Optimization-inspired Barriers in Nested Sampling

Farin Lippmann

13.01.2025

Friedrich-Schiller-University Jena

Supervisor: Prof. Dr. Michael Habeck

Table of Contents

Background

Nested Sampling - Overview

Introduced by Skilling [1].

Input:

- likelihood and
- prior of a parameterized generative statistical model
- data from the process that is being modeled

Output:

- samples from the posterior distribution of parameter values
- an estimate of the evidence (useful for model comparison)

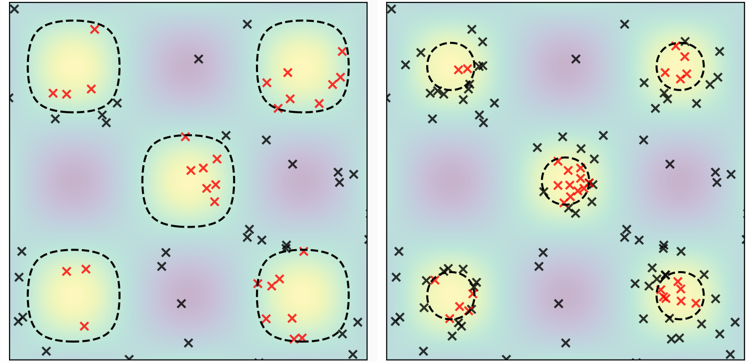
Nested Sampling - Bayes' Theorem

$$\overbrace{P(\theta \mid D)}^{\text{Posterior}} = \frac{\overbrace{P(D \mid \theta)}^{\text{Likelihood}}\ \overbrace{P(\theta)}^{\text{Prior}}}{\underbrace{P(D)}_{\text{Evidence}}} = \frac{L(\theta)\,\pi(\theta)}{Z}$$

Nested Sampling - Algorithm

- generate samples iteratively and slowly raise a minimum likelihood samples must satisfy
- use an ensemble of points ("live points") $\theta_1 \ldots \theta_n$
- in each iteration:
    - remove the lowest-likelihood point θ_{min}
    - estimate its contribution to the evidence
    - set $L^* \leftarrow L(\theta)$ as the new likelihood constraint
    - sample a new point from the constrained prior $\theta_{\mathrm{new}} \sim \pi(\theta)\mathbb{1}_{\{L(\theta)>L^*\}}(\theta)$ and add it to the live points
- repeat until some stopping criteria are fulfilled

Nested Sampling - A Visualization

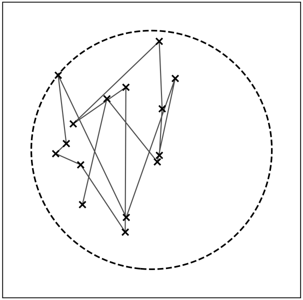


Nested Sampling - Likelihood-Restricted Prior Sampling

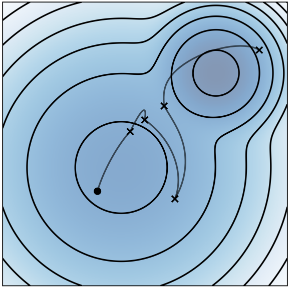
Sampling from the constrained prior $\theta_{\text{new}} \sim \pi(\theta)\mathbb{1}_{\{L(\theta)>L^*\}}(\theta)$

Skilling: “such points will usually be found by some MCMC approximation, ...” [1, p. 6]

Metropolis-Hastings



Hamiltonian Monte Carlo



The Barrier Method - The Problem

Solves optimization problems of the form

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && f_i(x) \leq 0 \quad \text{for } i = 1, ..., m \end{aligned}$$

with f and all f_i being convex and twice continuously differentiable [2, p. 561].

The Barrier Method - Exact Barriers

Incorporate constraints into the objective function:

$$\text{minimize} \qquad f(x) + \sum_{i=1}^{m} \mathcal{I}(f_i(x))$$

with:

$$\mathcal{I}(u) = \begin{cases} 0, & \text{if } u \leq 0 \\ \infty, & \text{if } u > 0 \end{cases}$$

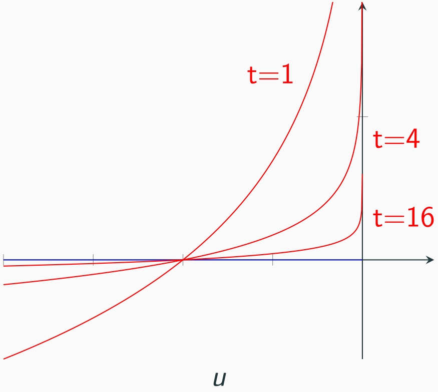
Accurately represents the problem, but is not differentiable.

The Barrier Method - Log Barriers

Log barrier function:

$$\hat{\mathcal{I}}(u) = -\frac{1}{t}\log(-u)$$

with parameter t.



The Barrier Method - Approximate Optimization Problem

$$\text{minimize} \qquad f(x) + \sum_{i=1}^{m} -\frac{1}{t}\log(-f_i(x))$$

The Barrier Method - Approximate Optimization Problem

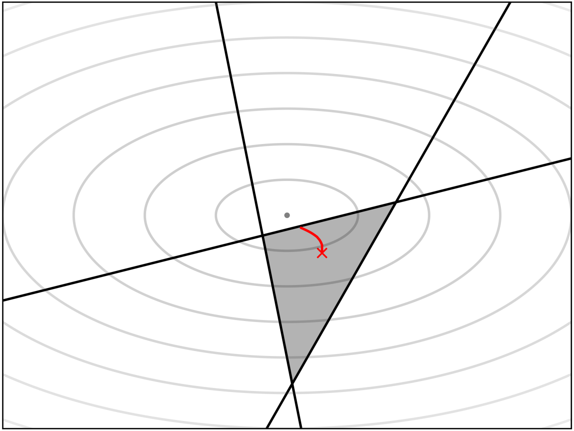
$$\text{minimize} \qquad f(x) + \sum_{i=1}^{m} -\frac{1}{t}\log(-f_i(x))$$

Multiply objective function by t to obtain:

$$\text{minimize} \qquad t \cdot f(x) + \sum_{i=1}^{m} -\log(-f_i(x))$$

There is a different solution $x^*(t)$ for each value of t. As $t \to \infty$, $x^*(t)$ approaches the real solution [2, p. 561ff.].

