

# CONVERGENCE RATE OF MCMC AND SIMULATED ANNEALING WITH APPLICATION TO CLIENT-SERVER ASSIGNMENT PROBLEM

THUAN DUONG-BA, THINH NGUYEN, BELLA BOSE

*School of Electrical Engineering and Computer Science,  
Oregon State University, USA*

Email address: {duongba, thinhq, bose}@eecs.oregonstate.edu

## Abstract

Simulated annealing (SA) algorithms can be modeled as time-inhomogeneous Markov chains. Much work on the convergence rate of simulated annealing algorithms has been well-studied. In this paper, we propose an adiabatic framework for studying simulated annealing algorithm behavior. Specifically, we focus on the problem of simulated annealing algorithms that start from an initial temperature  $T_0$  and evolve to  $T_{final}$  which are pre-specified, and remain at the final temperature so that the solution will be adaptive to the dynamical changes of the system.

*Keywords:* Distributed optimization, distributed graph partitioning, simulated annealing, adiabatic time analysis.

## 1. Introduction

Proposed by Kirkpatrick , Gelatt and Vecchi [1], Simulated annealing (SA) is a probabilistic optimization method [2, 1, 3]. Initially, starting at a random state at a high temperature, SA process changes to a neighbor state which has lower energy, to higher energy neighbors or with Boltzmann's law of probability. Temperature often decreases over time such that when the algorithm reaches a low enough temperature,

---

\* Postal address:

1148 Kelley Engineering Center  
Oregon State University  
Corvallis, OR 97331, USA

the system is frozen at an expected distribution so that the system will remain at the lowest energy states with the highest probabilities. The Markov chain associated with an SA algorithm is actually in-homogeneous during the period that temperature is decreasing. At a fixed temperature, i.e frozen at a (low) temperature, SA exhibits as a time-homogeneous Markov chain.

The homogeneous Markov chain convergence property has been well investigated in [4]. The *mixing time*, the soonest time that the homogeneous Markov chains converge to a unique stationary distribution, was well studied in [5, 6]. However, *adiabatic time* or the time taken to converge, of in-homogeneous Markov chains, i.e Simulated annealing algorithm, has not been deeply analyzed. In [7], the authors proposed an analysis framework for Simulated Annealing algorithms which exhibits ergodic property focusing on the distance to the optimal distribution. On the other hand, our work focuses on the distance to stationary distribution at a specified target temperature. There are some applications of simulated annealing algorithm that stops decreasing temperature but continue running so that the system can continue to optimize and adapt to the dynamical changes.

Adiabatic analysis frameworks for continuous time-inhomogeneous Markov Chain have been proposed in several works [8, 9, 10]. However, author in these works focused on the family of in-homogeneous Markov Chain that evolve with the following rule:

$$P(t) = \phi(t)P(0) + (1 - \phi(t))P(T_f) \quad (1.1)$$

where  $P(0), P(T_f)$  are initial and final transition matrices respectively;  $\phi(t) : [0, \infty) \rightarrow [0, 1]$  is the evolution function and monotonically decreasing. On the other hand, the transition matrices associated with Simulated Annealing algorithms do not exhibit such evolution property. In fact, evolution of each element in the transition matrix of a Simulated Annealing algorithm is exponential with its own parameter.

Adiabatic time characterizes the in-homogeneous Markov chain convergence which is an analog to the mixing time of the time-homogeneous Markov chain. It measures the time taken to converge, i.e the total variance distance of the distribution to the final stationary distribution. In this paper, we concentrate on the convergence properties of SA algorithm based on Ergodic Coefficient with application to the Client-Server

assignment problems. Specifically, we will evaluate the adiabatic time of SA algorithms with decreasing temperatures to a specified final temperature.

The structure of the paper is organized as follows: Section 2 lists the background related to Markov chain and general Simulated annealing algorithms. Section 3 concentrates on analyzing the adiabatic time of the SA algorithms. Section 4 formulates the Client-Server assignment problem and proposes a simulated annealing based algorithm as well as derives the adiabatic bounds for different cooling schemes based on the framework presented in Section 3. Section 5 shows some simulation results. Finally, section 6 concludes the paper.

## 2. Preliminaries

This section introduces a general simulated annealing algorithm and summarizes some Markov chain properties that are needed for the main focus of the paper.

### 2.1. Simulated Annealing algorithm

Given a finite set of states  $\Omega$  and a real-valued cost function  $F$  defined on  $\Omega$  and a transition matrix  $Q$  ( $Q_{ij} \geq 0$  and  $\sum_j Q_{ij} = 1$ ). Each state  $i \in \Omega$  has a set of neighbor states  $\mathcal{N}(i)$ , i.e  $Q_{ij} > 0, \forall j \in \mathcal{N}(i)$  and  $Q_{ij} = 0$  otherwise. The simulated annealing algorithm is listed in algorithm 1.

---

#### Algorithm 1 Simulated annealing

---

- 1: Set  $T_b = T_0$  - an initial temperature
  - 2: Set  $X = X_0$  - an initial state
  - 3: **repeat**
  - 4:     Select  $Y \in \mathcal{N}(X)$
  - 5:     **if**  $F(Y) < F(X)$  **then**
  - 6:         Set  $X \leftarrow Y$
  - 7:     **else**
  - 8:         Set  $X \leftarrow Y$  with probability  $\exp[-\frac{F(Y)-F(X)}{T_b}]$
  - 9:     **end if**
  - 10:    Decrease  $T_b$
  - 11: **until** Freeze
-

The kernel of simulated annealing algorithms is the probabilistic transition from one state to another which is controlled by a parameter ( $T_b$ ) (a metaphor of controlling temperature). Initially, simulated annealing will choose an initial (normally arbitrary) state, and an initial temperature which is chosen empirically depending on problems. At each iteration, a neighbor state is selected to be considered whether or not being accepted as the new state. The transition probability depends on the different energy (cost function value) at the current state  $F$ . If the selected state has lower energy, the transition is surely taken; otherwise, the transition will happen with a probability of  $\exp(-\frac{\Delta F}{T_b})$ . The temperature  $T_b$  will decrease over time and the system will freeze when  $T_b$  is small enough.

At each specific value of temperature  $T_b$ , there is an associated transition matrix which exhibits as a homogeneous Markov Chain. When  $T_b$  decreases over time, transition matrices will evolve toward the final transition matrix corresponding to the frozen states of the algorithm. This transition exhibits properties of an in-homogeneous Markov Chain.

## 2.2. Markov Chain properties

A finite Markov chain is a process which changes its internal state (configurations) within a set of finite states with a fixed probability distribution and the transition probabilities depend only on the current state [5]. Let  $X_t \in \Omega, t = 0, 1, 2, \dots$  be a chain of states at time  $t$  and  $P$  be the transition matrix; the Markov chain can be represented by the following Markov property:

$$P\{X_{t+1} = y_{t+1} | X_t = y_t, X_{t-1} = y_{t-1}, \dots, X_0 = y_0\} = P\{X_{t+1} = y_{t+1} | X_t = y_t\} \quad (2.1)$$

where  $y_i \in \Omega$  and  $t$  is discrete time.

If  $p_{ij} = P\{X_{t+1} = i | X_t = j\}$  is independent of  $t$ , the Markov chain is said to be time-homogeneous. A matrix  $P$  whose elements are  $p_{ij}$  is the transition matrix.

