Exploiting Symmetries to Construct Efficient MCMC Algorithms With an Application to SLAM

Roshan Shariff, András György, Csaba Szepesvári {roshan.shariff, gyorgy, szepesva}@ualberta.ca

UNIVERSITY OF ALBERTA
DEPARTMENT OF COMPUTING SCIENCE

Overview

Goal: Design efficient sampling methods

- Studied in CS, operations research ("Monte Carlo"), statistics, etc.
- Markov Chain Monte Carlo: a distribution-independent method
  - Gibbs sampling, rejection sampling, slice sampling, etc.
  - The Metropolis-Hastings (MH) algorithm

Here: Metropolis-Hastings moves based on group actions

- can take advantage of (approximate) symmetries
- a powerful, elegant, convenient, efficient family of algorithms

Markov Chain Monte Carlo (MCMC)

To sample from probability distribution $p(x)\,dx$ using MCMC:

- Random walk $X_0, X_1, X_2, \dots$
- $X_0 \sim p_0(x)\,dx$ *(initial distribution)*
- $X_{i+1} \sim k(x' \mid X_i)\,dx'$ *(transition kernel)*

Choose transition kernel to make X_i converge in distribution to P

Metropolis-Hastings (MH)

MCMC transition kernel that adds a *rejection* step to a given *proposal kernel* q:

1. Propose $X'_{i+1} \sim q(x' \mid X_i)\,dx'$
2. Accept $X_{i+1} := X'_{i+1}$ with probability $\alpha(X_i, X'_{i+1})$
3. Reject otherwise: $X_{i+1} := X_i$

This is the "textbook" algorithm, and with acceptance probability

$$\alpha(x, x') := \min\left\{1, \frac{p(x')\,q(x \mid x')}{p(x)\,q(x' \mid x)}\right\}$$

produces a *reversible* Markov transition kernel k:

$$p(x)\,k(x' \mid x)\,dx\,dx' = p(x')\,k(x \mid x')\,dx\,dx'$$

A More General MH Kernel

Unlike the textbook case, target distribution $P(dx)$ and proposal kernel $Q(dx' \mid x)$ may not have densities w.r.t. a common reference measure

Theorem (Tierney, 1998, Theorem 2)
For any $Q(dx' \mid x)$ and $P(dx)$, with an appropriate α, the MH algorithm gives a Markov kernel K satisfying

$$P(dx)\,K(dx' \mid x) = P(dx')\,K(dx \mid x')$$

An appropriate acceptance probability:

- Define $\mu(dx, dx') = P(dx)\,Q(dx' \mid x)$ and $\mu^T(dx, dx') = \mu(dx', dx)$
- Find $R \subset \mathcal{X} \times \mathcal{X}$ on which μ and μ^T are mutually absolutely continuous, and outside which they are mutually singular
- Find the Radon-Nikodym derivative

$$r(x, x') = \frac{d\mu_R}{d\mu_R^T}(x, x')$$

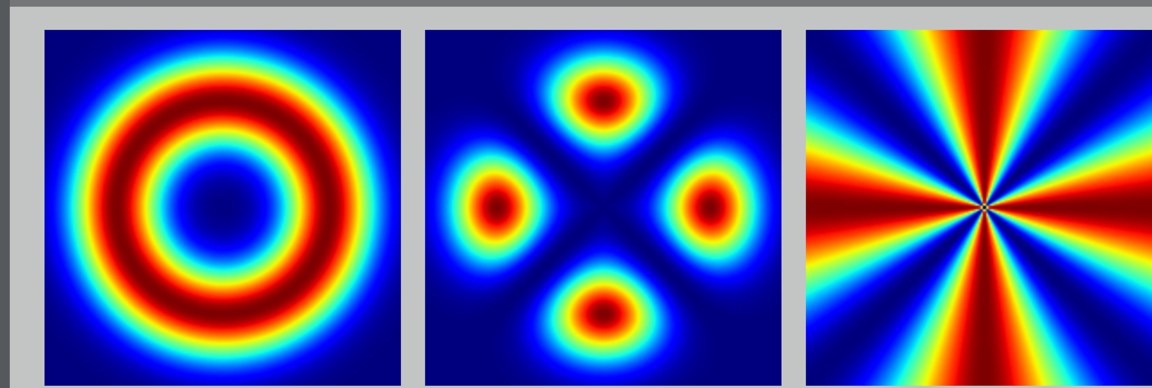
such that $0 < r(x, x') < \infty$ and $r(x, x') = 1/r(x', x)$ for all $x, y \in \mathcal{X}$