The Barrier Method - Central Path



The Barrier Method - The Idea

$x^*(t)$ for a high value of t is a good estimation, but has extreme derivatives.

The Barrier Method - The Idea

$x^*(t)$ for a high value of t is a good estimation, but has extreme derivatives.

Stabilize by using a good starting point, a solution using a lower value of t. $\rightarrow$ Bootstrapping.

Nested Sampling with Barriers

Nested Sampling with Barriers - Motivation

- LRPS is difficult

Nested Sampling with Barriers - Motivation

- LRPS is difficult
- Ratio of the valid area to the area of the whole sample space decreases steadily, finding samples becomes harder.

Nested Sampling with Barriers - Motivation

- LRPS is difficult
- Ratio of the valid area to the area of the whole sample space decreases steadily, finding samples becomes harder.
- Result: low acceptance rates or (if using an adaptive step size) highly correlated samples.

Nested Sampling with Barriers - Motivation

- LRPS is difficult
- Ratio of the valid area to the area of the whole sample space decreases steadily, finding samples becomes harder.
- Result: low acceptance rates or (if using an adaptive step size) highly correlated samples.
- These methods only notice the likelihood constraint boundary once they cross it.

Nested Sampling with Barriers - Motivation

- LRPS is difficult
- Ratio of the valid area to the area of the whole sample space decreases steadily, finding samples becomes harder.
- Result: low acceptance rates or (if using an adaptive step size) highly correlated samples.
- These methods only notice the likelihood constraint boundary once they cross it.
- $\implies$ Would like an approach that feels the boundary earlier and pushes the samples away from it.

Nested Sampling with Barriers - Motivation

- LRPS is difficult
- Ratio of the valid area to the area of the whole sample space decreases steadily, finding samples becomes harder.
- Result: low acceptance rates or (if using an adaptive step size) highly correlated samples.
- These methods only notice the likelihood constraint boundary once they cross it.
- $\implies$ Would like an approach that feels the boundary earlier and pushes the samples away from it.
- The log barrier lends itself to this idea.

Derivation - Pretend Optimization

We pretend that likelihood-restricted prior sampling (LRPS) is an optimization problem.

$$\text{maximize} \qquad \pi(\theta) \qquad \text{subject to} \quad L(\theta) > L^*$$

Derivation - Pretend Optimization

We pretend that likelihood-restricted prior sampling (LRPS) is an optimization problem.

$$
\begin{aligned}
& \text{maximize} && \pi(\theta) && \text{subject to} && L(\theta) > L^* \\
\iff \quad & \text{minimize} && -\log(\pi(\theta)) && \text{subject to} && l(\theta) > l^*
\end{aligned}
$$

Derivation - Pretend Optimization

We pretend that likelihood-restricted prior sampling (LRPS) is an optimization problem.

$$\begin{array}{lllll} & \text{maximize} & \pi(\theta) & \text{subject to} & L(\theta) > L^* \\ \iff & \text{minimize} & -\log(\pi(\theta)) & \text{subject to} & l(\theta) > l^* \\ \iff & \text{minimize} & -\log(\pi(\theta)) & \text{subject to} & l^* - l(\theta) < 0 \end{array}$$

Incorporate constraints into objective function:

$$\text{minimize} \quad -\log(\pi(\theta)) + \mathcal{I}(l^* - l(\theta))$$

Derivation - The Log Barrier Term

Associated approximate problem using log barriers:

$$\begin{aligned} &\text{minimize} \quad -\log(\pi(\theta)) + \hat{\mathcal{I}}(I^* - I(\theta)) \\ \iff &\text{minimize} \quad -\log(\pi(\theta)) - \frac{1}{t}\log(I(\theta) - I^*) \end{aligned}$$

Derivation - The Log Barrier Term

Associated approximate problem using log barriers:

$$\begin{aligned} &\text{minimize} \quad -\log(\pi(\theta)) + \hat{\mathcal{I}}(l^* - l(\theta)) \\ \iff &\text{minimize} \quad -\log(\pi(\theta)) - \frac{1}{t}\log(l(\theta) - l^*) \end{aligned}$$

Negate and exponentiate to return to a maximization problem in probability space:

$$\begin{aligned} &\text{maximize} \quad \exp\left(-\left(-\log(\pi(\theta)) - \frac{1}{t}\log(l(\theta) - l^*)\right)\right) \\ \iff &\text{maximize} \quad \pi(\theta)\left(l(\theta) - l^*\right)^{\frac{1}{t}} \end{aligned}$$

Derivation - Introducing q

Expand sampling space:

$$\theta \to (\theta, q)$$

Derivation - Introducing q

Expand sampling space:

$$\theta \to (\theta, q)$$

Start by assuming independence:

$$P(\theta, q) = P(\theta)\, P(q)$$

Derivation - Introducing q

Expand sampling space:

$$\theta \to (\theta, q)$$

Start by assuming independence:

$$P(\theta, q) = P(\theta)\, P(q)$$

This independence follows through to the likelihood, prior and evidence:

$$L(\theta, q) = L_\theta(\theta)\, L_q(q)$$
$$\pi(\theta, q) = \pi_\theta(\theta)\, \pi_q(q)$$
$$Z_{\theta,q} = Z_\theta\, Z_q$$

