Combinatorics, Probability and Computing (2001) 10, 29–40. Printed in the United Kingdom
© 2001 Cambridge University Press

Simulated Annealing and Graph Colouring†

ANDREAS NOLTE and RAINER SCHRADER‡

Universität zu Köln, Weyertal 80, 50931 Köln, Germany
(e-mail: a.nolte@allianz.de, rs@zpr.uni-koeln.de)

Received 27 April 1999; revised 18 January 2000

Simulated annealing is a very successful heuristic for various problems in combinatorial optimization. In this paper an application of simulated annealing to the 3-colouring problem is considered. In contrast to many good empirical results we will show for a certain class of graphs that the expected first hitting time of a proper colouring, given an arbitrary cooling scheme, is of exponential size.

These results are complementary to those in [13], where we prove the convergence of simulated annealing to an optimal solution in exponential time.

1. Introduction

Simulated annealing has been a very successful general algorithm for the solution of large combinatorial optimization problems. It is a random local search heuristic that has received much attention since it was first introduced in [8].

In this paper we consider the simulated annealing algorithm applied to the 3-colouring problem, which is known to be NP-complete [2]. Let $G = (V, E)$ be an undirected graph with n vertices. A 3-colouring of G is a mapping $f : V \to \{1, 2, 3\}$. f is a proper 3-colouring, if $f(u) \neq f(v)$ for all $(u, v) \in E$. The graph 3-colouring problem is to decide whether a given graph G has a proper 3-colouring. According to the experimental results in [7], simulated annealing is a very successful heuristic for this problem. In contrast to these practical observations we will show the existence of two classes of graphs for which the expected first hitting time of a proper colouring is of exponential size. For the first class of graphs we assume a fixed cooling temperature for the simulated annealing algorithm. We can show for the first time an exponential first hitting time nearly without any limitations on the starting state. In the second class there is one natural starting state given (the empty colouring). Here we prove an exponential first hitting time of

† Extended version of [14].
‡ Supported by DFG grant (Schr 390/2-1).

https://doi.org/10.1017/S0963548300004557 Published online by Cambridge University Press

the maximum colouring even for an arbitrary cooling schedule. There are only a few negative results concerning the efficiency of simulated annealing applied to problems in combinatorial optimization. Observations of a similar flavour have been established by Jerrum [6] for the clique problem on random graphs and by Sasaki and Hajek [16] for the matching problem.

This shows that the exponential running time bounds for simulated annealing proved by the authors in [13] are essentially tight.

2. A homogeneous approach

First, we consider a special case of simulated annealing called the Metropolis algorithm [10]. It performs the simulated annealing strategy with a fixed temperature $T \in \mathbb{R}^+$. We describe one natural way to apply the algorithm to the 3-colouring problem.

To start the algorithm, let an arbitrary colouring f be given; $c(f)$ denotes the number of 'bad' edges (u,v) with $f(u) = f(v)$. We choose uniformly at random a vertex $v \in V$ and a colour $j \in \{1,2,3\}$ and colour the vertex with j. We move to the resulting colouring f' if $c(f') \leq c(f)$. If $c(f') > c(f)$, we move to f' with probability $\exp((c(f) - c(f'))/T)$ and stay at f with probability $1 - \exp((c(f) - c(f'))/T)$. We call the neighbourhood structure defined by all possible transitions N. The sequence of states visited by the algorithm forms a Markov chain $(X_t)_{t\in\mathbb{N}}$ on the state space Ω of all colourings. It is easily seen that this Markov chain is irreducible and aperiodic, hence ergodic. Therefore there exists a unique probability vector π^T with $\pi^T P = \pi^T$, where P is the transition matrix of the chain with

$$p_{fg} = \begin{cases} \min\left\{1, e^{\frac{c(f)-c(g)}{T}}\right\} \frac{1}{3n}, & \text{for } f \neq g,\ (f,g) \in N, \\ 0, & \text{for } f \neq g,\ (f,g) \notin N, \\ 1 - \sum_{h\neq f} p_{fh}, & \text{for } f = g, \end{cases}$$

for all $(f,g) \in \Omega^2$. The important fact is that the distribution of the chain $P(X_t = .)$ tends to π^T as $t \to \infty$. It is easily verified that the Gibbs distribution

$$\pi_f^T = \frac{e^{\frac{-c(f)}{T}}}{\sum_{h\in\Omega} e^{\frac{-c(h)}{T}}}$$

is the stationary distribution of the chain. By letting $T \to 0$ we can see that π^T converges to the uniform distribution concentrated on the optimal solutions. Therefore we could expect a good asymptotic performance of the chain if we choose a very 'deep' temperature T that is only slightly larger than 0 [9].

