Necessary and Sufficient Conditions for Success of the Metropolis Algorithm for Optimization

Swagato Sanyal Department of Computer Science and Engineering Indian Institute of Technology Kanpur 208016 swagatos@cse.iitk.ac.in Raja S Department of Computer Science and Engineering Indian Institute of Technology Kanpur 208016 raja@cse.iitk.ac.in

Somenath Biswas Department of Computer Science and Engineering Indian Institute of Technology Kanpur 208016 sb@cse.iitk.ac.in

ABSTRACT

This paper focusses on the performance of the Metropolis algorithm when employed for solving combinatorial optimization problems. One finds in the literature two notions of success for the Metropolis algorithm in the context of such problems. First, we show that both these notions are equivalent. Next, we provide two characterizations, or in other words, necessary and sufficient conditions, for the success of the algorithm, both characterizations being conditions on the family of Markov chains which the Metropolis algorithm gives rise to when applied to an optimization problem. The first characterization is that the Metropolis algorithm is successful iff in every chain, for every set A of states not containing the optimum, the ratio of the ergodic flow out of Ato the capacity of A is high. The second characterization is that in every chain the stationary probability of the optimum is high and that the family of chains mixes rapidly. We illustrate the applicability of our results by giving alternative proofs of certain known results.

Categories and Subject Descriptors

G.1.6 [**Optimization**]: simulated annealing; G.3 [**Probability And Statistics**]: Probabilistic algorithms (including Monte Carlo)

General Terms

Theory, Algorithms

Keywords

Metropolis algorithm, optimization, success of Markov chain families, rapid mixing of Markov chains

1. INTRODUCTION

The Metropolis algorithm [2] is a very simple-to-implement randomized search heuristic which can be used to locate an

GECCO'10, July 7-11, 2010, Portland, Oregon, USA.

Copyright 2010 ACM 978-1-4503-0072-8/10/07 ...\$10.00.

optimum point in a large search space. Although the algorithm is widely used, our theoretical understanding of the performance of the algorithm is somewhat limited. While it is known that the Metropolis algorithm cannot work efficiently on all instances of even problems like the maximum bipartite matching [4] or the minimum spanning tree [6], yet it has been proved to be efficient on random instances of even some NP-hard problems, e.g., the graph bisection problem [1]. We provide in this paper two characterizations (i.e., necessary and sufficient conditions) for the success of the Metropolis algorithm for optimization. As our characterizations relate certain notions (e.g., rapid mixing of a family of Markov chains) which have been very well studied in related but different contexts, we hope that our results will help to understand better why and when the Metropolis algorithm will work efficiently for optimization.

Given an optimization problem, the Metropolis algorithm defines a Markov chain for each instance of the problem, the state space of the chain being the set of feasible solutions of the instance. There is also a *neighborhood structure* defined on the set of feasible solutions where the set N(s) of neighbors for a solution s is the set of all those feasible solutions which can be reached from s by one inexpensive local move. The algorithm also defines a probability distribution on N(s) for every s. The Markov chain when at s at a step will move in its next step to one element of N(s) as per the defined probability distribution on N(s). In this manner the chain traverses the space of feasible solutions with the aim of encountering an optimum solution sooner or later.

The probability distribution on N(s) is so defined that it will favor an element with a value better than the one for the current element s, the extent of the preference depending on a parameter called *temperature* which is defined for each specific chain. One would then consider the Metropolis algorithm to be *successful* for an optimization problem Π if for each instance of Π , there is a temperature such that the corresponding Markov chain is guaranteed to encounter an optimum solution efficiently, i.e., in a number of steps bounded by a fixed polynomial in the instance size, irrespective of the initial solution the chain starts at. One finds in the literature two somewhat different formalizations of this notion of efficient behavior: one in terms of the expected number of steps to reach an optimum, the other in terms of the probability that the number of steps to reach an optimum is bounded by a fixed polynomial in the instance size. We show both of these formalizations of success to be, not surprisingly, equivalent to each other. We then provide a

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

necessary and sufficient condition for success, which is that for every set A of search points, A not containing an optimum, the ratio $\Phi(A)$ of the ergodic flow out of A to the capacity of A is high.

As each of the Markov chains defined by the Metropolis algorithm for a problem Π is ergodic, each chain is guaranteed to reach close to its unique stationary distribution when run for sufficiently many steps. It is easy to see, therefore, that a sufficient condition for the success of the Metropolis algorithm for Π is that (a) the Markov chain for each instance reaches a distribution close to its stationary distribution in steps polynomial in instance size (in other words, the family of chains is rapidly mixing) and (b) the stationary distribution probability of an optimum solution is high. Because, in such a case, if we run the chain corresponding to an instance for polynomially many steps, the probability of encountering an optimum is high.

Our other result is that the above sufficient condition for success is actually also a necessary one. We prove this by showing that our first characterization is equivalent to this new condition. Thus we prove that rapid mixing is necessary for the success of the Metropolis algorithm for optimization.

We then show that our characterizations can be used in a straightforward and easy manner to provide both positive and negative results about the success of the Metropolis algorithm by providing alternative proofs of two results on the behavior of the Metropolis algorithm on certain classes of inputs for the minimum spanning tree problems, which were originally proved in [6]. We also provide a necessary condition on the density of states (a notion defined in [3]) needed for the success of the Metropolis algorithm, making use of our characterization of success. The result is similar to the one given in [3].

2. THE METROPOLIS ALGORITHM, NO-TIONS OF SUCCESS

2.1 The Metropolis Algorithm

Let Π be an optimization problem, and without loss of generality, we will consider it to be a minimization problem. Let I denote a set of instances of Π . For $i \in I$, let n_i denote the size of the instance i. We denote by $S^{(i)}$ the set of all feasible solutions of the instance i. (Typically, $|S^{(i)}|$ is exponential in n_i .) Without loss of generality, we make the following assumptions:

- 1. For every instance i in I, the neighborhood structure $N^{(i)}$ on $S^{(i)}$, the set of feasible solutions of i, is a connected, undirected graph. We use $d^{(i)}$ to denote the largest degree of the graph. We assume that $d^{(i)}$ is bounded above by a fixed polynomial in n_i . Thus, the neighbourhood structure graph is symmetric and typically, the degree of any vertex in the graph is small compared to the total number of vertices.
- 2. There is an easy to evaluate cost function $c(\cdot)$ which, given a feasible solution r, gives the cost c(r) of r, and this cost is a non-negative real number.
- 3. Every instance has a unique optimum. (The case of multiple optima can be reduced to the unique one case by a reduction given in [3]).

