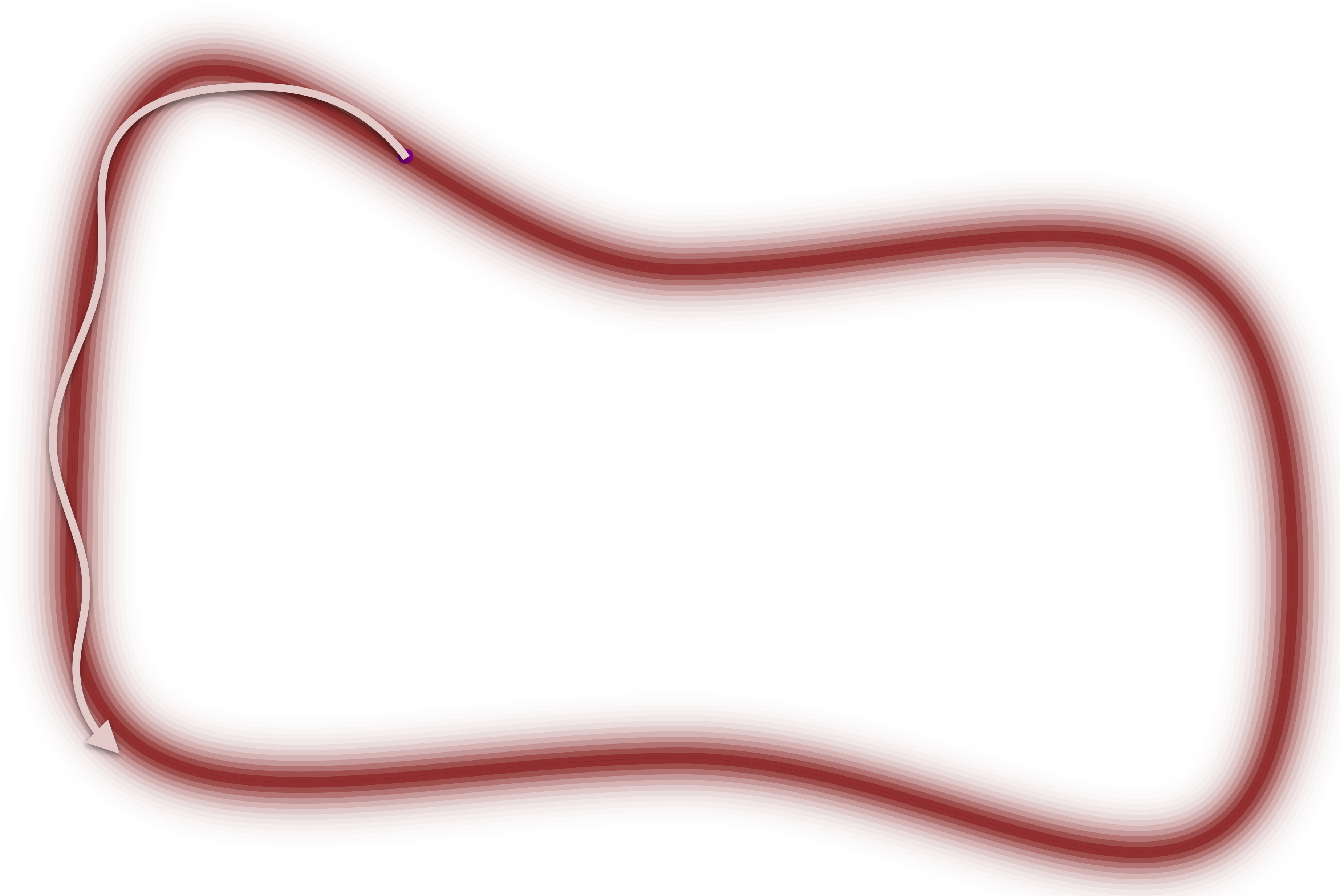


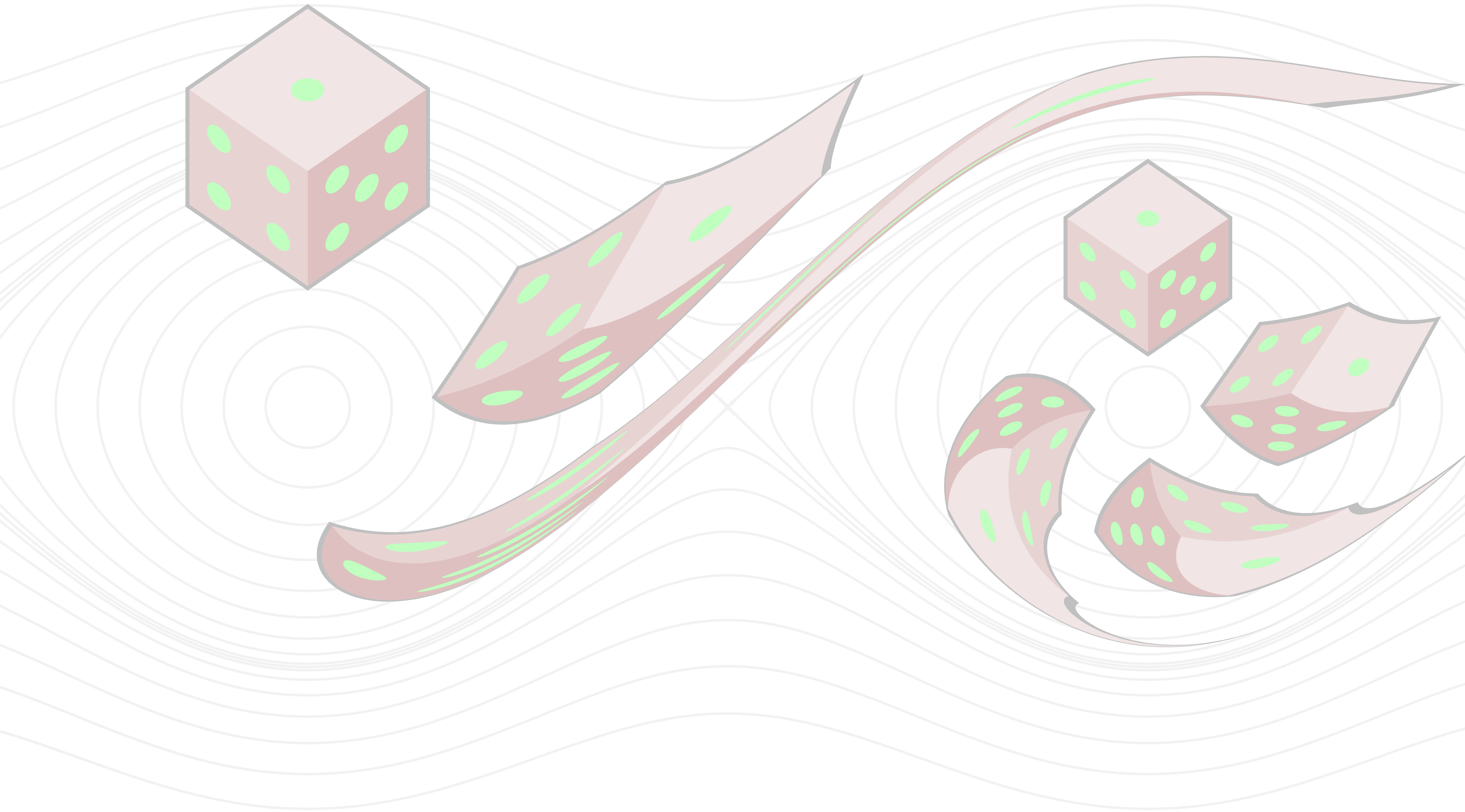
$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N} f(q_n) \to \mathbb{E}_{\pi}[f]$$

In order to scale to high-dimensional target distributions, however, we need *efficient* exploration of the typical set.



In order to scale to high-dimensional target distributions, however, we need *efficient* exploration of the typical set.





Hamiltonian Monte Carlo

Hamiltonian Monte Carlo uses auxiliary momenta and density gradients to generate coherent exploration.

$$q \to (q, p)$$

Hamiltonian Monte Carlo uses auxiliary momenta and density gradients to generate coherent exploration.

$$q \rightarrow (q, p)$$

$$\pi(q) \rightarrow \pi(q, p)$$

Hamiltonian Monte Carlo uses auxiliary momenta and density gradients to generate coherent exploration.

$$q \to (q, p)$$

$$\pi(q) \to \pi(q, p) = \pi(p \mid q) \, \pi(q)$$

The addition of momenta defines a *Hamiltonian* that decomposes into a *potential energy* and *kinetic energy*.

$$H(p, q) = -\log \pi(p|q)\,\pi(q)$$

The addition of momenta defines a *Hamiltonian* that decomposes into a *potential energy* and *kinetic energy*.

$$H(p, q) = -\log \pi(p|q)\, \pi(q)$$

$$= -\log \pi(p|q) - \log \pi(q)$$

The addition of momenta defines a *Hamiltonian* that decomposes into a *potential energy* and *kinetic energy*.

$$H(p, q) = -\log \pi(p|q)\, \pi(q)$$

$$= -\log \pi(p|q) - \log \pi(q)$$

$$T$$

The addition of momenta defines a *Hamiltonian* that decomposes into a *potential energy* and *kinetic energy*.

$$H(p, q) = -\log \pi(p|q)\, \pi(q)$$

$$= -\log \pi(p|q) - \log \pi(q)$$

$$\qquad\qquad T \qquad\qquad\qquad V$$

The Hamiltonian defines a vector field aligned with the typical set from which we can generate exploration.

$$\frac{\mathrm{d}q}{\mathrm{d}t} = \frac{\partial T}{\partial p}$$

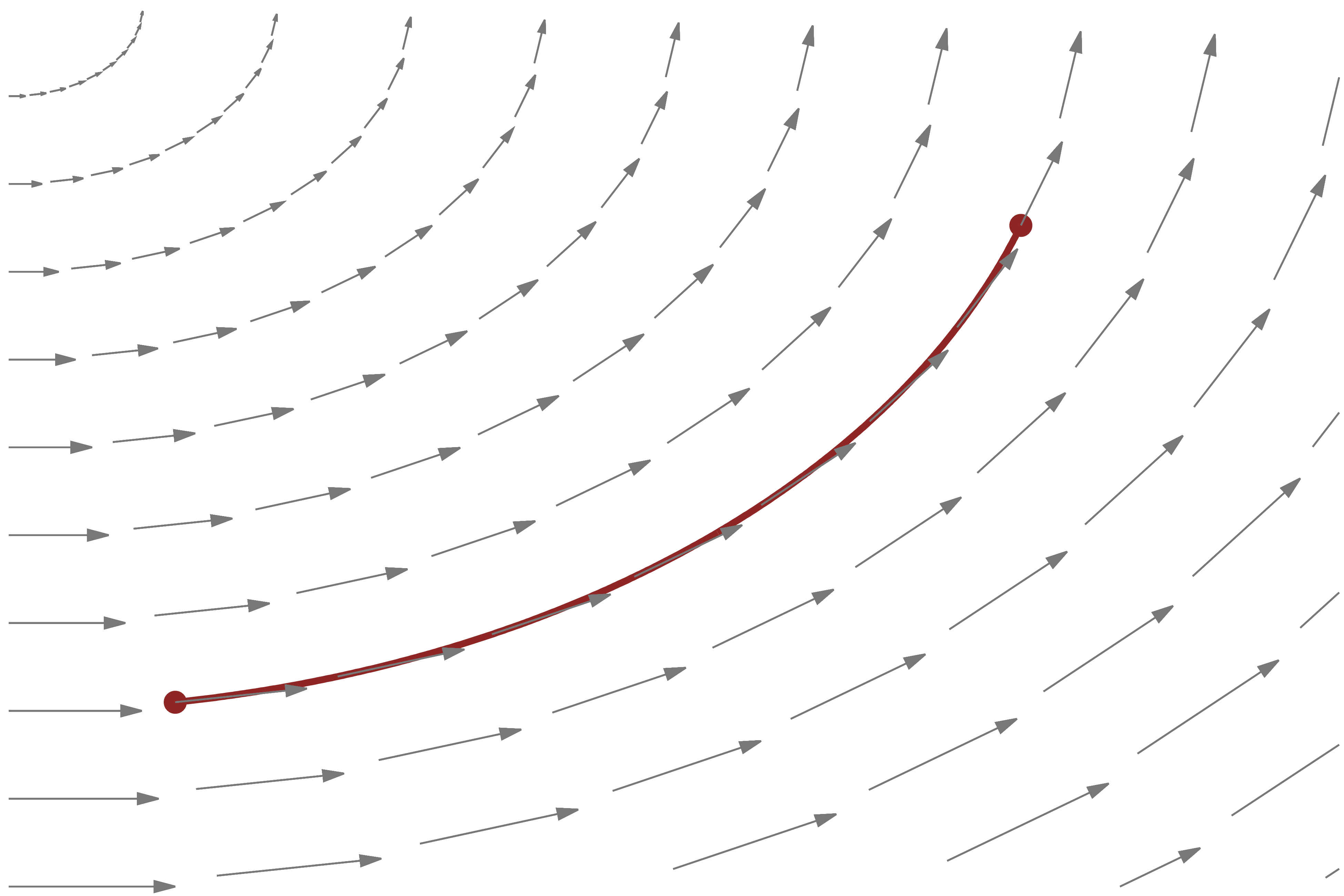
$$\frac{\mathrm{d}p}{\mathrm{d}t} = -\frac{\partial T}{\partial q} - \frac{\partial V}{\partial q}$$

The Hamiltonian defines a vector field aligned with the typical set from which we can generate exploration.

