MALA and Barker's Proposal: Gradient-Based MCMC Methods

- Background: From RWM to gradient-based methods
- Langevin dynamics and discretization
- Metropolis-Adjusted Langevin Algorithm (MALA)
- Optimal scaling theory
- Barker's Proposal: An alternative approach
- Comparison and practical considerations

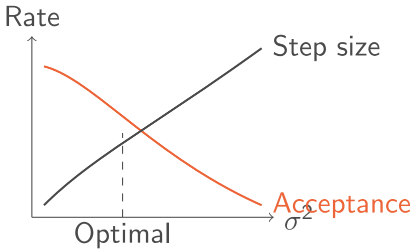
Random Walk Metropolis: The Challenge

Random Walk Metropolis (RWM):

$$q^* = q + \sigma W, \quad W \sim N(0, I_d)$$

Fundamental Trade-off:

- Large σ: Low acceptance
- Small σ: Slow exploration
- Optimal: $\sigma = \mathcal{O}(d^{-1})$



Problem: In high dimensions, RWM becomes inefficient

- Optimal acceptance rate: 0.234
- Curse of dimensionality: step size $\propto 1/d$

From Langevin Diffusion to MALA

Continuous Langevin Diffusion:

$$dX_t = \frac{1}{2}\nabla \log \pi(X_t)dt + dB_t$$

- Has π as stationary distribution
- Gradient provides drift toward high-probability regions

Euler-Maruyama Discretization (ULA):

$$X^{(t)} = X^{(t-1)} + \frac{\epsilon}{2}\nabla \log \pi(X^{(t-1)}) + \sqrt{\epsilon}W$$

Problem π is **not** the invariant distribution of ULA!
Solution: Add Metropolis-Hastings correction $\Rightarrow$ MALA

Metropolis-Adjusted Langevin Algorithm

Algorithm 1 MALA

Input: Initial $X^{(0)}$, step size ϵ, target π, proposal q

for $t = 1, 2, \ldots$ **do**

  Propose: $X^* = X^{(t-1)} + \frac{\epsilon}{2}\nabla \log \pi(X^{(t-1)}) + \sqrt{\epsilon}W$

  Compute acceptance ratio:

$$\alpha = \min\left\{1, \frac{\pi(X^*)q(X^{(t-1)}|X^*)}{\pi(X^{(t-1)})q(X^*|X^{(t-1)})}\right\}$$

  Accept $X^{(t)} = X^*$ with probability α, else $X^{(t)} = X^{(t-1)}$

end for

Optimal Scaling Theory

Maximizing Expected Squared Jump Distance (ESJD)

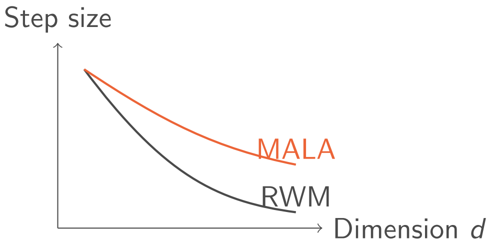
$$\mathbb{E}\left[\|X^{(t+1)} - X^{(t)}\|^2\right]$$

Dimension Scaling:

- RWM: $\sigma = \mathcal{O}(d^{-1})$
- MALA: $\sigma = \mathcal{O}(d^{-1/3})$

Optimal Acceptance:

- RWM: 0.234
- MALA: 0.574



Implication: MALA maintains larger step sizes in high dimensions

- Better exploration efficiency
- Faster convergence to target distribution
- **Catch** - requires gradient computation

Local-Balanced Proposals

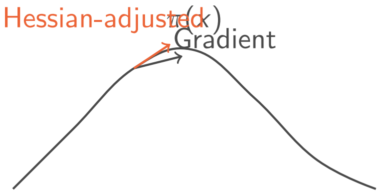
General Framework: Use local information about π

First-order (MALA):

$$X^* = X^{(t-1)} + \frac{\epsilon}{2}\nabla \log \pi(X^{(t-1)}) + \sqrt{\epsilon}W$$

Second-order:

$$X^* = X^{(t-1)} + \frac{\epsilon}{2}[\nabla^2 \log \pi(X^{(t-1)})]^{-1}\nabla \log \pi(X^{(t-1)}) + \sqrt{\epsilon}W$$



Higher-order methods better approximate local geometry

Barker's Proposal: An Alternative Approach

Key Idea: Use gradient to stochastically bias proposal direction

Proposal Density: $Q_B(x, dy) = \frac{2}{1+e^{-\nabla \log \pi(x)^T (y-x)}} K(x, dy)$
where $K(x, dy)$ is a base kernel (e.g., Gaussian)

Algorithm 2 1D case with Gaussian kernel

Sample $Z \sim N(0, \sigma^2)$
Calculate $p(x, z) = 1/(1 + \exp(-Z^T \nabla \log \pi(x)))$:
Set $b(x, z) = 1$ with probability $p(x, z)$, else $b(x, z) = -1$
Propose $Y = x + b(x, z)Z$
Apply Metropolis-Hastings acceptance

MALA vs Barker's Proposal

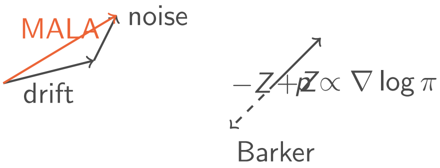
Both use gradient information, but differently

MALA:

- Deterministic drift
- $X^* = X + \frac{\epsilon}{2}\nabla \log \pi + \text{noise}$
- Gradient always adds to proposal
- Well-studied optimal scaling
- Proven efficiency in high dimensions

Barker:

- Stochastic direction choice
- Probability depends on gradient
- May flip proposal direction
- More recent theoretical development
- Potentially better for certain targets



Summary and Practical Considerations

Key Takeaways:

- Gradient information dramatically improves MCMC efficiency
- MALA: Proven workhorse with $O(d^{-1/3})$ scaling
- Barker: Promising alternative with different mixing properties
- Both methods correct discretization bias via Metropolis step

When to use which?

Choose MALA when:

- High-dimensional problems
- Gradients are cheap
- Well-conditioned targets
- Need proven reliability

Consider Barker when:

- Exploring alternatives
- Specific target structure
- Research applications
- Robustness needed

Both methods: Major improvements over RWM!