**Definition 2.1.** (Stationary distribution.)

*If there exist a distribution  $\pi$  over state space  $\Omega$  such that*

$$\pi^T = \pi^T P \quad (2.2)$$

*then  $\pi$  is called a stationary distribution.*

**Definition 2.2.** (Total variance distance.)

Let  $\mu$  and  $\nu$  be any two distributions over state space  $\Omega$ , the total variance distance between  $\mu$  and  $\nu$  is defined as:

$$\|\mu - \nu\|_{TV} = \sup_{A \subset \Omega} |\mu(A) - \nu(A)| = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)| \quad (2.3)$$

In other words, the total variance distance quantifies the difference between two probability distributions over the same state space.

**Definition 2.3.** Separation [4](p.219)

The separation of  $\mu$  from  $\nu$ , denoted by  $s(\mu; \nu)$ , is defined by

$$s(\mu; \nu) = \max_{i \in \Omega} \left(1 - \frac{\mu(i)}{\nu(i)}\right) \quad (2.4)$$

**Observation 2.1.** [4]

$$0 \leq s(\mu; \nu) \leq 1 \quad (2.5)$$

**Proposition 2.1.** [4]

$$\|\mu - \nu\|_{TV} \leq s(\mu; \nu) \quad (2.6)$$

**Definition 2.4.** The Ergodic Coefficient [4]

Let  $P$  be a stochastic matrix indexed by  $\Omega \times \Omega$ . Its Dobrushin's ergodic coefficient  $\tau(P)$  is defined by:

$$\tau(P) = \frac{1}{2} \sup_{i, j \in \Omega} \sum_{k \in \Omega} |p_{ik} - p_{jk}| = \sup_{i, j \in \Omega} \|p_i - p_j\|_{TV}$$

The ergodic coefficient characterizes the long-term behavior of dynamical systems.

**Observation 2.2.** [4]

$$0 \leq \tau(P) = 1 - \inf_{i, j \in \Omega} \sum_{k \in \Omega} \min\{p_{ik}, p_{jk}\} \leq 1$$

**Proposition 2.2.** [4]

$$\tau(P_1 P_2) \leq \tau(P_1) \tau(P_2) \quad (2.7)$$

**Proposition 2.3.** [4] Let  $P$  be a stochastic matrix indexed by  $E$ , and let  $\mu$  and  $\nu$  be two probability distributions on  $E$ , then

$$\|\mu P^n - \nu P^n\|_{TV} \leq \tau(P)^n \|\mu - \nu\|_{TV} \quad (2.8)$$

### 3. Analysis Framework for Simulated annealing

**Definition 3.1.** (Cooling scheme.) *A sequence of temperatures  $\{T_k > 0\}, k = [0, 1, 2, \dots]$  is called a cooling scheme if it is strictly decreasing and goes to 0. Mathematically,*

$$\begin{cases} T_k > T_{k+1} > 0 \\ \lim_{k \rightarrow \infty} T_k = 0 \end{cases}$$

**Definition 3.2.** (Transition Matrix Evolution.) *Given an initial temperature  $T_0$ , a target temperature  $T_{final}$  and a slow cooling scheme as defined in Definition 3.1 consisting of  $K$  steps such that  $T_K \leq T_{final} < T_{K-1}$  are used as parameters of a simulated annealing algorithm. Let  $P_k$  be the transition matrix of the corresponding Markov chain at time  $k$ . The sequence of matrices  $\{P_k\}, 0 \leq k \leq K$  is the evolution of the in-homogeneous Markov chain's transition matrix.*

**Definition 3.3.** (Adiabatic time.)

*The existence of stationary distribution  $\pi_k$  of the corresponding Markov chain at time step  $k$  has been proved [7]. Given  $\epsilon > 0$ , the adiabatic time of the in-homogeneous Markov chain is defined as follows:*

$$K(\epsilon) = \inf \left\{ K : \max_{\nu} \|\nu P_0 P_1 \dots P_K - \pi_K\|_{TV} \leq \epsilon \right\} \quad (3.1)$$

*with  $\nu$  is any initial probability distribution over  $\Omega$  and  $P_t$  is the corresponding transition matrix evolution defined in Definition 3.2.*

The adiabatic time presents how gradually the transition matrix of a in-homogeneous Markov chain evolves from  $\nu_0$  (initial distribution) to  $\nu_K$  (final distribution) such that the final distribution is close to the stationary distribution of the final transition matrix. In simulated annealing optimization problem, the adiabatic time  $K(\epsilon)$  always converges since the transition matrices converge to the final transition matrix.

**Proposition 3.1.** *For any cooling scheme,  $\forall k_1, k_2$  such that  $0 < k_1 < k_2 \leq K$  and  $F_L, F_U$  are lower bound and upper bound of the cost function, the total variant distance between two corresponding stationary distributions is bounded by:*

$$\|\pi_{k_1} - \pi_{k_2}\|_{TV} \leq \Delta F \left( \frac{1}{T_{k_2}} - \frac{1}{T_{k_1}} \right) \quad (3.2)$$

*Proof.*

$$\begin{aligned}
\|\pi_{k_1} - \pi_{k_2}\|_{TV} &\leq s(\pi_{k_1}, \pi_{k_2}) \\
&\triangleq \max_i \left\{ 1 - \frac{\pi_{k_1}(i)}{\pi_{k_2}(i)} \right\} \\
&= 1 - \min_i \left( \frac{\pi_{k_1}(i)}{\pi_{k_2}(i)} \right) \\
&= 1 - \frac{\pi_{k_1}(i^*)}{\pi_{k_2}(i^*)} \leq \epsilon_1
\end{aligned} \tag{3.3}$$

where

$$i^* = \operatorname{argmin}_i F(i) \tag{3.4}$$

We also have

$$\begin{aligned}
\frac{\pi_{k_1}(i^*)}{\pi_{k_2}(i^*)} &= \frac{C_{k_2}}{C_{k_1}} \times \frac{e^{-F(i^*)/T_{k_1}}}{e^{-F(i^*)/T_{k_2}}} \\
&\geq \frac{\sum_i e^{-F_U/T_{k_2}}}{\sum_i e^{-F_U/T_{k_1}}} \times \frac{e^{-F_L/T_{k_1}}}{e^{-F_L/T_{k_2}}} \\
&= e^{(F_L - F_U)(\frac{1}{T_{k_2}} - \frac{1}{T_{k_1}})}
\end{aligned} \tag{3.5}$$

hence

$$\|\pi_{k_1} - \pi_{k_2}\|_{TV} \leq 1 - e^{-\Delta F(\frac{1}{T_{k_2}} - \frac{1}{T_{k_1}})}$$

which finishes the proof (by Taylor expansion of the exponential function).

**Proposition 3.2.** Denote  $P_L(l) = \prod_{k=l-L+1}^l P(k)$ ,  $l \geq 1$ . If the graph associated with the Markov Chain at a given temperature  $T_k$  is a connected graph, there exists  $L_0$  such that

$$\begin{aligned}
\tau(P_L(l)) &= 1 \text{ if } L < L_0 \\
\text{And } \tau(P_L(l)) &< 1 - \gamma_L^l \text{ if } L \geq L_0
\end{aligned} \tag{3.6}$$

Where  $\gamma_L^l = \frac{1}{(\max_{i \in \Omega} |\mathcal{N}(i)|)^L} e^{-\frac{L\Delta F}{T_i}}$  and  $\Delta F = F_U - F_L$ ,  $\mathcal{N}(i)$  is the set of state  $i$ 's neighboring states.

