Effect of increasing the energy gap between the two lowest energy states on the mixing time of the Metropolis algorithm

Apurv Nakade and Somenath Biswas*

Department of Computer Science and Engineering
IIT Kanpur, Kanpur 208016, India

Abstract. In order to understand what makes natural proteins fold rapidly, Sali, Shakhnovich and Karplus [SSK94a], [SSK94b] had used the Metropolis algorithm to search for the minimum energy conformations of chains of beads in the lattice model of protein folding. Based on their computational experiments, they concluded that the Metropolis algorithm would find the minimum energy conformation of a chain of beads within an acceptable time scale if and only if there is a large gap between the energies of the minimum energy conformation and that of the second minimum. Clote [Cl99] attempted to support this conclusion by a proof that the mixing time of the underlying Markov chain would decrease as the gap in energies of the minimum energy conformation and that of the second minimum increased. He was able to show that an *upper bound* on the mixing time does indeed decrease as the energy gap increases. We show in this paper that the mixing time *itself*, however, is a nondecreasing function of the value of the energy gap. Therefore, our result contradicts what Clote had attempted to prove.

Keywords: Metropolis algorithm, Markov chain mixing time, protein folding, randomized algorithms.

1 Introduction

One of the most outstanding open issues in biology is the rapid folding of proteins: in spite of the existence of exponentially many possible configurations, a protein, after its formation as a chain of arbitrary shape, manages to fold, i.e., to reach its unique lowest free-energy state, very quickly.[1] In order to understand the factors responsible for the rapid folding of proteins, Sali, Shakhnovich, and Karplus [SSK94a], [SSK94b] considered the *lattice model of protein folding*. In this model, one considers a chain of n beads, $a_1, a_2, \ldots, a_n$, as an idealization of the polymeric chain of amino acids that defines the primary sequence of a protein. A *conformation* of the chain of beads is a placement of the beads in the

* E-mail address for correspondence on this work: sb@cse.iitk.ac.in

[1] This phenomenon is known as Levinthal's paradox in the literature.

three-dimensional lattice, a placement that is required to satisfy two constraints: that beads a_i and $a_{i+1}, 1 \leq i < n$, will occupy adjacent lattice positions, and that no two distinct beads will occupy the same lattice position. (In other words, the chain is folded in such a manner that each bead occupies a lattice position and this folding is, so called, self-avoiding). Sali et. al. defines the total energy E of a conformation to be $E = \Sigma_{i<j}\Delta(r_i, r_j)B_{ij}$, where r_i's denote positions of the ith bead a_i in the lattice, $\Delta(r_i, r_j) = 1$ if $|i - j| > 1$ and r_i, r_j are two adjacent positions in the lattice, $\Delta(r_i, r_j)$ is 0 otherwise. The term B_{ij} is the attractive force between the bead pair a_i, a_j, and is defined suitably to model the attractive forces between pairs of amino acids. The chain of beads is considered to be *folded* if it is in a conformation with minimum total energy.

To study the kinetics of folding, Sali, Shakhnovich and Karplus used the Metropolis algorithm to locate the minimum energy conformation, after identifying a suitable notion of neighbourhood amongst the set of conformations. A search of the space of conformations through the Metropolis algorithm is indeed an acceptable approximation for the molecular dynamics reponsible for folding, as explained in [SK91]. They considered 200 instances, each with 27 beads, with varying $B_{i,j}$'s. They found that the Metropolis algorithm was able to locate the global minimum within an acceptable timescale in 30 of the 200 instances. On examining the set of instances, Sali, Shakhnovich and Karplus found that what distinguished the 30 rapidly folding instances from the rest was that all those 30 instances had *pronounced gobal minima*, which the rest did not. A configuration is said to have pronounced global minimum if there is a large gap between the energies of the minimum energy conformation and that of the second minimum. On the basis of this finding they made the conjecture that the *necessary and sufficient* condition for a chain to fold rapidly is that it should have a pronounced global minimum.