Derivation - LRPS on the Expanded Space

Likelihood-restricted prior sampling on the expanded space:

$$
\begin{aligned}
(\theta, q) &\sim \bar{\pi}(\theta, q \mid L^*) \\
&\bar{\pi}(\theta, q \mid L^*) = \pi(\theta, q)\, \mathbb{1}_{\{L(\theta,q) > L^*\}}(\theta, q)
\end{aligned}
$$

Derivation - LRPS on the Expanded Space

Likelihood-restricted prior sampling on the expanded space:

$$(\theta, q) \sim \bar{\pi}(\theta, q \mid L^*)$$
$$\bar{\pi}(\theta, q \mid L^*) = \pi(\theta, q)\, \mathbb{1}_{\{L(\theta,q) > L^*\}}(\theta, q)$$

Use collapsed Gibbs sampling [3]:

$$\theta \sim \bar{\pi}(\theta \mid L^*)$$
$$q \sim \bar{\pi}(q \mid L^*, \theta)$$

Derivation - LRPS on the Expanded Space

Define:

$$L_q(q) = \frac{1}{q}$$

Derivation - LRPS on the Expanded Space

Define:

$$L_q(q) = \frac{1}{q}$$

Now we can reshape the likelihood-restricted priors:

$$\begin{aligned}
\bar{\pi}(\theta \mid L^*) &= \int_{q_{\min}}^{q_{\max}} \bar{\pi}(\theta, q \mid L^*)\, dq \\
&= \int_{q_{\min}}^{q_{\max}} \pi_\theta(\theta)\, \pi_q(q)\, \mathbb{1}_{\{L(\theta,q) > L^*\}}(\theta, q)\, dq \\
&= \pi_\theta(\theta) \int_{q_{\min}}^{L^*/L_\theta(\theta)} \pi_q(q)\, dq \\
&= \pi_\theta(\theta)\, \Pi_q(L_\theta(\theta)/L^*) \\
\bar{\pi}(q \mid L^*, \theta) &= \pi_q(q)\, \mathbb{1}_{\{q < L_\theta(\theta)/L^*\}}(q)
\end{aligned}$$

Derivation - LRPS on the expanded Space

Add the log barrier term via the CDF of q:

$$\Pi_q(L_\theta(\theta)/L^*) \quad \stackrel{!}{=} \quad (l_\theta(\theta) - l^*)^{\frac{1}{t}}$$

Derivation - LRPS on the expanded Space

Add the log barrier term via the CDF of q:

$$\begin{aligned} \Pi_q(L_\theta(\theta)/L^*) &\stackrel{!}{=} (l_\theta(\theta) - l^*)^{\frac{1}{t}} \\ \Pi_q(L_\theta(\theta)/L^*) &= \log(L_\theta(\theta)/L^*)^{\frac{1}{t}} \end{aligned}$$

Derivation - LRPS on the expanded Space

Add the log barrier term via the CDF of q:

$$\begin{aligned} \Pi_q(L_\theta(\theta)/L^*) &\stackrel{!}{=} (l_\theta(\theta) - l^*)^{\frac{1}{t}} \\ \Pi_q(L_\theta(\theta)/L^*) &= \log(L_\theta(\theta)/L^*)^{\frac{1}{t}} \\ \text{let } q &= L_\theta(\theta)/L^* \end{aligned}$$

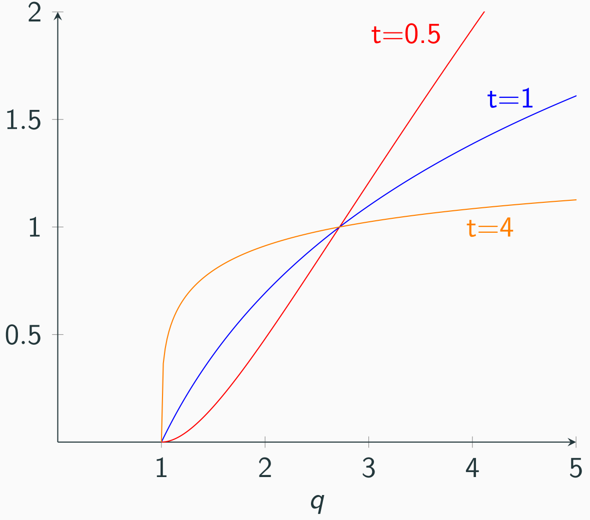
Derivation - LRPS on the expanded Space

Add the log barrier term via the CDF of q:

$$\begin{aligned}
\Pi_q(L_\theta(\theta)/L^*) &\stackrel{!}{=} (l_\theta(\theta) - l^*)^{\frac{1}{t}} \\
\Pi_q(L_\theta(\theta)/L^*) &= \log(L_\theta(\theta)/L^*)^{\frac{1}{t}} \\
\text{let } q &= L_\theta(\theta)/L^* \\
\Pi_q(q) &= \log(q)^{\frac{1}{t}}
\end{aligned}$$

Also implicitly defined $q_{\min} = 1$.

Derivation - Unnormalized CDF of q



Derivation - Nested Sampling with Barriers

With likelihood and prior of q defined and the log barrier term incorporated, we have a new algorithm.

Derivation - Nested Sampling with Barriers

With likelihood and prior of q defined and the log barrier term incorporated, we have a new algorithm.

Changes from the original Nested Sampling:

1. Set new parameters: $t > 0$ and $q_{\mathsf{max}} > 1$.

Derivation - Nested Sampling with Barriers

With likelihood and prior of q defined and the log barrier term incorporated, we have a new algorithm.