The Metropolis algorithm on instance i runs a Markov chain $X^{(i)} = (X_0^{(i)}, X_1^{(i)}, \ldots)$, using the temperature parameter as some $T^{(i)}$. The state space of the chain is the set $S^{(i)}$ of the feasible solutions of i, and the transition probabilities are defined as below. (u and v denote any two feasible solutions).

$$\mathbf{P}[X_{k+1} = v | X_k = u] = \begin{cases} 0 & \text{(if } u \neq v \text{ and } v \notin N(u)) \\ \frac{1}{2d^{(i)}} e^{-\frac{c(v) - c(u)}{T^{(i)}}} & \text{(if } c(v) > c(u) \text{ and } v \in N(u)) \\ \frac{1}{2d^{(i)}} & \text{(if } c(v) \leq c(u) \text{ and } v \in N(u)) \\ 1 - \sum_{j' \neq u} \mathbf{P}[X_{k+1} = j' | X_k = u] & \text{(if } u = v) \end{cases}$$

Let $P^{(i)} = (p_{u,v}^{(i)} : u, v \in S^{(i)})$ denote the transition matrix of the Markov chain $X^{(i)}$ where for all $u, v \in S^{(i)}, p_{u,v}^{(i)} = P[X_{k+1} = v | X_k = u]$. As per [3], for $T = 0, \delta \ge 0$, we define

$$e^{-\frac{\delta}{T}} = \begin{cases} 0 & \text{if } \delta > 0\\ 1 & \text{if } \delta = 0 \end{cases}$$

The above chain can easily be seen to be ergodic and timereversible, and its unique stationary distribution $\pi = (\pi_j : j \in S^{(i)})$ is given by:

$$e_{j} = \frac{e^{-\frac{c(j)}{T(i)}}}{\sum_{v \in S(i)} e^{-\frac{c(v)}{T(i)}}}$$

π

Our chain is a so called *lazy Markov chain* because it satisfies the condition that for every state $j \in S^{(i)}, p_{jj}^{(i)} \geq \frac{1}{2}$. This condition is necessary for the applicability of the Conductance Theorem [5]. We note that every Markov chain can be easily modified into a lazy one without incurring a slowdown of its rate of convergence by more than a constant factor.

2.2 Notions of Success

One finds in the literature two notions of success, one in terms of the worst expected number of steps to reach the optimum as we vary the starting state, and the other in terms of the minimum of the probabilities to reach the optimum within steps bounded by a fixed polynomial, the minimization being done over all starting states. (See, e.g., [3] and [6].) We recall that for the chain $X^{(i)}$, $S^{(i)}$ denotes its state space, $P^{(i)} = (p_{j,k}^{(i)} : j, k \in S^{(i)})$ denotes its transition probability matrix and $(\pi_j^{(i)} : j \in S^{(i)})$ denotes its unique stationary probability distribution. Let $s_{opt}^{(i)}$ denote the state in $S^{(i)}$ with the maximum stationary probability(which corresponds to minimum cost). We now define formally the two notions of success.

DEFINITION 1 (S-SUCCESS). The family $X = \{X^{(i)} | i \in I\}$ is defined to be S-successful if there exist constants $k, n_0 > 0$ such that $\forall i \in I$ such that $n_i \ge n_0$,

$$\max_{s \in S^{(i)}} E[\min\{t \ge 0 | X_t^{(i)} = s_{opt}^{(i)}\} \mid X_0^{(i)} = s] \le n_i^k$$

DEFINITION 2 (W-SUCCESS). The family $X = \{X^{(i)} | i \in I\}$ is defined to be W-successful if there exist constants $k_1, k_2, n_0 > I\}$

0 such that $\forall i \in I$ such that $n_i \geq n_0$,

$$\min_{s \in S^{(i)}} \mathbf{P}[\min\{t \ge 0 | X_t^{(i)} = s_{opt}^{(i)}\} \le n_i^{k_1} \mid X_0^{(i)} = s] \ge n_i^{-k_2}$$

These definitions for success may perhaps be considered too stringent as they require the Metropolis algorithm to do well in the worst case over all possible start states. This is because, in practice, one may start the algorithm either from a state uniformly selected at random, or from a state which we have reasons to believe to be a good start state. However, in the general case with which we are concerned in this paper, selecting a state uniformly at random is in itself a difficult problem, and we may not have any a priori idea as to which of the states would be good as start state(s).

The following Theorem shows that the two definitions are actually equivalent.

THEOREM 1. The family $X = \{X^{(i)} | i \in I\}$ is S-successful if and only if it is W-successful.

We defer the proof of the above to Appendix A.

From the above theorem we have that we can use either of the two definitions of success to argue that a family of Markov chains is successful.

DEFINITION 3 (SUCCESS). A family X of Markov chains is successful if it is W-successful, or equivalently, S-successful.

3. TWO CHARACTERIZATIONS OF SUC-CESS

In this section we give two different characterizations of success of a family of Markov chains.

DEFINITION 4 (JERRUM AND SINCLAIR, [5]). Let X be an ergodic Markov chain with state space S, transition matrix $P = (p_{i,j} : i, j \in S)$ and stationary distribution $\pi = (\pi_i : i \in S)$. Then for all non empty proper subsets A of S,

$$\Phi(A) = \frac{\sum_{i \in A, j \in \overline{A}} \pi_i p_{ij}}{\sum_{i \in A} \pi_i}$$

The conductance of the chain is defined as

$$\Phi(X) = \min_{A \subset S, A \neq \phi, cap(A) \le \frac{1}{2}} \Phi(A)$$

where $cap(A) = \sum_{i \in A} \pi_i$.

Let I be a set of problem instances for a combinatorial optimization problem. For each i in I with temperature of i as T_i , let $X^{(i)} = (X_0^{(i)}, X_1^{(i)}, \ldots)$ be the Markov chain i gives rise to. We assume each $X^{(i)}$ to satisfy the assumptions listed in Section 2.1. As before, let n_i denote the size of the instance i, and let $S^{(i)}$ and $P^{(i)} = (p_{j,k}^{(i)} : j, k \in S^{(i)})$ denote the state space and the state transition matrix of $X^{(i)}$ respectively. Let $s_{opt}^{(i)}$ be the state in $S^{(i)}$ with the minimum cost. Let the stationary probability distribution of $X^{(i)}$ be $(\pi_j^{(i)} : j \in S^{(i)})$. We assume that there is a polynomial $q(\cdot)$ such that for all i in I, $\pi_i^{(i)} \geq 2^{-q(n_i)}$ for every j in $S^{(i)}$.

THEOREM 2. The following three statements are equivalent:

(a) The family X of Markov chains is successful.

(b) There exist constants $k, n_0 > 0$, such that for all i in I

with $n_i \ge n_0$, for all non-empty subsets A of $S^{(i)} - \{s_{opt}^{(i)}\}\$, the condition

$$\Phi(A) \ge n_i^{-k}$$

is satisfied.

(c) The family X of Markov chains mixes rapidly¹, and there exist constants $n_0, k > 0$ such that for all i in I with $n_i \ge n_0$ we have that $\pi_o^{(i)} \ge n_i^{-k}$ where o is $s_{opt}^{(i)}$.

