

An Efficient Simulated Annealing Schedule: Derivation

Jimmy Lam and Jean-Marc Delosme

Report 8816

September 1988

Department of Computer Science, Yale University.
Department of Electrical Engineering, Yale University.

Abstract

The popularity of simulated annealing comes from its ability to find close to optimal solutions for NP-hard combinatorial optimization problems. Unfortunately, current implementations of the method usually require massive computation time. This paper presents a general annealing schedule which speeds up simulated annealing significantly when compared with general schedules available in the literature. The annealing schedule, which lowers the temperature at every step and keeps the system in quasi-equilibrium at all times, is derived from a new quasi-equilibrium criterion. For a given move generation strategy, this schedule is shown to give the minimum final average cost among all schedules that maintain the system in quasi-equilibrium. An alternate form of the schedule is derived that explicitly relates move generation to temperature decrement. This leads to a control method for the move generation strategy that further minimizes the final average cost.

This research was supported by the Army Research Office under contract DAAL03-86-K-0158, by the National Science Foundation under grant ECS-8314750, and by the Office of Naval Research under contracts N00014-84-K-0092 and N00014-85-K-0461.

1. Introduction

The ground states of a solid can be reached by heating it up to some high temperature and then cooling it down slowly. This process, called annealing, lets the system settle into a low energy state without getting trapped in a local minimum. By identifying the cost function in a combinatorial optimization problem with the energy of a physical system and the solution space with the system state space, close to optimal solutions can be found via a *simulated annealing* of the associated system. This method, proposed by Kirkpatrick *et al.* [9], explores the solution space of the optimization problem by a controlled hill climbing search whose control parameter, T , is the analogue of the temperature. By slowly lowering T towards zero according to a properly chosen schedule, one can show that the set of globally optimal solutions is approached asymptotically, see e.g. [6].

In a typical implementation of simulated annealing [9], the initial temperature is set sufficiently high. A new state (or solution) is generated incrementally from the current state by randomly selecting and proposing a move from a set of predefined moves. A move is a perturbation whose application to the current state leads to a new state. We shall call the move selection process the *move generation strategy*. Let the energy (or cost) of the current state be x and the energy of the proposed new state be x_p . The probability that a proposed move is accepted or rejected is determined by the Metropolis criterion [13]:

$$p(\Delta x) = \min(1, e^{-\frac{\Delta x}{T}}), \quad (1)$$

where $\Delta x = x_p - x$ is the proposed energy change and $p(\Delta x)$ denotes the probability of accepting that move. If a proposed move is accepted, time is incremented and the proposed state becomes the current state. If a proposed move is rejected, time is also incremented but the current state remains unchanged. By controlling the temperature T , we control the probability of accepting a hill climbing move (a move that results in a positive Δx) and, therefore, the exploration of the state space. This process of selecting and proposing a move is repeated until the system is considered in thermal equilibrium. Then, the temperature is reduced according to a temperature schedule called *simulated annealing schedule* (or annealing schedule in short), and the system is allowed to reach thermal equilibrium again. The system is considered frozen and the process is terminated when no significant improvement is expected by further lowering the temperature. At this point, the current state of the system is the solution to the optimization problem. Note that as the temperature is lowered, the exploration of the state space is changed gradually from random—all proposed moves are accepted at $T = \infty$ —to greedy—only downhill moves are accepted at $T = 0$.

Since its introduction, simulated annealing has been applied to optimization problems in diverse areas, spanning computer-aided IC design [15], image restoration [5] and code design [4]. This popularity comes from its ability to find close to optimal solutions for NP-hard combinatorial optimization problems. However, its massive requirement of computation time still hinders its application. This paper tackles this problem by optimizing the simulated annealing schedule.

In the remainder of this section, we discuss the main approaches employed to develop annealing schedules. (Refer to [10, Chapter 5] for an extensive survey of different annealing schedules.) For ease of presentation, we define the *inverse temperature* s as $s \equiv 1/T$. Time (or step) is discrete and is incremented whenever a proposed move is either accepted or rejected. Furthermore, the inverse temperature is assumed to be a non-decreasing function of time,

$$s_{n+k} \geq s_n, \quad k \geq 1, \quad (2)$$

where s_n is the inverse temperature at step n .

1.1. Logarithmic schedule

The analysis of simulated annealing based on time-inhomogeneous Markov chains has been carried out in [5], [6] and [14] among others. This analysis determines constraints on s_n , $n \geq 0$, that ensure convergence to globally optimum solutions. A particular schedule is defined by the sequence of inverse temperatures that satisfy these constraints tightly.