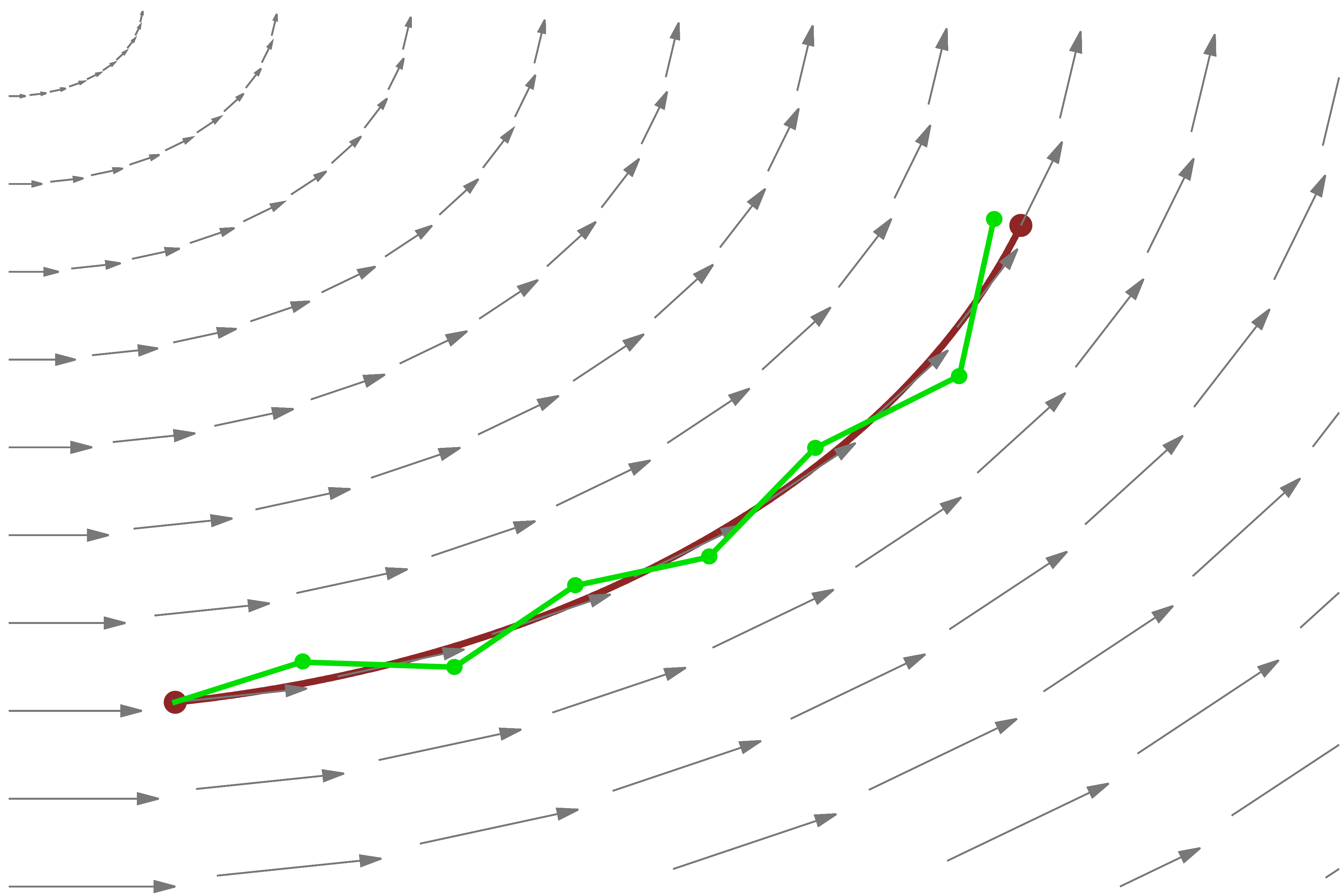
$$\frac{\mathrm{d}q}{\mathrm{d}t} = \frac{\partial T}{\partial p}$$

$$\frac{\mathrm{d}p}{\mathrm{d}t} = -\frac{\partial T}{\partial q} - \frac{\partial V}{\partial q}$$

In practice we integrate along this vector field only approximately, using powerful *symplectic integrators*.



In practice we integrate along this vector field only approximately, using powerful *symplectic integrators*.



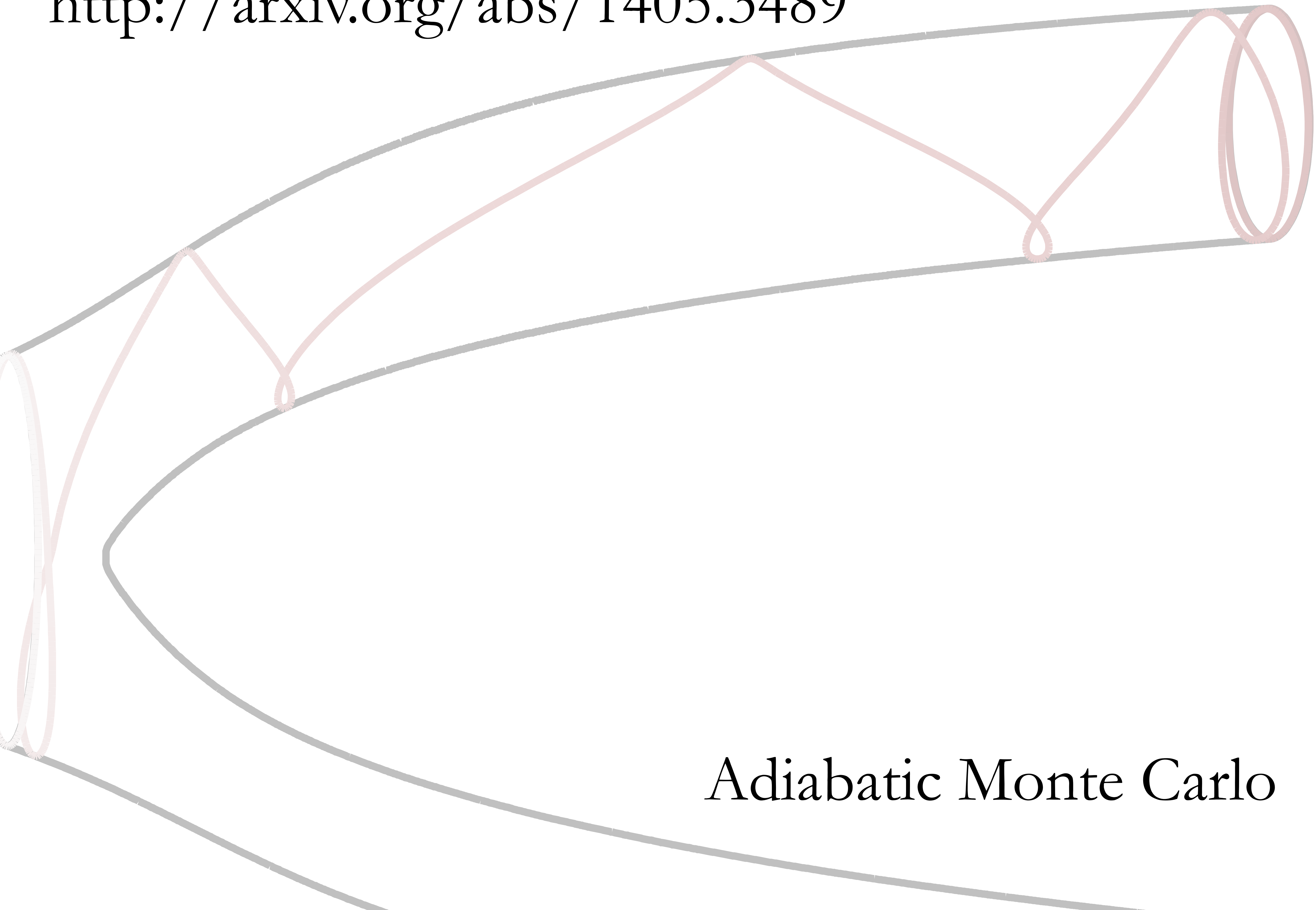
The numerical error introduced by the integrator can be eliminated with a careful Metropolis correction.

$$q \to q + \epsilon \frac{\partial T}{\partial p}$$

$$p \to p - \epsilon \left(\frac{\partial T}{\partial q} + \frac{\partial V}{\partial q} \right)$$

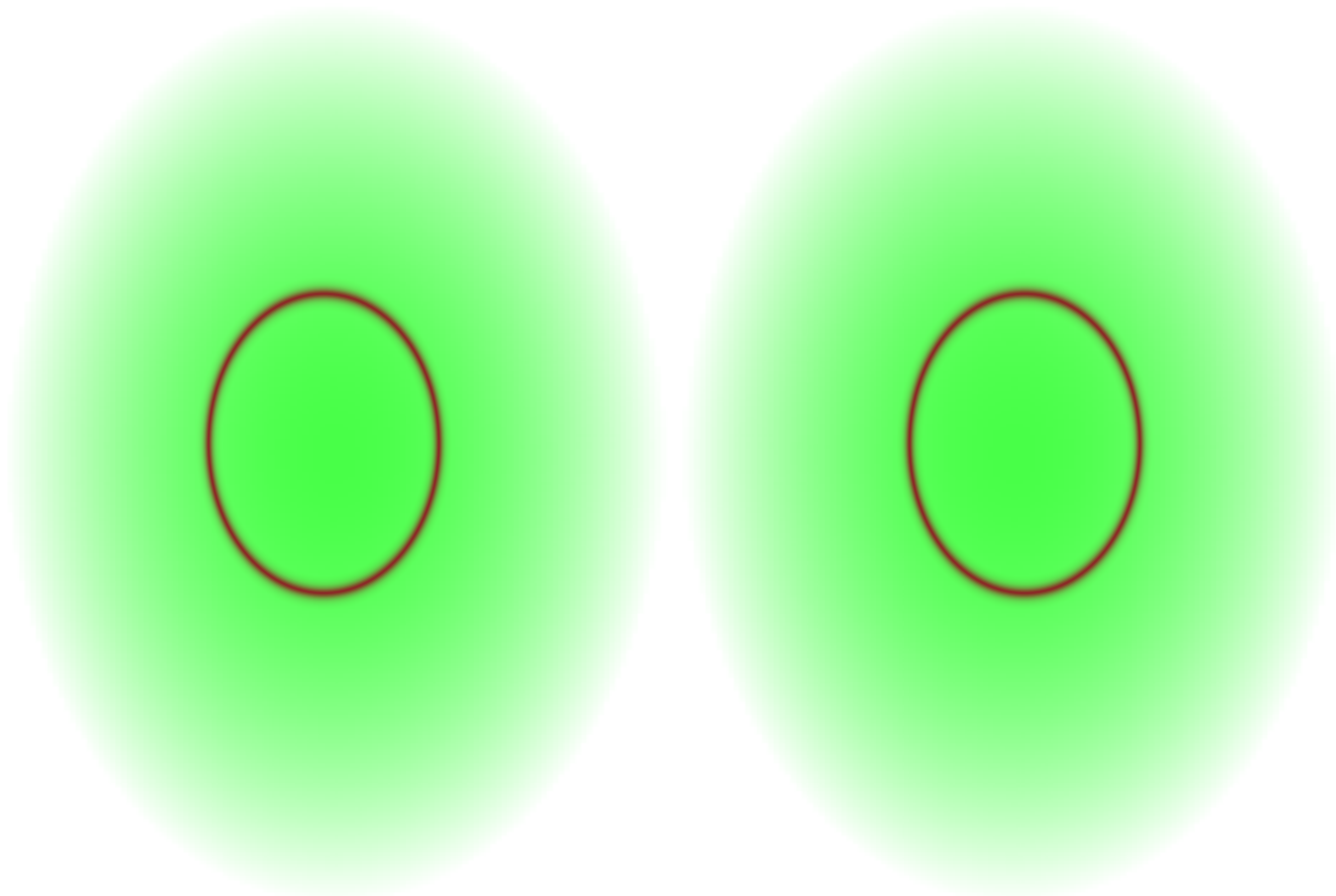
$$\pi(\text{accept}) = \min \left(1, \frac{\pi(\Phi_{\tau}(p, q))}{\pi(p, q)} \right)$$

http://arxiv.org/abs/1405.3489

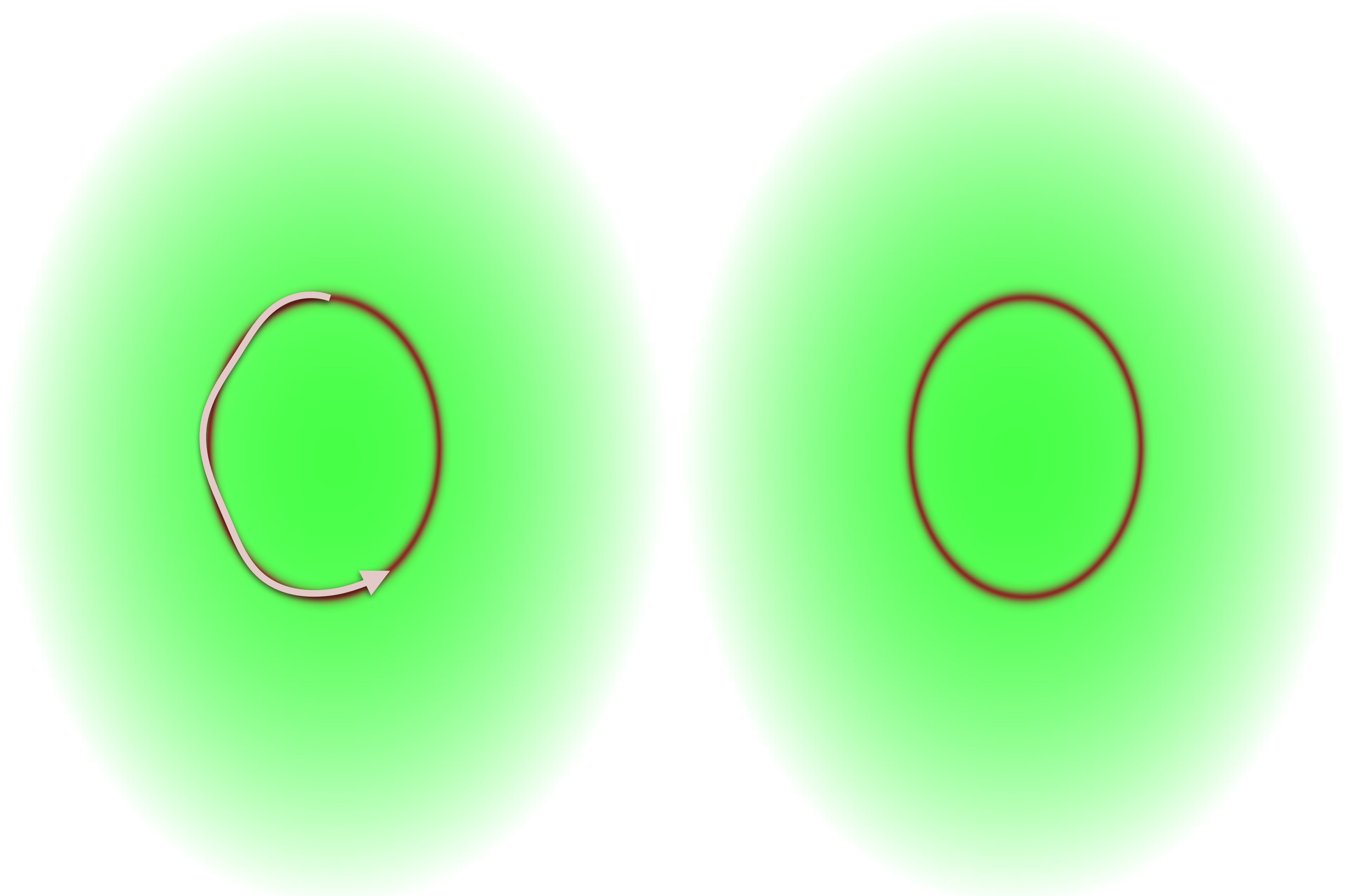


Adiabatic Monte Carlo

Like any MCMC algorithm, Hamiltonian Monte Carlo struggles to explore multimodal target distributions.



Like any MCMC algorithm, Hamiltonian Monte Carlo struggles to explore multimodal target distributions.