PROOF. ((a) *implies* (b)) Let X be successful and hence W-successful. Thus there exist constants $k_1, k_2, n_0 > 0$ such that $\forall i \in I$ with $n_i \geq n_0$,

$$\min_{\in S^{(i)}} \mathbf{P}[\min\{t \ge 0 | X_t^{(i)} = s_{opt}^{(i)}\} \le n_i^{k_1} \mid X_0^{(i)} = s] \ge n_i^{-k_2}$$

For each *i* such that $n_i \geq n_0$, suppose that the start state is chosen from $S^{(i)}$ according to some probability distribution $(f_s^{(i)}: s \in S^{(i)})$ on states. Then $\mathbf{P}[\min\{t \geq 0 | X_t^{(i)} = s_{opt}^{(i)}\} \leq n_i^{k_1}]$

$$= \sum_{s \in S^{(i)}} \mathbf{P}[\min\{t \ge 0 | X_t^{(i)} = s_{opt}^{(i)}\} \le n_i^{k_1} | X_0^{(i)} = s] \mathbf{P}[X_0^{(i)} = s]$$
$$\ge n_i^{-k_2} \sum_{s \in S^{(i)}} f_s^{(i)}$$
$$= n_i^{-k_2} \tag{1}$$

Our proof is obtained by showing that we reach a contradiction if we assume that the successful family X does not satisfy the condition of (b).

Let us assume that for X it holds that for all constants $k, m > 0, \exists i \in I$ with $n_i > m$, and an $A \subseteq S^{(i)} - \{s_{opt}^{(i)}\}$ such that

$$\Phi(A) < n_i^{-k} \tag{2}$$

We fix constants k', m' > 0. Then $\exists i' \in I$ with $n_{i'} > m'$, and an $A \subseteq S^{(i')} - \{s_{opt}^{(i')}\}$, such that

$$\Phi(A) < n_{i'}^{-k'} \tag{3}$$

For the chain of i', using the A satisfying 3, we define the following initial distribution $f_s^{(i')} : s \in S^{(i')}$:

$$f_s^{(i')} = \begin{cases} \frac{\pi_s^{(i')}}{cap(A)} & \text{if } s \in A; \\ 0 & \text{otherwise} \end{cases}$$

With this as the initial distribution, we derive a contradiction to 1 above for the chain $X^{(i')}$ of i'. First, we prove by induction on t that the probability that $X^{(i')}$ makes a transition from some state in A to some state in \overline{A} for the first time in the t-th step, with initial distribution as $f_s^{(i')}$, is less than $n_{i'}^{-k'}$. Let $P^{(i')} = (p_{j,k}^{(i')} : j, k \in S^{(i')})$ be the transition matrix of $X^{(i')}$.

Base step(t=1): The probability that the 1st step of $X^{(i')}$ is from some state in A to some state in \overline{A} is clearly

$$\sum_{j \in A, k \in \overline{A}} f_j^{(i')} p_{j,k}^{(i')} = \Phi(A) < n_{i'}^{-k'} (\text{from 3})$$

¹Informally, a family of chains mixes rapidly means that within steps bounded by a fixed polynomial in instance size, each chain will come very close to its stationary distribution, irrespective of the initial distribution. For the definition, see [5].

Induction step: We assume the induction hypothesis to be true for all $t \leq t'$. The probability that $X^{(i')}$ makes a move from some state in A to some state in \overline{A} for the first time in (t'+1)-th step is $\mathcal{T}_{t'+1}/cap(A)$, where

$$\begin{aligned} \mathcal{T}_{t'+1} &= \sum_{s_1, \dots, s_{t'+1} \in A, s_{t'+2} \in \overline{A}} \pi_{s_1}^{(i')} p_{s_1, s_2}^{(i')} \dots p_{s_{t'+1}, s_{t'+2}}^{(i')} \\ &= \sum_{s_1, \dots, s_{t'+1} \in A, s_{t'+2} \in \overline{A}} \pi_{s_2}^{(i')} p_{s_2, s_1}^{(i')} \dots p_{s_{t'+1}, s_{t'+2}}^{(i')} \end{aligned}$$

(using the time reversibility of $X^{(i')}$.)

$$= \sum_{s_2,\dots,s_{t'+1}\in A, s_{t'+2}\in\overline{A}} \left(\pi_{s_2}^{(i')} p_{s_2,s_3}^{(i')} \dots p_{s_{t'+1},s_{t'+2}}^{(i')} \sum_{s_1\in A} p_{s_2,s_1}^{(i')} \right)$$
$$\leq \sum_{s_2,\dots,s_{t'+1}\in A, s_{t'+2}\in\overline{A}} \pi_{s_2}^{(i')} p_{s_2,s_3}^{(i')} \dots p_{s_{t'+1},s_{t'+2}}^{(i')}$$
$$= \mathcal{T}_{t'}$$

Thus, the probability that $X^{(i')}$ makes a move from some state in \overline{A} to some state in \overline{A} for the first time in the (t'+1)-th step

$$= \mathcal{T}_{t'+1}/cap(A)$$

$$\leq \mathcal{T}_{t'}/cap(A)$$

This is the probability that $X^{(i')}$ makes a move from some state in \overline{A} to some state in \overline{A} for the first time in the t'-th step and this probability is less than $n_{i'}^{-k'}$ (from the induction hypothesis). Using the union bound, again for the same initial distribution, $\forall k'' > 0$, $\mathbf{P}[\min\{t \ge 0|X_t^{(i')} = s_{opt}^{(i)}\} \le n_{i'}^{k''}]$

$$\leq \mathbf{P}[\min\{t \geq 0 | X_t^{(i')} \in \overline{A}\} \leq n_{i'}^{k''}]$$

$$< n_{i'}^{k''} \cdot n_{i'}^{-k'}$$

$$= n_{i'}^{-(k'-k'')}$$

Since we are free to choose k', k'' and m', we have, therefore, proved that $\forall c, c', m'' > 0, \exists i \in I$ with $n_i > m''$ and an initial distribution, such that

$$\begin{split} \mathbf{P}[\min\{t\geq 0|X_t^{(i)}=s_{opt}^{(i)}\}\leq n_i^c] < n^{-c'}\\ \text{under that initial distribution. This contradicts 1, hence} \end{split}$$

the family X could not have been successful and we have a contradiction.

((b) *implies* (c)) Let there exist constants $k, n_0 > 0$, such that $\forall i$ such that $n_i \geq n_0$,

$$\min_{A \subseteq S^{(i)} - \{s_{opt}^{(i)}\}, A \neq \phi} \Phi(A) \ge n_i^{-k}$$

We first show that this implies at least inverse polynomial conductance and therefore rapid mixing.