Let x_i be the energy of state i , $w_i(n)$ be the probability that the system is at state i at step n , and $g_{i,j}$ be the probability of proposing a move from state i to state j . Since the probability that a move from state i to state j is accepted is determined by the Metropolis criterion,

$$p(x_j - x_i) = \min(1, e^{-s(x_j - x_i)}),$$

the one-step transition probability from state i to state j at step t is

$$p_{i,j}(t) = g_{i,j} \min(1, e^{-s_t(x_j - x_i)}), \quad i \neq j. \quad (3)$$

If a proposed move is rejected, the system remains in the same state. Therefore, the probability of staying at state i is

$$p_{i,i}(t) = 1 - \sum_{j \neq i} p_{i,j}(t). \quad (4)$$

In matrix notation, the probability vector of the system at step n is

$$\mathbf{W}(n) = \mathbf{P}(n)\mathbf{P}(n-1) \cdots \mathbf{P}(1)\mathbf{W}(0), \quad (5)$$

where $\mathbf{W}(n)$ is a column matrix with elements $w_i(n)$ and $\mathbf{P}(t)$ is a square matrix with elements $p_{i,j}(t)$. The preceding relation for a *time-inhomogeneous* Markov chain ($\mathbf{P}(t)$ is a function of time) is a precise mathematical model for simulated annealing.

Using this model, the system may be shown to converge asymptotically with probability 1 to the set of globally optimal solutions if the inverse temperature is changed according to the following schedule:

$$s_n = \lambda \log(n+1), \quad n = 1, 2, \dots, \quad (6)$$

where $\lambda \leq \lambda^*$, and λ^* is a parameter that depends on the structure of the optimization problem. We shall call this annealing schedule the *logarithmic schedule*. Note that temperature changes at every step and the physical notion of thermal equilibrium is not relevant to this mathematical analysis.

Since the schedule must guarantee convergence even in the worst case of stepping into a deep local minimum, no matter how unlikely such an event is to occur, the schedule has to raise the inverse temperature quite slowly. The time required to get close to a global optimum would be too long to be practical. Given a fixed amount of time, one could retain the functional form of the schedule and select the value of λ that gives the best quality of solution. Still, since the schedule is based on very little information about the system, it would not perform well. To achieve better performance the following schedules, based on approximate equilibrium criteria,

exploit increasing amounts of statistical information about the system.

1.2. Aarts's schedule

The first schedule is by Aarts and van Laarhoven [1]. Their analysis is also based on Markov chain theory but they consider time-homogeneous Markov chains, in which the inverse temperature s_n is fixed at a particular value for k steps (hence $\mathbf{P}(t)$ is constant for that period). Moreover, instead of deriving an annealing schedule that slowly converges to the globally optimal solutions, they determine a schedule that satisfies an approximate equilibrium criterion. They propose that the inverse temperature be raised after every k steps with the new inverse temperature, s_{n+k} , chosen such that the following quasi-equilibrium criterion is satisfied:

$$\|\mathbf{\Pi}(s_n) - \mathbf{\Pi}(s_{n+k})\| < \epsilon, \quad (7)$$

where $\mathbf{\Pi}(s_n)$ is the stationary probability vector for the time-homogeneous Markov chain at inverse temperature s_n , and $\|\mathbf{\Pi}\|$ is the 1-norm of vector $\mathbf{\Pi}$. Based on this quasi-equilibrium criterion and under some assumptions, they derive an annealing schedule of the form

$$s_{n+k} = s_n + \frac{\lambda}{\sigma(s_n)}, \quad (8)$$

where λ is a user specified constant that is related to the ϵ in (7) and $\sigma(s_n)$ is the equilibrium standard deviation of the energy at inverse temperature s_n . Furthermore, they argue that k , the number of steps before the temperature is changed, should be set to the maximum number of states that may be reached from any state in one step.

We know from (5) that the probability vector of a system at step n , $\mathbf{W}(n)$, depends on the transition probability matrix $\mathbf{P}(t)$ for $t = 1, 2, \dots, n$. Therefore, the rate of convergence of a time-homogeneous Markov chain towards its stationary probability vector also depends on the transition probability matrix. Now, according to (3), the transition probability matrix depends on the inverse temperature and $g_{i,j}$. Since $g_{i,j}$ depends on the move generation strategy, the rate of convergence has to depend on the move generation strategy also. However, the logarithmic schedule and Aarts's schedule do not account for either the dependency on the inverse temperature or the dependency on the move generation strategy and, therefore, do not decrease the temperature fast enough during the periods when the rate of convergence is highest.