In Bayesian settings the marginal likelihood is difficult to estimate from the Markov chain output.

$$\pi(\mathcal{D}) = \int \pi(\mathcal{D} \mid q)\, \pi(q)\, \mathrm{d}q$$

In Bayesian settings the marginal likelihood is difficult to estimate from the Markov chain output.

$$\pi(\mathcal{D}) = \int \pi(\mathcal{D} \mid q) \, \pi(q) \, \mathrm{d}q$$

$$\pi(\mathcal{D}) = \mathbb{E}_{\mathrm{prior}}[\pi(\mathcal{D} \mid q)]$$

In Bayesian settings the marginal likelihood is difficult to estimate from the Markov chain output.

$$\pi(\mathcal{D}) = \int \pi(\mathcal{D} \mid q) \, \pi(q) \, \mathrm{d}q$$

$$\pi(\mathcal{D}) = \mathbb{E}_{\mathrm{prior}}[\pi(\mathcal{D} \mid q)]$$

$$\pi(\mathcal{D}) = \mathbb{E}_{\mathrm{post}}\left[(\pi(\mathcal{D} \mid q))^{-1}\right]$$

Both of these problems are facilitated by interpolating between the target and an auxiliary, unimodal distribution.

$$\pi_{\beta}(q) = \frac{1}{Z(\beta)} \left(\Delta\pi(q) \right)^{\beta} \pi_{B}(q)$$

Both of these problems are facilitated by interpolating between the target and an auxiliary, unimodal distribution.

$$\pi_{\beta}(q) = \frac{1}{Z(\beta)} \left(\Delta\pi(q) \right)^{\beta} \pi_{B}(q)$$

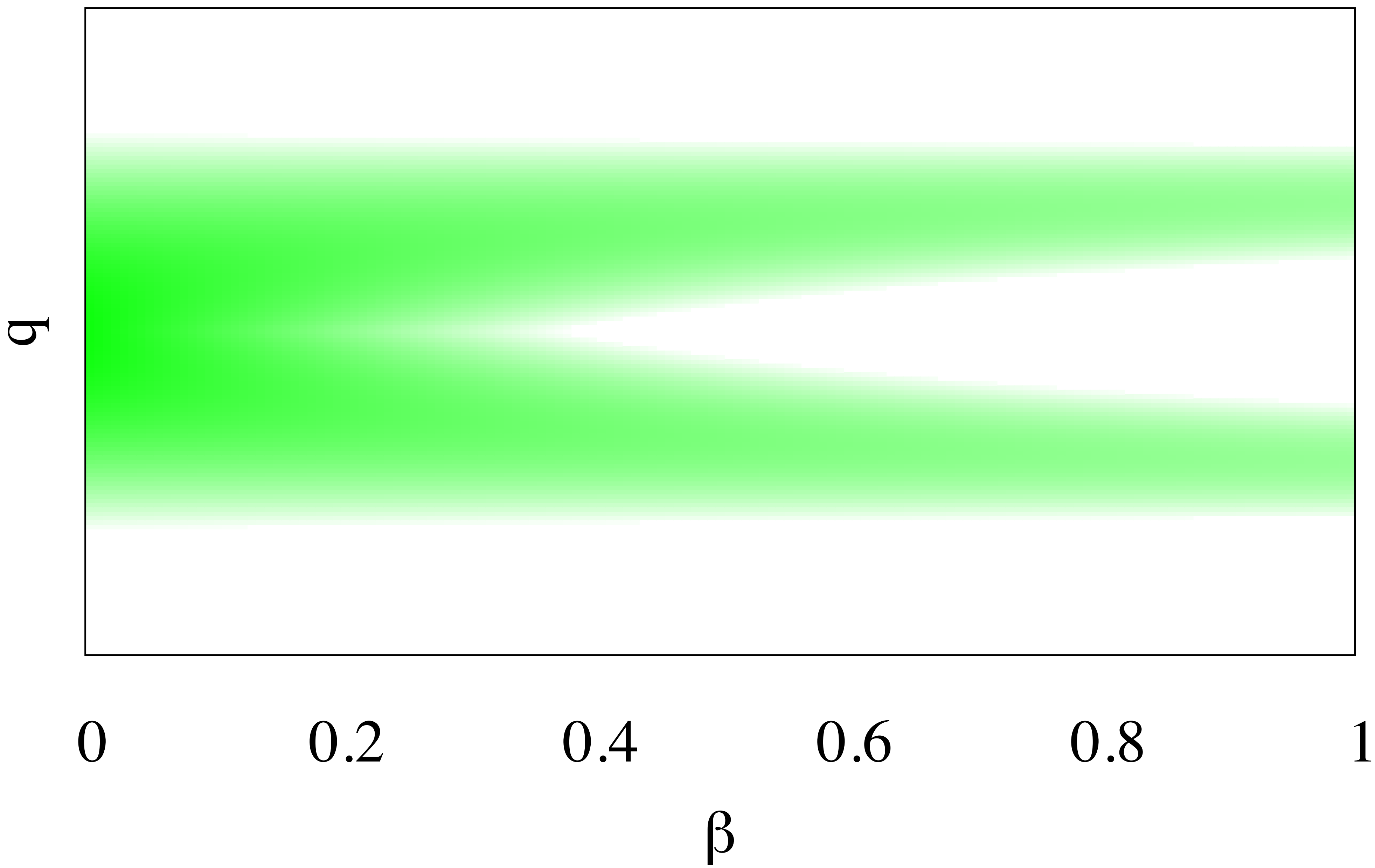
Both of these problems are facilitated by interpolating between the target and an auxiliary, unimodal distribution.

$$\pi_\beta(q) = \frac{1}{Z(\beta)} \left(\Delta\pi(q)\right)^\beta \pi_B(q)$$

Both of these problems are facilitated by interpolating between the target and an auxiliary, unimodal distribution.

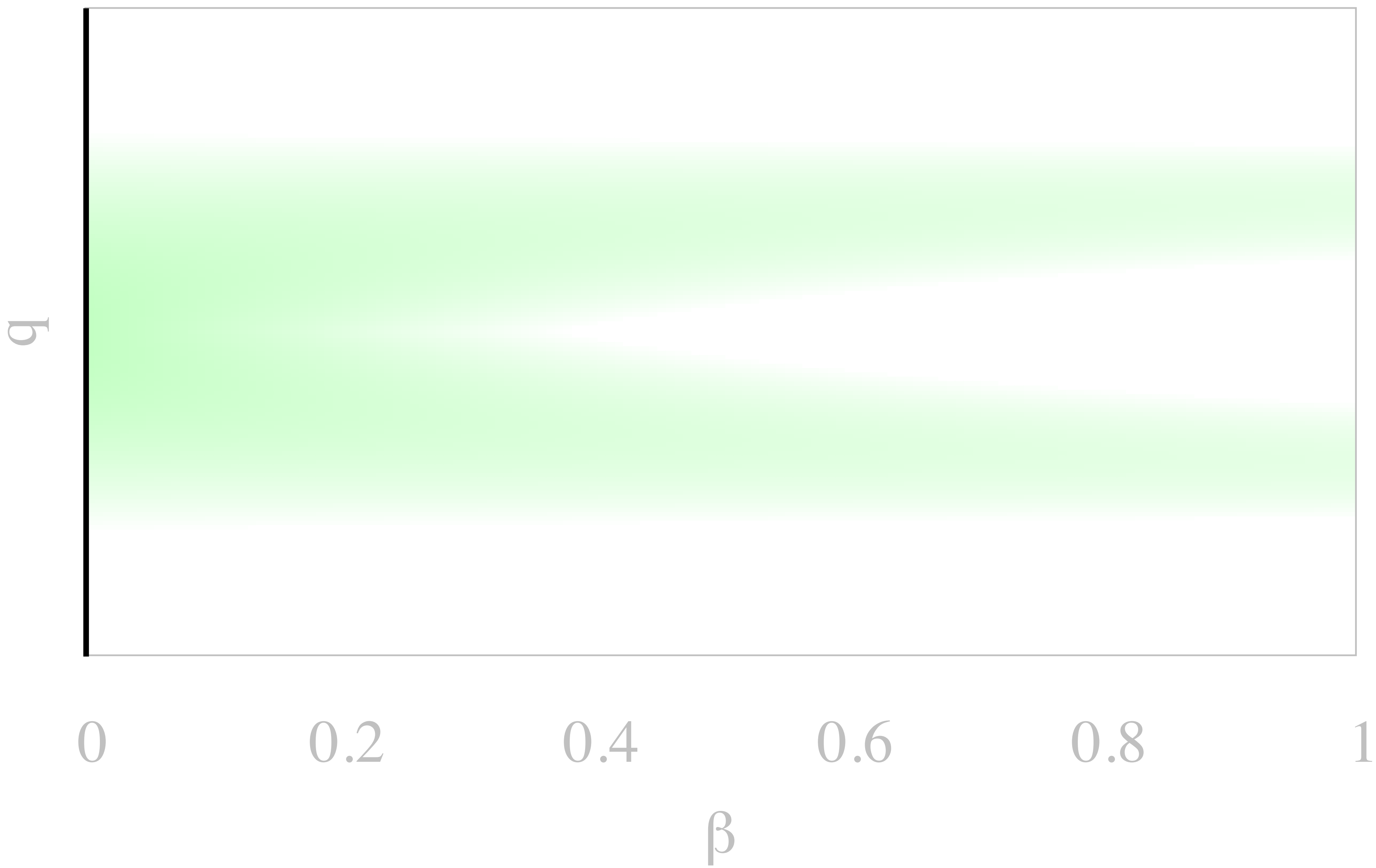
$$\pi_{\beta}(q) = \frac{1}{Z(\beta)} \left(\Delta\pi(q) \right)^{\beta} \pi_{B}(q)$$

Both of these problems are facilitated by interpolating between the target and an auxiliary, unimodal distribution.

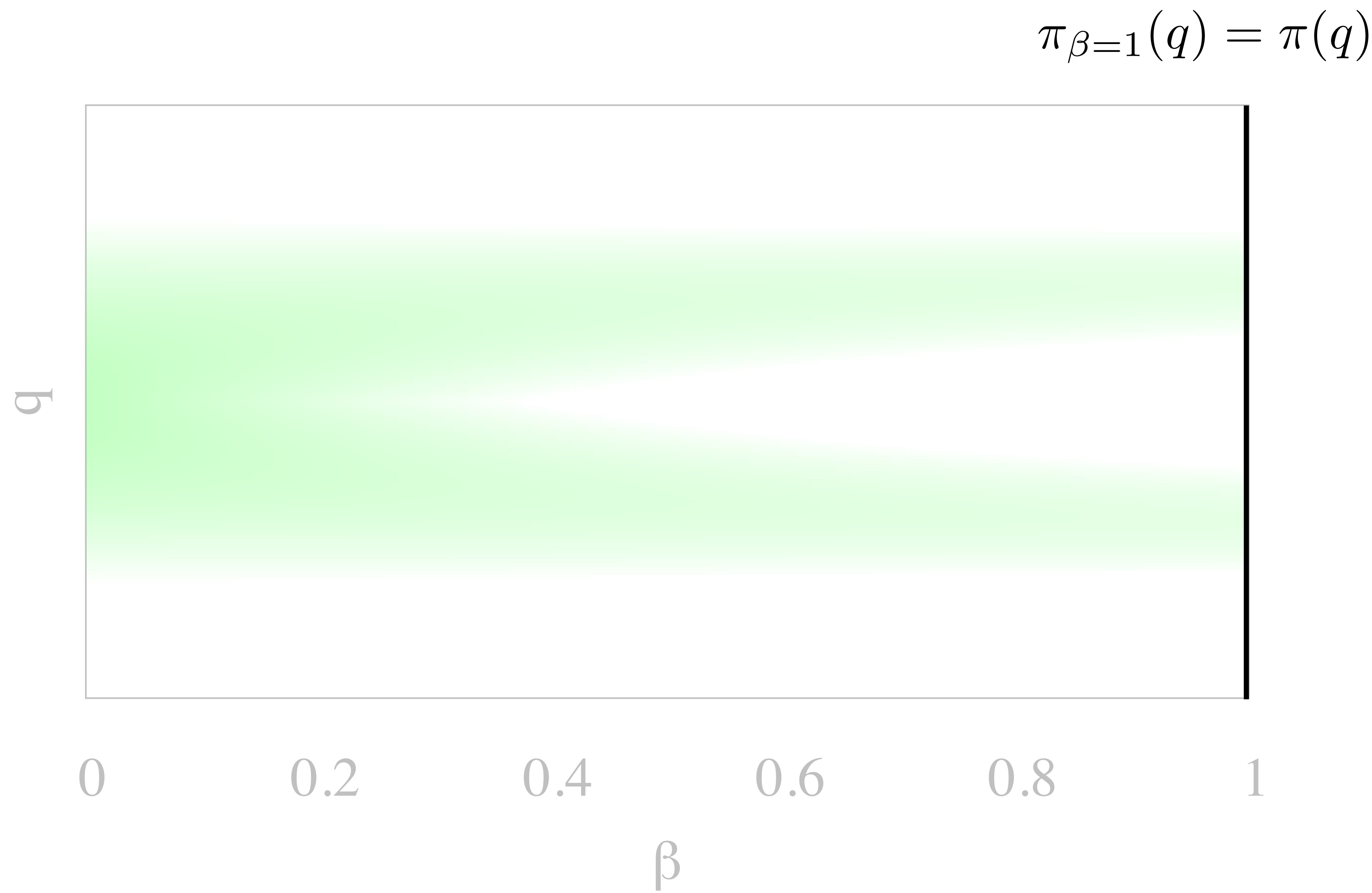


Both of these problems are facilitated by interpolating between the target and an auxiliary, unimodal distribution.

$$\pi_{\beta=0}(q) = \pi_B(q)$$

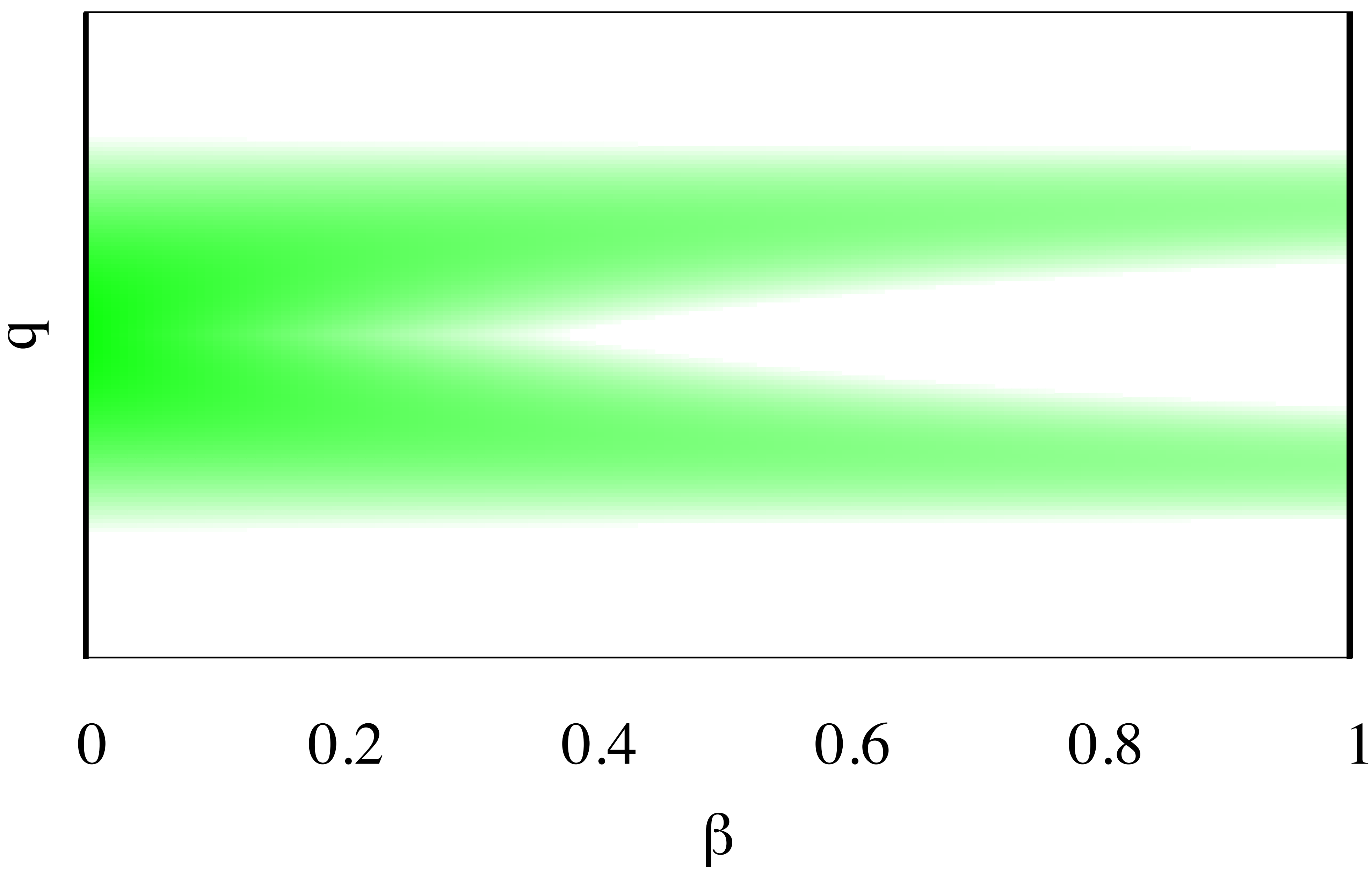


Both of these problems are facilitated by interpolating between the target and an auxiliary, unimodal distribution.