Changes from the original Nested Sampling:

1. Set new parameters: $t > 0$ and $q_{\text{max}} > 1$.
2. For each starting θ, also sample a q.

Derivation - Nested Sampling with Barriers

With likelihood and prior of q defined and the log barrier term incorporated, we have a new algorithm.

Changes from the original Nested Sampling:

1. Set new parameters: $t > 0$ and $q_{\text{max}} > 1$.
2. For each starting θ, also sample a q.
3. Use the joint likelihoods of θ and q, $L(\theta, q) = \frac{L_\theta(\theta)}{q}$, instead of just the likelihood of θ.

Derivation - Nested Sampling with Barriers

With likelihood and prior of q defined and the log barrier term incorporated, we have a new algorithm.

Changes from the original Nested Sampling:

1. Set new parameters: $t > 0$ and $q_{\mathsf{max}} > 1$.
2. For each starting θ, also sample a q.
3. Use the joint likelihoods of θ and q, $L(\theta, q) = \frac{L_\theta(\theta)}{q}$, instead of just the likelihood of θ.
4. Sample θ_{new} from the modified constrained prior $\pi_\theta(\theta)\,\Pi_q(l_\theta(\theta) - l^*)$.

Derivation - Nested Sampling with Barriers

With likelihood and prior of q defined and the log barrier term incorporated, we have a new algorithm.

Changes from the original Nested Sampling:

1. Set new parameters: $t > 0$ and $q_{\text{max}} > 1$.
2. For each starting θ, also sample a q.
3. Use the joint likelihoods of θ and q, $L(\theta, q) = \frac{L_\theta(\theta)}{q}$, instead of just the likelihood of θ.
4. Sample θ_{new} from the modified constrained prior $\pi_\theta(\theta)\, \Pi_q(l_\theta(\theta) - l^*)$.
5. Sample q_{new} from the constrained prior $\pi_q(q)\, \mathbb{1}_{\{q < L_\theta(\theta_{\text{new}})/L^*\}}(q)$ dependent on the sampled θ_{new}.

Derivation - Nested Sampling with Barriers

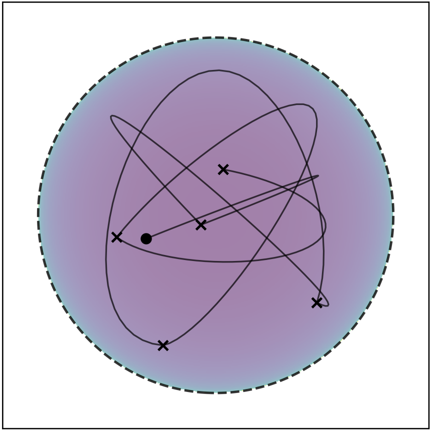
With likelihood and prior of q defined and the log barrier term incorporated, we have a new algorithm.

Changes from the original Nested Sampling:

1. Set new parameters: $t > 0$ and $q_{\text{max}} > 1$.
2. For each starting θ, also sample a q.
3. Use the joint likelihoods of θ and q, $L(\theta, q) = \frac{L_\theta(\theta)}{q}$, instead of just the likelihood of θ.
4. Sample θ_{new} from the modified constrained prior $\pi_\theta(\theta)\, \Pi_q(l_\theta(\theta) - l^*)$.
5. Sample q_{new} from the constrained prior $\pi_q(q)\, \mathbb{1}_{\{q < L_\theta(\theta_{\text{new}})/L^*\}}(q)$ dependent on the sampled θ_{new}.
6. Divide out $Z_q(t, q_{\text{max}})$ from the evidence before returning it.

Derivation - A Visualization

LRPS using HMC in Nested Sampling with Barriers



Interpretation - q

- Each θ is paired with a q.

Interpretation - q

- Each θ is paired with a q.
- $L(\theta, q) = \frac{L_\theta(\theta)}{q}$ and $q > 1$.

Interpretation - q

- Each θ is paired with a q.
- $L(\theta, q) = \frac{L_\theta(\theta)}{q}$ and $q > 1$.
- $\implies$ The likelihood of each θ is reduced by its q.

Interpretation - q

- Each θ is paired with a q.
- $L(\theta, q) = \frac{L_\theta(\theta)}{q}$ and $q > 1$.
- $\implies$ The likelihood of each θ is reduced by its q.
- Since Z_q is divided out, this only influences sampling.

Interpretation - q

- Each θ is paired with a q.
- $L(\theta, q) = \frac{L_\theta(\theta)}{q}$ and $q > 1$.
- $\implies$ The likelihood of each θ is reduced by its q.
- Since Z_q is divided out, this only influences sampling.
- q is sampled from $\pi_q(q)\, \mathbb{1}_{\{q<L_\theta(\theta)/L^*\}}(q)$

Interpretation - q

- Each θ is paired with a q.
- $L(\theta, q) = \frac{L_\theta(\theta)}{q}$ and $q > 1$.
- $\implies$ The likelihood of each θ is reduced by its q.
- Since Z_q is divided out, this only influences sampling.
- q is sampled from $\pi_q(q)\, \mathbb{1}_{\{q < L_\theta(\theta)/L^*\}}(q)$
- $\implies$ θ further away from the constraint barrier tend to have higher q, reducing their likelihood

Interpretation - q

- Each θ is paired with a q.
- $L(\theta, q) = \frac{L_\theta(\theta)}{q}$ and $q > 1$.
- $\implies$ The likelihood of each θ is reduced by its q.
- Since Z_q is divided out, this only influences sampling.
- q is sampled from $\pi_q(q)\, \mathbb{1}_{\{q<L_\theta(\theta)/L^*\}}(q)$
- $\implies$ θ further away from the constraint barrier tend to have higher q, reducing their likelihood
- Interpretation: the lowered likelihood makes up for the log barrier term encouraging samples further away from the constraint