1.3. Huang's schedule

While the two preceding annealing schedules are based on a Markov chain analysis applied to an exact state space representation, the schedule proposed by Huang *et al.* [7] and our annealing schedule are derived following White's approach [16], built upon a model of the energy of the system. White suggests that the energy values be partitioned into small non-overlapping intervals of length δx , that all values in a given interval be considered identical, and that the states be grouped according to their energy values in that coarser scale. Let $P(x)$ be the *energy density function* defined as

$$P(x) \equiv \frac{\text{number of states with energy in the interval containing } x}{\delta x \cdot \text{total number of states}},$$

and let $P_s(x)$ be defined as

$$P_s(x) \equiv \frac{P(x)e^{-sx}}{Z(s)}, \text{ with } Z(s) = \int_{-\infty}^{\infty} P(x)e^{-sx} dx. \quad (9)$$

Using these definitions, it is a simple matter to verify that the equilibrium (stationary) average energy and the equilibrium variance of the energy are given, respectively, by

$$\mu(s) = -\frac{1}{Z(s)} \frac{\partial Z(s)}{\partial s}, \text{ and } \sigma^2(s) = -\frac{\partial \mu(s)}{\partial s}, \quad (10)$$

with negligible error if δx is sufficiently small. White also proposes to model the energy density function by a normal density function. With that model, $\sigma(s)$ is a constant, σ .

To maintain quasi-equilibrium, Huang *et al.* require that the new inverse temperature, s_{n+k} , be chosen such that

$$\mu(s_n) - \mu(s_{n+k}) = \lambda \sigma(s_n) \quad (11)$$

with $\lambda \leq 1$. From this criterion and an expression for the slope of the annealing curve—a curve of the equilibrium average energy, $\mu(s)$, versus the logarithm of the inverse temperature, $\log(s)$ —they obtain the annealing schedule

$$s_{n+k} = s_n e^{\frac{\lambda}{s_n \sigma(s_n)}}. \quad (12)$$

For small values of $\lambda / (s_n \sigma(s_n))$ the schedule becomes

$$s_{n+k} = s_n + \frac{\lambda}{\sigma(s_n)}$$

which seems equivalent to Aarts's annealing schedule in (8). However there is a significant difference: the selection of k , the number of steps before the inverse temperature is changed, is *dynamic* in Huang's schedule and determined by the following equilibrium detection mechanism.

A normal density function is assumed. Thermal equilibrium is detected by keeping track of two counters for the number of accepted moves that result in energy values within and without the interval $[\bar{X}(s) - 0.5\sigma, \bar{X}(s) + 0.5\sigma]$, where $\bar{X}(s)$ is the measured average energy at inverse temperature s . If the number of accepted moves within the interval reaches its target first, the system is considered in thermal equilibrium. If the number of accepted moves without the interval reaches its target first, both counters are reset to zero and the counting is initiated again. Expressions for the target values are determined from the normal density function. This *dynamic* equilibrium detection scheme is an improvement over Aarts's proposal of changing the inverse temperature after a constant number of steps, since it accounts for the dependency of the time to reach thermal equilibrium on the inverse temperature and the move generation strategy.

Although Huang's equilibrium detection scheme accounts for the dynamics of the system, his quasi-equilibrium criterion, like the one of Aarts, does not. Yet the size of the inverse temperature increments should clearly be function of the response time of the system to temperature changes. In Section 2 we introduce a quasi-equilibrium criterion that incorporates the dynamics of system and enables (like in the logarithmic schedule) a change of the temperature at each step, a desirable feature since the best schedule for a given problem must clearly possess that property. In contrast to Aarts's and Huang's criteria which depend only on the stationary statistics of the system, our quasi-equilibrium criterion also depends on the average energy of the system which, in turn, depends on the move generation strategy. By selecting a good move

generation strategy that lets the system reach quasi-equilibrium faster, we can change the inverse temperature faster or in larger increments. Since the inverse temperature is raised *at every step* while keeping the system in quasi-equilibrium at all times, an equilibrium detection scheme is no longer needed. We also derive in Section 2 an intermediate form of our annealing schedule and show that, for a given move generation strategy, this schedule gives the minimum final average energy among all schedules that maintain the system in quasi-equilibrium at all times. Clearly, the freedom of choice of move generation strategy should be exploited. In order to show explicitly the dependence on the move generation strategy and to select among the available strategies the one that maximizes the speed of simulated annealing while remaining in quasi-equilibrium, a final form of the schedule is derived in Section 4 based on models for move generation and the energy distribution of the states introduced in Section 3. Implementation and experimental evaluation of the schedule are presented in a companion paper [11].