- Use the acceptance probability

$$\alpha(x, x') = \begin{cases} \min\{1, r(x, x')\}, & \text{if } (x, x') \in R \\ 0 & \text{otherwise} \end{cases}$$

… a powerful result, but not straightforward to use

Motivating Example



Sampling from a probability density $p(x, y)$ (center) on $\mathbb{R}^2 \setminus \{(0,0)\}$ having factors p_1 (left) and p_2 (right)

MH with normally distributed proposal?

- $q(x', y' \mid x, y) = \mathcal{N}((x, y), (\sigma_x, \sigma_y))$
- Hard to choose σ_x, σ_y: if too small, chain doesn't move between modes; otherwise, high rejection rate

Gibbs sampling?

- At each step, modify either x or y with equal probability: either $x' \sim p(\cdot\,, y)$ or $y' \sim p(x, \cdot)$
- Doesn't move between $\pm X$ modes and $\pm Y$ modes
- Transition kernel concentrated on X, Y axes: no density
- Only works when moving along linear subspaces

Change of variables?

- Use polar coordinates (r, θ)
- p_1 depends only on r, and p_2 only on θ
- Gibbs sampling gives fast and no-reject MCMC kernel

What if re-parametrization is non-obvious? Can we directly use *symmetries*? (p_1 symmetric under rotation, p_2 under scaling)

Groups and Actions

A group G is a set of elements, with

- A binary operation, written $a \cdot b$ or ab with $a, b, ab \in G$
- A unit element e, with $eg = ge = g$ for all $g \in G$
- An inverse g^{-1} for every g, with $g^{-1}g = gg^{-1} = e$

It *acts* on (state) space $\mathcal{X}$ with action $T : G \times \mathcal{X} \to \mathcal{X}$ if:

- $T(e, x) = x$ for $x \in \mathcal{X}$
- $T(gh, x) = T(g(T(h, x)))$ for $g, h \in G$, $x \in \mathcal{X}$

Then G elements are *invertible transformations* of $\mathcal{X}$

- unit is identity transformation: $ex := T(e, x) = x$
- group operation is composition: $ghx := (gh)x = g(hx)$

Invariant Measures

Theorem (Haar)
Every locally compact topological group has a unique (up to positive scale factor) left-invariant measure μ:

$$\mu(gA) := \mu(\{ga \mid a \in A\}) = \mu(A), \quad g \in G, A \subset G$$

Examples:

- Lebesgue measure on $\mathbb{R}^n$ is invariant under vector addition
- $\mu(dx) = x^{-1}\,dx$ is invariant on $\mathbb{R}_+^\times$
- Rotations of $\mathbb{R}^2$ can be represented as an angle in $[0, 2\pi)$; the Lebesgue measure is invariant

μ is a *χ-relatively invariant* measure if $\mu(gA) = \chi(g)\mu(A)$ for $\chi : G \to \mathbb{R}_+$: then $\chi(e) = 1$ and $\chi(gh) = \chi(g)\chi(h)$

- Example: Haar measure μ under *right* multiplication: $\mu(Ag) = \Delta_r^G(g)\,\mu(A)$ with *right modulus* $\Delta_r^G : G \to \mathbb{R}_+$

Problem Statement

Sample from probability distribution P on $\mathcal{X}$ having density p w.r.t. measure λ: $P(dx) = p(x)\,\lambda(dx)$. For $i = 1, \dots, m$,

- group G_i acts on $\mathcal{X}$
- Haar measure μ_i on G_i
- proposal kernel on G_i: $q_i(g \mid x)\,\mu_i(dg)$ ($g \in G_i$, $x \in \mathcal{X}$)
- λ is $\chi_i : G_i \to \mathbb{R}_+$-relatively invariant under G_i