Clote [Cl99] attempted to provide a theoretical justification for the above conjecture by considering the effect of increasing the gap in energies of the minimum and the second minimum conformations on the mixing time of the Markov chain that the Metropolis algorithm uses. If the conjecture is true then it is reasonable to expect that the mixing time will decrease with increase in the gap. What Clote was able to show is that *an upper bound* on the mixing time does, indeed, decrease as the energy gap increases. However, as Clote himself had pointed out, the result does not prove anything about the mixing time *per se*.

Our work here actually *contradicts* what Clote had set out to prove. In particular, we show that the second largest eigenvalue of the underlying Markov chain, as a function of the energy gap between the minimum and and second minimum energy configurations, is non-decreasing. As the mixing time decreases with the decrease in the second-largest eigenvalue, our result proves that increasing the gap in energies will not reduce the mixing time.

Our approach to the problem is different from that of Clote's. Clote used the notion of canonical paths [S93] to argue about an upper bound on the conductance of the underlying graph of the Markov chain, whereas we directly work

with the second largest eigenvalue. Further, our proof is not entirely algebraic, it also uses certain facts from analysis, e.g., that the eigenvalues of a matrix are continuous functions of its entries.

2 Preliminaries

To find the minimum energy conformation, the Metropolis algorithm runs a Markov chain. The state space of the chain is the set of conformations. There is a neighbourhood structure associated with the set of conformations: one conformation c_1 is a neighbour of another, c_2, if due to a local move c_2 can result from c_1. We assume that the neighbourhood structure is an undirected, connected d-regular graph.[2] Let $f(c)$ denote the energy of the conformation c. As per the Metropolis algorithm, if the Markov chain is in the state (i.e., current conformation is) c_i at a certain instant of time, and c_j is a neighbour of c_i, then the chain will move to c_j with probability $\min\{1/2d, (1/2d) \times (e^{-f(c_j)/T}/e^{-f(c_i)/T})$. ($T$ is the ambient temperature in which the modeled physical process is situated; for our purpose, it is simply a constant.) Our goal in this paper is to study how the mixing time of the chain would change, as we change the f value of the minimum energy conformation, leaving the f values of the other conformations unchanged.

In order to keep the notations simple, we study the following equivalent problem. Let G be a d-regular undirected connected graph with vertex set $\{0, 1, \ldots, n-1\}$. Associated with each vertex i, is the *value* of the vertex: V_i. We consider the Metropolis algorithm that aims to identify the vertex with the *highest* value. We assume without loss of generality that $V_0 > V_1 > \ldots > V_{n-1} \geq 0$. ($V_i$'s are proportional to the stationary probabilities of the respective i's, hence negative values for V_i's are not considered).

The Metropolis algorithm runs a Markov chain to find the vertex having the largest value; the state space of the chain is the set of vertices of G, and its transition probability matrix $P = \{p_{ij}\}$ is as defined below: (We recall that $V_0 > V_1 > \ldots > V_{n-1}$.)

$$p_{ij} = \begin{cases} 0 & \text{if } i \neq j \text{ and } i \notin Neighbour(j) \\ \frac{1}{2d}\frac{V_j}{V_i} & \text{if } j \in Neighbour(i) \text{ and } j > i \\ \frac{1}{2d} & \text{if } j \in Neighbour(i) \text{ and } j < i \\ 1 - \Sigma_{i \neq k} p_{ik} & \text{if } i = j \end{cases}$$

It is well-known (see, e.g., [S93]) that matrices such as P have n real eigenvalues in $[0, 1]$;[3] further, as the Markov chain using P is ergodic,[4] the chain has a unique stationary distribution corresponding to the unique largest eigenvalue 1.

[2] The regularity condition is not necessary, but the connectedness of the neighbourhood structure is necessary for the chain to have a stationary distribution.

[3] The self loop probability, $p_{ii} \geq 1/2$ ensures that there is no negative eigenvalue.

[4] The connectedness of the neighbourhood structure ensures irreducibility of the chain, and as $p_{ii} > 0$, the aperiodicity of the chain is guaranteed.