2. Efficient annealing schedule - basic form

The simulated annealing method is based on the observation that annealing is successful if the system is kept close to thermal equilibrium as the temperature is lowered. However, to keep the system in equilibrium at all times requires that the temperature decrements be infinitesimal. A practical annealing schedule must, therefore, achieve a compromise between final average energy and computation time. To determine by how much temperature may be lowered at a given time while maintaining the system sufficiently close to equilibrium, we need an approximate equilibrium criterion.

First we have to expound the relationship between equilibrium and stationary process. The energy, $X(s_n)$, of a system is a stochastic process whose statistics depend on step n as well as the inverse temperature s_n . Thus, given a system, there exists a stochastic process corresponding to each annealing schedule. Let us consider a special class of annealing schedules where the inverse temperature, s_n , is constant. As $n \rightarrow \infty$, the system approaches equilibrium, and the process $X(s_n)$ becomes stationary. We denote this kind of stationary process by $\underline{X}(s_n)$ and its mean and variance by $\mu(s_n)$ and $\sigma^2(s_n)$, respectively. The stationary process $\underline{X}(s_n)$ is of interest because comparison with its statistics provides information on how far the system is away from equilibrium at any inverse temperature.

2.1. Quasi-stationarity criterion

We say a process, $X(s_{n-1})$, is *quasi-stationary at inverse temperature s_n* if it satisfies the *quasi-stationarity* (or quasi-equilibrium) criterion:

$$\left| \bar{X}(s_{n-1}) - \mu(s_n) \right| \leq \lambda \sigma(s_n), \quad (13)$$

where $\bar{X}(s_{n-1})$ is the average energy of the system at step $n-1$ (after a proposed move is either accepted or rejected), and λ is a user specified parameter. Note that $\bar{X}(s_{n-1})$ depends on s_{n-1} , not s_n , and that the criterion is invariant with respect to translation and scaling of the energy, as it should be. A *quasi-stationary process* is a process that satisfies the quasi-stationarity criterion at *all* inverse temperatures in an annealing schedule; a schedule that gives rise to such a process is called a λ -*schedule*. The λ -schedules are important, for they result in carefully controlled executions of simulated annealing. Note that, assuming the move generation strategy is fixed, a unique quasi-stationary process is associated with each λ -schedule. A λ -schedule is n -*step*

efficient if it minimizes the final average energy at step n among all λ -schedules with the same move generation strategy. Moreover, if a λ -schedule is n -step efficient for all $n > 0$, we call it the *efficient λ -schedule* for that move generation strategy. Our goal in this section is to show that for any move generation strategy the associated efficient λ -schedule exists and to construct it by studying the behavior of the associated quasi-stationary process.

The parameter λ , which can be made as small as desired to ensure a good approximation of equilibrium, realizes the compromise between the quality of the final average energy and the computation time: the smaller the λ ; the better the final average energy; the longer the computation time.

2.2. Autoregressive process

In order to derive the efficient λ -schedule associated to any move generation strategy, we need to know how the average energy of the system evolves. We model the stationary process $\underline{X}(s)$ as a first order autoregressive process:

$$\underline{X}_n(s) = r(s)(\underline{X}_{n-1}(s) - \mu(s)) + \mu(s) + \mathcal{N}_n, \quad (14)$$

where $r(s)$, which depends on the move generation strategy, is the first autocorrelation coefficient of the stationary process $\underline{X}(s)$, and the random fluctuation \mathcal{N}_n is uncorrelated (white noise). Note that the value of the first autocorrelation coefficient is close to 1 because the energy difference between moves is usually small when compared with the standard deviation, $\sigma(s)$, of the energy, as the proposed solution is obtained by slightly perturbing the current solution.

There are two reasons behind the model in (14). First, given the current state (not energy) of a system, the next state of the system does not depend on its previous states; it only depends on the current state, the move generation strategy, and the inverse temperature. This Markov property follows directly from the simulated annealing method. The grouping of states according to the coarse values of their energy (with a resolution of δx , see Section 1.3), however, may not preserve this property: given the current value of the energy of a system, the next value may depend on the previous values. Yet the dependency must be slight: since the states with a given energy are *numerous*, the probability is small that

1. a sequence of previous energy values favors a particular group of states, and
2. the next energy for this particular group of states is different from the expected next energy of the system given the current energy.