In the following we are interested in the performance of the chain depending on the problem size, that is, convergence bounds of the chain after a finite number of steps. We will show in this section that, at least for general graphs, good performance bounds do not exist. We do this by defining a class of graphs where the expected first hitting time of a proper colouring will be of exponential size, even when we assume that the starting state could be chosen uniformly at random from all possible starting states. Our example gives stronger evidence of the sometimes bad performance of simulated annealing than those in [6] and [16], where only the existence of one starting state with exponential first hitting time of the optimum is shown.

https://doi.org/10.1017/S0963548300004557 Published online by Cambridge University Press

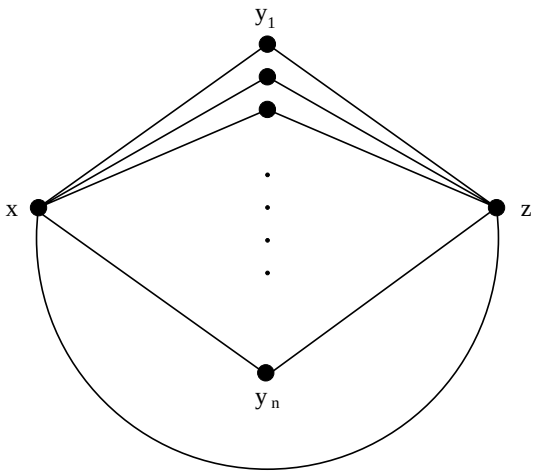


Figure 1 Graph G_n

Let $G_n = (V_n, E_n)$ with $V_n = \{x, z, y_1, \ldots, y_n\}$ and $E_n = \{(x, y_i), (y_i, z), i = 1, \ldots, n\} \cup \{(x, z)\}$ (Figure 1). Obviously the graph G_n has only one proper partition in 3 parts and therefore only 6 proper colourings.

Theorem 2.1. *Almost all colourings f of G_n have an expected first hitting time H_f of a proper 3-colouring of G_n of exponential size, when f is selected as a starting state.*

Before we prove the theorem we want to present the idea of the proof and give some helpful lemmas. In the following we will define a mapping $g : \Omega \to \mathbb{N}$ from the set of all possible colourings to the natural numbers. A proper colouring f will have the value $g(f) = 0$ and nearly all others will have positive values. A transition of the Markov chain will change the value of g by at most 1. We will consider the process $g(X_t)$, which is no longer necessarily Markovian, and show under the hypothesis $X_t = f$ with certain colourings f that the expected value after one transition step is greater than $g(X_t) + \epsilon, \epsilon > 0$. With the help of a result in [4] on hitting time bounds implied by drift analysis, we are able to prove that, given a certain starting state, the expected first hitting time of a colouring $X_t = f$ with $g(f) = 0$ is of exponential size. By specializing Theorem 2.3 of [4] to the discrete state space we get the following.

Theorem 2.2 (Hajek). *Let $(Y_j)_{j \in \mathbb{N}}$ be a sequence of random variables with values in $\mathbb{N}$ and $a, b \in \mathbb{R}, b < a$. The conditions*

C1 $E(|Y_{j+1} - Y_j| \, | Y_j = i_j, \ldots, Y_0 = i_0) \leq s$

C2 $E(Y_{j+1} - Y_j - \epsilon, Y_j < a, H_b > j | Y_j = i_j, \ldots, Y_0 = i_0) \geq 0$
with $\epsilon > 0, H_b = \min\{j | Y_j = b\}$ (*first hitting time of b*)

imply

$$\exists D, \eta \in \mathbb{R} \quad \forall i_0 \geq a, t \in \mathbb{N} \qquad P(H_b = t | Y_0 = i_o) \leq D e^{\eta(b-a)}$$

(D, η *depend only on s and* ϵ, *not on a or b*). □

https://doi.org/10.1017/S0963548300004557 Published online by Cambridge University Press

Suppose f is a colouring of G_n with 3 colours. Then the set $\{y_1, \dots y_n\}$ of nodes can be divided into at most 3 parts according to the colouring. Let $g(f) : \Omega \to \mathbb{N}$ be the size of the second-largest part.

Lemma 2.3. *Let*

$$c = \frac{1}{4}\frac{e^{-2/T}}{2 + e^{-2/T}}, \quad \epsilon = \frac{e^{-2/T}}{12}$$

and let f be a colouring of G_n with $0 < g(f) \leq cn$. Let $X_t = f$. Then $E(g(X_{t+1}) - g(X_t)) \geq \epsilon$ follows.