Instead of working with the matrix P, it is more convenient to work with another matrix N defined as

$$N = 2d(I_{n\times n} - P)$$

λ is an eigenvalue of P iff $2d(1-\lambda)$ is an eigenvalue of N. With respect to N our goal, therefore, is to show that as V_0 increases, while the other V_i's remain the same, the second smallest eigenvalue is non-increasing.

We make the substitution y for $1/V_0$ in the matrix N; let $N(y)$ denote the matrix N with this substitution, and let $\lambda(y)$ denote the second smallest eigenvalue of $N(y)$. Our goal is proved by showing that $\lambda(y)$ is a non-decreasing function in y.

Let D be the diagonal matrix $diag(V_0^{1/2}, V_1^{1/2}, \ldots, V_{n-1}^{1/2})$. Let the matrix M be defined as $M = DND^{-1}$. M is similar to N, and therefore, it has the same set of eigenvalues as N does. M is also a symmetric matrix and therefore has eigenvectors which are pairwise orthogonal.

Notations: We use the following notations in the rest of the paper:
$N(y), M(y)$: The matrices N and M with V_0 substituted as $1/y$,
$\lambda(y)$: The second smallest eigenvalue of $N(y)$, therefore, of $M(y)$,
$M_{ij}(y)$: The i,jth entry of the matrix $M(y)$.

3 Results

First, we show that there is an interval starting at 0 where $\lambda(y) > \lambda(0)$. Using this, we prove later that $\lambda(y)$ is a non-decreasing function in y.

Lemma 1. *There is a $\delta, \delta > 0$ such that for all $y, 0 < y \leq \delta, \lambda(0) < \lambda(y)$.*

Proof. We consider two cases:
Case 1: $\lambda(0) = 0$
Consider $y > 0$, and suppose that $\lambda(y) = 0$. In this case, 0 is the value of the two smallest eigenvalues of $N(y)$, therefore, of $M(y)$. As $M(y)$ is symmetric, there will exist two orthogonal eigenvectors for these two minimum eigenvalues, which in turn means that corresponding to these two 0 eigenvalues, $N(y)$ will have two distinct eigenvectors. This will imply that $P(y)$ will have two distinct stationary distributions, which is not possible as $P(y)$ is ergodic for every $y > 0$. Therefore, $\lambda(y) > 0$ for $y > 0$, which proves the Lemma for the present case.

Case 2: $\lambda(0) > 0$
First we prove the following:

Claim. Let $z = (z_0, z_1, \ldots, z_{n-1})$, with $\Sigma z_i^2 = 1$, be the eigenvector corresponding to the second smallest eigenvalue, $\lambda(y)$, of $M(y)$, with the assumption that $y > 0$. Let $z' = (0, z_1, z_2, \ldots, z_{n-1})$. Then

$$M_{00}(y) < \lambda(y) \Rightarrow \frac{z'.M(0).z'^T}{z'.z'^T} < \lambda(y).$$

Proof.