*Proof.* Since the graph associated with the Markov Chain at a given temperature  $T_k$  is not a fully connected graph, there exists a state  $s_0$  such that there is no direct transition from all other states into  $s_0$ . Hence from Observation 2.2 we have  $\tau(P) = 1$ .

Let  $L_0$  be the diameter of the Markov Chain graph at a given temperature  $T_k$ , i.e

$$L_0 = \min_{x,y \in \Omega} \max |\gamma_{xy}|. \quad \forall L \geq L_0 \text{ we have:}$$

$$\tau(P_L(l)) = 1 - \inf_{i,j} \sum_k \min\{P_L(l)_{[i,k]}, P_L(l)_{[j,k]}\} \quad (3.7)$$

$$\leq 1 - \prod_{k=l-L+1}^l \min_{i,j \in \Omega} P_{ij}(k) \quad (3.8)$$

$$\leq 1 - \prod_{k=l-L+1}^l \frac{1}{\max_{i \in \Omega} |\mathcal{N}(i)|} e^{-\Delta F/T_k} \quad (3.9)$$

$$\leq 1 - \frac{1}{(\max_{i \in \Omega} |\mathcal{N}(i)|)^L} \prod_{k=l-L+1}^l e^{-\Delta F/T_k} \quad (3.10)$$

$$\leq 1 - \frac{1}{(\max_{i \in \Omega} |\mathcal{N}(i)|)^L} e^{-L\Delta F/T_l} \quad (3.11)$$

$$= 1 - \gamma_L^l \quad (3.12)$$

Where (3.7) follows from Definition of  $P_L(l)$  and Observation 2.2, (3.8) follows from expanding  $P_L(l)$ , (3.9) follows from how to select transition probability in Simulated Annealing, (3.10) is trivial.

**Note:** If  $L$  is chosen such that the aggregation matrix  $P_L(l)$  has all positive columns we can obtain the following value of  $\gamma_L(l)$ :

$$\gamma_L^l = \frac{|\Omega|}{(\max_{i \in \Omega} |\mathcal{N}(i)|)^L} e^{-\frac{L\Delta F}{T_l}} \quad (3.13)$$

**Corollary 3.1.** For any  $0 < L < L_0, l \geq 0$

$$\|\pi_l P_l \dots P_{l+L-1} - \pi_{l+L}\|_{TV} \leq \sum_{k=l}^{l+L-1} \|\pi_k - \pi_{k+1}\|_{TV} \quad (3.14)$$

---

Positive column has all positive elements [11]



*Proof.*

$$\begin{aligned} \|\pi_l P_l \dots P_{l+L-1} - \pi_{l+L}\|_{TV} &\leq \|\pi_l P_l \dots P_{l+L-1} - \pi_{L-1}\|_{TV} \\ &\quad + \|\pi_{l+L-1} - \pi_{l+L}\|_{TV} \end{aligned} \quad (3.15)$$

$$\begin{aligned} &\leq \|\pi_l P_l \dots P_{l+L-2} - \pi_{L-1}\|_{TV} \tau(P_{l+L-1}) \\ &\quad + \|\pi_{l+L-1} - \pi_{l+L}\|_{TV} \end{aligned} \quad (3.16)$$

$$\begin{aligned} &\leq \|\pi_l P_l \dots P_{l+L-2} - \pi_{L-1}\|_{TV} \\ &\quad + \|\pi_{l+L-1} - \pi_{l+L}\|_{TV} \end{aligned} \quad (3.17)$$

$$\leq \sum_{k=l}^{l+L-1} \|\pi_k - \pi_{k+1}\|_{TV} \quad (3.18)$$

Where (3.15) follows from triangle inequality, (3.16) and (3.17) follow from Proposition 2.3, and (3.18) follows from expanding (3.17) recursively.

**Corollary 3.2.** *For any cooling scheme, and  $0 < L < K$  we have:*

$$\sum_{l=K-L}^{K-1} \|\pi_l - \pi_{l+1}\|_{TV} \leq \Delta F \left( \frac{1}{T_{l+L}} - \frac{1}{T_l} \right) \quad (3.19)$$

*Proof.* The Corollary follows directly Corollary 3.1 and Proposition 3.1.

**Theorem 3.1.** *Let  $\nu_K, \pi_K$  be the distribution of the in-homogeneous Markov chain and the stationary distribution at time  $K$  respectively. Given  $T_0, T_K$ , and a cooling scheme, i.e  $T_0 \geq T_k > T_{k+1} \geq T_K, 0 \leq k < K$ , if there exists two non-increasing functions  $\alpha_1(k), \alpha_2(k)$  satisfying  $0 < \alpha_1(K) < \gamma_L^K; 0 < \alpha_2(K) < 1$  and  $\sum_{l=K-L}^{K-1} \|\pi_l - \pi_{l+1}\|_{TV} \leq \alpha_1(K)\alpha_2(K)$ , specifically  $\alpha_2(K)$  is strictly decreasing, the total variance distance between  $\nu_K$  and  $\pi_K$  is bounded by:*

$$\|\nu_K - \pi_K\|_{TV} \leq \max \left\{ \prod_{k=1}^{\frac{K-K_0}{L}} (1 - \gamma_L^{kM} + \alpha_1(K)) \|\nu_{K_0} - \pi_{K_0}\|_{TV}; \alpha_2(K)(1 + \alpha_1(K)) \right\}$$

where  $K_0$  is any reference time step such that  $K - K_0 \equiv 0 \pmod{L}$ .

*Proof.*

$$\begin{aligned} \|\nu_K - \pi_K\|_{TV} &\leq \|\nu_{k-L} P_{k-L+1} \dots P_K - \pi_{K-L} P_{K-L+1} \dots P_K\|_{TV} \\ &\quad + \|\pi_{K-L} P_{K-L+1} \dots P_K - \pi_K\|_{TV} \end{aligned} \quad (3.20)$$

$$\begin{aligned} &\leq \tau(P_L(K)) \|\nu_{K-L} - \pi_{K-L}\|_{TV} \\ &\quad + \sum_{l=K-L}^{K-1} \|\pi_l - \pi_{l+1}\|_{TV} \end{aligned} \quad (3.21)$$

$$\begin{aligned} &\leq (1 - \gamma_L^K) \|\nu_{K-L} - \pi_{K-L}\|_{TV} \\ &\quad + \sum_{l=K-L}^{K-1} \|\pi_l - \pi_{l+1}\|_{TV} \end{aligned} \quad (3.22)$$

The theorem is finished by plugging the following cases into (3.22).

**Case 1:**  $\|\nu_{K-L} - \pi_{K-L}\|_{TV} > \alpha_2(K)$ , we have

$$\sum_{l=K-L}^{K-1} \|\pi_l - \pi_{l+1}\|_{TV} \leq \alpha_1(K) \|\nu_{K-L} - \pi_{K-L}\|_{TV}$$

Plugging into (3.22) we obtain,

$$\|\nu_K - \pi_K\|_{TV} \leq (1 - \gamma_L^K + \alpha_1(K)) \|\nu_{K-L} - \pi_{K-L}\|_{TV}$$

Continuing recursively we obtain the 1<sup>st</sup> term of the Theorem.