Let *m* be a constant such that $\forall x \ge m, x^k + 1 \le x^{k+1}$. Let $i \in I$ be an instance where $n_i \ge \max\{n_0, m\}$. From Definition 4 and the time reversibility of X_i , we have

tion 4 and the time reversibility of X_i , we have $\Phi(X^{(i)}) = \min_{A \subseteq S^{(i)}, A \neq \phi} \max\{\Phi(A), \Phi(\overline{A})\}$. As one of Aand \overline{A} does not contain $s_{opt}^{(i)}, \Phi(X^{(i)})$ is at least n_i^{-k} . Together with the assumption that $\forall j \in S^{(i)}, \pi_j^{(i)} \ge 2^{-q(n_i)}$ for a fixed polynomial $q(\cdot)$, and the assumptions of Section 2.1, it follows from the *Conductance Theorem* (see Corollary 2.8 of [5]) that X is a rapidly mixing family of Markov chains. Next we prove that the optimum element has a high stationary distribution probability. Let $A' = S^{(i)} - \{s_{opt}^{(i)}\}$ and let *o* denote $s_{opt}^{(i)}$. Thus,

$$\Phi(A') = \frac{\sum_{j \in S^{(i)} - \{o\}} \pi_j^{(i)} p_{j,o}^{(i)}}{\sum_{j \in S^{(i)} - \{o\}} \pi_j^{(i)}}$$

$$= \frac{\sum_{j \in S^{(i)} - \{o\}} \pi_o^{(i)} p_{o,j}^{(i)}}{1 - \pi_o^{(i)}} \quad (\text{From time reversibility of } X^{(i)})$$

$$= \frac{\pi_o^{(i)} \sum_{j \in S^{(i)} - \{o\}} p_{o,j}^{(i)}}{1 - \pi_o^{(i)}}$$

$$\leq \frac{\pi_o^{(i)}}{1 - \pi_o^{(i)}}$$

Also $\Phi(A') \ge n_i^{-k}$. Thus, we have

$$n_i^{-k} \le \frac{\pi_o^{(i)}}{1 - \pi_o^{(i)}}$$
$$\implies 1 - \pi_o^{(i)} \le n_i^k \pi_o^{(i)}$$
$$\implies \pi_o^{(i)}(1 + n_i^k) \ge 1$$
$$\implies \pi_o^{(i)} \ge \frac{1}{1 + n_i^k} \ge \frac{1}{n_i^{k+1}} \quad (\text{as } n_i \ge m)$$

((c) implies (a)) Let X be a rapidly mixing family of Markov chains. Let $\forall i \in I$, such that n_i is large enough, $s_{opt}^{(i)} \geq n_i^{-k_1}$ for some constant $k_1 > 0$. Let $k_2 > 0$ be a constant such that if $(f_j^{(i)} : j \in S^{(i)})$ be the distribution of X_i after n^{k_2} steps, and, $\forall j \in S^{(i)}, f_j^{(i)} \geq \frac{\pi_j^{(i)}}{2}$, for any start state (since X is a rapidly mixing family of Markov chains, there exists such a constant). Then the expected number of steps till the chain hits the optimal state for the first time is at most $2n_i^{k_1+k_2}$ (in other words, on expectation, the chain requires at most $2n_i^{k_1}$ blocks, each of $n_i^{k_2}$ steps, to reach $s_{opt}^{(i)}$ for the first time). This shows that X is S-successful and hence successful. \Box

4. APPLICATIONS

We have shown that the performance of the Metropolis algorithm is closely related to the mixing time of the underlying family of Markov chains. More specifically, a family of Markov chains generated by instances from a set satisfying certain properties, is successful *if and only if* the family is a rapidly mixing one and the stationary probability of the goal state is at least some inverse polynomial in the size of the input instance. These results provide a systematic way of proving both positive and negative results about the success of the Metropolis algorithm. In this section we provide alternative proofs of certain known results making use of our characterizations of success.

4.1 Minimum Spanning Tree Problem on Connected Triangles

To illustrate the applicability of our success characterizations, we consider the *Minimum Spanning Tree* problem. The instances and the heuristic are the same as those considered in [6] except for a self loop with probability at least $\frac{1}{2}$ added to each state to ensure strong aperiodicity. The problem is, given an undirected weighted graph G = (V, E, w), wbeing the *weight function*, to compute the minimum spanning tree of G. The instances we consider are *connected triangles* as shown in figure 1.

The state space here is the set of all connected subgraphs of



Figure 1: Connected triangles

the input graph. The cost c of a subgraph is the sum of the weights of its edges. Clearly, the state corresponding to the minimum cost is the minimum spanning tree which we denote by MST. We call two states s_1 and s_2 neighbors of each other iff s_2 can be obtained from s_1 either by including an edge not present in s_1 , or by removing an edge in s_1 (i.e. by *flipping* an edge as we call it). Thus the relation defined by neighborhood is symmetric. We denote the set of all neighbors of a state s by N(s), the number of edges by m, and the current state by s. We take the number of triangles as the size of an instance. The maximum of the degrees of nodes of the graph which underlies the Markov chain the Metropolis algorithm simulates, is clearly m. At any given instant, the Metropolis algorithm stays at its current state s with probability $\frac{1}{2}$. With probability $\frac{1}{2}$ it flips an edge selected uniformly at random and, if the resultant graph t is connected, accept the flip with probability $\min\{1, e^{-\frac{c(t)-c(s)}{\tau}}\}$, τ being the temperature.

4.1.1 Example where the Metropolis Algorithm Fails at Every Temperature

We here consider the same set of instances as considered by Wegener in [6] where the Metropolis algorithm fails to efficiently compute the minimum spanning tree. The set of instances I_1 is the set of all connected triangles consisting of 2n triangles, so that the number of vertices is 4n + 1 and number of edges is m = 6n, for each n > 0. There are *n* light triangles, the edge weights of each of which are 1, 1 and m. The edge weights of each of the remaining *n* triangles (heavy triangles) are m^2, m^2 and m^3 . The unique MST consists of all the edges with weight 1 or m^2 . Let T be a function mapping each instance $i \in I_1$ to a temperature $T^{(i)}$. For each $i \in I_1$, the algorithm runs an ergodic Markov chain $Y_{T^{(i)}}^{(i)}$ with state space $S^{(i)}$ (which is the set of all connected sub-graphs of i), and transition probability matrix $P_{T^{(i)}}^{(i)}$ $(p_{j,k}^{(i,T^{(i)})}: j, k \in S^{(i)})$ where

$$p_{s,t}^{(i,T^{(i)})} = \begin{cases} 0 & \text{(if } s \notin N(t) \text{ and } s \neq t) \\ \frac{1}{2m} \min\left\{1, e^{-\frac{c(t)-c(s)}{T^{(i)}}}\right\} & \text{(if } s \in N(t)) \\ 1 - \sum_{t' \in S^{(i)} - \{s\}} (1 - p_{s,t'}^{(i,T^{(i)})}) & \text{(if } s = t) \end{cases}$$

Each $Y_{T^{(i)}}^{(i)}$ is time-reversible, with Gibb's distribution as its stationary distribution. It is easy to verify that each node has stationary probability at least some inverse exponential in n, so that we can use Conductance Theorem. Let $\pi_{T^{(i)}}^{(i)} = (\pi_j^{(i,T^{(i)})}: j \in S^{(i)})$ be the stationary distribution of $Y_{(T^{(i)})}^{(i)}$. Let $Y_T = (Y_{T^{(i)}}^{(i)}: i \in I_1)$.