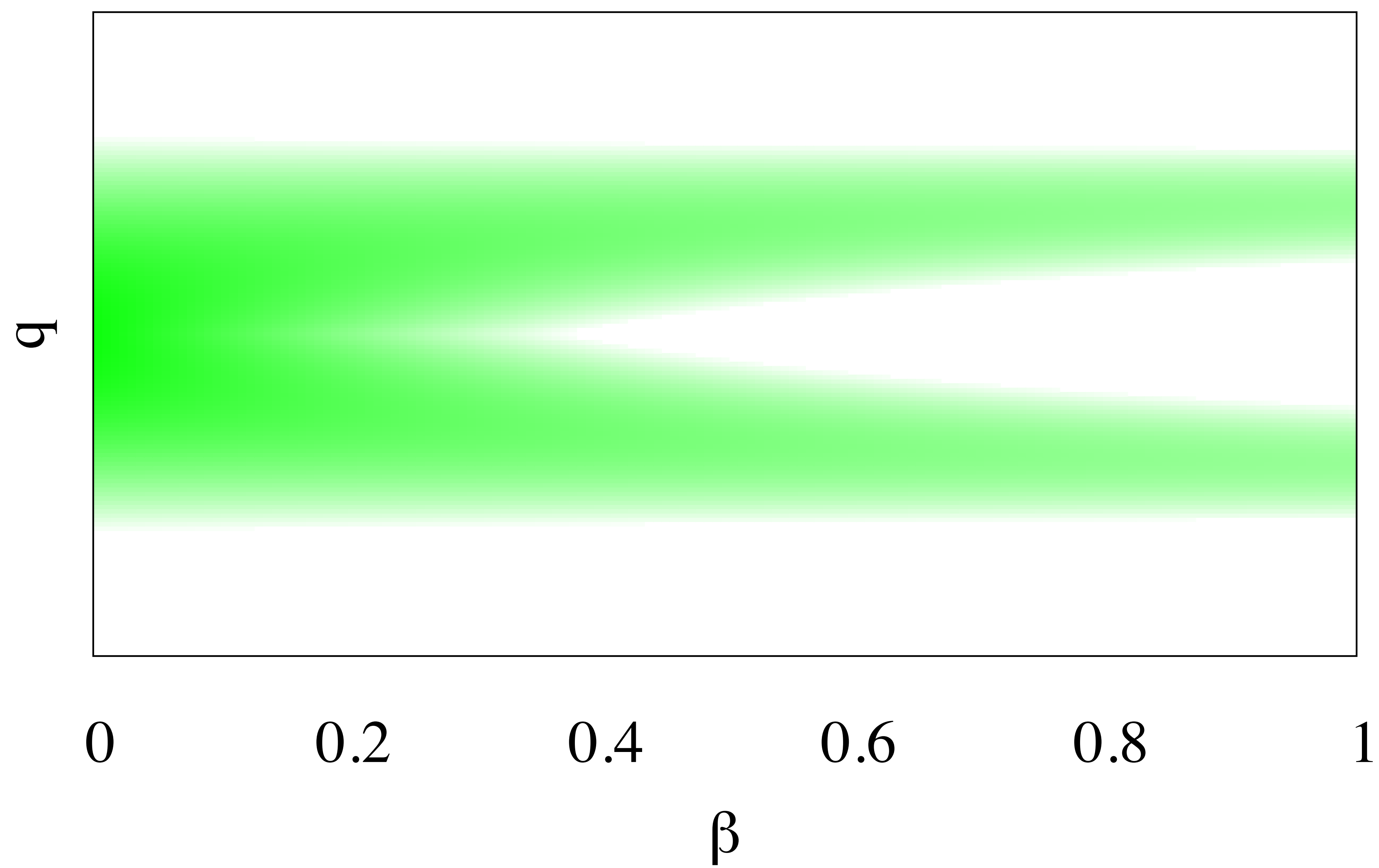


Both of these problems are facilitated by interpolating between the target and an auxiliary, unimodal distribution.

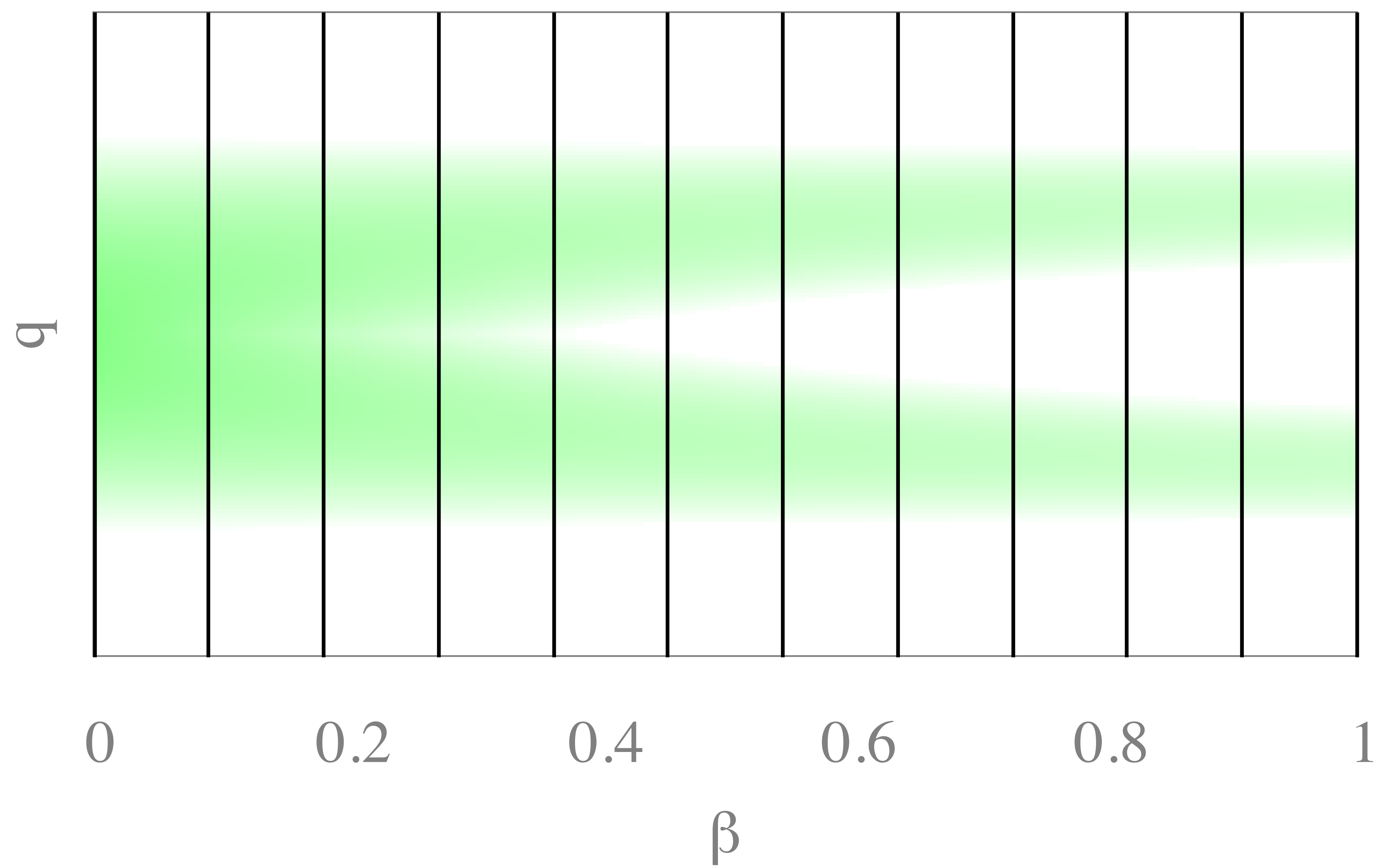
$\pi_{\beta=0}(q) = \pi_B(q)$ $\pi_{\beta=1}(q) = \pi(q)$



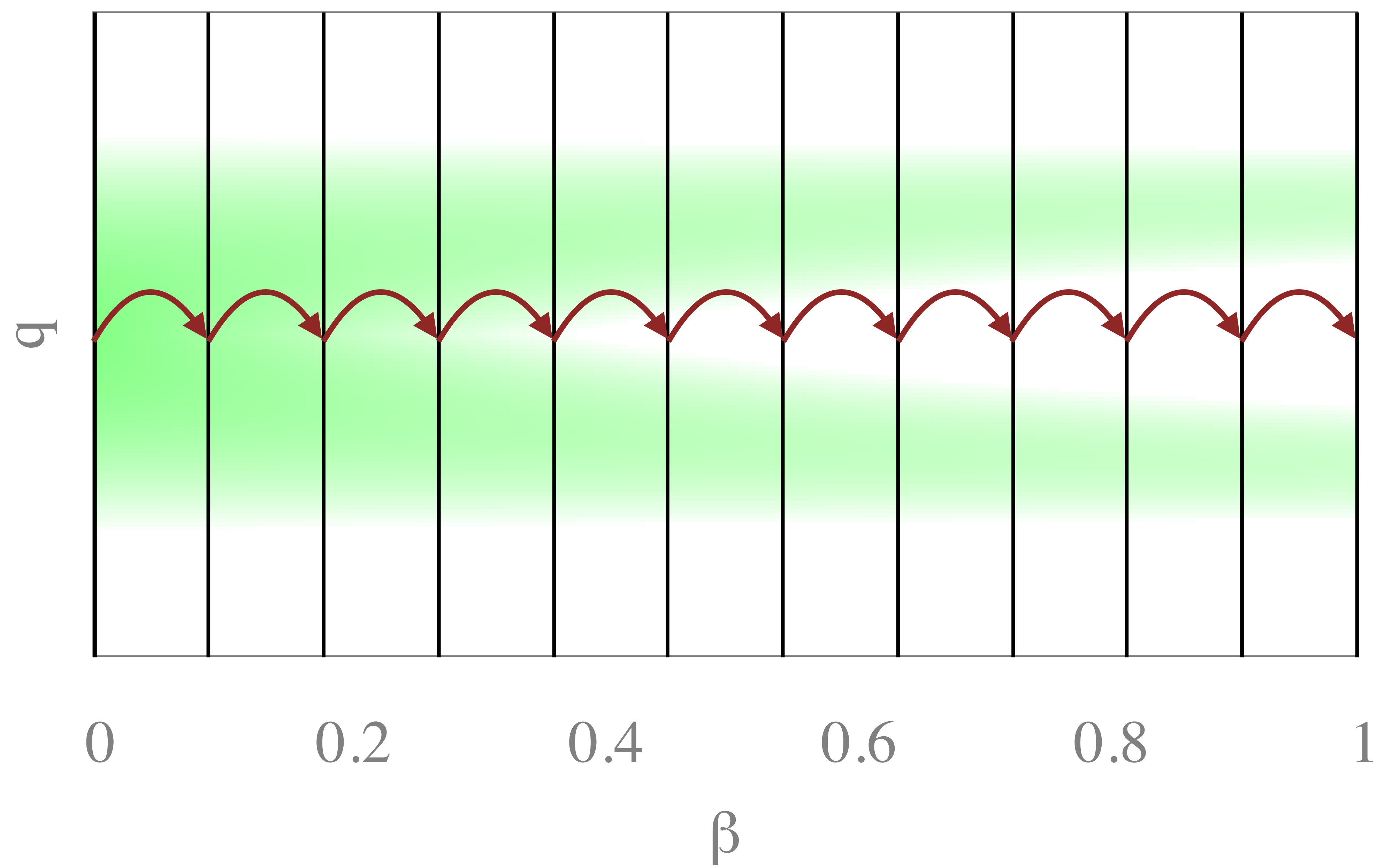
To move along the interpolation in practice, however, we need to impose a discrete partition of the interpolation.