Interpretation - q

- Each θ is paired with a q.
- $L(\theta, q) = \frac{L_\theta(\theta)}{q}$ and $q > 1$.
- $\implies$ The likelihood of each θ is reduced by its q.
- Since Z_q is divided out, this only influences sampling.
- q is sampled from $\pi_q(q)\, \mathbb{1}_{\{q < L_\theta(\theta)/L^*\}}(q)$
- $\implies$ θ further away from the constraint barrier tend to have higher q, reducing their likelihood
- Interpretation: the lowered likelihood makes up for the log barrier term encouraging samples further away from the constraint
- If the likelihoods were not lowered, the constraint would rise much more quickly than in standard Nested Sampling

Interpretation - t and q_{max}

- t and q_{max} influence the shape and support of q's distribution

Interpretation - t and q_{max}

- t and q_{max} influence the shape and support of q's distribution
- Since the log barrier term is incorporated using q's CDF, they also influence how the log barrier term acts.

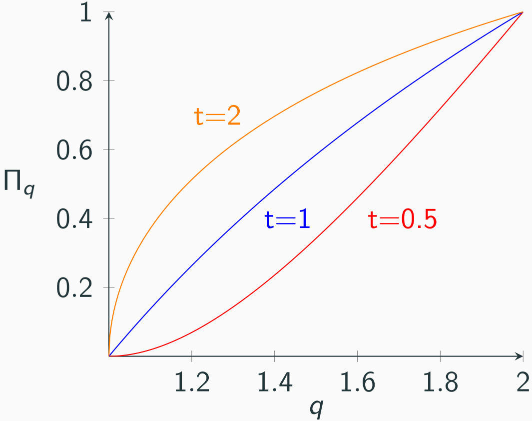
Interpretation - t and q_{max}

- t and q_{max} influence the shape and support of q's distribution
- Since the log barrier term is incorporated using q's CDF, they also influence how the log barrier term acts.
- Larger $q_{\text{max}} \implies$ the log barrier term affects sampling earlier

Interpretation - t and q_{max}

- t and q_{max} influence the shape and support of q's distribution
- Since the log barrier term is incorporated using q's CDF, they also influence how the log barrier term acts.
- Larger $q_{\text{max}} \implies$ the log barrier term affects sampling earlier
- Larger $t \implies$ the log barrier becomes steeper, affecting only values closer to the boundary

Interpretation - t and $q_{\max}$



Results

Setup - Example Problems

Gaussian mixtures with two components, one slab and a spike:

- $L_1(\theta) = 100\,\mathcal{N}(\theta \mid \mu = 0, \Sigma = uI_{20}) + \mathcal{N}(\theta \mid \mu = 0, \Sigma = vI_{20})$
- $L_2(\theta) = 100\,\mathcal{N}(\theta \mid \mu = (0.2, 0.2, ..., 0.2), \Sigma = uI_{20}) + \mathcal{N}(\theta \mid \mu = 0, \Sigma = vI_{20})$

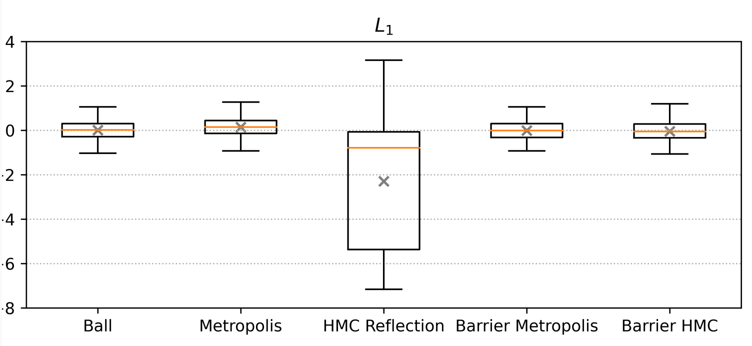
with $u = 0.01$ and $v = 0.1$.

Setup - Configurations

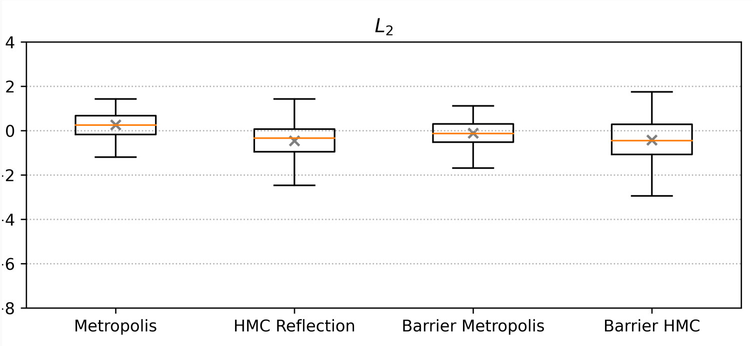
These configurations were tested:

- Metropolis Algorithm (with simple adaptive step size)
- HMC with reflection (only for standard Nested Sampling, with simple adaptive step size) [4]
- HMC (only for Nested Sampling with Barriers, with simple adaptive step size)
- an optimal method for sampling inside a d-ball (only for the L_1-problem with Standard Nested Sampling; labeled “Ball” in the figures)