**Case 2:**  $\|\nu_{K-L} - \pi_{K-L}\|_{TV} \leq \alpha_2(K)$ . Plugging into (3.22) we obtain the 2<sup>nd</sup> term of the Theorem.

$$\|\nu_K - \pi_K\|_{TV} \leq (1 - \gamma_L^K) \alpha_2(K) + \sum_{l=K-L}^{K-1} \|\pi_l - \pi_{l+1}\|_{TV} \quad (3.23)$$

$$\leq \alpha_2(K) + \alpha_1(K) \alpha_2(K) \quad (3.24)$$

$$= \alpha_2(K) (1 + \alpha_1(K)) \quad (3.25)$$

**Corollary 3.3.** *The adiabatic time can be bounded by deriving from the above Theorem depending on which cooling scheme is chosen.*

$$K_A(\epsilon) \leq \max\{K_1(\epsilon), K_2(\epsilon)\}$$

where

$$K_1(\epsilon) = \inf\left\{K : \prod_{k=1}^{\frac{K-K_0}{L}} (1 - \gamma_L^{kM} + \alpha_1(K)) \|\nu_{K_0} - \pi_{K_0}\|_{TV} \leq \epsilon\right\}$$

and

$$K_2(\epsilon) = \inf\{K : \alpha_2(K)(1 + \alpha_1(K)) \leq \epsilon\}$$

*Proof.* This Corollary comes directly from letting the upper bound in Theorem 3.1 be smaller or equal to a small positive number  $\epsilon$ .

**Note:** The quality of the upper bound on the adiabatic time depends on how the two parameters  $\alpha_1$  and  $\alpha_2$  are chosen and bounded.

**Theorem 3.2.** (Convergence rate at fixed temperature  $T_K$ .) *Let  $\epsilon$  be an upper bound of  $\|\nu_K - \pi_K\|_{TV}$ . and let  $\delta$  be a positive number such that  $0 < \delta < \epsilon$ , and  $\tau_L^B$  be an upper bound of  $\tau(P^L)$ . The time when the Markov Chain corresponding to the Simulated Annealing at a fixed temperature  $T_K$  converges, i.e  $\|\nu_K P^{K_M} - \nu_K\|_{TV} \leq \delta$ , is bounded by:*

$$K_M(\epsilon, \delta) \leq L \frac{\log \delta - \log \epsilon}{\log \tau_L^B} \quad (3.26)$$

*Proof.* Without loss of generality, assume that  $K_M \equiv 0 \pmod{L}$

$$\|\nu_K - \pi\|_{TV} \leq \tau(P^L) \|\nu_K - \pi_K\|_{TV} \quad (3.27)$$

$$\leq \tau(P^L)^{K_M/L} \|\nu_K - \pi_K\|_{TV} \quad (3.28)$$

$$\leq \epsilon (\tau_L^B)^{K_M/L} \quad (3.29)$$

$$\leq \delta \quad (3.30)$$

The theorem follows.

## 4. Application to Client-Server Assignment

### 4.1. Problem formulation

In this section, we will formulate the Client-Server assignment problem mathematically. Assume that there are  $M$  users needing to be assigned to  $N$  servers. The communication pattern of users is represented by a graph  $G = (V, E)$  where  $V$  denotes the set of  $M$  users ( $|V| = M$ ) and  $E$  denotes the friendship among users. If two users have messages exchanged, it will form an edge in the graph  $G$ . Let matrix  $A$  be the adjacent matrix corresponding to the user graph  $G$ . Elements in  $A$  depict the friendship between users having strictly positive values, otherwise their values are 0.

The processing load at servers as well as the inter-server communication load incurred by messages sent from one user to another and vice versa are identical, hence  $A$  is symmetric. Now we define a matrix  $X$  to represent a valid assignment scheme. Each user will associate to a row in  $X$  with columns' values representing which server is the primary server of that user. Specifically,  $X[u, i] = 1$  if server  $i$  is the primary server of user  $u$  and 0 otherwise. We also assume that each user must have exactly one primary server, i.e  $\sum_{i=1}^N X[u, i] = 1, \forall 0 \leq u \leq M, 0 \leq i \leq N$ .

The global total system load  $S(X)$  for an assignment  $X$  is computed as follow:

$$S(X) = \|X^T A X\|_1 - \frac{1}{2} \mathbf{Tr}(X^T A X) \quad (4.1)$$

where  $\|\cdot\|_1$  denotes the entry-wise  $l1$ -matrix norm, i.e., the sum of all the entries in the matrix.

For the purpose of evaluating servers' imbalance, we first define the average load of all the servers as:

$$\bar{S}(X) = \frac{S(X)}{N} = \frac{\|X^T A X\|_1 - \frac{1}{2} \mathbf{Tr}(X^T A X)}{N} \quad (4.2)$$

The load imbalance at server  $i$  is defined as :

$$\Delta S_i(X) \equiv S_i(X) - \bar{S}(X), \quad (4.3)$$

where  $S_i(X)$  denotes the total load at server  $i$ .

We now propose the global objective function depending on the global total load and the maximum server's imbalance:

$$F(X) = \alpha S(X) + \beta \max_i \Delta S_i(X), \quad (4.4)$$

where  $\alpha$  and  $\beta$  are pre-specified weighted positive coefficients.

The Client-Server assignment problem can be stated as the following optimization problem:

$$\begin{aligned} \text{Minimize:} \quad & F(X) \\ \text{Subject to:} \quad & X_{ui} \in \{0, 1\} \\ & \sum_i X_{ui} = 1 \\ & u \in \{1, 2, \dots, M\} \\ & i \in \{1, 2, \dots, N\} \end{aligned}$$

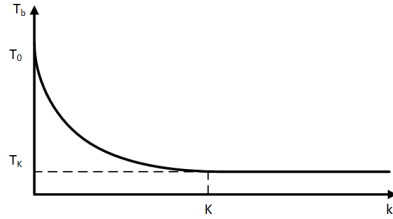


FIGURE 1: Temperature decreasing scheme for simulated annealing based algorithm

We note that minimizing  $F(X)$  implies minimizing the total load and the load imbalance. However, as previously discussed, to obtain the lowest total load (putting all the users in one server), it is necessary that the load imbalance increase. Therefore, tuning coefficients  $\alpha$  and  $\beta$  allow for a trade-off between the total load and the load balance. At one extremity, setting  $\alpha = 0$  implies minimizing the load balance regardless of the total load, while setting  $\beta = 0$  implies minimizing the total load regardless of the load balance. Details on the Client-Server assignment problem can be found at [12].

#### 4.2. Simulated Annealing based Algorithm

Based on computations above, we propose an algorithm for assigning users to servers appropriately based on simulated annealing as shown in Algorithm 2.

Initially, an arbitrary assignment scheme (state) is selected at time step  $k = 0$  with corresponding temperature  $T_0$ . At each iteration, select a neighbor of current assignment scheme with uniform probability. If the selected state has lower objective function's value, accept immediately; otherwise accept with probability of  $\exp(-\frac{\Delta F}{T_k})$ . Increment  $k$  and decrease  $T_k$  (if  $T_k > T_{final}$ ). Continue with a new iteration. We keep the system running forever at temperature  $T_k$  so that the dynamic changes, which are reflected in changes of  $F(X)$ , can be taken into account.

##### 4.2.1. in-homogeneous Markov chain

1. State space: Let  $\Omega$  be the set of all possible assignment schemes.  $|\Omega| = N^M$ . An assignment scheme is called a neighbor of another scheme if their matrices differ in exactly 1 row. There are  $N^M$  assignment schemes.