$$\begin{aligned}
z'.M(0).z'^T &= \Sigma_{i=0}^{n-1}\Sigma_{j=0}^{n-1} z'_i M_{ij}(0) z'_j \\
&= \Sigma_{i=1}^{n-1}\Sigma_{j=1}^{n-1} z'_i M_{ij}(0) z'_j, \text{ as } M_{ij} = 0 \text{ for } i = 0 \text{ or } j = 0 \\
. &= \Sigma_{i=1}^{n-1}\Sigma_{j=1}^{n-1} z'_i M_{ij}(y) z'_j \text{ as } M_{ij}(0) = M_{ij} \text{ when neither } i \text{ nor } j = 0, \\
&= \Sigma_{i=1}^{n-1}\Sigma_{j=1}^{n-1} z_i M_{ij}(y) z_j \text{ as } z_k = z'_k \text{ when } k \neq 0, \\
&= \Sigma_{i=1}^{n-1} z_i \Sigma_{j=1}^{n-1} M_{ij}(y) z_j \\
&= \Sigma_{i=1}^{n-1} z_i (\Sigma_{j=0}^{n-1} M_{ij}(y) z_j - M_{i0}(y) z_0) \\
&= \Sigma_{i=1}^{n-1} z_i (\lambda(y) z_i - M_{i0}(y) z_0) \\
&= \lambda(y)\Sigma_{i=1}^{n-1} z_i^2 - z_0 \Sigma_{i=1}^{n-1} z_i M_{i0}(y) \\
&= \lambda(y)(1 - z_0^2) - z_0(\Sigma_{i=0}^{n-1} z_i M_{i0}(y) - z_0 M_{00}(y)) \\
&= \lambda(y)(1 - z_0^2) - z_0(\Sigma_{i=0}^{n-1} z_i M_{0i}(y) - z_0 M_{00}(y)) \text{ as } M \text{ is symmetric}, \\
&= \lambda(y)(1 - z_0^2) - z_0(\lambda(y) z_0 - z_0 M_{00}(y))
\end{aligned}$$

Now,

$$\begin{aligned}
z'.z'^T &= \Sigma_{i=1}^{n-1} z'_i z'_i \\
&= \Sigma_{i=1}^{n-1} z_i^2 \\
&= 1 - z_0^2
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{z'.M(0).z'^T}{z'.z'^T} &= \frac{\lambda(y)(1 - z_0^2) - z_0(\lambda(y) z_0 - z_0 M_{00}(y))}{1 - z_0^2} \\
&= \lambda(y) + \frac{z_0^2}{1 - z_0^2}(M_{00}(y) - \lambda(y))
\end{aligned}$$

Whenever $y > 0$, we have $z_0^2 < 1$. Because, otherwise, z becomes $(1, 0, \ldots, 0)$. However, such a z cannot be an eigenvector of $M(y)$. Reason: let k be a vertex adjacent to vertex 0. Such a k must exist as the underlying Markov chain is irreducible. As $y > 0$, $M_{k0}(y) \neq 0$. So, denoting $M_k(y)$ as the kth row of $M(y)$, $M_k(y)(1, 0, \ldots, 0)^T \neq 0$. Therefore, $(1, 0, \ldots, 0)$ is not an eigenvector of $M(y)$.

As we have $z_0^2 < 1$, the Claim follows from the last equality.

Claim. There is a $\delta, \delta > 0$ such that for all $y, 0 \leq y \leq \delta, M_{00}(y) < \lambda(y)$.

Proof. By definition, $M_{00}(y) = y(\Sigma_{i \in S} V_i)$, where S is the set of neighbouring vertices of the vertex 0. Hence, $\lim_{y \to 0} M_{00}(y) = 0$, but $\lim_{y \to 0} \lambda(y) = \lambda(0)$ which is greater than 0 by assumption in the case under consideration. Hence, the proposition follows.

Using Claims 3 and 3, we now can complete the proof for Case 2 of Lemma 1.

We consider the interval $(0, \delta), \delta$ as in Claim 3 and any y in this interval. $(1, 0, \ldots, 0)$ is an eigenvector of $M(0)$ corresponding to its smallest eigenvalue 0. As $M(0)$ is a symmetric matrix, we can use the variational characterization for $\lambda(0)$, its second smallest eigenvalue, to get

$$\begin{aligned}\lambda(0) &= \min_{t \perp (1,0,\cdots,0)} \frac{t.M(0).t^T}{t.t^T} \\ &\leq \frac{z'.M(0).z'^T}{z'.z'^T} \text{ (here } z' \text{ is as in Claim 3)} \\ &< \lambda(y)\end{aligned}$$

for all y, $0 < y \leq \delta$, where δ is as in 3.

The first inequality is because z' is orthogonal to $(1, 0, \ldots, 0)$ and the next inequality follows then from the Claim 3.

This completes the proof of Case 2, and hence that of Lemma 1.