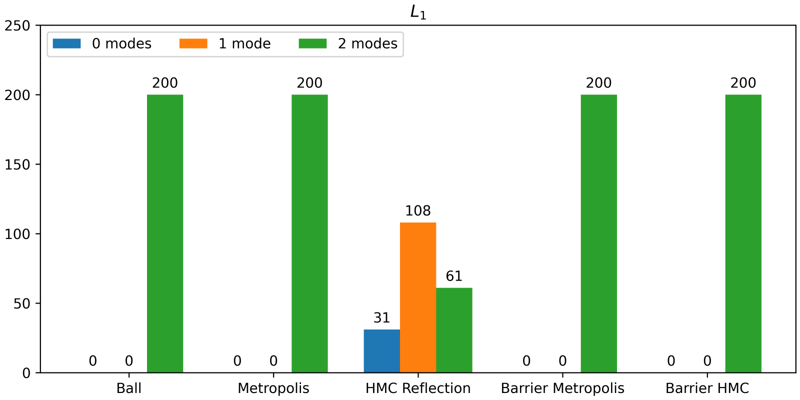
Nested Sampling with and without Barriers - Evidence Estimates



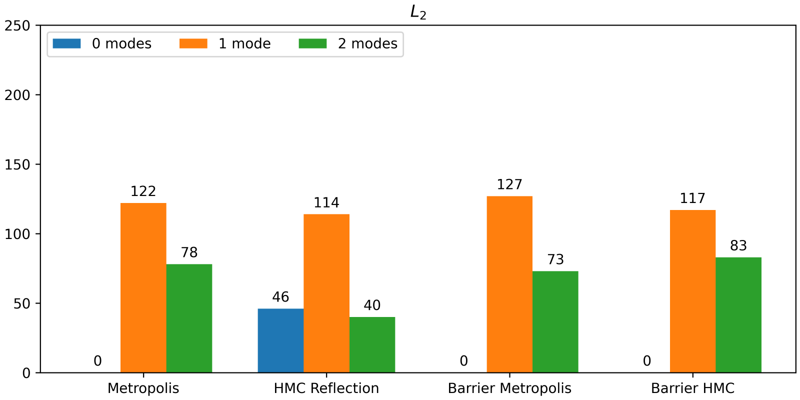
Nested Sampling with and without Barriers - Evidence Estimates



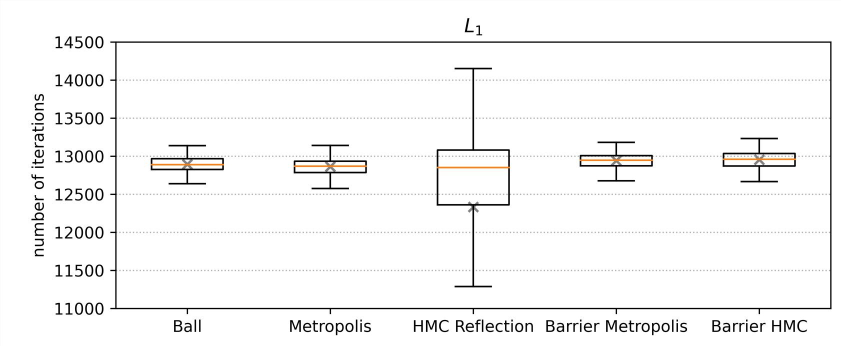
Nested Sampling with and without Barriers - Posterior Modes Found



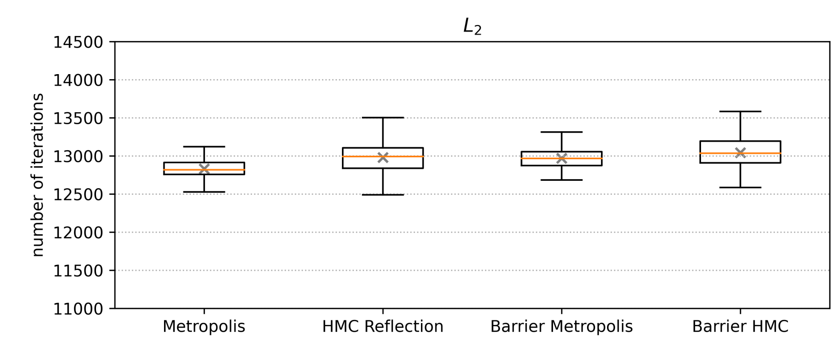
Nested Sampling with and without Barriers - Posterior Modes Found



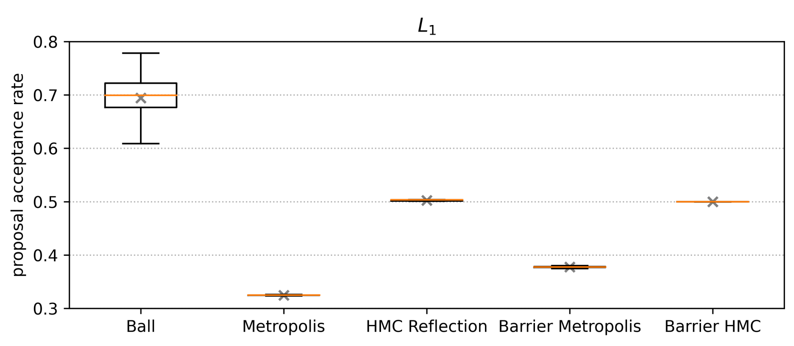
Nested Sampling with and without Barriers - Iterations



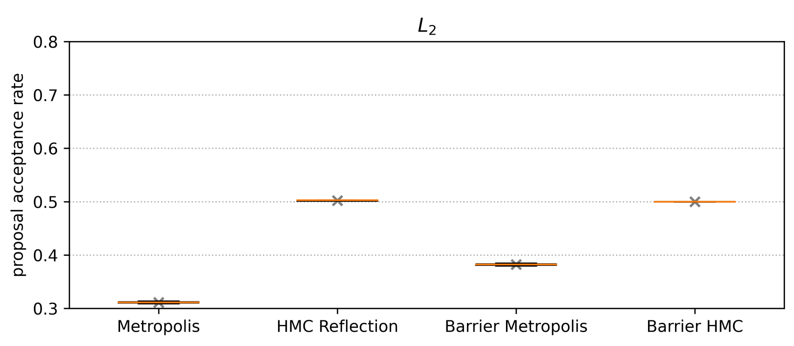
Nested Sampling with and without Barriers - Iterations