Consequently, the previous values of the energy provide little extra information, and we assume that the next value of the energy does not depend on the previous values given the current value, that is, $\underline{X}_n(s)$ is a function of $\underline{X}_{n-1}(s)$, but not $\underline{X}_{n-2}(s)$, $\underline{X}_{n-3}(s)$,.... Such Markov property at two different levels is often encountered in the physical sciences [8, p. 79]. Second, we assume that the next energy depends linearly on the current energy. This assumption, leading to the model in (14), is imposed for simplicity. Furthermore, as the following experiments illustrate, a nonlinear dependency is unlikely to improve the accuracy of our model significantly.

To assess the quality of the model, we manipulate (14) to get an expression for the random fluctuation:

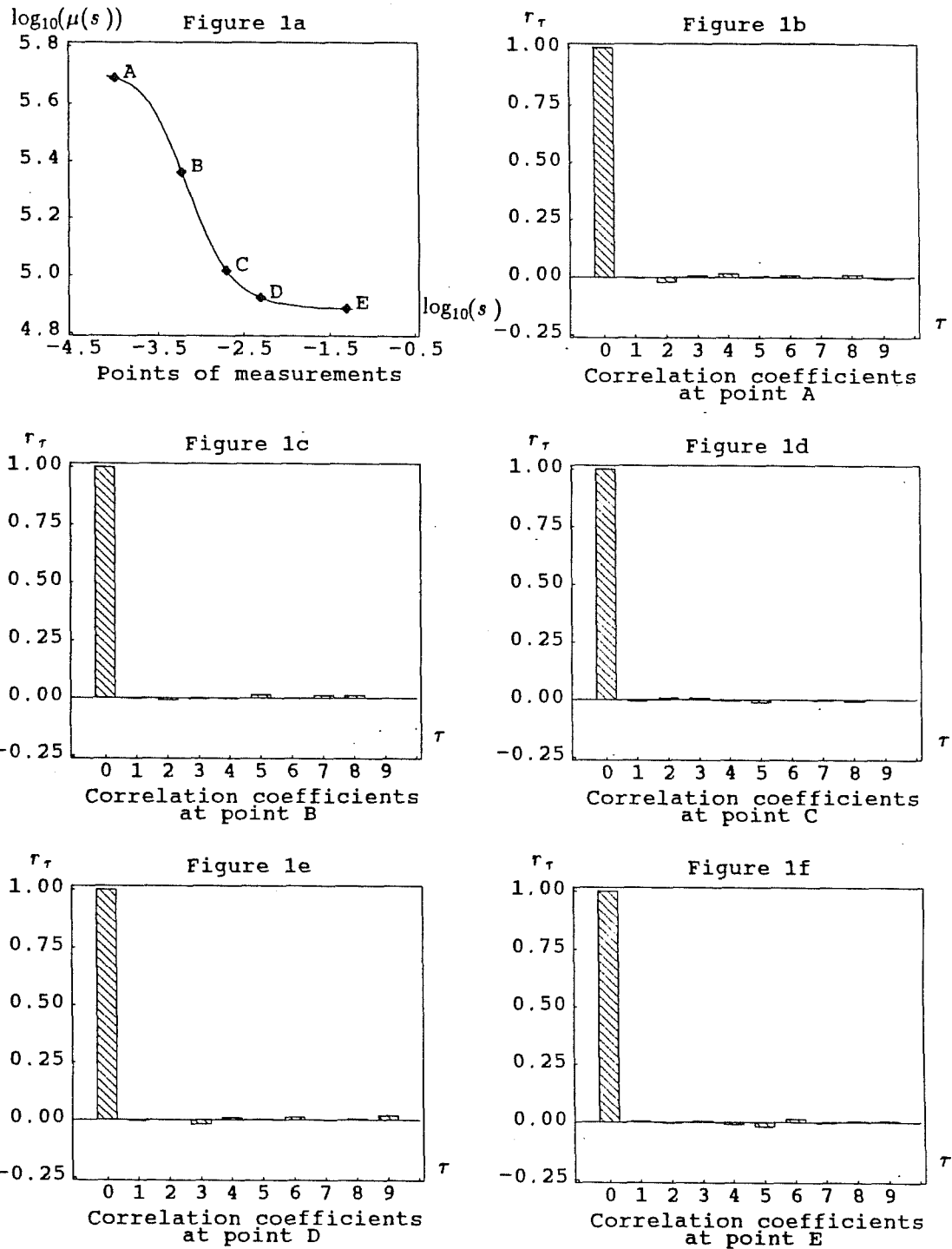


Figure 1: Autocorrelation coefficients of the random fluctuations for a 100-city traveling salesman problem. The inverse temperatures at which the tests were performed are indicated on the annealing curve in Figure 1a while the autocorrelation coefficients are shown in Figure 1b to Figure 1f.