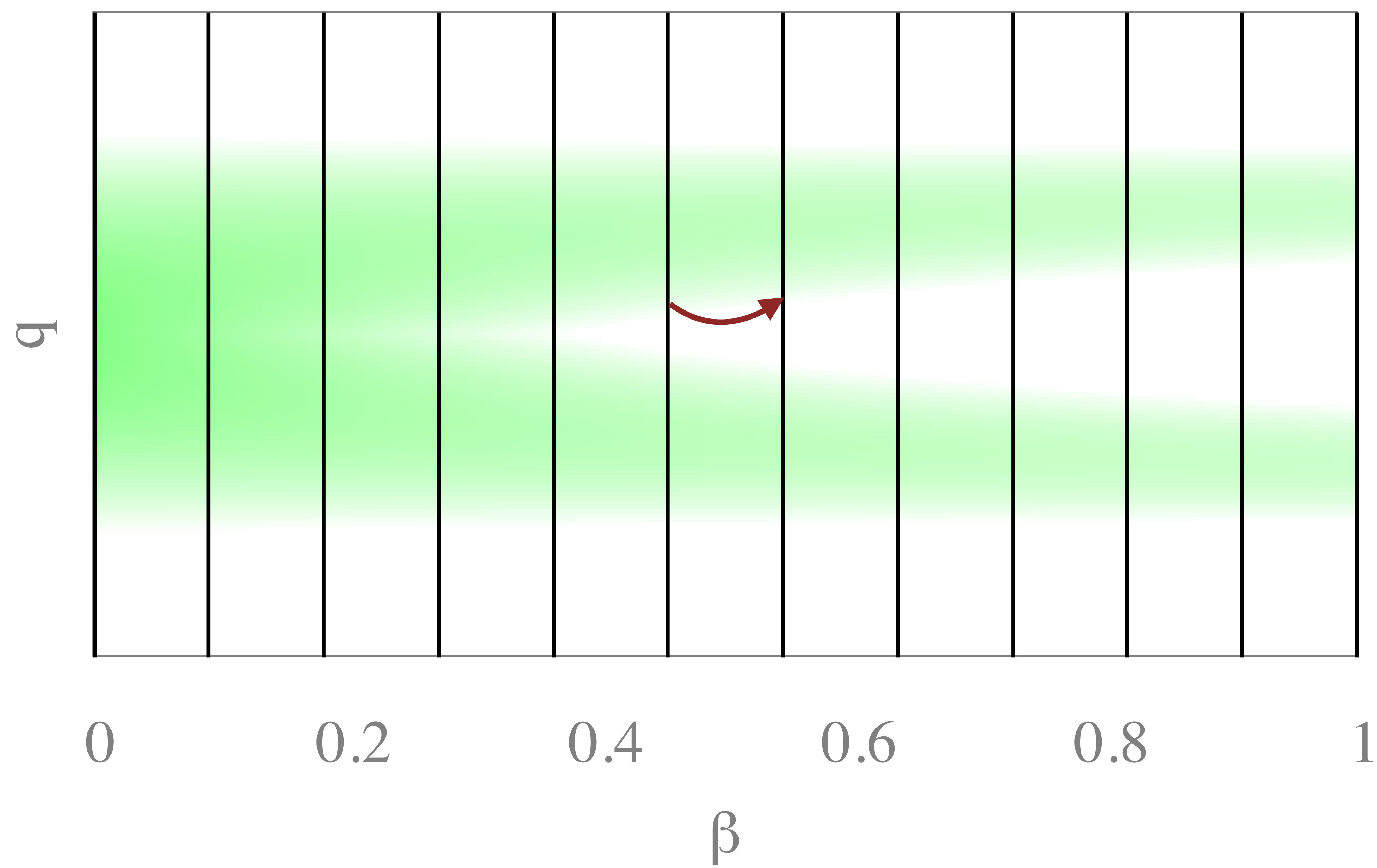
To move along the interpolation in practice, however, we need to impose a discrete partition of the interpolation.



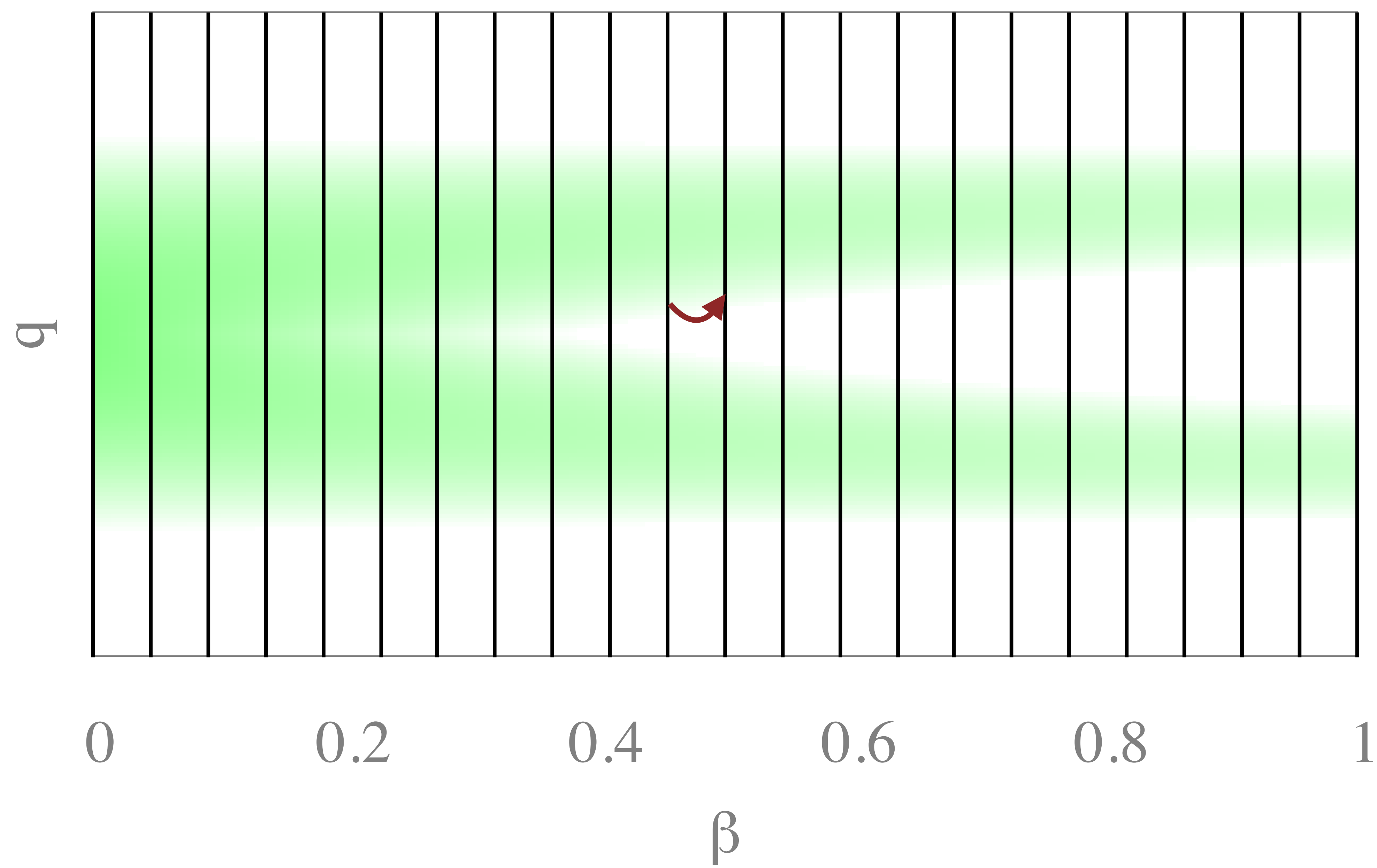
To move along the interpolation in practice, however, we need to impose a discrete partition of the interpolation.



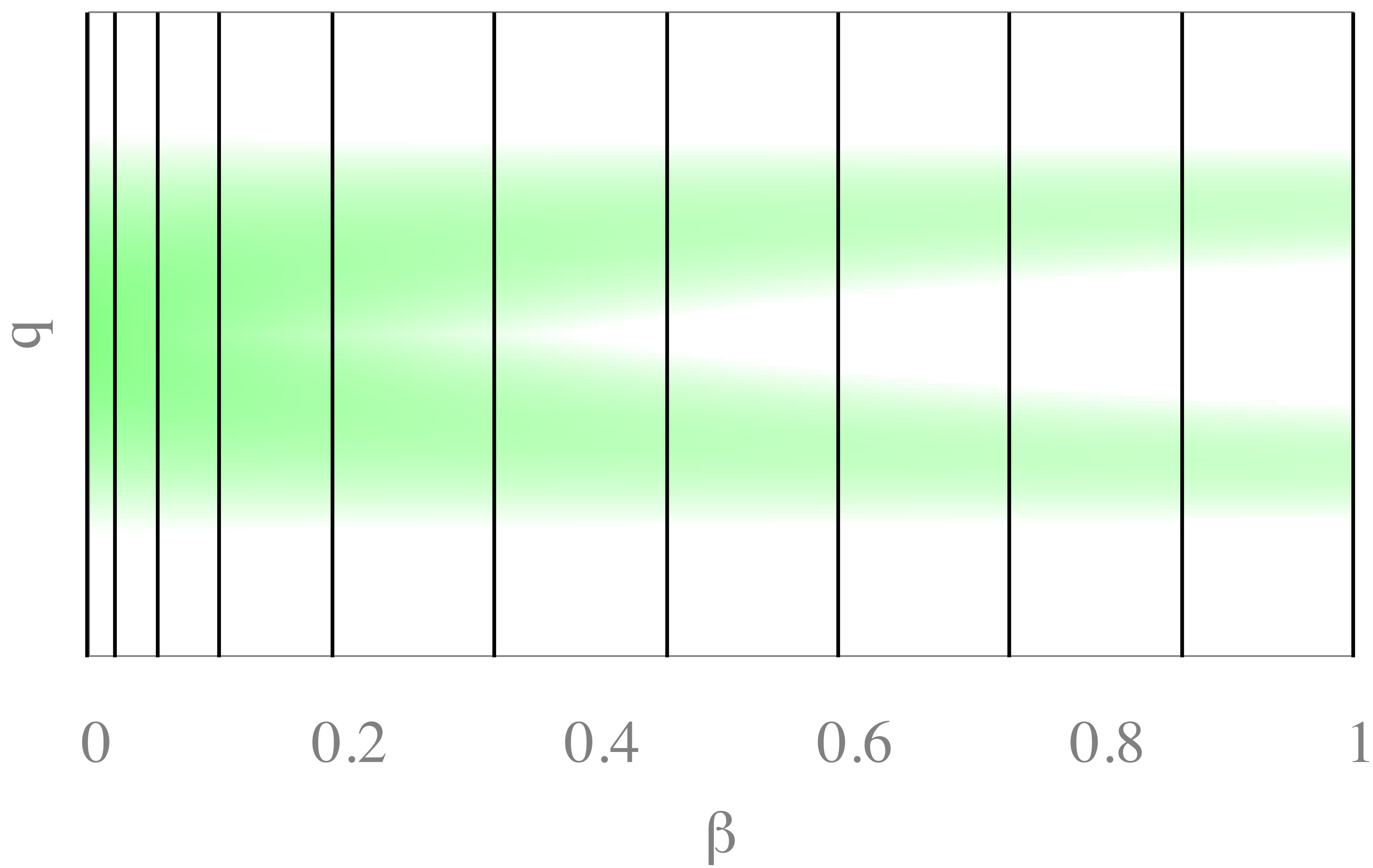
If the partition is chosen poorly then transitions will be difficult, and we are left with a delicate tuning problem.



If the partition is chosen poorly then transitions will be difficult, and we are left with a delicate tuning problem.



If the partition is chosen poorly then transitions will be difficult, and we are left with a delicate tuning problem.



Mirroring Hamiltonian Monte Carlo, *Adiabatic Monte Carlo* generates optimal transitions by making the perk dynamic.

$$q \to (q, p, \beta)$$

Mirroring Hamiltonian Monte Carlo, *Adiabatic Monte Carlo* generates optimal transitions by making the perk dynamic.

$$q \to (q, p, \beta)$$

$$\pi(q) \to \pi_{\beta}(q, p) = \pi(p \mid q) \, \pi_{\beta}(q)$$

Introducing auxiliary momenta to the interpolating distributions defines a *contact Hamiltonian*.

$$H(q, p, \beta) = -\log \pi(p \mid q)\, \pi_{\beta}(q)$$

Introducing auxiliary momenta to the interpolating distributions defines a *contact Hamiltonian*.

$$H(q, p, \beta) = -\log \pi(p \mid q) \, \pi_{\beta}(q)$$

$$= -\log \pi(p \mid q) - \log \pi_{B}(q)$$

$$-\beta \log \Delta\pi(q) + \log Z(\beta) + H_0$$

Introducing auxiliary momenta to the interpolating distributions defines a *contact Hamiltonian*.

$$H(q, p, \beta) = -\log \pi(p \mid q) \, \pi_\beta(q)$$

$$= -\log \pi(p \mid q) - \log \pi_B(q)$$

$$T$$

$$-\beta \log \Delta\pi(q) + \log Z(\beta) + H_0$$

Introducing auxiliary momenta to the interpolating distributions defines a *contact Hamiltonian*.

$$H(q, p, \beta) = -\log \pi(p \mid q)\, \pi_{\beta}(q)$$

$$= -\log \pi(p \mid q) - \log \pi_B(q)$$

$$V_B$$

$$-\beta \log \Delta\pi(q) + \log Z(\beta) + H_0$$

Introducing auxiliary momenta to the interpolating distributions defines a *contact Hamiltonian*.

$$H(q, p, \beta) = -\log \pi(p \mid q) \, \pi_{\beta}(q)$$

$$= -\log \pi(p \mid q) - \log \pi_{B}(q)$$

$$-\beta \log \Delta\pi(q) + \log Z(\beta) + H_0$$

$$-\Delta V$$

Introducing auxiliary momenta to the interpolating distributions defines a *contact Hamiltonian*.

$$H(q, p, \beta) = -\log \pi(p \mid q)\, \pi_\beta(q)$$

$$= -\log \pi(p \mid q) - \log \pi_B(q)$$

$$-\beta \log \Delta\pi(q) + \log Z(\beta) + H_0$$

The contact Hamiltonian defines a vector field that generates efficient motion along the interpolation.

$$\frac{\mathrm{d}q}{\mathrm{d}t} = \frac{\partial T}{\partial p}$$

$$\frac{\mathrm{d}p}{\mathrm{d}t} = -\frac{\partial T}{\partial q} - \frac{\partial V_\beta}{\partial q} + \left(\Delta V - \mathbb{E}_{\pi_\beta}[\Delta V]\right) p$$

$$\frac{\mathrm{d}\beta}{\mathrm{d}t} = -p \frac{\partial T}{\partial p}$$

The contact Hamiltonian defines a vector field that generates efficient motion along the interpolation.

$$\frac{\mathrm{d}q}{\mathrm{d}t} = \frac{\partial T}{\partial p}$$

$$\frac{\mathrm{d}p}{\mathrm{d}t} = -\frac{\partial T}{\partial q} - \frac{\partial V_\beta}{\partial q} + \left(\Delta V - \mathbb{E}_{\pi_\beta}[\Delta V]\right) p$$

$$\frac{\mathrm{d}\beta}{\mathrm{d}t} = -p \frac{\partial T}{\partial p}$$

The contact Hamiltonian defines a vector field that generates efficient motion along the interpolation.

$$\frac{\mathrm{d}q}{\mathrm{d}t} = \frac{\partial T}{\partial p}$$

$$\frac{\mathrm{d}p}{\mathrm{d}t} = -\frac{\partial T}{\partial q} - \frac{\partial V_\beta}{\partial q} + \left(\Delta V - \mathbb{E}_{\pi_\beta}[\Delta V]\right) p$$

$$\frac{\mathrm{d}\beta}{\mathrm{d}t} = -p \frac{\partial T}{\partial p}$$

The contact Hamiltonian defines a vector field that generates efficient motion along the interpolation.

$$\frac{\mathrm{d}q}{\mathrm{d}t} = \frac{\partial T}{\partial p}$$

$$\frac{\mathrm{d}p}{\mathrm{d}t} = -\frac{\partial T}{\partial q} - \frac{\partial V_\beta}{\partial q} + \left(\Delta V - \mathbb{E}_{\pi_\beta}[\Delta V]\right) p$$

$$\frac{\mathrm{d}\beta}{\mathrm{d}t} = -p \frac{\partial T}{\partial p}$$

Because the contact Hamiltonian is invariant to this motion, we can also recover the normalizing constant.

$$H(q, p, \beta) = -\log \pi(p \mid q) - \log \pi_B(q)$$

$$-\beta \log \Delta\pi(q) + \log Z(\beta) + H_0$$

Because the contact Hamiltonian is invariant to this motion, we can also recover the normalizing constant.

$$H(q, p, \beta) = -\log \pi(p \mid q) - \log \pi_B(q)$$

$$-\beta \log \Delta\pi(q) + \log Z(\beta) + H_0$$

Because the contact Hamiltonian is invariant to this motion, we can also recover the normalizing constant.

$$H(q, p, \beta) = -\log \pi(p \mid q) - \log \pi_B(q)$$

$$-\beta \log \Delta\pi(q) + \log Z(\beta) + H_0$$

$$\log Z(\beta) = \Delta H(q, p, \beta)$$

Adiabatic Monte Carlo dynamically transitions from the base distribution to the target distribution.