We are now in a position to prove our main result which is about the behaviour of $\lambda(y)$ as y increases from 0 onwards. (Recall that y is the inverse of V_0, and is greater than 0).

Theorem 1. *With $y \geq 0$, $\lambda(y)$, the second smallest eigenvalue of $N(y)$, hence that of $M(y)$, is a non-decreasing function in y.*

Proof. We consider the characteristic polynomial $p(x, y)$ of $N(y)$. The polynomial $p(x, y)$, being $det(N(y) - xI_{n \times n})$, is a degree n polynomial in x, and is linear in y. It can be easily seen that for each y,

$$\lim_{x \to -\infty} p(x, y) = \infty$$

We recall that all the roots of $p(x, y)$ are non-negative. Therefore, $x < 0 \Rightarrow p(x, y) > 0$ and $0 < x < \lambda(y) \Rightarrow p(x, y) < 0$.

Suppose the theorem is false, let

$$(\exists y_a, y_b)[0 < y_a < y_b \text{ and } \lambda(y_a) > \lambda(y_b)]$$

(As a witness of the contradiction, the assumption that $y_a > 0$ is justified because if $\lambda(0) > r$, for some r, then for a z within a small neighbourhood of 0, $\lambda(z) > r$, as the function $\lambda(y)$ is continuous in y). Further, we assume that the δ in Lemma 1 is such that

$$0 < \delta < y_a < y_b \tag{1}$$

This condition can be made true by making δ take a value smaller than y_a, if so needed.

Consider $x = \lambda(0)$. From Lemma 1, we have that $p(\lambda(0), 0) > p(\lambda(0), \delta)$. Now, the fact that $p(x, y)$ is linear in y gives us

$$p(\lambda(0), 0) > p(\lambda(0), \delta) > p(\lambda(0), y_a) > p(\lambda(0), y_b) \tag{2}$$

However, as $\lambda(y_b) < \lambda(y_a)$, the two curves $p(x, y_a)$ and $p(x, y_b)$ will intersect at some value of x, say x^*, with $x^* < \lambda(y_b)$, i.e., $p(x^*, y_a) = p(x^*, y_b) = b$ (say). Again, the fact that $p(x, y)$ is linear in y implies that for every value, say y_1, that y takes, the resultant curve $p(x, y_1)$ will also pass through that point of intersection. In particular, $p(x^*, 0)$ will also be b. We consider two cases (i)$\lambda(0) < x^* < \lambda(y_b)$ and (ii) $\lambda(0) > x^* > \lambda(y_b)$

Case (i): $\lambda(0) < x^* < \lambda(y_b)$

Since $0 < x^* < \lambda(y_b)$, $p(x^*, y_b) < 0$. Therefore the number of intersections of $p(x, y_b)$ between $x = 0$ and $x = x^*$, both inclusive, with the x-axis is exactly 1.

But $x^* > \lambda(0)$ and $p(x^*, 0) = p(x^*, y_b) < 0$, therefore, the curve $p(x, 0)$ will intersect the x-axis at some $x_2, \lambda(0) < x_2 < x^*$. This implies that there is some y_1 such that $p(x, y_1)$ is tangent to the x-axis at some point between $\lambda(0)$ and x_2. The reason is as follows.

For the sake of convenience, let x_1 denote $\lambda(0)$. We recall that the function $p(x, y)$ is of the form $r(x) + ys(x)$. We show that there exist $x_m, x_1 < x_m < x_2$, and some y_1 such that $r(x_m) + y_1 s(x_m) = 0$ as well as $r'(x_m) + y_1 s'(x_m) = 0$. ($r'(x)$ and $s'(x)$ denote the derivatives of $r(x)$ and $s(x)$.) For this to happen,

$$\frac{r(x_m)}{s(x_m)} = \frac{r'(x_m)}{s'(x_m)} = -y_1$$

will hold.