THEOREM 3. For no T is the family Y_T successful.

PROOF. We consider an instance $i \in I_1$ having 2n connected triangles, and m = 6n edges. case 1: $T^{(i)} > m$

We show that, when $T^{(i)}$ is at least m, the stationary probability of MST is less than any inverse polynomial in n. From Theorem 2, it implies that the family Y_T is not successful if, for arbitrarily large instances $i, T^{(i)} \geq m$.

In each connected subgraph of the input, there may be 2 or 3 edges of each triangle. 2 edges can be chosen out of 3 in 3 ways. Thus each triangle can be in one of four different configurations. Thus the search space $S^{(i)}$ contains 4^{2n} states. We partition $S^{(i)}$ into 4^n classes of 4^n states each. In a class, the configuration of each light triangle is fixed. We can fix the configurations of all light triangles in 4^n ways and thus we have 4^n classes. Each class has 4^n states for 4^n different combinations of configurations of the heavy triangles. We show that the sum of the stationary probabilities of the states in the class containing the optimal point is not more than inverse exponential in n. Let $\mathcal{C}_H^{(i)}$ and $\mathcal{C}_L^{(i)}$ denote the sets of all possible combinations of configurations of heavy and light triangles respectively . $\forall u \in \mathcal{C}_H^{(i)} \bigcup \mathcal{C}_L^{(i)}$, let $\mathcal{W}_{T^{(i)}}^{(u)}$ denote the ratio of the sum of the weights of all

edges present in u, to $T^{(i)}$. Let $\mathcal{S}um_{T^{(i)}}^{(i)} = \sum_{u \in \mathcal{C}_H^{(i)}} e^{-\mathcal{W}_{T^{(i)}}^{(u)}}$. The sum of the stationary probabilities is minimum for the class corresponding to $g \in \mathcal{C}_L^{(i)}$ where all the edges of all light triangles are present, and maximum for the class corresponding to $b \in \mathcal{C}_L^{(i)}$ where only the edges of light triangles with weight 1 are present. The ratio of these two sums is

$$\begin{aligned} &\frac{\mathcal{S}um_{T^{(i)}}^{(i)}.e^{-\mathcal{W}_{T^{(i)}}^{(g)}} / \sum_{t \in S_{T^{(i)}}^{(i)}} e^{-\frac{c(t)}{T^{(i)}}}}{\mathcal{S}um_{T^{(i)}}^{(i)}.e^{-\mathcal{W}_{T^{(i)}}^{(b)}} / \sum_{t \in S_{T^{(i)}}^{(i)}} e^{-\frac{c(t)}{T^{(i)}}}} \\ &= e^{-\frac{2n-n(m+2)}{T^{(i)}}} \\ &\leq e^{n} \text{ (as } T^{(i)} > m) \end{aligned}$$

Observing that there are 4^n classes, we can conclude that the maximum sum of the stationary probabilities of all states in a class cannot be more than $\frac{1}{4^n}e^n \leq 1.2^{-n} < n^{-k}$ for all constant k for all large enough n, which is also an upper bound on the stationary probability of the MST. case 2: $T^{(i)} < m$

For $i \in I_1$ we define a set $A \subseteq S^{(i)} - \{MST\}$, such that $\Phi(A)$ is not more than inverse exponential in n. We take A to be the set of all states where each heavy triangle have one edge of weight m^2 missing. Clearly A does not contain the goal state which is the MST. If $s_1 \in A$ and $s_2 \in \overline{A}$ are neighbors, then

$$p_{s_1,s_2}^{(i,T^{(i)})} = \frac{1}{2m} e^{\frac{-m^2}{T^{(i)}}} < \frac{1}{2m} e^{-m} \text{ (as } T^{(i)} < m) = \frac{1}{12n} e^{-6r}$$

(because we have to include the missing edge of weight m^2 of some triangle to come out of A) Thus we have

$$\Phi(A) = \frac{\sum_{j \in A, k \in \overline{A}} \pi_j^{(i, T^{(i)})} p_{j, k}^{(i, T^{(i)})}}{\sum_{i \in A} \pi_j^{(i, T^{(i)})}} < \frac{1}{12n} e^{-6n} < n^{-k}$$

for all constant k for large enough n. From Theorem 2 it follows that for Y_T to be successful, $T^{(i)}$ cannot be less than m for arbitrarily large instances $i \in I$.

Thus for all temperatures, the Metropolis algorithm fails to efficiently compute the minimum spanning tree for the instances in I_1 . \Box

4.1.2 Example where the Metropolis Algorithm Succeeds at Some Temperature

Now we consider a set of instances considered by Wegener in [6] for which the Metropolis algorithm provably computes the minimum spanning tree efficiently at some temperature. Each instance of our next instance set I_2 consists of n connected triangles, for each n > 0. The weights of the sides of each of them are m^2, m^2 and m^3 , where m = 3n is the number of edges. We take the temperature $T^{(i)}$ to be $m^{\frac{5}{2}}$ for the instance i with m edges. $\forall i \in I_2$, in each connected subgraph of i, each triangle is in one of the three configurations: good(two edges each of weight m^2 are present), bad (one edge of weight m^2 and one edge of weight m^3 are present), complete(all three edges are present). We here view the algorithm, when run on $i \in I_2$ at temperature $m^{\frac{5}{2}}$, to simulate the ergodic Markov chain $Z^{(i)}$, with state space $S_1^{(i)}$ and transition probability matrix $Q^{(i)} = (q_{j,k}^{(i)} : j, k \in S_1^{(i)})$.

$$S_1^{(i)} = \{(g, b, c) : 0 \le g, b, c \le n, g + b + c = n\}$$

where (g, b, c) stands for the state in which number of good, bad and complete triangles are g, b and c respectively. Clearly the goal state is (n, 0, 0).