$\beta = 4.58862\text{e-}05$

Adiabatic Monte Carlo dynamically transitions from the base distribution to the target distribution.

$\beta = 4.58862\text{e-}05$

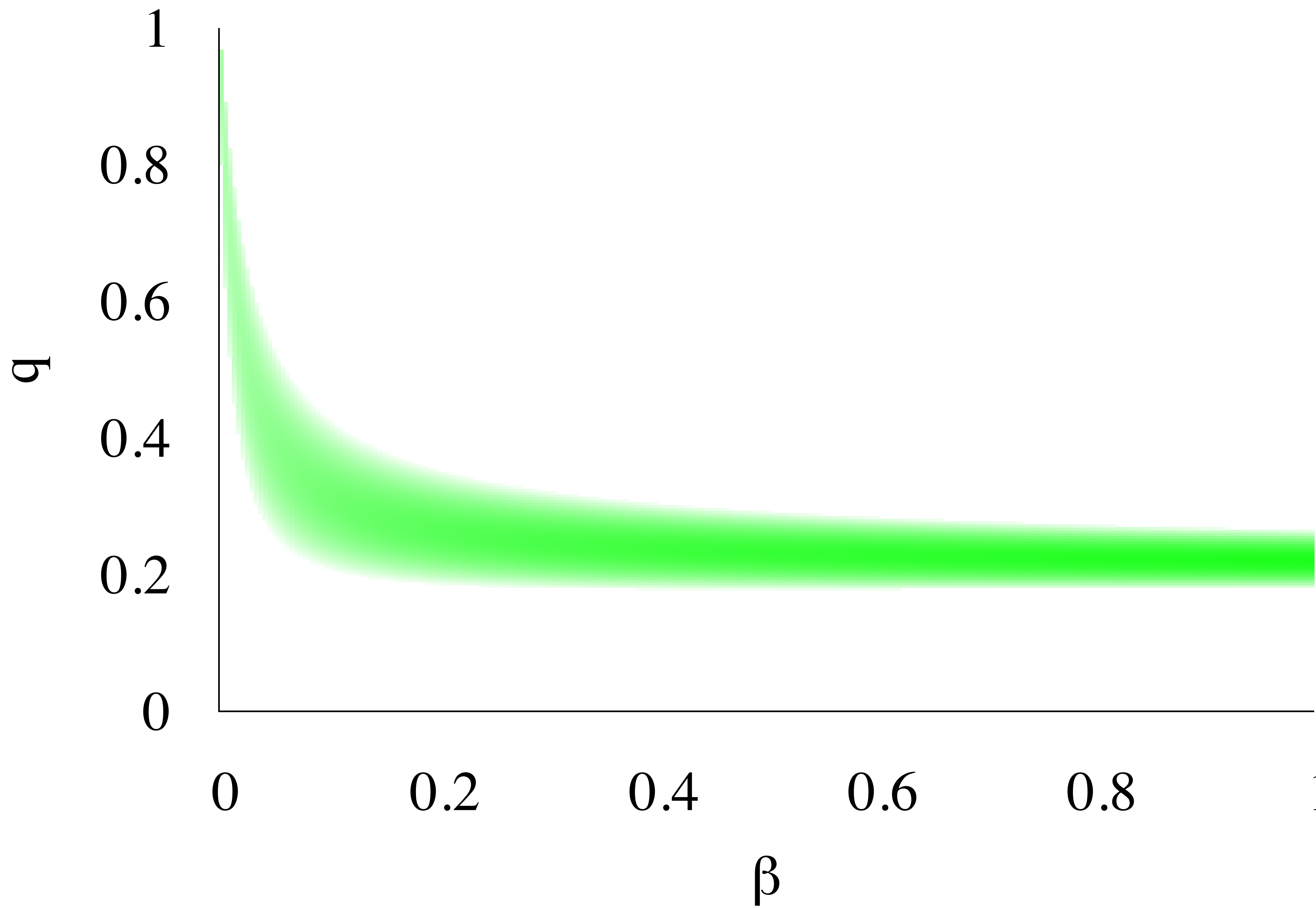
Adiabatic Monte Carlo dynamically transitions from the base distribution to the target distribution.

$\beta = 4.58897\text{e-}05$

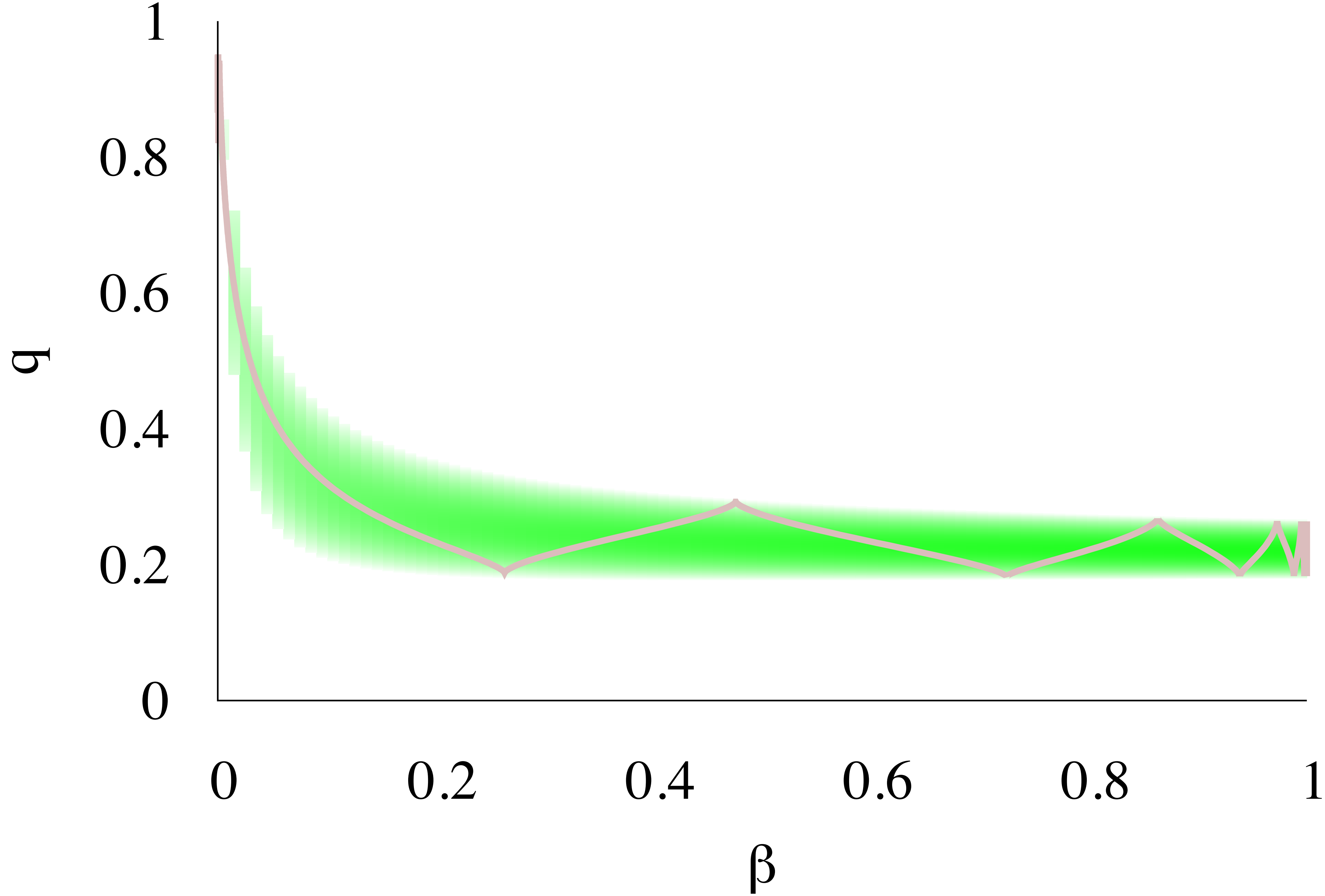
Adiabatic Monte Carlo dynamically transitions from the base distribution to the target distribution.

$\beta = 4.58897\text{e-}05$

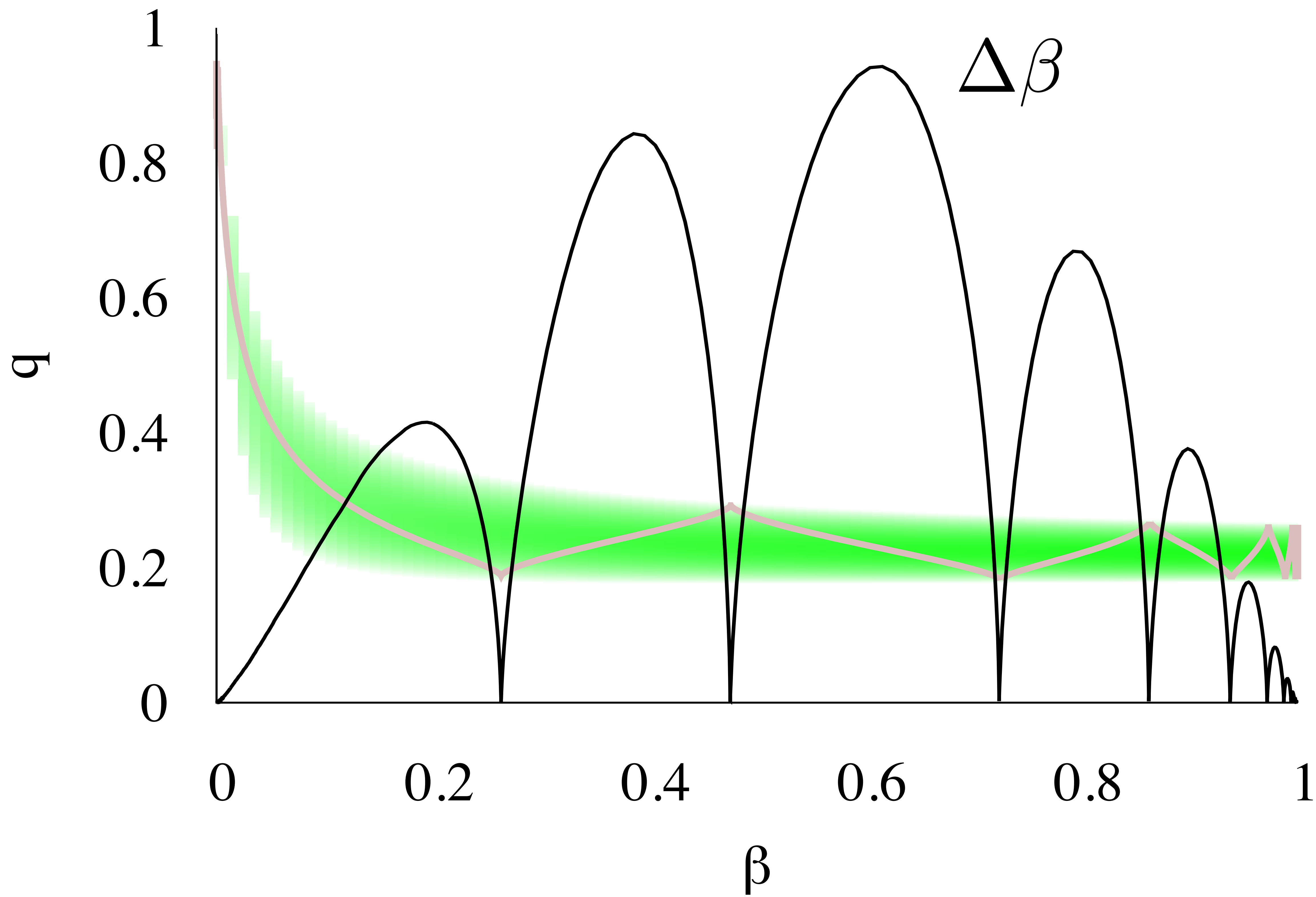
To see the optimality of adiabatic transitions, consider the interpolation of a unidimensional distribution.



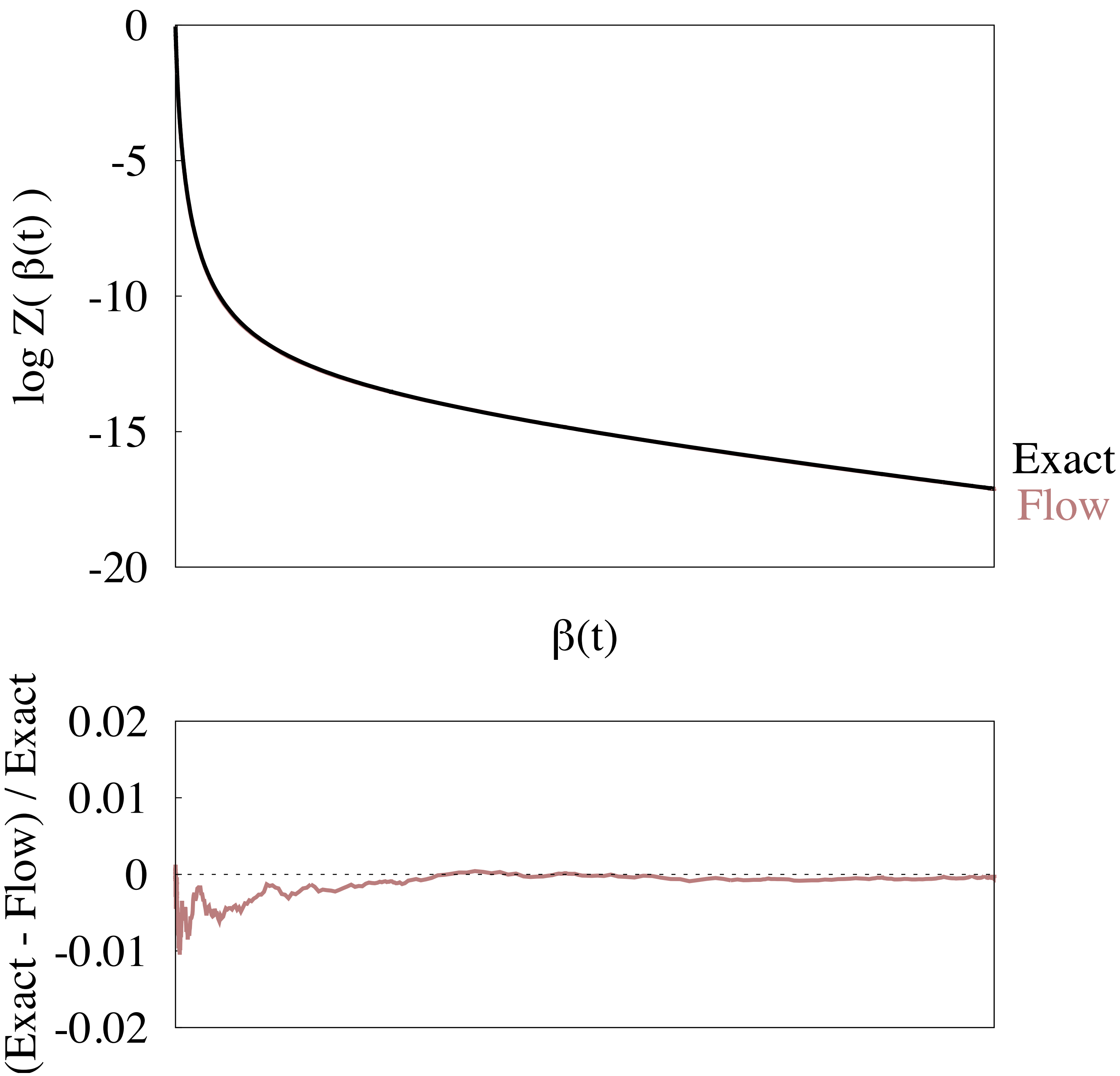
Adiabatic transitions automatically equilibrate, implicitly generating an optimal interpolation partition.