Proof. Let $g(f) = k$. By considering the transition probabilities of the graph we get

$$P(g(X_{t+1}) = k + 1) \geq \frac{e^{-2/T}}{3(n+2)}(n-k)$$

and

$$P(g(X_{t+1}) = k - 1) \leq \frac{2}{3(n+2)}k.$$

Therefore we can conclude

$$\begin{aligned} & E(g(X_{t+1}) - g(X_t)) \\ \geq & \left(\frac{e^{-2/T}}{3(n+2)}(n-k) - \frac{2}{3(n+2)}k\right) \\ \geq & \frac{e^{-2/T}}{12} \\ = & \epsilon \end{aligned}$$

as required. □

To count those colourings with $g(f) > cn$, the following version of the Chernoff bound [12] is helpful. Let $B(n, p)$ denote the binomial distribution. By definition, if $x \in B(n, p)$, then $P(x = k) = \binom{n}{k} p^k (1-p)^{n-k}$.

Proposition 2.4. *If $x \in B(n, p)$, then for all α with $0 < \alpha < 1$ we have $P(x \leq (1-\alpha)np) < e^{-\alpha^2 np/2}$ and $P(x \geq (1+\alpha)np) < e^{-\alpha^2 np/3}$.*

Lemma 2.5. *Let c be as in Lemma 2.3. All but an exponentially small fraction of the colourings of G_n satisfy $g(f) > cn$.*

Proof. Let z and y be coloured arbitrarily and let Y_r be the set of vertices y_i that are coloured red. By assuming that every colouring of $\{y_1, \dots, y_n\}$ has the same probability $1/3^n$, we get that the colour red is $B(n, 1/3)$ distributed. Therefore the last proposition implies

$$P(|Y_r| < cn) < e^{-(1-3c)^2 n/6} < e^{-n/96}.$$

By applying the same argument to the other colours, we get

$$P(g(f) < cn) \leq 3e^{-n/96}$$

as required. □

https://doi.org/10.1017/S0963548300004557 Published online by Cambridge University Press

Now we are able to prove the main theorem.

Proof. Let $X_0 = f_0$ with $g(f_0) > cn$ be the starting state of the Metropolis process. We consider only a subprocess of the actual Markov chain. Let Z_j be the points in time at which the value g of the actual colouring changes: $Z_0 = 0$ and, for all $j \in \mathbb{N}$, $Z_j = \min\{j > Z_{j-1} | g(X_j) \neq g(X_{Z_{j-1}})\}$. Suppose f is a colouring with $0 < g(f) < cn$. With Lemma 2.3 we get

$$E(g(X_{Z_{j+1}}) - g(X_{Z_j}) | X_{Z_{j+1}-1} = f, X_{Z_{j+1}-2}, \ldots, X_0) \geq \epsilon,$$

for all $j \in \mathbb{N}$. Summing over the disjoint given events we get

$$E(g(X_{Z_{j+1}}) - g(X_{Z_j}) - \epsilon | 0 < g(X_{Z_j}) < cn, X_{Z_j}, \ldots, X_{Z_0}) \geq 0,$$

and hence

$$E(g(X_{Z_{j+1}}) - g(X_{Z_j}) - \epsilon | 0 < g(X_{Z_j}) < cn, g(X_{Z_j}) = i_j, \ldots, g(X_{z_0}) = i_0) \geq 0.$$

With $Y_j = g(X_{Z_j}), a = cn$ and $H_0 = \min\{t | Y_t = 0\}$ we obtain

$$E(Y_{j+1} - Y_j - \epsilon, Y_j < a, H_0 > j | Y_j = i_j, \ldots, Y_0 = i_0) \geq 0,$$

as required for condition C1 of Theorem 2.2. Since $|Y_{j+1} - Y_j| \leq 1$, condition C2 is also fulfilled, and so

$$\exists D, \eta \in \mathbb{R} \quad \forall i_0 \geq a, k \in \mathbb{N} \qquad P(H_0 = k | Y_0 = i_o) \leq De^{\eta(-a)} =: \delta.$$

Using the fact that $X_0 = f_0$, we get

$$P(H_0 = t | Y_0 = g(f_0)) = P(H_0 = t | X_0 = f_0),$$

and

$$P(H_0 > t | X_0 = f_0) = 1 - \sum_{k=1}^{t} P(H_0 = k | X_0 = f_0) \geq \max\{0, 1 - t\delta\}.$$

Combining the last results yields

$$E(H_0 | X_0 = f_0) = \sum_{t=0}^{\infty} P(H_0 > t | X_0 = f_0) \geq \sum_{t=0}^{\infty} \max\{0, 1 - t\delta\} \geq \frac{1}{2\delta}.$$