Of course $q_{s,t}^{(i)} = 0$ if $s \neq t$ and $s \notin N(t)$. The remaining transition probabilities are as follows²:

$$q_{(g,b,c),(g,b-1,c+1)}^{(i)} = \frac{b}{2m} e^{\left(\frac{-m^2}{m^2}\right)} = \frac{b}{2m} e^{-m^{-\frac{1}{2}}} \quad (b \neq 0)$$
$$q_{(g,b,c),(g,b+1,c-1)}^{(i)} = \frac{2c}{2m} = \frac{c}{m} \quad (c \neq 0)$$

 $\frac{q_{(g,b,c),(g-1,b,c+1)}^{(i)} = \frac{g}{2m}e^{\left(\frac{m}{m^{\frac{1}{2}}}\right)} = \frac{g}{2m}e^{-m^{\frac{1}{2}}} \quad (g \neq 0)$ ²The transition probabilities are well defined because

²The transition probabilities are well defined because $q_{(g,b,c),(g',b',c')}^{(i)}$ depends only on g, b, c, g', b', c' and not on which of the *n* triangles are good, bad or complete.

$$\begin{split} q^{(i)}_{(g,b,c),(g+1,b,c-1)} &= \frac{c}{2m} \quad (c \neq 0) \\ q^{(i)}_{(g,b,c),(g,b,c)} &= 1 - \sum_{t \in S_1^{(i)} - \{(g,b,c)\}} q^{(i)}_{(g,b,c),t} \geq \frac{1}{2} \end{split}$$

By detailed balance equations we observe that $Z^{(i)}$ is a time reversible Markov chain with stationary distribution $\sigma^{(i)} = (\sigma_i^{(i)} : j \in S_1^{(i)})$ of $Z^{(i)}$ is the following:

$$\sigma_{(g,b,c)}^{(i)} = \frac{\sum_{t \in S_{(g,b,c)}^{(i)}} e^{\left(-\frac{c(t)}{m^{\frac{5}{2}}}\right)}}{\sum_{t \in S^{(i)}} e^{\left(-\frac{c(t)}{m^{\frac{5}{2}}}\right)}}$$

where $S_{(g,b,c)}^{(i)}$ is set of all connected subgraphs of *i* with g,b and c good, bad and complete triangles respectively, and $S^{(i)}$, as in last subsection, is the set of all connected subgraphs of *i*. Here again we see that all stationary probabilities are at least some inverse exponential of *n*.

ties are at least some inverse exponential of n. We now consider another Markov chain $W^{(i)}$ with state space $S_1^{(i)}$ and transition matrix $R^{(i)} = (r_{j,k}^{(i)} : j, k \in S_1^{(i)}) =$ $(Q^{(i)})^2$. Thus each pair of steps of $Z^{(i)}$ simulates a single step of $W^{(i)}$. $W^{(i)}$ is easily seen to be an ergodic timereversible Markov chain with the same stationary distribution $\sigma^{(i)}$ as $Z^{(i)}$. If some eigenvalue of $W^{(i)}$ happens to be negative, we can modify the chain by replacing $R^{(i)}$ by $\frac{1}{2}(I_N + R^{(i)})$, where N is the cardinality of $S_1^{(i)}$ and I_N is the N cross N identity matrix, without changing its stationary distribution or slowing down its convergence too much. Clearly to establish that the Metropolis algorithm performs well on instances from I_2 at temperature $m^{\frac{5}{2}}$, it is sufficient to show that the family $W = (W^{(i)} : i \in I_2)$ is successful.

DEFINITION 5. $\forall i \in I_2, \forall (g, b, c) \in S_1^{(i)}$ such that $(g, b, c) \neq (n, 0, 0)$, next(s) is the state in $S_1^{(i)}$ defined as follows:

$$next((g,b,c)) = \begin{cases} (g+1,b-1,c) & (if \ b \neq 0) \\ (g+1,b,c-1) & (otherwise) \end{cases}$$

Intuitively next(s) is one step closer to MST than s. In one iteration of our algorithm, a bad triangle cannot become good. But a bad triangle can become good in a single move of $W^{(i)}$, as each move of $W^{(i)}$ represents a pair of iterations of the algorithm. Thus $\forall s \in S_1^{(i)} - \{MST\}, r_{s,next(s)}^{(i)} > 0, r_{next(s),s}^{(i)} > 0$.

DEFINITION 6. For every $s \in S_1^{(i)}$, We define path(s) to be a sequence $u_1u_2 \ldots u_p$ where $u_1 = s, u_p = MST, \forall j, 1 \leq j \leq p, u_j \in S_1^{(i)}, \forall j, 1 \leq j \leq p-1, u_{j+1} = next(u_j)$. From definition 5, such a path is sure to exist for every state s.

 $\begin{array}{l} \text{LEMMA 1. For large enough } n, \\ (a) \forall j \in S_1^{(i)} - \{MST\}, \sigma_j^{(i)} \leq \sigma_{next(j)}^{(i)}. \\ (b) \forall j \in S_1^{(i)} - \{MST\}, r_{j,next(j)}^{(i)} \geq \frac{1}{4m^2e} \end{array}$

PROOF. (a) Let j be
$$(g, b, c)$$

case 1: $b \neq 0$

In the algorithm, One way to go from (g, b, c) to (g+1, b-1, c) in two steps is via (g, b-1, c+1). Thus

$$r_{(g,b,c),(g+1,b-1,c)}^{(i)} \geq q_{(g,b,c),(g,b-1,c+1)}^{(i)} \cdot q_{(g,b-1,c+1),(g+1,b-1,c)}^{(i)}$$

$$= \frac{b}{2m}e^{-m^{-\frac{1}{2}}} \cdot \frac{c+1}{2m} = \frac{b(c+1)}{4m^2}e^{-m^{-\frac{1}{2}}}$$
(4)

To go from (g + 1, b - 1, c) to (g, b, c), a good triangle has to become complete. Even if the correct edge of a triangle (the one with weight m^3) is selected for flip, the probability that the flip is accepted is $e^{-m^{\frac{1}{2}}}$. Thus

 $r_{(g,b,c),(g+1,b-1,c)}^{(i)} \leq e^{-m^{\frac{1}{2}}}$ From the time-reversibility of $W^{(i)}$ we have $\sigma_{(g,b,c)}^{(i)} \frac{1}{4m^2} e^{-m^{-\frac{1}{2}}}$

$$\leq \sigma_{(g,b,c)}^{(i)} \frac{b(c+1)}{4m^2} e^{-m^{-\frac{1}{2}}} \\ \leq \sigma_{(g,b,c)}^{(i)} r_{(g,b,c),(g+1,b-1,c)}^{(i)} \\ = \sigma_{(g+1,b-1,c)}^{(i)} r_{(g+1,b-1,c),(g,b,c)}^{(i)} \\ \leq \sigma_{(g+1,b-1,c)}^{(i)} e^{-m^{\frac{1}{2}}}$$

Thus we have $\sigma_{(g,b,c)}^{(i)} \leq \sigma_{(g+1,b-1,c)}^{(i)}$ for large enough m.