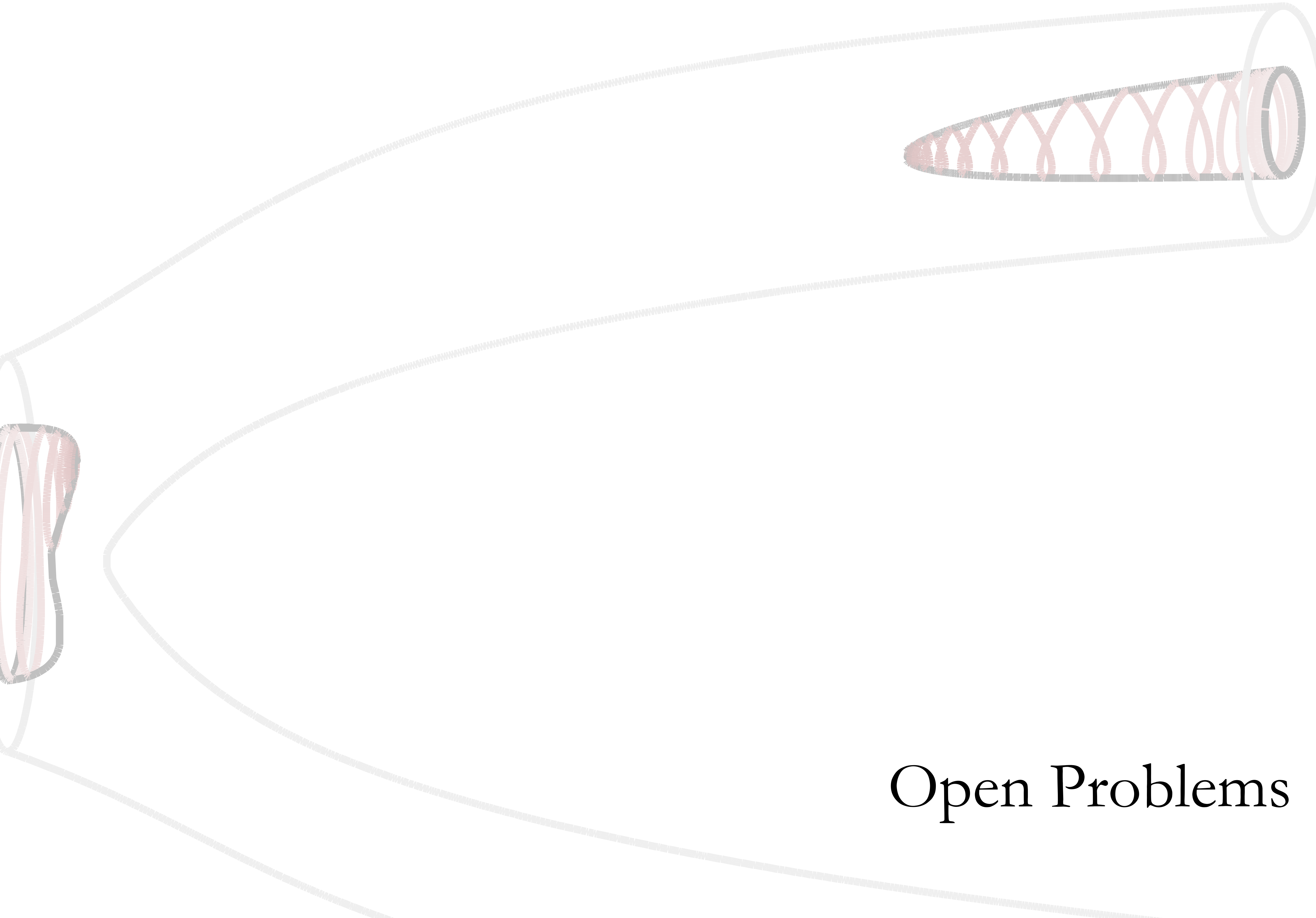


Adiabatic transitions automatically equilibrate, implicitly generating an optimal interpolation partition.



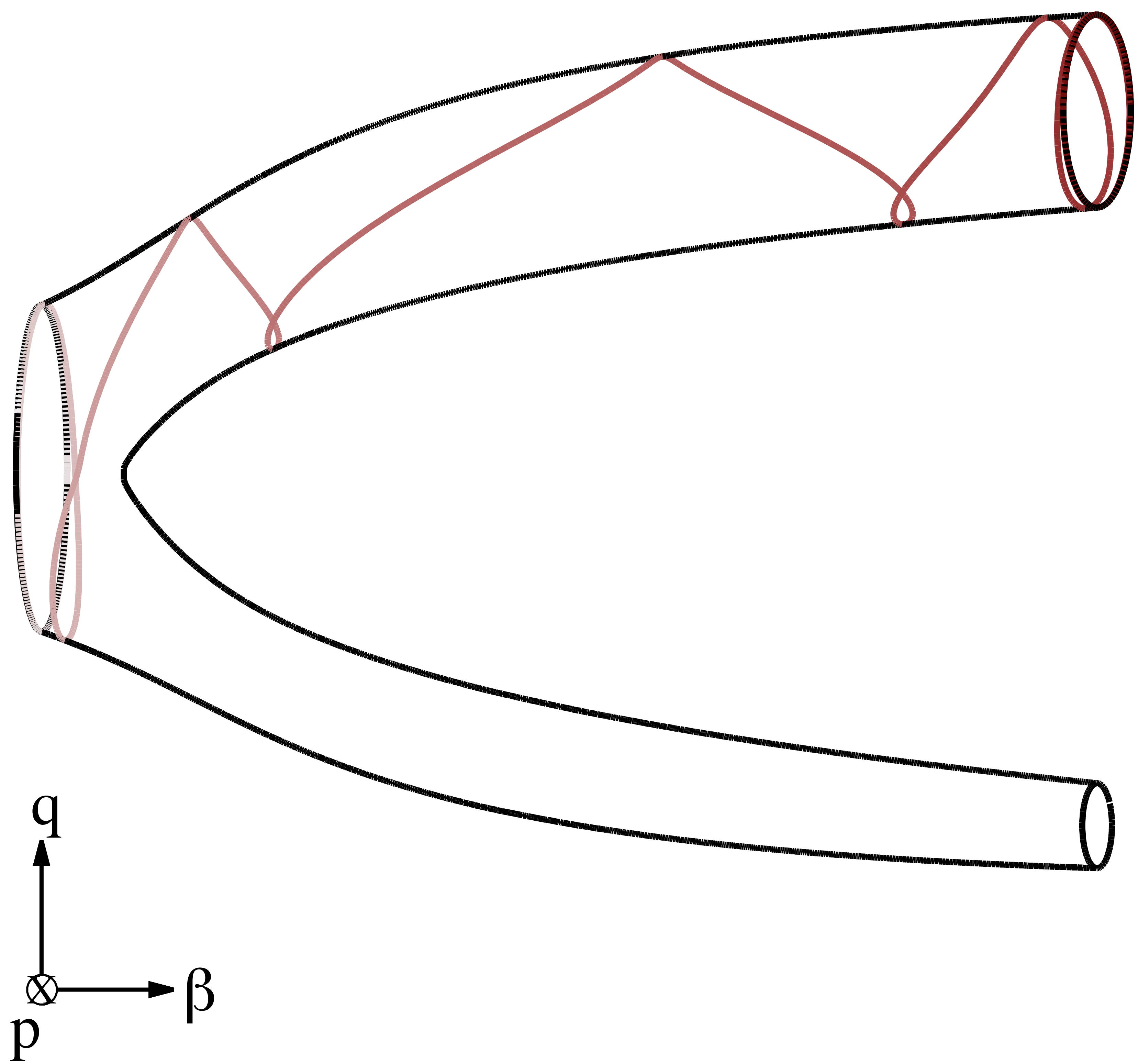
In theory we can recover the normalizing constant exactly. In practice we can recover it incredibly accurately.



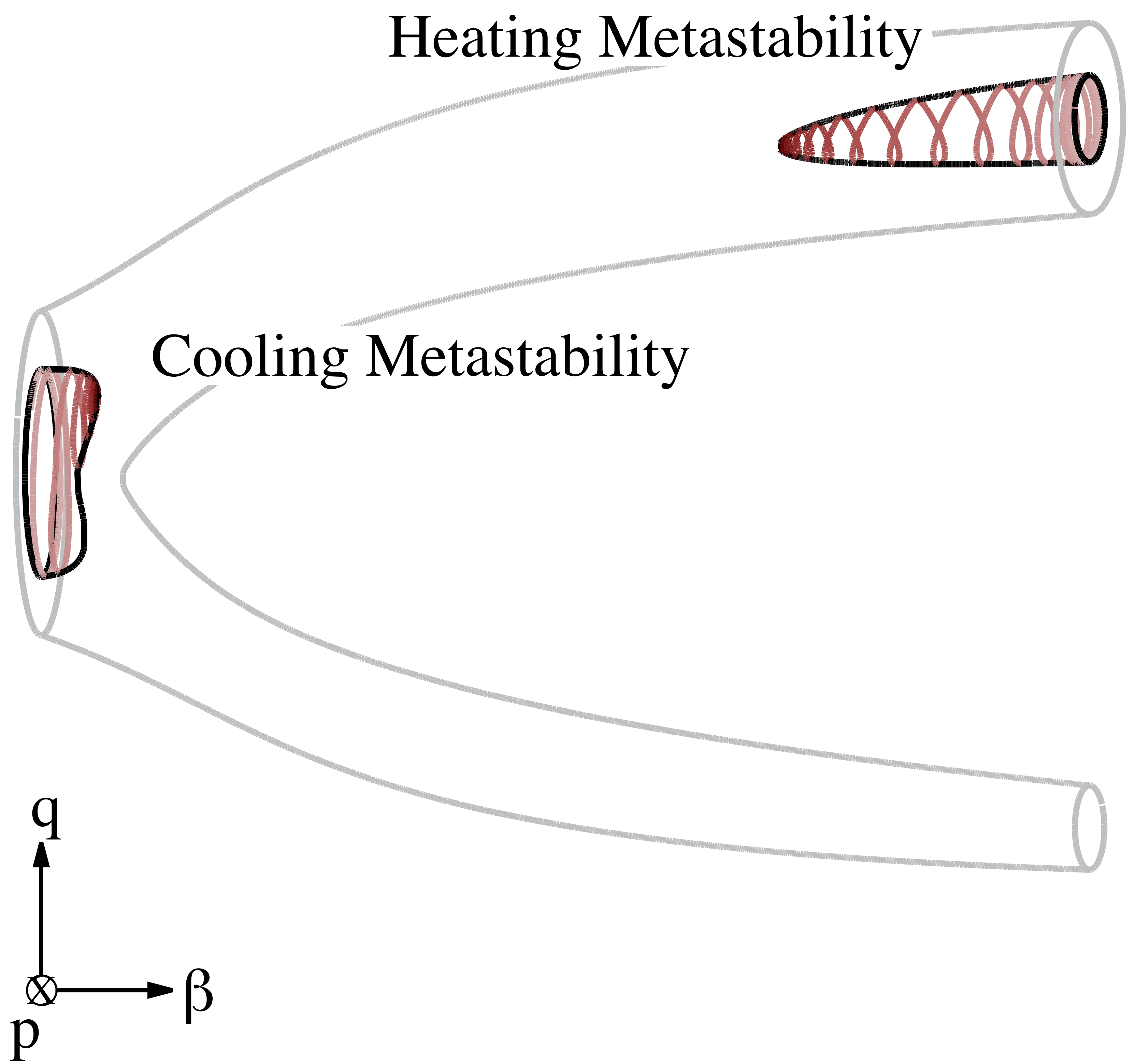


Open Problems

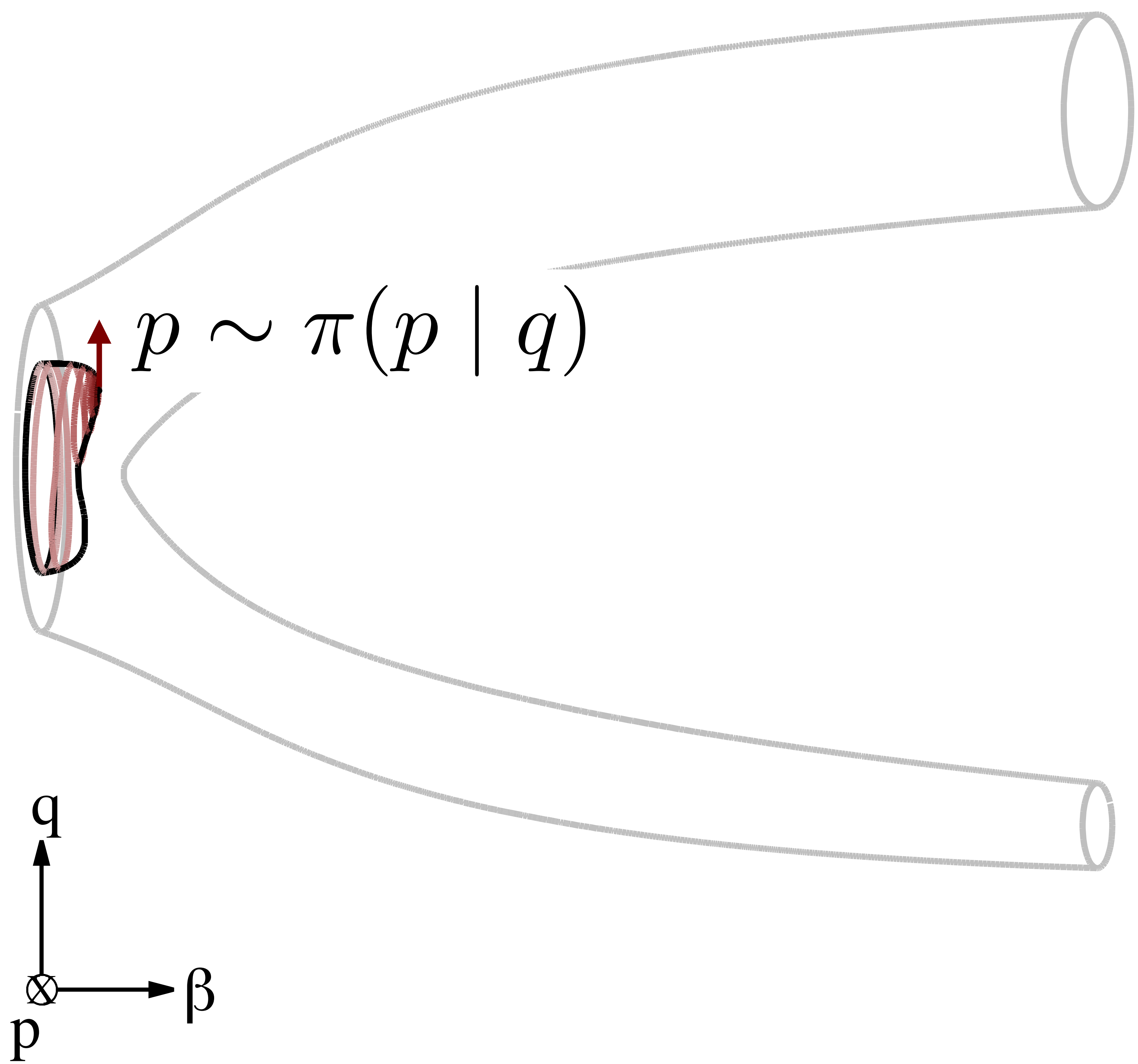
The immediate problem with adiabatic transitions is that *metastabilities* prevent them from being isomorphisms.