Since $X_t =$ (proper 3-colouring) implies $g(X_t) = 0$, we conclude

$$E(H_0 | X_0 = f_0) \geq \frac{1}{2\delta} = \frac{De^{\eta cn}}{2} = e^{\Omega(n)}. \qquad \square$$

Remark. We suspect that one can establish an exponential first hitting time even for an arbitrary time-dependent sequence T_t of temperatures. Unfortunately we have not been able to prove this inhomogeneous case for X_t up to now, but in the next section we establish this for another application of the simulated annealing algorithm to the 3-colouring problem.

https://doi.org/10.1017/S0963548300004557 Published online by Cambridge University Press

3. A different approach

In this section we consider another approach for solving the 3-colouring problem. According to some positive experimental results [7] we try to find a maximum 3-colourable induced subgraph of a given graph $G = (V, E)$ (of maximal cardinality). Obviously the problem is also NP-hard.

Therefore the state space Ω of the corresponding Markov chain $(X_t)_{t\in\mathbb{N}}$ consists of all proper 3-colourings of arbitrary subgraphs of G. We no longer consider the homogeneous case and let $(T_t)_{t\in\mathbb{N}}$ be an arbitrary sequence of temperatures. Given a state $f \in \Omega$ at time t, we choose uniformly at random a vertex $v \in V$. If v is already coloured, we remove v from the colouring with probability $\exp(-1/T_t)$ and stay at f with probability $1 - \exp(-1/T_t)$. If v is not yet coloured, we choose one of the 3 colours uniformly at random and try to colour v. If we get a proper colouring, we move to this colouring. Otherwise we stay at f.

3.1. Convergence to the optimal solution

In this subsection we briefly describe the known theoretical bounds of the convergence of the distribution of the Markov chain to a distribution that is concentrated on the optimal states (*i.e.*, maximal 3-colourable subgraphs). We start with a description of the underlying inhomogeneous Markov chain.

According to the above-defined strongly connected neighbourhood structure $N \subset \Omega^2$ of the 3-colourings of subgraphs of G, we get the following transition probabilities. For technical reasons we also assume that we stay at each state with probability at least $1/2$, so that we obtain, for all $(f, g) \in \Omega^2$,

$$p_{fg}^{(t)} = \begin{cases} \min\left\{1, e^{\frac{|g|-|f|}{T_t}}\right\}\frac{1}{2n}, & \text{for } f \neq g,\ (f,g) \in N, \\ 0, & \text{for } f \neq g,\ (f,g) \notin N, \\ 1 - \sum_{f'\neq f} p_{ff'}, & \text{for } f = g, \end{cases}$$

where $|f| = \sharp\{v|v \text{ is coloured by f}\}$. Clearly $p_{ff}^{(t)} \geq 1/2$, so that the chain is aperiodic. Assuming fixed transition probabilities p_{fg}^t (fixed temperature T_t), the resulting homogeneous chain is also ergodic and

$$\pi(t)_f = \frac{e^{\frac{|f|}{T_t}}}{\sum_g e^{\frac{|g|}{T_t}}} = \frac{t^{\frac{|f|}{\gamma}}}{\sum_g t^{\frac{|g|}{\gamma}}}$$

is the stationary distribution. Let $c_{\max} = \max\{|f|\ |f \in \Omega\}$. Obviously

$$\pi^* = \lim_{t\to\infty} \pi(t)_f = \lim_{t\to\infty} \frac{1}{\sum_g t^{\frac{|g|-|f|}{\gamma}}} = \begin{cases} 0, & \text{if } |f| < c_{\max}, \\ \frac{1}{\sharp\{f||f|=c_{\max}\}}, & \text{if } |f| = c_{\max}. \end{cases}$$

The limit distribution is therefore a constant probability vector in which probability charges are concentrated on the global optima of the solution space.

Almost all results in the inhomogeneous case assume a logarithmic cooling schedule $T_t = \gamma/\ln(t)$, where γ is a problem-dependent parameter. Asymptotic results have been published by Hajek [5] and Tsitsiklis [17] who were even able to find necessary and

https://doi.org/10.1017/S0963548300004557 Published online by Cambridge University Press

sufficient conditions for the convergence of simulated annealing. However, since we are concerned with finite time bounds, these asymptotic results are not helpful for us.

Anily and Federgruen [1], Gidas [3] and Mitra, Romeo and Sangiovanni-Vincentelli [11] have obtained independently similar deterministic upper bounds for the proximity of the probability distribution of the state space after the generation of t transitions to the uniform distribution on the set of optimal states π^*. Applied to the 3-colouring problem, the bound of [11], for example, would yield

$$t \geq c_1 \frac{1}{\epsilon^{n(3n)^n}} \quad \Rightarrow \quad \sum_f |P(X(t) = f) - \pi_f^*| \leq \epsilon$$

for any $c_1 \in \mathbb{N}$. In [13] we were able to improve these running time bounds by an application of the theory of rapidly mixing Markov chains. We can guarantee the same inequality with a considerably smaller bound $t \geq \epsilon^{-n} \exp(n^{c_2})$ with $c_2 \in \mathbb{N}$.