MH With Group Actions

Given, for $i = 1, \dots, m$, mixture coefficients $a(i, x)$ with $\sum_i a(i \mid x) = 1$ (for all $x \in \mathcal{X}$):

1. Sample $i \sim a(\cdot \mid x)$ and $g \sim q_i(g \mid x)\,\mu_i(dg)$
2. Calculate

$$\alpha(i, x, g) := \frac{\chi_i(g)\,a(i \mid gx)\,p(gx)\,\bar{q}_i(g^{-1} \mid gx)}{\Delta_r^{G_i}(g)\,a(i \mid x)\,p(x)\,\bar{q}_i(g \mid x)}$$

3. Accept $x' := gx$ with probability $\min\{1, \alpha(i, x, g)\}$
4. Reject otherwise: $x' = x$

where

- $\bar{q}_i(g \mid x) := \int_{G_{i,x}} q_i(gh \mid x)\,\beta_{i,x}(dh)$
- $G_{i,x} := \{g \in G_i \mid gx = x\}$ is the *stabilizer subgroup* of G_i at $x \in \mathcal{X}$
- $\beta_{i,x}$ is a Haar measure on $G_{i,x}$ with $\beta_{i,x}(G_{i,x}) = 1$

Exploiting Symmetries

Suppose $p(x) \propto \prod_{i=1}^k p_i(x)$ for $x \in \mathcal{X}$

- Groups H_i ($i = 1, \dots, k$) act on $\mathcal{X}$
- $p_i(hx) = p_i(x)$ for $h \in H_i$

Then each group H_i is a *symmetry* of factor p_i; $\alpha(i, x, g)$ simplifies:

- If $G_i \subset H_j$ for some j, then $p_j(gx) = p_j(x)$ for $g \in G_i$ and the p_j factor cancels
- If $q_i(g \mid x) \propto \chi_i(g)\,p_i(gx)$ then the χ_i, $\Delta_r^{G_i}$, and p_i factors cancel

Ideally, only non-symmetric terms contribute to α

- Lower rejection rate
- Faster to compute

For approximate symmetries, improves convergence but not computation

Simultaneous Localization and Mapping (SLAM)

A robot

- moves (with noisy control mechanisms)
- observes landmarks (with noisy sensors)

and wants to know (for $t = 0, \dots, T$ and $i = 1, \dots, N$)

- where it is: its trajectory $X := (X_0, \dots, X_T)$
- where the landmarks are: the map $Y := (Y_1, \dots, Y_N)$

assuming it has

- observations $Z := (Z_t^i)_{t,i}$ of landmark i at time t
- control models $p_{X_t \mid X_{t-1}, Z_{<t}}$
- sensor models $p_{Z_t^i \mid X_t, Y_i}$

Bayesian posterior (under natural independence assumptions):

$$p_{X,Y \mid Z}(x, y, z) = c(z) \cdot p_Y(y) \cdot \prod_{t=0}^{T} p_{X_t \mid X_{t-1}, Z_{<t}}(\dots) \prod_{i=1}^{N} p_{Z_t^i \mid X_t, Y_i}(\dots)$$

Symmetries of SLAM

In two dimensions, the environment is invariant under the *special Euclidean group* $\mathrm{SE}(2)$ of rigid transformations of $\mathbb{R}^2$, which acts by a rotation followed by a translation:

- Landmark locations $Y_i \in \mathbb{R}^2$
- Robot pose $X_t \in \mathrm{SE}(2)$ transforms body coordinates to global
- $p_{X_t \mid X_{t-1}, Z_{<t}}$ depends on X_{t-1} and X_t only through their *relative movement*: $X_{t-1}^{-1} X_t = (gX_{t-1})^{-1}(gX_t)$
- $p_{Z_t^i \mid X_t, Y_i}$ depends on X_t and Y_i only through their *relative position*: $X_t{}^{-1} Y_i = (gX_t)^{-1}(gY_i)$
- p_Y is uniform (improper) map prior: invariant under $\mathrm{SE}(2)$