 $case \ 2: \ b=0$

In the algorithm, (g+1, b, c-1) may be reached from (g, b, c)in two steps in the following way: In the first step, a complete triangle becomes good. In the next step, the state remains unchanged. Thus,

$$r_{(g,b,c),(g+1,b,c-1)}^{(i)} \ge \frac{c}{2m}\frac{1}{2} = \frac{c}{4m}$$
(5)

To go from (g+1, b, c-1) to (g, b, c), a good triangle has to become complete. Thus from similar argument as in case 1, we have

$$\begin{aligned} r_{(g+1,b,c-1),(g,b,c)}^{(i)} &\leq e^{-m^{\frac{1}{2}}} \\ \text{From the time reversibility of } W^{(i)}, \\ \sigma_{(g,b,c)}^{(i)} \frac{1}{4m} \\ &\leq \sigma_{(g,b,c)}^{(i)} \frac{c}{4m} \text{ (as } b = 0 \text{ and } (g,b,c) \neq (n,0,0), c \geq 1) \\ &\leq \sigma_{(g,b,c)} r_{(g,b,c),(g+1,b,c-1)}^{(i)} \\ &= \sigma_{(g+1,b,c-1)} r_{(g+1,b,c-1),(g,b,c)}^{(i)} \\ &\leq \sigma_{(g+1,b,c-1)} e^{-m^{\frac{1}{2}}} \end{aligned}$$

Giving us $\sigma_{(g,b,c)}^{(i)} \leq \sigma_{(g+1,b,c-1)}^{(i)}$ for large enough m. (b) Follows from (4) and (5) in the proof of part (a). \Box

Before proving the success of the family W, we prove another Lemma.

LEMMA 2. Let A be a non-empty subset of $S_1^{(i)} - \{MST\}$. Then are states $u \in A, v \in \overline{A}$ such that $\sigma_u^{(i)} = \max\{\sigma_j^{(i)} : j \in S_1^{(i)}\}$, and $r_{u,v}^{(i)} \ge \frac{1}{4m^2e}$.

PROOF. Let $t \in A$ and $\sigma_t^{(i)} = \max\{\sigma_j^{(i)} : j \in S_1^{(i)}\}$. From Lemma 1(*a*), path(t) is s sequence of states with non decreasing stationary probabilities. Let *u* be the first vertex on path(t) such that $v = next(u) \in \overline{A}$. Thus $\sigma_t^{(i)} \leq \sigma_u^{(i)}$. Since $\sigma_t^{(i)} = \max\{\sigma_j^{(i)} : j \in S_1^{(i)}\}$, we have $\sigma_u^{(i)} = \sigma_t^{(i)} = \max\{\sigma_j^{(i)} : j \in S_1^{(i)}\}$. From Lemma 1(*b*) the claim follows. \Box

THEOREM 4. The Markov chain family $W = (W^{(i)} : i \in I_2)$ is successful.

PROOF. Let A be a non-empty subset of $S_1^{(i)} - \{MST\}$. From Lemma 2, $\exists u \in A, v \in \overline{A}$ such that $\sigma_u^{(i)} = \max\{\sigma_j^{(i)} : j \in S_1^{(i)}\}$ and $r_{u,v}^{(i)} \geq \frac{1}{4m^2e}$. Thus,

$$\begin{split} \Phi(A) &= \frac{\sum_{j \in A, k \in \overline{A}} \sigma_{j}^{(i)} r_{j,k}^{(i)}}{\sum_{j \in A} \sigma_{j}^{(i)}} \\ &\geq \frac{\sigma_{u}^{(i)} r_{u,v}^{(i)}}{(n+1)^{3} . \sigma_{u}^{(i)}} \\ (\text{Since } |S_{1}^{(i)}| < (n+1)^{3} \text{ and } \forall t \in A, \sigma_{u}^{(i)} \ge \sigma_{t}^{(i)}) \\ &\geq \frac{1}{4m^{2}e(n+1)^{3}} \\ &= \frac{1}{36e(n+1)^{5}} \ge n^{-6} \quad \text{(for large enough } n\text{)te :P} \end{split}$$

Since A is any non-empty subset of $S_1^{(i)} - \{MST\}$, from Theorem 2, the claim follows.

4.2 Relation of Success of a Family of Markov Chains to Density of States

Our result enables us to arrive at a result which is very similar to that of Sasaki (see Proposition 1 of [3]). For some problem II, for a set of instances I, let $X = \{X^{(i)} : i \in I\}$ be a family of Markov chains which arise out of applying the Metropolis algorithm on instances in I. Let $d^{(i)}$ be the maximum of the degrees of all nodes of the underlying graph of $X^{(i)}$. The state space of $X^{(i)}$, the cost function, the stationary distribution of $X^{(i)}$, the neighborhood and size of instance i have usual notations. The transition probability matrix $P^{(i)} = (p_{j,k}^{(i)} : j, k \in S^{(i)})$ is same as the one given in Section 2.1. Let $c_{min}^{(i)}$ and $c_{max}^{(i)}$ be the minimum and maximum costs of the states in $S^{(i)}$ respectively. For each $r, c_{min}^{(i)} \leq r < c_{max}^{(i)}$, let $S_r^{(i)} = \{s \in S^{(i)} : c(s) > r\}$ and $W_r^{(i)} = \{s \in S_r^{(i)} : \sum_{j \in S_r^{(i)}} p_{s,j}^{(i)} > 0\}$. Let $\eta^{(S_r^{(i)})}$ be the number of states with maximum stationary probability $\pi_{max}^{(S_r^{(i)})}$ in $S_r^{(i)}$. Now,

$$\begin{split} \Phi(S_r) &= \frac{\sum_{j \in S_r^{(i)}, k \in \overline{S_r^{(i)}}} \pi_j^{(i)} \pi_j^{(i)}}{\sum_{j \in S_r^{(i)}} \pi_j^{(i)}} \\ &\leq \frac{\pi_{max}^{(S_r^{(i)})} d_{max}^{(i)} |W_r^{(i)}| \frac{1}{2d_{max}^{(i)}}}{\sum_{j \in S_r^{(i)}} \pi_j^{(i)}} \\ (\text{since } \forall j \in S_r^{(i)}, k \in \overline{S_r^{(i)}}, p_{j,k}^{(i)} = \frac{1}{2d_{max}^{(i)}}) \\ &\leq \frac{\pi_{max}^{(S_r^{(i)})} |W_r^{(i)}|}{2\eta^{(S_r^{(i)})} \pi_{max}^{(S_r^{(i)})}} \\ &= \frac{|W_r^{(i)}|}{2\eta^{(S_r^{(i)})}} \end{split}$$

Thus from Theorem 2 we have the following Theorem:

THEOREM 5. The family X of Markov chains as defined above is not successful if for all constants k > 0, we have instances $i \in I$ of arbitrarily large size such that for some $r, c_{min}^{(i)} \leq r < c_{max}^{(i)}$,

$$\frac{|W_r^{(i)}|}{2\eta^{(S_r^{(i)})}} < n_i^{-k}$$

5. CONCLUDING REMARKS

We have provided in this paper two characterizations of the success of the Metropolis algorithm for combinatorial optimization problems. As Section 4 shows, our results can be used in a straightforward manner for deciding the success of the algorithm for specific problems. One of our characterizations implies that the rapid mixing of the underlying Markov chains is a necessary condition for the success of the Metropolis algorithm. Further, as fairly powerful techniques, e.g., canonical paths, resistance, etc., are already available to argue about rapid mixing, our results open the possibility of making use of such techniques in the context of the success of the Metropolis algorithm.