In the next section we will show that there is no hope of further improving this bound; in other words, we will prove that the exponential bound in [13] is essentially tight.

3.2. Negative results

In the following we give an example of a class of graphs for which the expected first hitting time of a maximal 3-colourable subgraph is of exponential size when we start with the empty colouring and allow arbitrary cooling schedules.

Consider the following family of graphs. Let $G_r = (R_1 \cup R_2, E_r)$ with $R_1 = \{r_i, 1 \leq i \leq n\}, R_{2_j} = \{r_{jk}, 1 \leq k \leq n, j \neq k\}, 1 \leq j \leq n$, $R_2 = \cup_{j=1}^n R_{2_j}$ and $E_r = \{\{r_i\} \times R_{2_i}\} \cup \{\cup_{1 \leq j < k \leq n} R_{2_j} \times R_{2_k}\} \cup \{\cup_{1 \leq i,j \leq n} \{r_j, r_{ij}\}\}\}$, where the $\times$ operation is understood to form a set of two element sets, thus creating a set of edges.

The structure of this graph is as follows. There are n vertices in the set R_1 and for each of these $r_i \in R_1$ there is an associated set of vertices R_{2_i}. This associated set has $n-1$ elements labelled r_{i1} to r_{in} with the diagonal element r_{ii} missing.

The edges come in three distinct families. The first joins every vertex $r_i \in R_1$ with every vertex r_{ij} in its associated set. The second turns the sets R_{2_j} into complete n-partite graphs by connecting all the associated sets with each other. The third and final collection of edges joins every element r_{jk} to a specific vertex in R_1, namely r_k.

By this construction we can see that each of the sets R_{2_j} is independent and of size $n-1$, and that R_1 is independent and of size n.

Now define G_b and G_g as copies of G_r. G_n is formed by considering G_r, G_b, G_g as subgraphs and adding all edges between vertices that are in different graphs G_*.

Lemma 3.1. *The subgraph induced by $R_1 \cup B_1 \cup G_1$ is the maximal 3-colourable subgraph of G_n.*

Proof. $G(R_1 \cup B_1 \cup G_1)$ is obviously 3-colourable. Since every set of vertices coloured with the same colour has to be independent, we look for independent sets of size at least n. As any two nodes in different subgraphs G_r, G_b and G_g are joined by an edge, every independent set is completely contained in one of these subgraphs. Because of symmetry it suffices to look at G_r.

https://doi.org/10.1017/S0963548300004557 Published online by Cambridge University Press

We prove here that R_1 is the largest independent set of vertices in G_r. Let U be an independent set in $G(R_1 \cup R_2)$. Clearly, if $R_2 \cap U = \emptyset$ then $U \subset R_1$ and the claim is obvious. Now suppose $R_2 \cap U \neq \emptyset$. We want to show that $|U| < n$.

Since the sets R_{2_i} form a complete n-partite graph the intersection of U with R_2 must be entirely in R_{2_j} for some j. Consider all edges of the form $\{r_i, r_{ji}\}$ and observe that at most one end-vertex of each of these edges can be in the set U. There are $n-1$ such edges and no other vertex of the graph can be in U. This completes the proof that R_1 is the largest independent set in G_r and the lemma follows. □

Similarly to the first section we will define a mapping $\omega : \Omega \to \mathbb{N}$ with $\omega(\emptyset) = n$ and $\omega(\text{maximal colouring}) = 0$. We will consider the process $\omega(X_t)$, which is no longer necessarily Markovian. Under the hypothesis $X_t = f$ with $\omega(X_t) < a$ we show that the expected value after one transition step is greater than $\omega(X_t) + \epsilon, \epsilon > 0$. Moreover, the difference $|\omega(X_t) - \omega(X_{t-1})|$ is bounded. Then we are able to prove with the help of Theorem 2.2 that the expected first hitting time of a colouring $X_t = f$ with $\omega(f) = 0$ is of exponential size, given a starting state h with $\omega(h) > a$.