$$\mathcal{N}_n = (\underline{X}_n(s) - \mu(s)) - r(s)(\underline{X}_{n-1}(s) - \mu(s)).$$

The average value of \mathcal{N}_n is zero by construction. We can test the quality of this model by measuring the autocorrelation coefficients of the random fluctuation. If the autocorrelation coefficient, r_τ , between \mathcal{N}_n and $\mathcal{N}_{n+\tau}$ is close to zero for $\tau \neq 0$, the model is good. Otherwise, the random fluctuation is not a white process and the model is poor. We performed the tests on a 100-city traveling salesman problem and a graph partition problem. The results for the traveling salesman problem are shown in Fig. 1. (The results for the graph partition problem are similar [12].) Since r_τ for $\tau \geq 1$ are close to zero, we conclude that the model is good.

The properties of the quasi-stationary process $X(s_{n-1})$ will be close to that of the stationary process $\underline{X}(s_n)$ if λ is made small enough. Consequently, we can use (14),

$$X(s_n) = r(s_n)(X(s_{n-1}) - \mu(s_n)) + \mu(s_n) + \mathcal{N}_n,$$

with \mathcal{N}_n a zero-mean white noise, to model the evolution of the quasi-stationary process $X(s_{n-1})$ as the inverse temperature is changed to s_n . From the preceding expression, we obtain, after taking expectations, an expression for *the evolution of the average energy*:

$$\bar{X}(s_n) = r(s_n)(\bar{X}(s_{n-1}) - \mu(s_n)) + \mu(s_n), \quad (15)$$

where $\bar{X}(s_{n-1})$ is the average energy of the quasi-stationary process $X(s_{n-1})$, which usually differs from $\mu(s_n)$, the average energy of the stationary process $\underline{X}(s_n)$. Assuming the process $X(s_{n-1})$ satisfies the quasi-stationarity criterion at inverse temperature s_n , equation (15) gives the average energy at step n . Hence equation (15) specifies the evolution of the average energy at all inverse temperatures in a λ -schedule. Furthermore, we can replace the quasi-stationarity criterion in (13) with the simpler expression

$$\bar{X}(s_{n-1}) - \mu(s_n) \leq \lambda\sigma(s_n). \quad (16)$$

Indeed, observe that the inverse temperature in an annealing schedule does not decrease with respect to time, and that $\partial\mu(s) / \partial s \leq 0$ from (10). These two conditions imply $\mu(s_n)$ is a non-increasing function of s_n , $\mu(s_n) \leq \mu(s_{n-1})$. If $\bar{X}(s_{n-2}) \geq \mu(s_{n-1})$, we deduce from (14) that $\bar{X}(s_{n-1}) \geq \mu(s_{n-1})$ and, hence, $\bar{X}(s_{n-1}) \geq \mu(s_n)$. If the system is in equilibrium to start with, $\bar{X}(s_o) = \mu(s_o)$, $\bar{X}(s_{n-1})$ will always be greater than or equal to $\mu(s_n)$ for any n . Consequently, we can replace (13) with (16).

2.3. Efficient λ -schedule

Now exploiting our formula for the evolution of the average energy, we can claim that, for a given move generation strategy, the n -step efficient λ -schedule consists of the $(n-1)$ -step efficient λ -schedule followed by the inverse temperature s_n that maximizes the decrement in average energy at step n subject to the quasi-stationarity constraint (16), that is, the n -step efficient λ -schedule is also t -step efficient for $t = 1, 2, \dots, n-1$. This property, which can be viewed as an instance of Bellman's principle of optimality, is essential, for it enables us to claim that the efficient λ -schedule exists and, moreover, to compute it recursively.