---

**Algorithm 2** Simulated annealing based algorithm
 

---

```

1: Initialize  $Tb$ 
2: while true do
3:   Select an user  $u$ 
4:   Select a new server  $\{s : X[u, s] = 0\}$  with probability  $\frac{1}{N-1}$ 
5:   Calculate future objective function's value  $F_{new}$ 
6:   if  $F_{new} < F$  then
7:     Switch user  $u$  to new server  $s$ 
8:   else
9:     Switch user  $u$  to new server  $s$ 
10:    with probability  $\exp(-\frac{F-F_{new}}{Tb})$ 
11:   end if
12:   if  $Tb > T_K$  then
13:     Decrease  $Tb$ 
14:   end if
15: end while

```

---

2. Neighboring state: Given an assignment scheme  $X$ , a new scheme is selected by choosing a user and changing its server. There are  $M$  users and  $N - 1$  new servers, hence each assignment scheme has  $M(N - 1)$  neighbors. Consequently, the probability of selecting a neighbor is  $p_{XY} = \frac{1}{M(N-1)}$ .

**Claim 4.1.** *At a fixed temperature  $T_k$ , the stationary distribution  $\pi_k$  of the corresponding Markov chain exists.*

*Proof.* The acceptance probability is  $\min\{1, \exp(-\frac{F(Y)-F(X)}{T_k})\}$ , hence the transition matrix  $P_k$  of the corresponding Markov chain at time  $T_k$  can be described as follows:

$$P_k(X, Y) = \begin{cases} \frac{1}{M(N-1)} & \text{if } Y \in \mathcal{N}(X) \wedge F(Y) \leq F(X) \\ \frac{1}{M(N-1)} \exp(-\frac{F(Y)-F(X)}{T_k}) & \text{if } Y \in \mathcal{N}(X) \wedge F(Y) > F(X) \\ 1 - \sum_{j \in \mathcal{N}(X)} P_k(X, j) & \text{if } X \equiv Y \\ 0 & \text{o.w} \end{cases}$$

Let  $\pi_k$  be an distribution defined by:

$$\pi_k(X) = C_k \exp\left(-\frac{F(X)}{T_k}\right)$$

where  $C_k = \sum_{x \in \Omega} \exp\left(-\frac{F(x)}{T_k}\right)$  For any two solution  $X, Y$ , without loss of generality assuming that  $F(X) \leq F(Y)$ , we have

$$\pi_k(X)P_k(X, Y) = C_k \exp\left(-\frac{F(X)}{T_k}\right) p_{XY} \exp\left(-\frac{F(Y) - F(X)}{T_k}\right) \quad (4.5)$$

$$= C_k \exp\left(-\frac{F(Y)}{T_k}\right) p_{XY} \quad (4.6)$$

$$= \pi_k(Y)P_k(Y, X) \quad (4.7)$$

hence the detailed balance equation satisfies and  $\pi_k$  is the stationary distribution of the corresponding Markov chain at time  $k$ .

**Claim 4.2.** (All positive columns.) *The aggregation matrix during  $M$  time steps has all positive columns. In other words, after  $M$  time steps any assignment scheme can be changed into any other new assignment scheme.*

*Proof.* Let  $X, Y$  be two arbitrary distinct assignment schemes. Let  $X(i, :)$  and  $Y(i, :)$  be the  $i^{\text{th}}$  row in  $X$  and  $Y$  respectively. A neighbor of  $X$  can be selected by changing one of its row, i.e change the column that has the value of 1 in one row.

Define  $\mathcal{D}(X, Y) = \{i : X(i, :) \neq Y(i, :)\}$ ,  $|\mathcal{D}(X, Y)|$  is the Hamming distance between two assignment schemes, i.e the necessary number of users in  $X$  need to be reassigned to achieve  $Y$ . We have  $0 \leq |\mathcal{D}(X, Y)| \leq M$ .

**Corollary 4.1.** Define  $\Delta F = \max_{X \in \Omega} \max_{Y \in \mathcal{N}(X)} |F(X) - F(Y)|$ .

$$\tau(P_M(l)) \leq 1 - \left(\frac{N}{M(N-1)}\right)^M \exp\left(-\frac{M\Delta F}{T_l}\right)$$

Or

$$\gamma_M^l = \left(\frac{N}{M(N-1)}\right)^M \exp\left(-\frac{M\Delta F}{T_l}\right)$$

*Proof.* Apply Proposition 3.2 and Claim 4.2.

**Observation:**  $\gamma_M^K = \left(\frac{N}{M(N-1)}\right)^M \exp\left(-\frac{M\Delta F}{T_{final}}\right)$  is independent of  $K$ .

## 4.2.2. Adiabatic time

## 1. Logarithmic cooling schemes

$$T_k = \frac{c}{k+d}$$

with  $c > 0, d > 1$  are chosen such that  $T_0 = c/\log(d)$  and  $T_{final} = c/\log(K+d)$ . The question here is how to choose  $K$  which is large enough so that the final distribution is close to the stationary distribution corresponding to a predefined final temperature.

**Claim 4.3.** Define  $K^* = \inf\{k : \frac{\Delta F}{\log k} < 1 \wedge K - k \equiv 0(\text{mod}M)\}$  and  $K^{**} = \inf\{K : \frac{\log(K-M)}{\log(K)} > 1 - \frac{T_K}{\Delta F} (\frac{N}{M(N-1)})^M \exp(-\frac{M\Delta F}{T_{final}})\}$ . With  $K > \max\{K^*, K^{**}\}$ , there exist two strictly decreasing functions  $\alpha_1(k), \alpha_2(k)$  such that  $\alpha_1(K) < \gamma_M^K$ ,  $0 < \alpha_2(K) < 1$  and  $\Delta F(\frac{1}{T_K} - \frac{1}{T_{K-M}}) \leq \alpha_1(K)\alpha_2(K)$

*Proof.*

$$\frac{\log(K-M)}{\log(K)} > 1 - \frac{T_K}{\Delta F} \gamma_M^K$$

$$\frac{\Delta F}{T_K} (1 - \frac{\log(K-M)}{\log(K)}) < \gamma_M^K$$

Let  $\alpha_1(K) = \frac{\Delta F}{\log K} \gamma_M^K$ .

and  $\alpha_2(K) = \frac{1}{T_K \gamma_M^K} \log \frac{K}{K-M}$

We have:

$$\Delta F(\frac{1}{T_K} - \frac{1}{T_{K-M}}) = \frac{\Delta F}{T_K} (1 - \frac{\log(K+d-M)}{\log(K+d)}) \quad (4.8)$$

$$< \frac{\Delta F}{T_K} (1 - \frac{\log(K-M)}{\log K}) \quad (4.9)$$

$$= \alpha_1(K)\alpha_2(K) \quad (4.10)$$

**Theorem 1.** With  $K \geq \max\{K^*, K^{**}\}$ , the total variance distance at time  $K$  is bounded by:

$$\|\nu_K - \pi_K\|_{TV} \leq \max \left\{ \begin{array}{l} [1 - (1 - \frac{\Delta F}{\log(K^*)}) \gamma_M^K]^{\frac{K-K^*}{M}} \|\nu_{K^*} - \pi_{K^*}\|_{TV}; \\ \frac{1}{T_K \gamma_M^K} \log(\frac{K}{K-M}) + \frac{\Delta F}{T_K} (1 - \frac{\log(K-M)}{\log(K)}) \end{array} \right\}$$



*Proof.* Without loss of generality,  $K$  is chosen such that  $K \equiv 0(\text{mod } M)$ . The theorem is proved by applying Theorem 3.1 and choosing  $\alpha_1(K), \alpha_2(K)$  as in Claim 4.3.