Suppose f is a proper 3-colouring of a subgraph of G_n and F is the set of the coloured vertices. Let

$$\omega_r(F) = 3\sharp F \cap R_2 + \max\{\sharp\{\text{uncoloured vertices in } R_1\} - 1, 0\}$$

and define $\omega_b(F)$ and $\omega_g(F)$ analogously. Let

$$\omega(F) = \max\{\omega_r(F), \omega_g(F), \omega_b(F)\}.$$

It is easily seen that $\omega(\emptyset) = n-1$ and $\omega(R_1 \cup G_1 \cup B_1) = 0$. Let $\omega(F) < n/2$. Then we can derive from the structure of G_n that all vertices coloured with the same colour must be contained in one of the sets $R_1 \cup R_2, G_1 \cup G_2, B_1 \cup B_2$. We now consider those transitions which colour an additional vertex. Let

$$\begin{aligned} M_\cup(F) &= \{v \in V_n | \omega(F \cup \{v\}) \neq \omega(F)\}, \\ M_\cup^+(F) &= \{v \in V_n | \omega(F \cup \{v\}) > \omega(F)\}, \\ M_\cup^-(F) &= \{v \in V_n | \omega(F \cup \{v\}) < \omega(F)\}, \end{aligned}$$

($F \cup \{v\}$ denotes the set of coloured vertices obtained by colouring the additional vertex v).

Lemma 3.2. *Suppose* $0 < \omega(F) < n/2$. *Then* $\sharp M_\cup^-(F) \leq \sharp M_\cup^+(F)$.

Proof. We assume that $\omega(F)$ is achieved only on $\omega_r(F)$. From the construction of G_r we observe that $M_\cup^- \subset R_1$ and $M_\cup^+ \subset R_2$. We distinguish two cases. In each case we show that for every vertex that can be added to R_1 there is a vertex that can be added to R_2 instead.

***F* ∩ *R*₂ = ∅.** Since $\omega_r(F) > 0$ we know that $M_\cap^-(F)$ must have at least two elements, say r_i and r_j. Then for every vertex r_k that can be added to R_1 we can enlarge R_2 by either r_{ik} (if $i \neq k$) or by r_{jk} otherwise.

https://doi.org/10.1017/S0963548300004557 Published online by Cambridge University Press

$\boldsymbol{F \cap R_2 \neq \emptyset}$**.** Since for $i \neq j$ every vertex of R_{2_i} is connected to every vertex of R_{2_j} we observe that $F \cap R_2 \subset R_{2_j}$ for some j. By the construction of G_r we can deduce that $r_j \notin F$. Consider all edges of the form $\{r_i, r_{ji}\}$ and observe again that at most one element of each of these edges can be in the set F. Let $r_i \in R_1$ be a vertex that can be added to F. Since $\{r_i, r_{ji}\}$ is an edge, we observe that vertex r_{ji} can be added to R_2.

This completes the proof. □

Now we consider the transitions that remove a vertex from a colouring. Let

$$\begin{aligned} M_{\setminus}(F) &= \{v \in V_n | \omega(F \setminus \{v\}) \neq \omega(F)\}, \\ M_{\setminus}^{+}(F) &= \{v \in V_n | \omega(F \setminus \{v\}) > \omega(F)\}, \\ M_{\setminus}^{-}(F) &= \{v \in V_n | \omega(F \setminus \{v\}) < \omega(F)\}, \end{aligned}$$

($F \setminus \{v\}$ denotes the set of coloured vertices obtained by uncolouring v in the given colouring f).

Lemma 3.3. *Suppose* $\omega(F) < n/12$. *Then* $M_{\setminus}^{-} \neq \emptyset$ *implies* $\sharp M_{\setminus}^{-} \leq n/36$ *and* $\sharp M_{\setminus}^{+} \geq 11/12n - 1$.

Proof. First we suppose $\omega(F) = \omega_r(F) \wedge \omega(F) > \omega_b(F) \wedge \omega(F) > \omega_g(F)$. Then $\sharp M_{\setminus}^{-} = |F \cap R_2| \leq n/36$. Moreover, we obtain $M_{\setminus}^{+} \supset F \cap R_1$ from the construction of G_n, and $\omega(F) < n/12$ yields $\sharp\{\text{uncoloured vertices in } R_1\} \leq n/12 + 1$. Then $\sharp M_{\setminus}^{+} \geq \sharp F \cap R_1 \geq 11/12n - 1$ as required. If $\omega(F) = \omega_r(F) = \omega_b(F)$ then $M_{\setminus}^{-} = \emptyset$ and the lemma follows trivially. The remaining cases follow by symmetry. □

Combining these two lemmas we will now show that, given a $\sharp F$-increasing or decreasing move, the expected value of the difference $\omega(X_{t+1}) - \omega(X_t)$ is separated from 0.