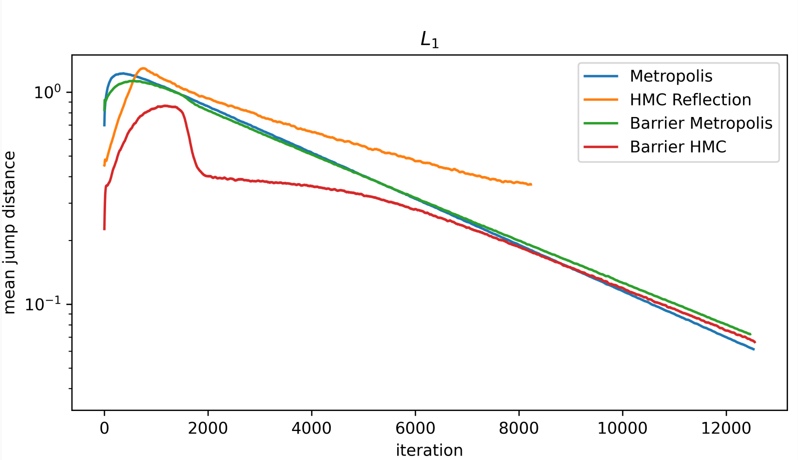
Nested Sampling with and without Barriers - Acceptance Rate



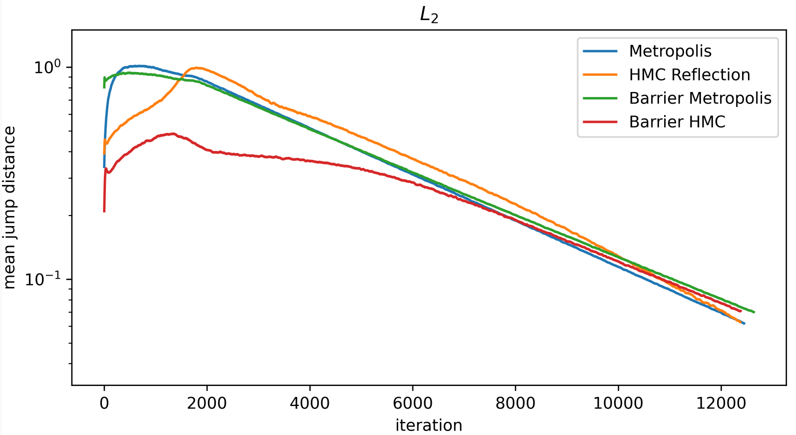
Nested Sampling with and without Barriers - Acceptance Rate



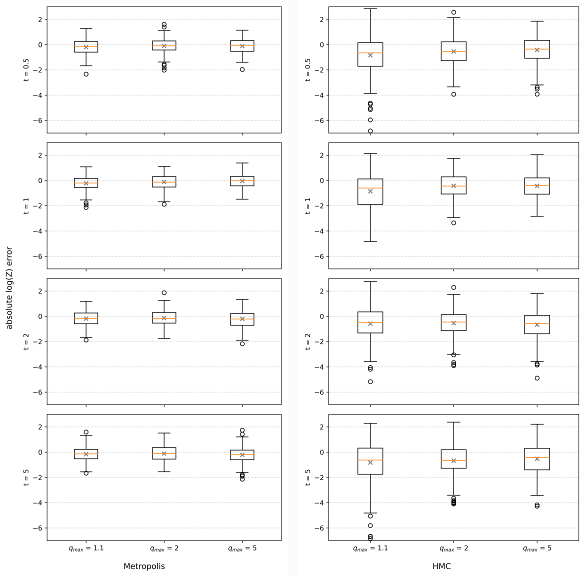
Nested Sampling with and without Barriers - Jump Distance



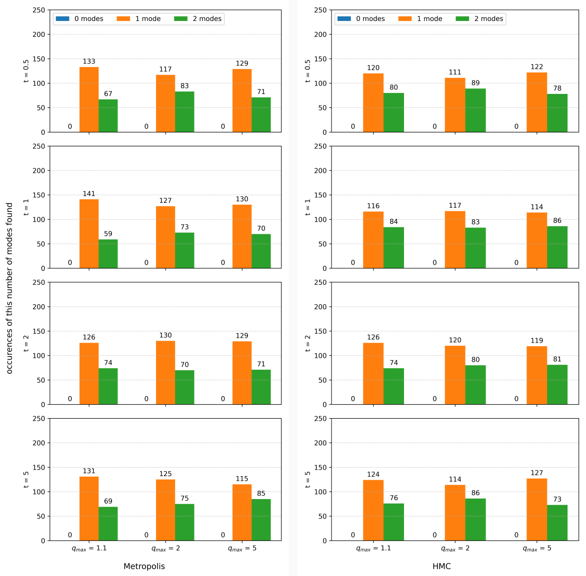
Nested Sampling with and without Barriers - Jump Distance



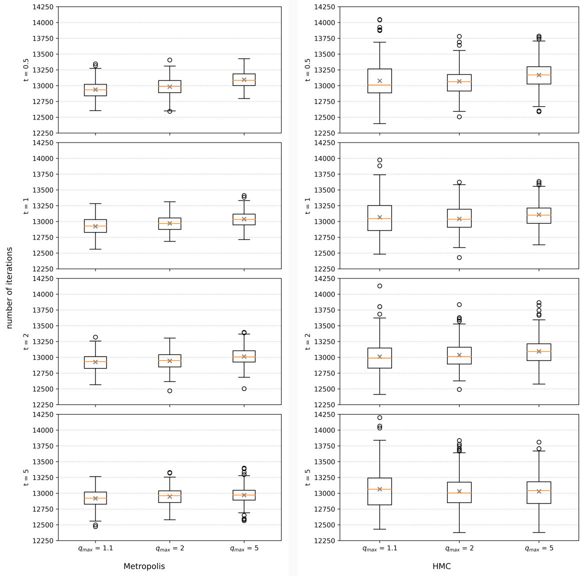
Influence of Parameters - Evidence Estimates