**Corollary 4.2.** *The adiabatic time is bounded by:*

$$K_A(\epsilon) \leq \max \left\{ \begin{array}{l} K^{**}; \\ K^* + M \frac{\log \epsilon - \log \|\nu_{K^*} - \pi_{K^*}\|_{TV}}{\log[1 - (1 - \frac{\Delta F}{\log K^*}) \gamma_M^K]}; \\ \frac{M \exp(\epsilon/2T_K \gamma_M^K)}{\exp(\epsilon/2T_K \gamma_M^K) - 1}; \\ \frac{2\Delta FM}{\epsilon T_K} \end{array} \right\}$$

*Proof.* Let the first elements in the 'max' operator of Theorem 1 be less than or equal to  $\epsilon$  and each term of the second element less than or equal to  $\epsilon/2$ , then the Corollary follows.

## 2. Exponential cooling schemes.

$$T_k = T_0 \left( \frac{T_{final}}{T_0} \right)^{k/K}$$

**Claim 4.4.** *Define  $K^{**} = \lceil \frac{M \log(T_{final}/T_0)}{\log[1 - (\frac{N}{M(N-1)})^M \exp(-\frac{M\Delta F}{T_{final}})(\frac{T_K}{T_0})^M \frac{T_K}{\Delta F}]} \rceil$ . With  $K > K^{**}$ , there exist two non-increasing functions  $\alpha_1(k), \alpha_2(k)$  such that  $\alpha_1(K) < \gamma_M^K$ ,  $0 < \alpha_2(K) < 1$  and  $\Delta F(\frac{1}{T_K} - \frac{1}{T_{K-M}}) = \alpha_1(K)\alpha_2(K)$*

*Proof.*

$$\gamma_M^K = \left( \frac{N}{M(N-1)} \right)^K \exp\left(-\frac{M\Delta F}{T_K}\right)$$

Let

$$\alpha_1(K) = \left( \frac{T_K}{T_0} \right) \left( \frac{N}{M(N-1)} \right)^M \exp\left(-\frac{M\Delta F}{T_K}\right)$$

and

$$\alpha_2(K) = \left( \frac{M(N-1)}{N} \right)^M \exp\left(\frac{M\Delta F}{T_K}\right) \left( \frac{T_0}{T_K} \right) \frac{\Delta F}{T_K} \left[ 1 - \left( \frac{T_K}{T_0} \right)^{M/K} \right]$$

We have:

$$\Delta F \left( \frac{1}{T_K} - \frac{1}{T_{K-M}} \right) = \frac{\Delta F}{T_K} \left[ 1 - \left( \frac{T_K}{T_0} \right)^{M/K} \right] \quad (4.11)$$

$$= \gamma_M^K \left( \frac{T_K}{T_0} \right) \frac{1}{\gamma_M^K} \left( \frac{T_0}{T_K} \right) \frac{\Delta F}{T_K} \left[ 1 - \left( \frac{T_K}{T_0} \right)^{M/K} \right] \quad (4.12)$$

$$= \alpha_1(K)\alpha_2(K) \quad (4.13)$$

Apparently,  $\alpha_1(K) < \gamma_M^K$  since  $T_K \leq T_{final} < T_0$ ; and with  $K > K^{**}$ ,  $\alpha_2(K) < 1$ .

**Theorem 4.1.**

$$\|\nu_K - \pi_K\|_{TV} \leq \max \left\{ \begin{array}{l} \|\nu_0 - \pi_0\|_{TV} \prod_{l=1}^{K/M} [1 - (1 - (\frac{T_K}{T_0}))\gamma_M^{lM}]; \\ \alpha_2(K)(1 + \alpha_1(K)) \end{array} \right\}$$

Where  $\alpha_1(K), \alpha_2(K)$  are defined in Claim 4.4

*Proof.* By choosing  $\alpha_1(K)$  and  $\alpha_1(K)$  as in Claim 4.4 and applying Theorem 3.1, the Theorem follows.

**Corollary 4.3.** *The bound on adiabatic time of the simulated annealing algorithms using exponential cooling schemes is:*

$$K_A(\epsilon) \leq \max \left\{ \begin{array}{l} K^{**}; \\ M \frac{\log \epsilon - \log(\|\nu_0 - \pi_0\|_{TV})}{\log(1 - (1 - \frac{T_K}{T_0})\gamma_M^K)}; \\ \frac{M \log(T_K/T_0)}{\log(1 - \frac{\epsilon T_K^2 \gamma_M^K}{\Delta F T_0 (1 + \frac{T_K}{T_0} \gamma_M^K)})} \end{array} \right\}$$

*Proof.*

### 3. Logarithmic cooling schemes with pre-specified parameters

$$T_k = \frac{c}{\ln(k+d)} \quad (4.14)$$

Where  $c, d$  satisfy  $T_0 = \frac{c}{\ln(d)}$ ,  $c > 0, d > 1$ .

In this type of cooling scheme, we fix parameter  $c, d$  so that  $T_0 = c/\log(d)$  and decide the upper bound of the adiabatic time. In other words, we will decide the final temperature  $T_K$  at which the total variance distance between the resulting distribution and the stationary distribution is small enough, i.e  $\|\nu_K - \pi_K\|_{TV} < \epsilon$ .

**Claim 4.5.** *The in-homogeneous Markov chain is ergdic if  $c > M\Delta F$*

*Proof.* Replacing  $T_k$  in Proposition 3.2 with the temperature decreasing rule in (4.14), we have:

$$\tau(P_M(l_0 M)) \leq 1 - \left(\frac{N}{M(N-1)}\right)^M (k+d)^{-\frac{M\Delta F}{c}} \quad (4.15)$$

With  $c > M\Delta F$  or equivalently  $M\Delta F/c < 1$ , we have:

$$\sum_{l=l_0}^{\infty} (1 - \tau(P_M(l_0 M))) = \infty \quad (4.16)$$

which finishes the proof according to [4].

**Proposition 4.1.** *There exists a temperature  $T^*$  such that  $\forall k \geq k^*$  where  $T_{k^*} = T^*$ ,*

$$\|\pi_k - \pi_{k+1}\|_{TV} > \|\pi_{k+1} - \pi_{k+2}\|_{TV}$$

*In other words, the total variance distance between stationary distributions of two consecutive time steps is decreasing.*

*Proof.* Let  $S^*$  be the set of states with global minimum objective values, i.e  $F(j) < F(i) \forall i \in S^* \wedge \forall j \in S \setminus S^*$ . There exist  $k_0 > 0$  such that  $\pi_k(i)$  is monotonically decreasing  $\forall i \notin S^*$  and  $\pi_k(i)$  is monotonically increasing if  $i \in S^*$  [7]. For  $\forall k > k_0$  we have

$$\|\pi_k - \pi_{k+1}\|_{TV} = \frac{1}{2} \sum_{i \in S^*} [\pi_{k+1}(i) - \pi_k(i)] + \frac{1}{2} \sum_{j \in S \setminus S^*} [\pi_k(j) - \pi_{k+1}(j)] \quad (4.17)$$

Therefore,

$$\sum_{k=k_0}^{\infty} \|\pi_k - \pi_{k+1}\|_{TV} \leq 1$$

Additionally, since  $\pi_k(j)$  is monotonically decreasing  $\forall j \notin S \setminus S^*$ , the  $2^{nd}$  term in (4.17) goes to 0 which finishes the proof.