We define $g(x)$ as

$$g(x) = r(x)s'(x) - r'(x)s(x)$$

As $p(x_1, 0) = p(x_2, 0) = 0$, we have $r(x_1) = r(x_2) = 0$. Further, as $p'(x_1, 0) > 0$ and $p'(x_2, 0) < 0$, we have $r'(x_1) > 0$ and $r'(x_2) < 0$. Since $p(x, y_a) < 0$ for $x_1 \leq x \leq x_2$, and $y_a > 0$, we have $s(x_1) < 0$ and $s(x_2) < 0$. These facts imply that $g(x_1) > 0$ and $g(x_2) < 0$. Therefore, there is an $x_m, x_1 < x_m < x_2$ such that $g(x_m) = 0$. Setting y_1 as $-\frac{r(x_m)}{s(x_m)}$, we have that $p(x, y_1)$ is tangent to the x-axis at x_m.

Applying the same reasoning after shifting the x-axis by a small δ, we see that there is an y_c in the neighbourhood of y_1 such that $p(x, y_c)$ either has a maximum below the x-axis, or a minimum above the x-axis.

We arrive at the same conclusion for Case (ii) as well.

But $p(x, y_c)$ is a degree n polynomial in x, therefore, it can have at most $(n-1)$ maxima/minima in total. Further, $p(x, y_c)$ has n real roots. So, between each pair of consecutive roots there will be exactly one maximum or one minimum, counting multiplicities, implying that it cannot have a maximum below the x-axis or a minimum above the x-axis. Hence, the assumption that $\lambda(y_b) < \lambda(y_a)$, even though $y_b > y_a$, is false.

4 Concluding remarks

We have investigated in the paper the effect of increasing the difference between the energies of the lowest and the second-lowest energy states on the mixing

time of the Markov chain of a Metropolis algorithm that seeks to find the lowest energy state. We show that such an increase in the difference will not reduce the mixing time of the chain; our result thus contradicts what [Cl99] had expected.

Besides the immediate context of the Sali, Shakhnovich and Karplus conjecture, our result may be considered of interest because ours is a statement about the performance of the Metropolis algorithm solely on an aspect of the stationary probabilities of the underlying Markov chain. The work of Sasaki [S91] is also of a similar kind, as it provides a lower bound on a performance measure of the Metropolis algorithm in terms of *density* of states, the density $D(\pi)$ being the ratio of the number of states with stationary probability π and the total number of states.

We would like to strengthen our result to show that the second smallest eigenvalue $\lambda(y)$ of $N(y)$ is an *increasing* function in y. To do so, besides what we have done here, we would also need to prove that there exist no $y_1, y_2, 0 \leq y_1 < y_2$ with $\lambda(y_1) = \lambda(y_2) = w$, (say). It would follow immediately then than for every value of y, w would be an eigenvalue of $N(y)$. The limited experimentation that we did suggest that such is not the case, however, we have not been able to provide a proof for it.

References

[Cl99] Peter Clote, Protein folding, the Levinthal paradox and rapidly mixing Markov Chains, *,Proceedings of 26th International Colloqium on Automata, Languages and Programming*, LNCS 1644 pp 240–249, 1999.

[S91] Galen Sasaki, The effect of the density of states on the Metropolis Algorithm, *Information Processing Letters*, Vol. 37, pp 159–163, 1991.

[S93] Alistair Sinclair, *Algorithms for Random Generation and Counting: A Markov Chain Approach*, Birkhäuser, Berlin, 1993.

[SSK94a] Andrej Šali,Eugene Shakhnovich and Martin Karplus, How does a protein fold? *Nature,* Vol. 369, pp 249–252, 1994.

[SSK94b] Andrej Šali,Eugene Shakhnovich and Martin Karplus, Kinetics of protein folding: A lattice model study of the requirement for folding to the native state, *Jl. Molecular Biology*, Vol. 235, pp 1614–1636, 1994.

[SK91] J. Skolnick and A. Kolinski, Dynamic Monte Carlo simulation of a new lattice model of globular protein folding, structure and dynamics, *Jl. Molecular Biology*, Vol. 221, pp 499–531, 1991.