An adaptive Metropolis Algorithm
Random walk Metropolis-Hasting, fundamental the choice of an effective proposal distribution:

- Run a certain number of iterations and tune the covariance matrix
- adaptation performed only at the time of recurrence to a specific atom (Gilks et al. 1998)

Adaptive Metropolis algorithms:

- AP algorithm (Haario et al. 1999), fixed number of previous states
- AM algorithm (Haario et al. 2001), all previous states. Idea: update the proposal distribution by using the knowledge we have so far acquired about the target distribution.

Note: no Markov chain, AM is a discrete-time stochastic process
Let $\pi(x)$ be the target distribution.

Assumptions:

- compact support $S \subset \mathbb{R}^d$ (can it be relaxed?), $\pi(x) \equiv 0$ outside $S$.
- $\pi(x)$ is bounded: for some $M < \infty$, $\pi(x) \leq M$ for $x \in S$
Algorithm

At time $t - 1$ we have sampled the states $X_0, ..., X_{t-1}$

$\pi(x)$ the target distribution.

$q_t(\cdot|X_0, ..., X_{t-1}) = N(X_{t-1}, C_t(X_0, ..., X_{t-1}))$ proposal

- candidate $Y$ sampled from $q_t(\cdot|X_0, ..., X_{t-1})$

- $Y$ accepted with probability

$$\alpha = \min \left(1, \frac{\pi(Y)}{\pi(X_{t-1})}\right)$$

  - depends only on $Y$ previous state $X_{t-1}$
  - cannot be justified with symmetry.

- In addition, only jumps in $S$ are accepted
Covariance matrix:

\[ C_t = C_0 \quad \text{for} \quad t \leq t_0 \]
\[ C_t = s_d \text{cov}(X_0, \ldots, X_{t-1}) + s_d \epsilon I_d \quad \text{for} \quad t > t_0 \]

- \( C_0 \) prior knowledge
- \( s_d \) depends on dimension \( d \) (\( s_d = 2.4^2/d \))
- \( \epsilon > 0 \) very small compared to \( S \), to have \( C_t \) positive definite.
Recursion formula for $t > t_0$:

$$
\bar{X}_t = \frac{t \bar{X}_{t-1} + X_t}{t + 1}
$$

$$
C_{t+1} = \frac{t - 1}{t} C_t + \frac{s_d}{t} (t \bar{X}_{t-1} \bar{X}_{t-1}^T - (t + 1) \bar{X}_t \bar{X}_t^T + X_t X_t^T + \epsilon \mathbb{I}_d).
$$

- $C_0$ initial covariance matrix, express prior knowledge
- $t_0$ express our confidence in $C_0$
- $s_d$ depends on dimension $d$ ($sd = 2.4/\sqrt(d)$)
- $\epsilon > 0$ very small compared to $S$ (strictly larger than 0 to prove ergodicity)
Properties

State space \((S, \mathcal{B}, m)\).

\(C\) positive definite covariance matrix on \(\mathbb{R}^d\). \(N_C\) density of the mean zero Gaussian distribution with covariance \(C\).

The Gaussian proposal transition probability corresponding to the covariance \(C\) satisfies \(Q_C(x; A) = \int_A N_C(y - x)dy\) where \(A \in \mathbb{R}\) is a Borel set.

\(Q_C(x; A)\) is m-symmetric.

Transition probability \(M_C\):

\[
M_c(x; A) = \int_A N_C(y - x) \min \left(1, \frac{\pi(Y)}{\pi(X_{t-1})}\right) m(dy) + \chi_A \int_{\mathbb{R}^d} N_C(y - x)(1 - \min \left(1, \frac{\pi(Y)}{\pi(X_{t-1})}\right)) m(dy)
\]

for \(A \in \mathcal{B}(S)\)
With the assumption above, initial covariance $C_0$ and $\epsilon > 0$. Define $C_n$ using the recursive formula for $n \geq 1$. For a given initial distribution $\mu_0$ the AM chain is a stochastic chain on $S$ defined by the sequence $\{K_n\}$ of generalized transition probabilities, where

$$K_n(x_0, \ldots, x_{n-1}; A) = M_{C_n(x_0, \ldots, x_{n-1})}(x_{n-1}, A)$$

for all $n \geq 1$, $x_i \in S$, and for subsets $A \in \mathcal{B}(S)$. 
Properties

Theorem

With the assumptions and definitions above.

Define the AM chain \( \{X_n\} \) by the generalized transition probabilities \( K_n(x_0, \ldots, x_{n-1}; A) \). Then the AM chain simulates properly the target distribution \( \pi \): for any bounded and measurable function \( f : S \in \mathbb{R} \), the equality

\[
\lim_{n \to \infty} \frac{1}{n+1}(f(X_0) + f(X_1) + \ldots + f(X_n)) = \int_S f(x) \pi(dx)
\]

holds almost surely
Corollary

With the assumptions and definitions above.

The covariance $C_t$ almost surely stabilizes during the algorithm. In fact, as $t \to \infty$ the covariance $C_t$ converges to $s_d \text{cov}(\pi) + \epsilon \mathbb{I}_d$ where $\text{cov}(\pi)$ denotes the covariance of the target distribution $\pi$.
Switch to R

Switch to R
Alternatives:

- greedy start, during initial period updating proposal using only the accepted states
- update the covariance every $n_0 > 1$ steps, using ALL history
- update using $\lfloor n/2 \rfloor$ previous iterations
Questions

- how restrictive the assumptions are?
- what about Metropolis-within-Gibbs, would it still work?
- is it convenient in low dimensionality?
- is it convenient for "nonlinear" shapes?