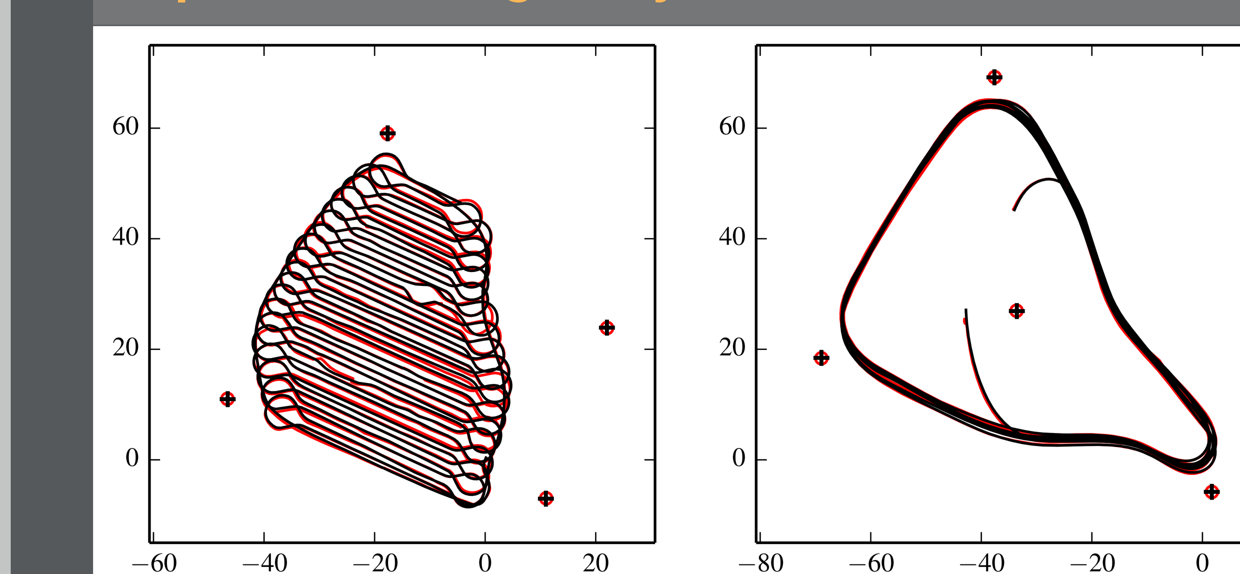
The MCMC-SLAM Algorithm

Without loss of generality, $X_0 = e$: robot starts at origin. Then state space is $U := \mathrm{SE}(2)^T \times \mathbb{R}^{2N}$, state $u = (x_t, y_i)_{t,i}$.

- Fix $b : \{1, \dots, N\} \to \{0, \dots, T\}$: landmark i is "anchored" to time $b(i)$
- Groups G_t ($t = 1, \dots, T$) act by transforming x_s (if $s \geq t$)
- Groups G^i ($i = 1, \dots, N$) act by transforming y_i
- MH algorithm with $m = T + N$ proposal groups
- Mixture coefficient inversely proportional to proposal likelihood: improve "bad" components first

See article in proceedings for details…

Experiments: Range-Only SLAM



Plaza 1 (left): 1.9 km trajectory, 9,657 steps, 3,529 observations.
Plaza 2 (right): 1.3 km trajectory, 4,091 steps, 1,816 observations.

Table: Comparison of Trajectory RMS Errors *(with running times)*.

Algorithm	Plaza 1	Plaza 2
Spectral	0.79 m *(0.73 s)*	0.35 m *(0.51 s)*
Spectral + Opt.	0.69 m *(9265 s)*	0.30 m *(2357 s)*
MCMC (10+1000)	0.32 ± 0.02 m *(13.8 s)*	0.54 ± 0.06 m *(2.8 s)*
MCMC (100+10000)	0.33 ± 0.04 m *(131 s)*	0.36 ± 0.03 m *(28 s)*

Conclusions and Future Work

Using group actions is a powerful extension to the standard Metropolis-Hastings algorithm. It can take advantage of the (possibly approximate) factor and symmetry structure of the target distribution, speeding up convergence to the steady state and improving computational efficiency.

Future work includes SLAM with data association and, generally, bringing more such "algebraic" techniques to MCMC.