Lemma 3.4. *Let* $0 < \omega(F) < n/12$. *Then*

(1) $\sum_{\cup}(F) := |M_{\cup}(F)|^{-1} \sum_{v \in M_{\cup}(F)} (\omega(F \cup \{v\}) - \omega(F)) \geq 1$, *if* $M_{\cup}(F) \neq \emptyset$,
(2) $\sum_{\setminus}(F) := |M_{\setminus}(F)|^{-1} \sum_{v \in M_{\setminus}(F)} (\omega(F \setminus \{v\}) - \omega(F)) \geq 3/4$, *if* $M_{\setminus}(F) \neq \emptyset$.

Proof. From straightforward calculation with the help of Lemmas 3.2 and 3.3, we have the following.

(1)

$$\begin{aligned} \sum_{\cup}(F) &= |M_{\cup}(F)|^{-1} \left(\sum_{v \in M_{\cup}^{-}(F)} (\omega(F \cup \{v\})) - \omega(F) + \sum_{v \in M_{\cup}^{+}(F)} (\omega(F \cup \{v\}) - \omega(F)) \right) \\ &\geq |M_{\cup}(F)|^{-1} \left((-1)|M_{\cup}^{-}(F)| + 3|M_{\cup}^{+}(F)| \right) \\ &\geq 1. \end{aligned}$$

(2) This is analogous to (1). □

https://doi.org/10.1017/S0963548300004557 Published online by Cambridge University Press

Finally we can prove the main theorem.

Theorem 3.5. *Let* $H_m = \min\{t \in \mathbb{N} | X_t = \text{maximal 3-colouring}\}$. *Then* $E(H_m | X_0 = \emptyset) = \exp(\Omega(n))$.

Proof. The argument is analogous to that in the proof of Theorem 2.1. First we define the points Z_j in time at which the value ω of the actual colouring changes: $Z_0 = 0$ and, for all $j \in \mathbb{N}$, $Z_j = \min\{j > Z_{j-1} | \omega(X_j) \neq \omega(X_{Z_{j-1}})\}$. Suppose f is a 3-colouring with $0 < \omega(F) < n/12$. Let $M_t = +$ if the tth transition colours an uncoloured vertex, $M_t = -$ if a vertex is removed from the colouring and $M_t = 0$ if the chain stays in the same state. From Lemma 3.4 we get, for all $j \in \mathbb{N}, f \in \Omega$ with $0 < \omega(F) < n/12$,

$$E(\omega(X_{Z_{j+1}}) - \omega(X_{Z_j}) | X_{Z_{j+1}-1} = f, M_{Z_{j+1}-1} = +, X_{Z_{j+1}-2}, \ldots, X_0) \geq 1,$$

and analogously

$$E(\omega(X_{Z_{j+1}}) - \omega(X_{Z_j}) | X_{Z_{j+1}-1} = f, M_{Z_{j+1}-1} = -, X_{Z_{j+1}-2}, \ldots, X_0) \geq 3/4.$$

Summing over the disjoint given events we get

$$E(\omega(X_{Z_{j+1}}) - \omega(X_{Z_j}) - 3/4 | 0 < \omega(X_{Z_j}) < n/12, X_{Z_j}, \ldots, X_{Z_0}) \geq 0,$$

owing to

$$E(f | \dot{\cup} A_i) = \frac{\sum_i P(A_i) E(f | A_i)}{P(\dot{\cup} A_i)},$$

and hence

$$E(\omega(X_{Z_{j+1}}) - \omega(X_{Z_j}) - 3/4 | 0 < \omega(X_{Z_j}) < n/12, \omega(X_{Z_j}) = i_j, \ldots, \omega(X_{z_0}) = i_0)$$

is not negative. With $Y_j = \omega(X_{Z_j}), a = n/12, \epsilon = 3/4$ and $H_0 = \min\{t | y_t = 0\}$ we obtain

$$E(Y_{j+1} - Y_j - \epsilon, Y_j < a, H_0 > j | Y_j = i_j, \ldots, Y_0 = i_0) \geq 0,$$

as required for condition C1 of Theorem 2.2. Since $|Y_{j+1} - Y_j| \leq 3$, condition C2 is also fulfilled and it follows that

$$\exists D, \eta \in \mathbb{R} \quad \forall i_0 \geq a, t \in \mathbb{N} \qquad P(H_0 = t | Y_0 = i_o) \leq D e^{\eta(-a)} =: \delta.$$

Using the fact that $\omega(\emptyset) = n - 1$ and $X_0 = \emptyset$, we get

$$P(H_0 = t | Y_0 = n - 1) = P(H_0 = t | X_0 = \emptyset)$$

and

$$P(H_0 > t | X_0 = \emptyset) = 1 - \sum_{k=1}^{t} P(H_0 = k | X_0 = \emptyset) \geq \max\{0, 1 - t\delta\}.$$