**Theorem 4.2.** *Let  $\nu_K$  be the distribution evolved from an initial distribution  $\nu_0$  at the final time step, and  $\pi_K$  be the stationary distribution. The bound on the total variance distance between the two distributions is:*

$$\|\nu_K - \pi_K\|_{TV} \leq \max \left\{ \|\nu_{K_0} - \pi_{K_0}\|_{TV} \prod_{l=1}^{\frac{K-K_0}{M}} \left[ 1 - \left(1 - \frac{\Delta F}{c}\right) \frac{\left(\frac{N}{M(N-1)}\right)^M}{(K_0 + lM + d)^{M\Delta F/c}} \right]; \right. \\ \left. \frac{M^M(N-1)^M}{N^M(K+d)^{1-M\Delta F/c}} + \frac{\Delta F}{c} \left[ \frac{1}{K-M+d} \right] \right\}$$

where  $K_0 = \inf\{k : \frac{c}{\log(k+d)} \leq T^* \wedge K - k \equiv 0 \pmod{M}\}$ ,  $T^*$  is defined in Proposition 4.1.

*Proof.* Let  $\alpha_1 = \left(\frac{N}{M(N-2)}\right)^M \frac{\Delta F}{c} \frac{1}{(K+d)^{M\Delta F/c}}$  and  $\alpha_2 = \left(\frac{M(N-1)}{N}\right)^M \frac{K+d}{(K-M+d)^{M\Delta F/c}}$ .

Applying Theorem 3.1 we obtain the following results:

**Case 1:** For  $\|\nu_{K-L} - \pi_{K-L}\| \geq \alpha_2$

$\Rightarrow \sum_{l=K-M}^{K-1} \|\pi_l - \pi_{l+1}\|_{TV} \leq \frac{\Delta F}{c} \frac{1}{(K-M+d)} < \alpha_1 \|\nu_{K-L} - \pi_{K-L}\|$ , Plugging into Theorem 3.1:

$$\|\nu_K - \pi_K\|_{TV} \leq \left[1 - \left(1 - \frac{\Delta F}{c}\right) \left(\frac{N}{M(N-1)}\right)^M \frac{1}{(K+d)^{M\Delta F/c}}\right] \|\nu_{K-L} - \pi_{K-L}\|_{TV}$$

Expand the bound recursively until  $K_0$ , we have

$$\|\nu_K - \pi_K\|_{TV} \leq \prod_{l=1}^{\frac{K-K_0}{M}} \left[1 - \left(1 - \frac{\Delta F}{c}\right) \left(\frac{N}{M(N-1)}\right)^M \frac{1}{(K_0 + lM + d)^{M\Delta F/c}}\right] \|\nu_{K_0} - \pi_{K_0}\|_{TV}$$

**Case 2:** For  $\|\nu_{K-L} - \pi_{K-L}\| < \alpha_2$ , Plugging into Theorem 3.1:

$$\|\nu_K - \pi_K\|_{TV} \leq (1 - \gamma_M^K) \frac{M^M (N-1)^M (K+d)^{M\Delta F/c}}{N^M (K-M+d)} + \frac{\Delta F}{c} \left[\frac{1}{K-M+d}\right]$$

**Corollary 4.4.** *The upper bound on the adiabatic time of the simulated annealing algorithm with the logarithmic cooling scheme is*

$$K_A(\epsilon) < \max \left\{ K_1(\epsilon), K_2(\epsilon) \right\}$$

where

$$K_1(\epsilon) = \left[ \frac{\|\nu_{K_0} - \pi_{K_0}\|_{TV}}{\epsilon} \right]^{(1-\Delta F/c)^{-1} \frac{M^M (N-1)^M}{N^M}} - d$$

end

$$K_2(\epsilon) = \frac{M^M (N-1)^M}{\epsilon N^M} + \frac{\Delta F}{\epsilon c} + M - d$$

*Proof.* This theorem is obtained directly from Theorem 4.2 by letting each term be less than or equal to  $\epsilon$ .

**Note:** The upper bound on adiabatic time depends on  $\Delta F$  - the highest difference in cost function of two neighboring states. This quantity depends on the communication pattern or the structure of user graphs. Additionally, the quality of the bound also depends on the total variance distance bound of stationary distributions of two consecutive time steps.

4.2.3. *Convergence rate at the final temperature:* At the fixed temperature  $T_K$ , the Simulated Annealing based algorithm can be viewed as a time homogeneous Markov Chain.

**Theorem 4.3.** *Given the bound on total variance distance between the distribution and the stationary distribution at time  $K$  is  $\epsilon$ . The temperature from time  $K$  will remain constant. Let  $\delta$  be a positive number such that  $\delta < \epsilon$ . The time of the homogeneous Markov chain corresponding to the simulated annealing algorithm at temperature  $T_K$  to converge is bounded by:*

$$K_M(\epsilon, \delta) \leq M \frac{\log \delta - \log \epsilon}{\log(1 - [\frac{N}{M(N-1)}]^M \frac{1}{(K+d)^{M\Delta F/c}})} \quad (4.18)$$

*Proof.* First note that from Proposition 3.2, if we choose  $L = M$ , the product  $P_K^M$  is an all-positive column matrix. Hence the upper bound on the value of  $\tau(P^M)$  is:

$$\tau(P^M) \leq 1 - [\frac{N}{M(N-1)}]^M \frac{1}{(K+d)^{M\Delta F/c}}$$

Apply Theorem 3.2 to finish the proof.

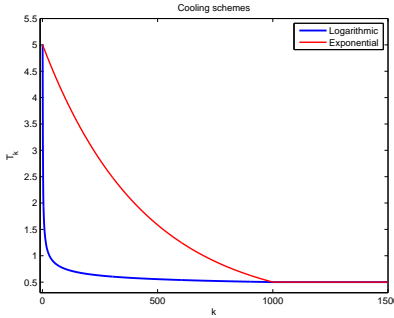
## 5. Simulation

In this section, we show the simulation results of a small user-server assignment problem. There are  $M = 4$  users with a communication pattern depicted in Figure 2a needed to be assigned into  $N = 2$  servers appropriately. The parameters  $\alpha$  and  $\beta$  in 4.4 are chosen equally, i.e.  $\alpha = \beta = 0.5$ . The state space has  $2^4 = 16$  states in total; each state has 4 neighboring states. Figure 2b shows an example of a state (a feasible assignment scheme). The initial temperature is set at  $T_0 = 5$ .

$$A = \begin{bmatrix} 0 & 0.4183 & 0.0892 & 0 \\ 0.4183 & 0 & 0 & 0.4924 \\ 0.0892 & 0 & 0 & 0 \\ 0 & 0.4924 & 0 & 0 \end{bmatrix} \quad X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

(a) Adjacent matrix                      (b) State example

FIGURE 2: Sample assignment problem



(a) Cooling schemes

$T_{final}$	Log.	Exp.
0.1	4.8517e+08	3.0370e+13
0.25	2.9795e+03	2.2861e+07
0.5	53	8.0454e+04
0.75	12	7.7624e+03
1.0	6	1.8995e+03

(b) Lower bound on adiabatic time

FIGURE 3: Cooling schemes and bounds on adiabatic times for different final  $T_{final}$ 

Table in Figure 3b shows the lower bound (the minimum number of time steps) that satisfies Claim 4.3 and Claim 4.4. It is necessary to choose  $K$  greater than the proposed lower bound.