6. **REFERENCES**

- M. Jerrum and G. B. Sorkin. Simulated Annealing for Graph Bisection. In Proc. Symp. on Foundations of Computer Science, pages 94–103, 1993.
- [2] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953.
- [3] G. Sasaki. The Effect of the Density of States on the Metropolis algorithm. *Information Processing Letters*, 37(5):158–163, 1991.
- [4] G. H. Sasaki and B. Hajek. The Time Complexity of Maximum Matching by Simulated Annealing. *Journal* of the ACM (JACM), 35(2):387–403, 1988.
- [5] A. Sinclair. Algorithms for Random Generation and Counting: A Markov Chain Approach. Birkháuser Verlag, 1993.
- [6] I. Wegener. Simulated Annealing Beats Metropolis in Combinatorial Optimization. In International Colloquium on Automata, Languages and Programming, 2005.

APPENDIX

A. PROOF OF THEOREM 1

We provide here a proof of Theorem 1.

Theorem 1 The family $X = \{X^{(i)} | i \in I\}$ be S-successful if and only if it is W-successful.

PROOF. (*if*)Let the family X is W-successful. Thus there exist constants $k_1, k_2, n_0 > 0$ such that $\forall i \in I$ with $n_i \geq n_0$,

$$\min_{s \in S^{(i)}} \mathbf{P}[\min\{t \ge 0 | X^{(i)}_t = s^{(i)}_{opt}\} \le n^{k_1}_i \mid X^{(i)}_0 = s] \ge n^{-k_2}_i$$

For each *i* such that $n_i \ge n_0$, we see the simulation of $X^{(i)}$ as a sequence of blocks, each of $n_i^{k_1}$ steps. Using the fact that the probability that the chain encounters $s_{opt}^{(i)}$ in some block is at least $n_i^{-k_2}$ for any starting state of the block, we fix a starting state *s* and bound Exp[l], where $s_{opt}^{(i)}$ is

reached at for the first time at *l*-th block. An upper bound on the expected number of steps before hitting $s_{opt}^{(i)}$ is then obtained by multiplying it by n^{k_1} . Thus, $\forall s \in S^{(i)}$,

$$\mathbf{E}[\{\min\{t \ge 0 | X_t^{(i)} = s_{opt}^{(i)}\}\} \mid X_0^{(i)} = s]$$

$$\leq n^{k_1} \sum_{j=1}^{\infty} j \cdot (1 - n_i^{-k_2})^{j-1} \cdot n_i^{-k_2}$$

$$= n^{k_1} \cdot n_i^{-k_2} \sum_{j=1}^{\infty} j \cdot (1 - n_i^{-k_2})^{j-1}$$

$$= n^{k_1} n_i^{-k_2} \cdot n_i^{2k_2}$$

$$= n_i^{k_1 + k_2}$$

Thus, $\forall i \in I$ with $n_i \geq n_0$,

$$\mathbf{E}[\max_{s \in S^{(i)}} \{\min\{t \geq 0 | X^{(i)}_t = s^{(i)}_{opt}\}\} \mid X^{(i)}_0 = s] \leq n^{k_1 + k_2}_i$$

Hence the family X is S-successful.

(only if)Let the family X is S-successful. Thus there exist constants $k, n_0 > 0$ such that $\forall i \in I$ such that $n_i \ge n_0$,

$$\mathbf{E}[\max_{s \in S^{(i)}} \{\min\{t \geq 0 | X^{(i)}_t = s^{(i)}_{opt}\}\} \mid X^{(i)}_0 = s] \leq n^k_i$$

For each *i* such that $n_i \geq n_0$, we consider a phase consisting of $2cn_i^k \lg n_i$ steps, where c > 0 is a constant. We fix a starting state $s_1 \in S^{(i)}$. We imagine the phase to be composed of $c \lg n_i$ blocks, each of $2n_i^k$ steps. Now by *Markov's inequality*, $\forall p, 1 \leq p \leq c \lg n_i, \forall s_p \in S$, the probability that $X^{(i)}$ does not reach $s_{opt}^{(i)}$ in *p*-th block, given that the start-

ing state of *p*-th block is s_p , is at most $\frac{n_i^k}{2n_i^k} = \frac{1}{2}$.

We introduce a family of indicator random variables $\{Y_j : 1 \le j \le c \lg n_i\}$, such that

$$Y_j = \begin{cases} 1, & \text{If } X^{(i)} \text{ encounters } s_{opt}^{(i)} \text{ in the } j\text{-th block}; \\ 0, & \text{otherwise.} \end{cases}$$

Let B_j denote the staring state of the *j*-th block. The probability that $s_{opt}^{(i)}$ is not encountered in the entire run is:

$$\begin{split} \mathbf{P}[\bigcap_{j=1}^{c \lg n} [Y_j = 0]] \\ &= \sum_{s_2, \dots, s_{c \lg n_i} \in S^{(i)}} \mathbf{P}[\bigcap_{j=1}^{c \lg n_i} [Y_j = 0] \bigcap_{j=2}^{c \lg n_i} [B_j = s_j]] \\ &= \sum_{s_2, \dots, s_{c \lg n_i} \in S^{(i)}} \mathbf{P}[\bigcap_{j=1}^{c \lg n_i} [Y_j = 0 \mid B_j = s_j]] \mathbf{P}[\bigcap_{j=2}^{c \lg n_i} [B_j = s_j]] \\ &= \sum_{s_2, \dots, s_{c \lg n_i} \in S^{(i)}} \prod_{j=1}^{c \lg n} \mathbf{P}[Y_j = 0 \mid B_j = s_j] \mathbf{P}[\bigcap_{j=2}^{c \lg n_i} [B_j = s_j]] \\ &\leq \sum_{s_2, \dots, s_{c \lg n_i} \in S^{(i)}} 2^{-c \lg n_i} \mathbf{P}[\bigcap_{j=2}^{c \lg n_i} [B_j = s_j]] \\ &= n_i^{-c} \end{split}$$

Thus for all starting state $s_1 \in S^{(i)}$, with probability at least $1 - n_i^{-c}$, $X^{(i)}$ reaches $s_{opt}^{(i)}$ within $2cn_i^k \lg n_i$ steps. Thus X is W-successful. \Box