To derive this property, we note that the Markov property and the separability of the objective function are necessary and sufficient conditions for the application of the principle of optimality [3, p. 38]. A system possesses the Markov property if after decision δ_n is made at step n , the state π_n that results from that decision depends only on state π_{n-1} and decision δ_n . This

condition is satisfied if we identify the average energy $\bar{X}(s_{n-1})$ as the state π_{n-1} and the inverse temperature s_n as the decision δ_n to be made at step n . The objective function is separable if for all k , the effect of the final k steps on the objective function of an n -step process depends only on state π_{n-k} and upon the final k decisions. In our case, the objective function is the average energy after n steps, $\bar{X}(s_n)$, and it is indeed separable. Consequently, the principle of optimality applies. (An alternate proof from first principles is given in [12, Chapter 2].)

To determine a recurrence equation for the efficient λ -schedule, we start with the $(n-1)$ -step efficient λ -schedule and find the inverse temperature s_n that maximizes the decrement in average energy at step n subject to the quasi-stationarity constraint (16). For ease of presentation, we denote s_{n-1} by s , s_n by s_+ , $\bar{X}(s_{n-1})$ by $\bar{X}(s)$, and $\bar{X}(s_n)$ by $\bar{X}(s_+)$. The decrement in average energy at step n can be obtained by subtracting $\bar{X}(s_+)$ from $\bar{X}(s)$ using (15):

$$\bar{X}(s) - \bar{X}(s_+) = (1 - r(s_+))(\bar{X}(s) - \mu(s_+)). \quad (17)$$

If the two constraints (to be checked in Section 4.1)

$$\sigma(s_+) \geq \frac{\lambda}{1 - r(s_+)} \left| \frac{\partial r(s_+)}{\partial s_+} \right| \quad (18)$$

and, for all ζ between the first and last inverse temperatures in the schedule,

$$\sigma^2(\zeta) \gg \lambda \left| \frac{\partial \sigma(\zeta)}{\partial \zeta} \right| \quad (19)$$

are satisfied, then we claim that the decrement in average energy is a non-decreasing function of s_+ ,

$$\frac{\partial}{\partial s_+} (\bar{X}(s) - \bar{X}(s_+)) \geq 0. \quad (20)$$

Consequently, maximizing the decrement in average energy subject to the quasi-stationarity constraint is equivalent to maximizing the inverse temperature, s_+ , subject to the same constraint. Furthermore, this maximum occurs when

$$\bar{X}(s) - \mu(s_+) = \lambda \sigma(s_+). \quad (21)$$

To show that the condition in (20) holds, we differentiate (17) and substitute $-\sigma^2(s_+)$ for $\partial \mu(s_+) / \partial s_+$ to obtain the equivalent condition

$$-\frac{\partial r(s_+)}{\partial s_+} (\bar{X}(s) - \mu(s_+)) + (1 - r(s_+)) \sigma^2(s_+) \geq 0$$

or

$$\sigma^2(s_+) \geq \frac{1}{1 - r(s_+)} \frac{\partial r(s_+)}{\partial s_+} (\bar{X}(s) - \mu(s_+)).$$

If $\partial r(s_+) / \partial s_+$ is negative, this inequality holds. If it is positive, since $\bar{X}(s) - \mu(s_+) \leq \lambda \sigma(s_+)$, the inequality is satisfied if

$$\sigma(s_+) \geq \frac{\lambda}{1-r(s_+)} \frac{\partial r(s_+)}{\partial s_+}, \quad (22)$$

which clearly holds if the constraint in (18) is satisfied.

To show that the maximum occurs when (21) is satisfied, we rewrite the quasi-stationarity constraint as

$$\bar{X}(s) \leq \mu(s_+) + \lambda\sigma(s_+).$$

The right hand side of this expression decreases monotonically with respect to s_+ since its first derivative with respect to s_+ is

$$-\sigma^2(s_+) + \lambda \frac{\partial \sigma(s_+)}{\partial s_+}$$

which is negative if (19) is satisfied. Therefore, the minimum value of $\mu(s_+) + \lambda\sigma(s_+)$ and, hence, the maximum value of s_+ that satisfies the quasi-stationarity constraint occur when

$$\bar{X}(s) = \mu(s_+) + \lambda\sigma(s_+).$$

Combining equations (17) and (21) we get

$$\bar{X}(s) - \bar{X}(s_+) = \lambda(1-r(s_+))\sigma(s_+) \quad (23)$$

which, upon replacement of $\bar{X}(s)$ by $\mu(s_+) + \lambda\sigma(s_+)$, leads to

$$\bar{X}(s_+) = \mu(s_+) + \lambda r(s_+)\sigma(s_+).$$

Therefore, we also have

$$\bar{X}(s) = \mu(s) + \lambda r(s)\sigma(s)$$

for all s in the schedule except the initial inverse temperature s_0 . Substituting the above expressions of $\bar{X}(s)$ and $\bar{X}(s_+)$ into (23), we get

$$\mu(s) - \mu(s_+) = \lambda \left[\frac{\sigma(s_+)}{\sigma(s)} - r(s) \right] \sigma(s). \quad (24)$$