The immediate problem with adiabatic transitions is that *metastabilities* prevent them from being isomorphisms.



Fortunately we can readily recover from a metastability by resampling the momenta, effectively reheating the system.



We also need to compute the intermediate expectations needed to generate each transition.

$$\frac{\mathrm{d}q}{\mathrm{d}t} = \frac{\partial T}{\partial p}$$

$$\frac{\mathrm{d}p}{\mathrm{d}t} = -\frac{\partial T}{\partial q} - \frac{\partial V_\beta}{\partial q} + \left(\Delta V - \mathbb{E}_{\pi_\beta}[\Delta V]\right) p$$

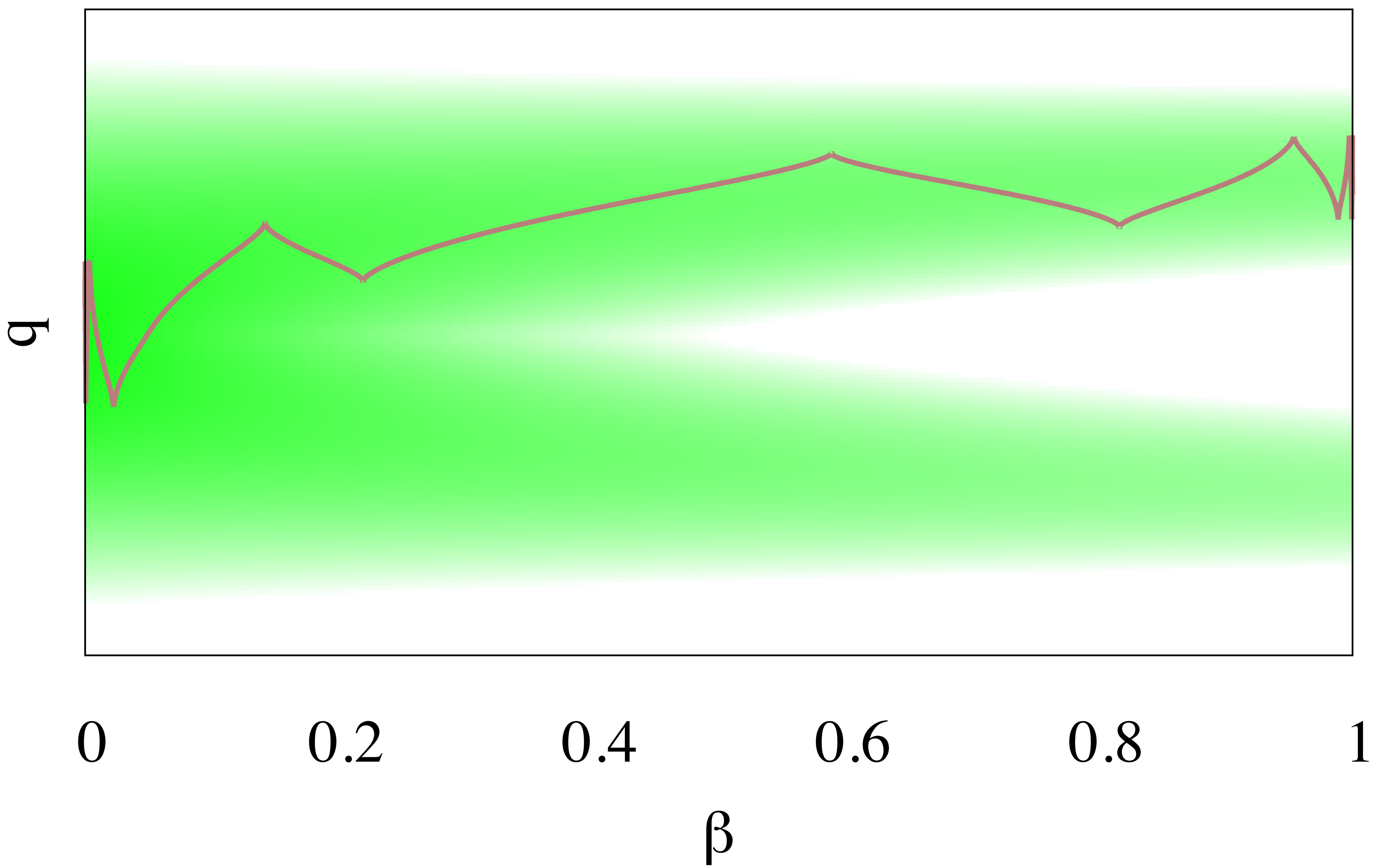
$$\frac{\mathrm{d}\beta}{\mathrm{d}t} = -p \frac{\partial T}{\partial p}$$

Hamiltonian Monte Carlo gives efficient local estimations, which can be aggregated together into a global estimator.

$$\mathbb{E}_{\pi_\beta}[\Delta V] \approx \frac{\sum_{n=1}^{N} \widehat{Z}_n \widehat{\Delta V}(\beta)}{\sum_{n=1}^{N} \widehat{Z}_n}$$

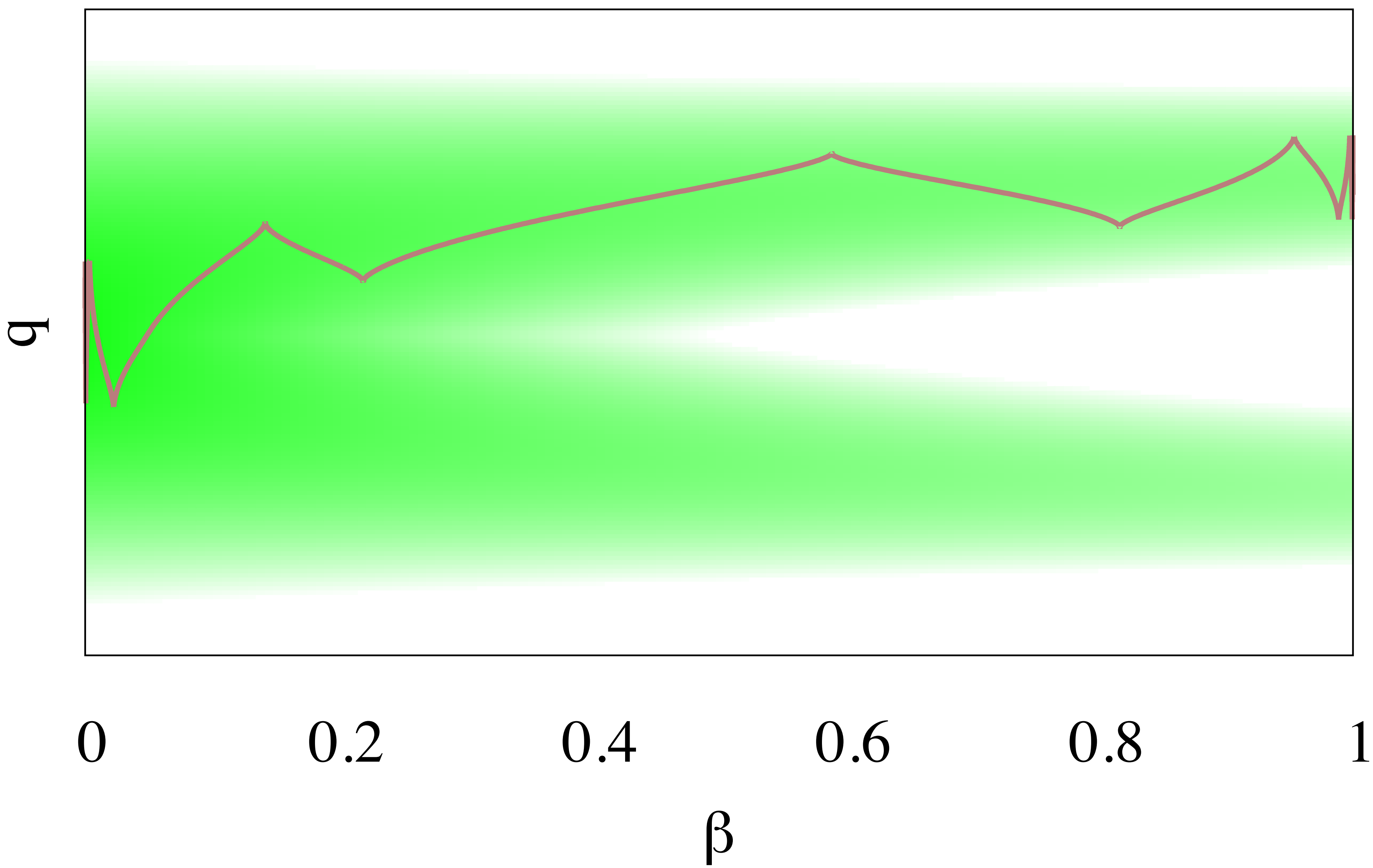
Finally, there is the problem of correcting for the error from numerical approximations to the exact transitions.

$(q_i, p_i) \sim \pi_{\beta=0}$ $(q_f, p_f) \sim \pi_{\beta=1}$



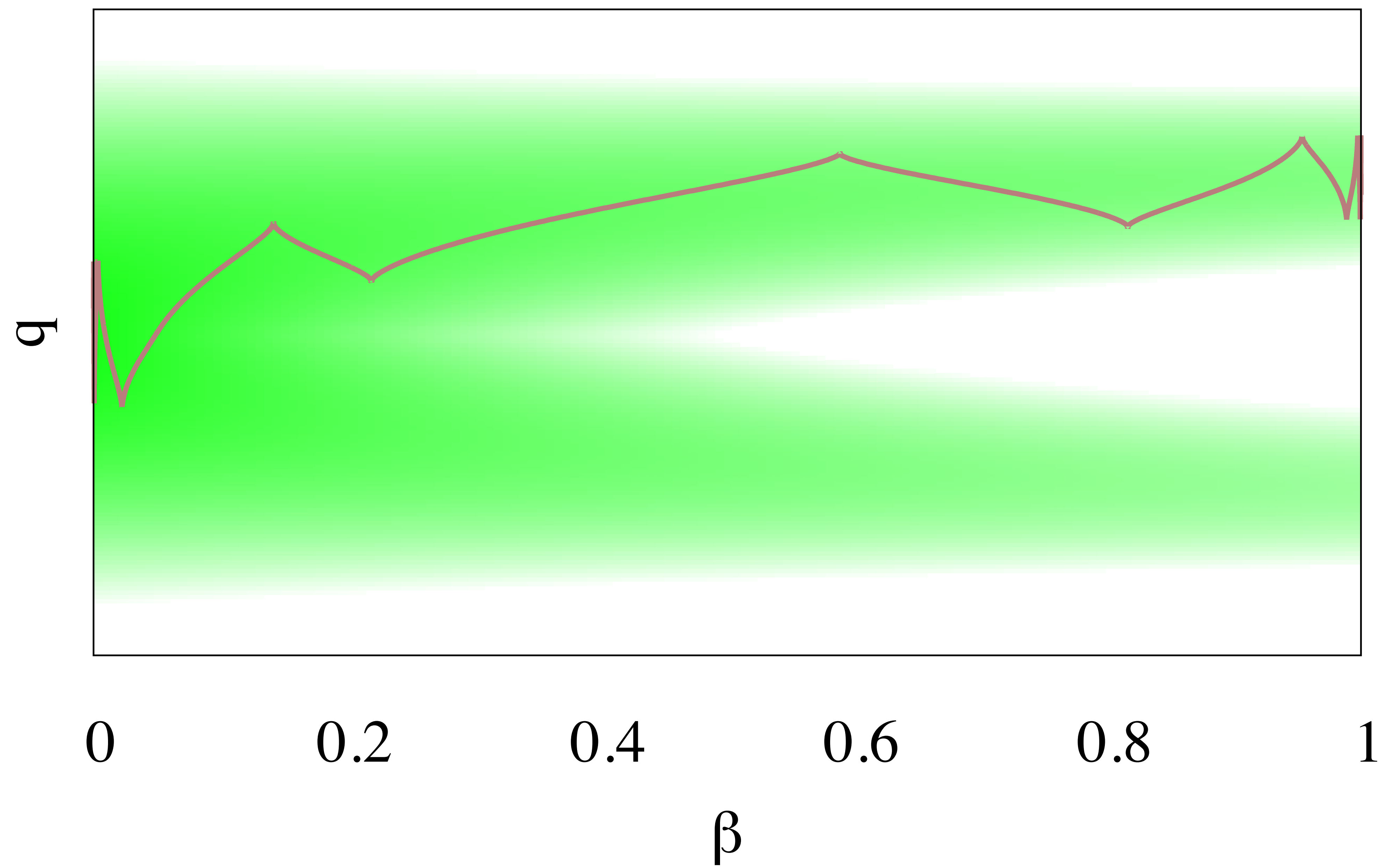
We can't apply a naive Metropolis correction, but perhaps we can apply a correction with a swap?

$$(q_i, p_i) \sim \pi_{\beta=0} \longleftarrow ? \longrightarrow (q_f, p_f) \sim \pi_{\beta=1}$$



Unfortunately, swapping states doesn't work because discretized perks will not, in general, be aligned.

$$(q_i, p_i) \sim \pi_{\beta \approx 0} \longleftarrow ? \longrightarrow (q_f, p_f) \sim \pi_{\beta \approx 1}$$



Unfortunately, swapping states doesn't work because discretized perks will not, in general, be aligned.

$$(q_i, p_i) \sim \pi_{\beta \approx 0} \quad \not\longleftrightarrow \quad (q_f, p_f) \sim \pi_{\beta \approx 1}$$