(We could assume $P(H_0 < \infty | X_0 = \emptyset) = 1$, because otherwise the proof of the claim would be trivial.) Combining the last results yields

$$E(H_0 | X_0 = \emptyset) = \sum_{t=0}^{\infty} P(H_0 > t | X_0 = \emptyset) \geq \sum_{t=0}^{\infty} \max\{0, 1 - t\delta\} \geq \frac{1}{2\delta}.$$

https://doi.org/10.1017/S0963548300004557 Published online by Cambridge University Press

Because $X_t =$ (maximal 3-colouring) implies $\omega(X_t) = 0$, we get

$$E(H_m | X_0 = \emptyset) \geq \frac{1}{2\delta} = De^{\eta n},$$

as required. □

4. Closing remarks

We have shown that there is no hope of further improving the exponential running time bounds for simulated annealing.

As already pointed out, we suspect that an exponential first hitting time can be established in the case of the first approach even for arbitrary cooling schedules and not only for the Metropolis process with constant temperature. A similar definition of ω with constant values on the energy levels as in the second approach does not seem to be applicable since the difference $c(f) - c(f')$ of two neighbouring colourings is not bounded by a constant.

Despite the negative results in this paper we believe that the performance of simulated annealing on certain random graphs is much better than on the graphs considered in this paper. The reason for this is the smooth structure of these graphs where the ratio of optimal to near-optimal solutions is not quite as bad as here. The authors have succeeded in showing convergence of the Metropolis process on certain random graphs with high probability [15] but the proof that annealing is beneficial for some combinatorial optimization problems is still missing.

Acknowledgement

The authors would like to thank an anonymous referee for careful proof-reading, for pointing out an error in an earlier version of the paper and for making a number of helpful suggestions.

References

[1] Anily, S. and Federgruen, A. (1987) Ergodicity in parametric nonstationary Markov chains: An application to simulated annealing methods. *Oper. Research* **35** 867–874.
[2] Garey, M. R. and Johnson, S. J. (1979) *Computers and Intractability*, Freeman.
[3] Gidas, B. (1985) Nonstationary Markov chains and convergence of the annealing algorithm. *J. Statist. Phys.* **39** 73–131.
[4] Hajek, B. (1982) Hitting and occupation time bounds implied by drift analysis with applications. *Adv. Appl. Probab.* **14** 502–525.
[5] Hajek, B. (1988) Cooling schedules for optimal annealing. *Math. Oper. Research* **13** 311–329.
[6] Jerrum, M. (1992) Large cliques elude the Metropolis process. *Random Struct. Alg.* **3** 347–360.
[7] Johnson, D. S., Aragon, C. R., McGeoch, L. A. and Schevon, C. (1991) Optimization by simulated annealing: An experimental evaluation. *Oper. Research* **39** 378–406.
[8] Kirkpatrick, S., Gelatt, C. D. and Vecchi, M. P. (1983) Optimization by simulated annealing. *Science* **220** 671–680.
[9] Laarhoven, P. J. M. and Aarts, E. H. L. (1989) *Simulated Annealing: Theory and Applications*, Kluwer.

https://doi.org/10.1017/S0963548300004557 Published online by Cambridge University Press

[10] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E. (1953) Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087–1092.
[11] Mitra, D., Romeo, F. and Sangiovanni-Vincentelli, A. L. (1986) Convergence and finite-time behavior of simulated annealing. *Adv. Appl. Probab.* **18** 747–771.
[12] Motwani, R. and Raghavan, P. (1995) *Randomized Algorithms*, Cambridge University Press.
[13] Nolte, A. and Schrader, R. (2000) A note on the finite time behaviour of simulated annealing. *Math. Oper. Research* **25** 476–484.
[14] Nolte, A. and Schrader, R. (1996) Simulated annealing and its problems to color graphs. In *Proceedings of the ESA 1996*, Vol. 1136 of *Springer Lecture Notes in Computer Science*, pp. 138–152.
[15] Nolte, A. and Schrader, R. (1997) Coloring in sublinear time. In *Proceedings of the ESA 1997*, Vol. 1284 of *Springer Lecture Notes in Computer Science*, pp. 388–401.
[16] Sasaki, G. H. and Hajek, B. (1988) The time complexity of maximum matching by simulated annealing. *J. Assoc. Comput. Mach.* **35** 387–403.
[17] Tsitsiklis, J. (1989) Markov chains with rare transitions and simulated annealing. *Math. Oper. Research* **14** 70–90.

https://doi.org/10.1017/S0963548300004557 Published online by Cambridge University Press