We set the final temperature  $T_{final} = 0.5$  and select randomly an initial assignment scheme. The cooling schemes are chosen such that temperature at each time step  $T_k$  is strictly decreasing and at the  $K^{th}$  time steps, the temperature  $T_K = T_{final}$ .

Figure 4 (resp. 5) shows the simulation results of the simulated annealing algorithm with logarithmic (resp. exponential) cooling schemes corresponding to Section 4.2.2-1 (resp. 4.2.2-2). From simulation results, we can see that simulated annealing algorithms with exponential cooling schemes converge much more slowly than ones with logarithmic cooling schemes. This observation can be explained by the temperature decreasing speed of the two cooling scheme families. Temperature of the logarithmic cooling schemes decrease quickly in the early time steps and extremely slow most of the remaining time steps (Figure 3a). If  $K$  is sufficiently large, the temperature difference can be infinitesimal. On the other hand, the temperature in exponential cooling schemes decrease evenly until reaching the final temperature.

Figure 6 shows the simulation results for the case that parameters of a logarithmic cooling scheme, i.e  $c, d$ , are pre-specified. We would like to know how long it takes for the simulated annealing algorithm to converge as described in Section 4.2.2(3). In other words, we will determine the final temperature at which the simulated annealing algorithms converge, i.e the distance between the final distribution and the stationary

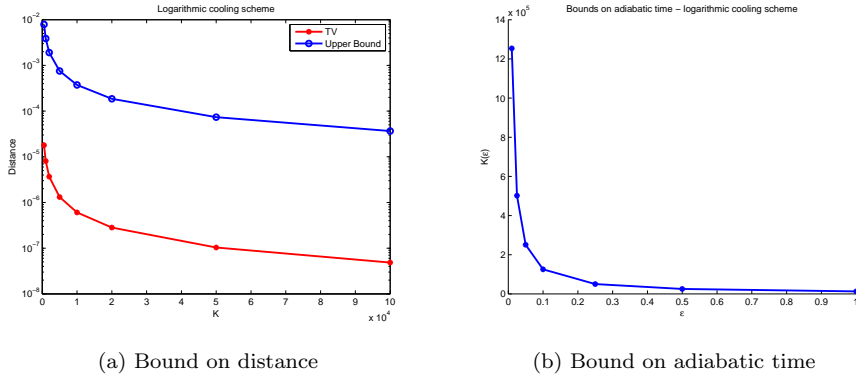


FIGURE 4: Logarithmic cooling scheme

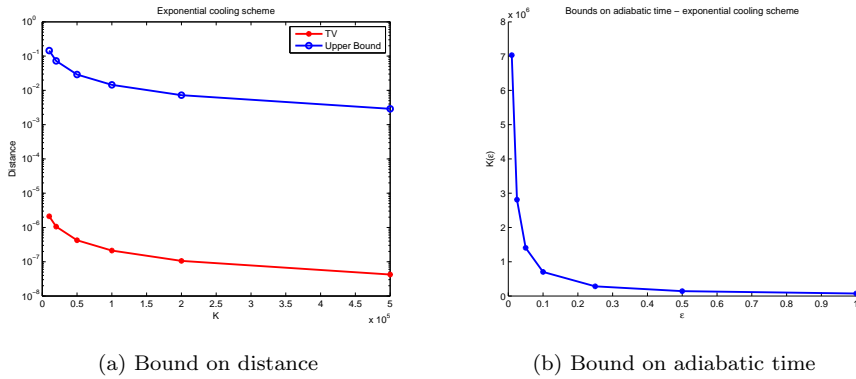


FIGURE 5: Exponential cooling scheme

distribution is bounded by a small value.

## 6. Conclusion

In this paper, we proposed a framework for analyzing convergence rates of simulated annealing based algorithms. We verified our framework by applying in the Client-Server assignment problem which appears in many large scale distributed settings such as social network applications. The quality of the bounds on adiabatic time depends on the optimization problems. Some specific problems for which we can derive a good bound on the stationary distributions of the two consecutive time steps, we will achieve a good bound on adiabatic time.

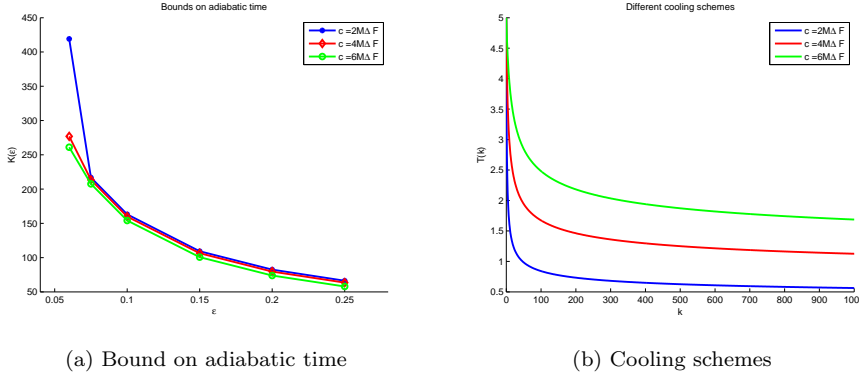


FIGURE 6: Logarithmic cooling scheme with specified parameters

## References

- [1] G. C. D. J. S. Kirkpatrick and M. P. Vecchi, “Optimization by simulated annealing,” May 1983.
- [2] L. M. D. S. Johnson, C. Aragon and C. Schevon, “Optimization by simulated annealing: An experimental evaluation,” *Operations Research*, 1989.
- [3] C. A. D.S. Johnson and L. MacGeoch, “Optimization by simulated annealing: An experimental evaluation part i: Graph partitioning,” *Operations research*, vol. 37, no. 6, Nove. 1989.
- [4] P. Bremaud, *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*. New York: Springer-Verlag, 1999.
- [5] D. Levin, Y. Peres, and E. Wilmer, *Markov chains and mixing times*. Amer Mathematical Society, 2009.
- [6] A. Sinclair, “Improved bounds for mixing rates of marcov chains and multicommodity flow,” *Combinatorics, Probability & Computing* 1 : 351-370, 1992.
- [7] D. Mitra, F. Romeo, and A. Sangiovanni-Vincentelli, “Convergence and finite-time behavior of simulated annealing,” *Advances in Applied Probability*, vol. 18, no. 3, pp. pp. 747–771.



- [8] L. Zacharias, T. Nguyen, Y. Kovchegov, and K. Bradford, “Analysis of adaptive queueing policies via adiabatic approach,” in *Computing, Networking and Communications (ICNC), 2013 International Conference on*, 2013, pp. 1053–1057.
- [9] T. Duong, D. Nguyen-Huu, and T. Nguyen, “Adiabatic markov decision process with application to queuing systems,” in *Information Sciences and Systems (CISS), 2013 47th Annual Conference on*, 2013, pp. 1–6.
- [10] K. Bradford and Y. Kovchegov, “Adiabatic times for markov chains and applications,” *Journal of Statistical Physics*, vol. 143, no. 5, pp. 955–969, 2011. [Online]. Available: <http://dx.doi.org/10.1007/s10955-011-0219-6>
- [11] D. Isaacson and R. Madsen, “Positive columns for stochastic matrices,” *Journal of Applied Probability*, vol. 11, no. 4, pp. pp. 829–835, 1974. [Online]. Available: <http://www.jstor.org/stable/3212567>
- [12] B. Thuan Duong-Ba, Thinh Nguyen and Duc A. Tran, “Distributed client-server assignment for online social network applications,” in *Submitted to IEEE Transactions on Emerging Topics in Computing*, 2013.