Motivating Example

Sampling from a probability density $p(x, y)$ (center) on $\mathbb{R}^2$ (1000)!
Having factors $p_1$ (left) and $p_2$ (right).

MH with normally distributed proposal distribution:
- $q(x', y' | x, y) \sim N(x, y, \sigma)$
- Hard to choose $\sigma$ (e.g., 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 $x' = p_1(x, y)$ or $y' = p_2(x, y)$
- Doesn’t move between $x$ modes and $y$ modes
- Transition kernel concentrated on $X_1, Y$ axes: no density
- Only works when moving along linear subspaces

Change of variables:
- Use polar coordinates $(r, \theta)$
- $p$ depends only on $r$, and $p_1$ only on $\theta$
- Gibbs sampling gives fast and no-reject MCMC kernel

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

Metropolis-Hastings (MH)

MCMC transition kernel that adds a rejection step to a given proposal kernel $q$:
1. Propose $X_1' \sim q(x' | x)$
2. Accept $X_1' \Rightarrow X_1$, with probability $a(X_1, X_1')$
3. Reject otherwise: $X_1 \Rightarrow X_1$

This is the "textbook" algorithm, and with acceptance probability $a(x, x') = \min(1, \frac{p(x')q(x | x')}{p(x)q(x' | x)})$ produces a reversible Markov transition kernel $K$:

$$K(x, y) = \min(1, \frac{p(x')q(x | x')}{p(x)q(x' | x)}) x' = K(x, y)$$

A More General MH Kernel

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

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

$$K(dx) \leq \alpha K(dx') K(dx)$$

An appropriate acceptance probability:
- Define $\mu(dx, dx') = Q(dx)Q(dx')/(dx)(dx') = \mu(dx', dx)$
- Find $R \subset X \times X$ on which $\mu$ and $\rho$ are mutually absolutely continuous, and outside which $\mu(dx, dx') = 0$
- Find the Radon-Nikodym derivative

$$\rho(dx | dx') = \frac{d\rho}{d\mu}(x, x') \begin{cases} 0 < r(x, x') &= \rho(dx | dx') & \text{and } r(x, x') = 1(r(x, x')) & \text{for all } x, y \in X \\ \text{Use the acceptance probability} a(x, x') = \min(1, r(x, x')) & \delta_{x, y} & \text{if } (x, x') \in R \\ 0 & \text{otherwise} \end{cases}$$

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 $X_t, X_{t+1}, X_{t+2}$...
- $X_t + u \cdot dx$ (initial distribution)
- $X_t \Rightarrow X_t'$ (transition kernel)
Choose transition kernel to make $X_t$ converge in distribution to $P$

Exploiting Symmetries to Construct Efficient MCMC Algorithms
With an Application to SLAM
Roshan Shariff, András György, Csaba Szepesvári
{roshan.shariff, gyorgy, szepesva}@ualberta.ca

Groups and Actions
A group $G$ is a set of elements, with:
- A binary operation, written $a \cdot b$ or $ab$ with $a, b \in G$
- 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 $X$ with action $\cdot G \times X \rightarrow X$ if:
- $T \cdot x = x$ for $x \in X$
- $T \cdot gh \cdot x = T \cdot gh(\cdot x)$ for $g, h \in G, x \in X$
Then $G$ elements are invertible transformations of $X$
- unit is identity transformation: $e \cdot X = X$, $x \cdot e = x$
- group operation is composition: $gh \cdot x = (gh)(\cdot x)$

Simultaneous Localization and Mapping (SLAM)
A robot
- moves (with noisy control mechanisms)
- observes landmarks (with noisy sensors)

And wants to know (for $t = 0, \ldots, T$ and $i = 1, \ld\ld, N$)
- where it is: its trajectory $X = (X_0, X_1, \ld, X_T)$
- where the landmarks are: the map $Y = (Y_0, Y_1, \ld)$

Assuming it has
- observations $Z_i : (Z_i)$ of landmark $i$ at time $t$
- control models $p_{X_{i+1}}(x_{i+1} | x_i)$
- sensor models $p_{Z_i | X_i}$
Bayesian posterior (under natural independence assumptions)

$$p_{X_0, X_1 \ld X_T, Y_0 \ld Y_N | Z_1 \ld Z_N} = \frac{p(z_i | x_{i+1}) p_{X_{i+1}}(x_{i+1} | x_i) p_{X_0} p_{X_1} \ld p_{X_T} p_{Y_0} \ld p_{Y_N}}{p(z_i \ld Z_N)}$$

The MCMC-SLAM Algorithm
Without loss of generality, $X_0 \sim c$: robot starts at origin. Then state space is $U = \{x(t) | x(0) = (0, 0, 0, 0)\}$

- $p(0, 1, \ld, T) = \{x(t) \in \mathbb{R}^3 | x(t) \}$: landmark is "anchored" to time $t$
- Groups $G_i (t = 1, \ld, T)$ act by transforming $x_i$ (if $i = t$)
- $G_0 (t = 1, \ld, T)$ act by transforming $x_0$

MH algorithm with $T = N$ proposal groups
Mixture coefficient inversely proportional to proposal likelihood:
- Improves "bad" components first

Experiments: Range-Only SLAM
Plaza 1 (left): 1.2 km trajectory, 9,657 steps, 3,529 observations.
Plaza 2 (right): 1.34 km trajectory, 4,091 steps, 1,836 observations.

Table: Comparison of Trajectory RMS Error (with running time)

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Plaza 1</th>
<th>Plaza 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral</td>
<td>0.79 m (0.73 s)</td>
<td>0.35 m (0.52 s)</td>
</tr>
<tr>
<td>Spectral + Opt.</td>
<td>0.61 m (0.26 s)</td>
<td>0.30 m (0.23 s)</td>
</tr>
<tr>
<td>MCMC (10-D)</td>
<td>0.32 ± 0.04 (13.8 s)</td>
<td>0.54 ± 0.05 (2.8 s)</td>
</tr>
<tr>
<td>MCMC (100-D)</td>
<td>0.31 ± 0.04 (13.2 s)</td>
<td>0.36 ± 0.01 (2.2 s)</td>
</tr>
</tbody>
</table>

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 structures 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.