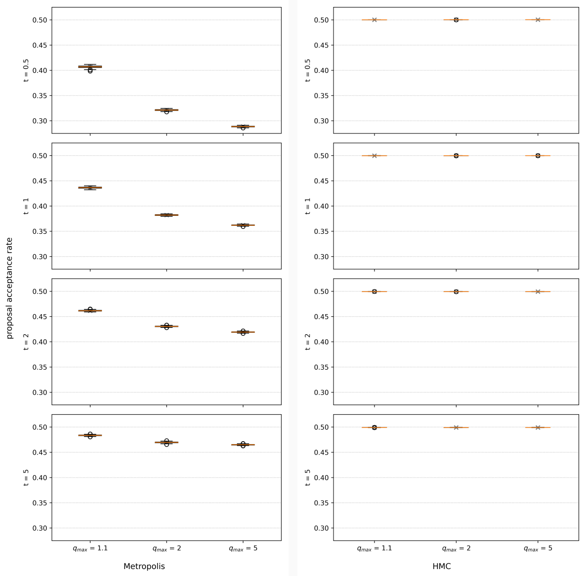
Influence of Parameters - Posterior Modes Found



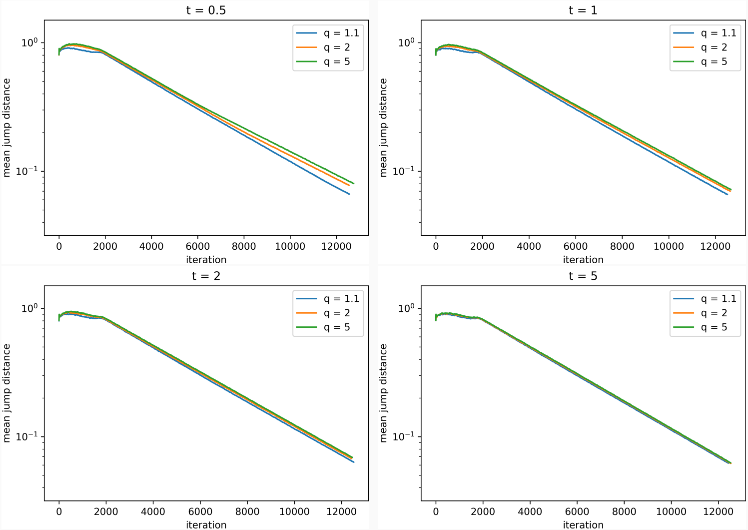
Influence of Parameters - Iteration



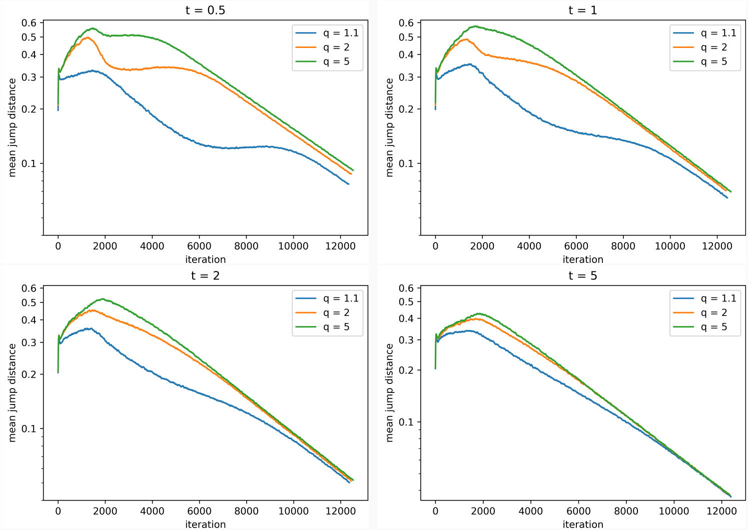
Influence of Parameters - Acceptance Rate



Influence of Parameters - Jump Distance Metropolis



Influence of Parameters - Jump Distance HMC



Conclusions

Conclusions

- Theoretical success, results are okay (but limited in scope).

Conclusions

- Theoretical success, results are okay (but limited in scope).
- Using Metropolis: Higher acceptance rates.

Conclusions

- Theoretical success, results are okay (but limited in scope).
- Using Metropolis: Higher acceptance rates.
- Using HMC: Lower jump distances early, higher jump distances late.

Conclusions

- Theoretical success, results are okay (but limited in scope).
- Using Metropolis: Higher acceptance rates.
- Using HMC: Lower jump distances early, higher jump distances late.
- Implementing more advanced Nested Sampling augmentations could be interesting.

The End

References

[1] John Skilling. "Nested sampling for general Bayesian computation". In: *Bayesian Analysis* 1.4 (2006), pp. 833–859. DOI: 10.1214/06-BA127.

[2] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[3] Jun S. Liu. "The Collapsed Gibbs Sampler in Bayesian Computations with Applications to a Gene Regulation Problem". In: *Journal of the American Statistical Association* 89.427 (1994), pp. 958–966. DOI: 10.1080/01621459.1994.10476829.

[4] Michael Betancourt et al. "Nested Sampling with Constrained Hamiltonian Monte Carlo". In: *AIP Conference Proceedings*. AIP, 2011. DOI: 10.1063/1.3573613.