(Compare this result with Huang's criterion in (11).) Identifying (24) with the Taylor expansion of $\mu(s_+)$ around s ,

$$\begin{aligned} \mu(s_+) &= \mu(s) + (s_+ - s) \frac{\partial \mu(s)}{\partial s} + \frac{(s_+ - s)^2}{2} \frac{\partial^2 \mu(\dot{s})}{\partial \dot{s}^2} \\ &= \mu(s) - (s_+ - s)\sigma^2(s) - (s_+ - s)^2\sigma(\dot{s}) \frac{\partial \sigma(\dot{s})}{\partial \dot{s}} \end{aligned}$$

where $s \leq \dot{s} \leq s_+$, and also noting that, from the Taylor expansion of $\sigma(s_+)$ around s ,

$$\frac{\sigma(s_+)}{\sigma(s)} = 1 + \frac{s_+ - s}{\sigma(s)} \frac{\partial \sigma(\bar{s})}{\partial \bar{s}}$$

where $s \leq \bar{s} \leq s_+$, we obtain

$$s_+ = s + \lambda \frac{1-r(s)}{\sigma(s)} + \frac{(s_+ - s)}{\sigma^2(s)} \left[\lambda \frac{\partial \sigma(\bar{s})}{\partial \bar{s}} - (s_+ - s) \sigma(\dot{s}) \frac{\partial \sigma(\dot{s})}{\partial \dot{s}} \right]. \quad (25)$$

To show that the last term is negligible when compared to $\lambda(1-r(s))/\sigma(s)$ and hence may be omitted, we assume that

$$s_+ - s \approx \lambda \frac{1-r(s)}{\sigma(s)} \quad (26)$$

so that we obtain the following approximate expression for the ratio of the last term to $\lambda(1-r(s))/\sigma(s)$:

$$\frac{1}{\sigma^2(s)} \left[\lambda \frac{\partial \sigma(\bar{s})}{\partial \bar{s}} - \lambda(1-r(s)) \frac{\sigma(\dot{s})}{\sigma(s)} \frac{\partial \sigma(\dot{s})}{\partial \dot{s}} \right]. \quad (27)$$

Now if we assume either $\sigma(s) \geq \sigma(\bar{s})$, a property usually satisfied for most of (if not all) the temperatures, or $\sigma(s) \approx \sigma(\bar{s})$, it follows from (19) that

$$\left| \frac{\lambda}{\sigma^2(s)} \frac{\partial \sigma(\bar{s})}{\partial \bar{s}} \right| \ll 1.$$

Similarly, if $\sigma(s) \geq \sigma(\dot{s})$ or $\sigma(s) \approx \sigma(\dot{s})$, (19) implies

$$\left| \frac{\lambda}{\sigma^2(s)} \frac{\sigma(\dot{s})}{\sigma(s)} \frac{\partial \sigma(\dot{s})}{\partial \dot{s}} \right| \ll 1$$

and also, since $|1-r(s)| < 1$,

$$\left| \frac{\lambda}{\sigma^2(s)} (1-r(s)) \frac{\sigma(\dot{s})}{\sigma(s)} \frac{\partial \sigma(\dot{s})}{\partial \dot{s}} \right| \ll 1.$$

Because both contributions to (27) are much smaller than 1, the approximation in (26) used for the proof was legitimate and the last term in (25) may be neglected. Therefore our update formula for the inverse temperature is

$$s_+ = s + \lambda \frac{1-r(s)}{\sigma(s)}. \quad (28)$$

Let $\rho_2(s)$ be the variance of the energy increment, $\rho_2(s) \equiv E \{ (\underline{X}(s_n) - \underline{X}(s_{n-1}))^2 \}$. ($E\{\cdot\}$ is the expectation operator. Since $\underline{X}(s_n)$ is a stationary process, $s_n = s_{n-1} = s$.) Replacing $r(s)$ in (28) by the expression

$$r(s) = 1 - \frac{\rho_2(s)}{2\sigma^2(s)}, \quad (29)$$

obtained by substituting the equality

$$E \{ \underline{X}(s_{n-1}) \underline{X}(s_n) \} = E \{ \underline{X}(s_n) \underline{X}(s_n) \} - \frac{1}{2} \rho_2(s)$$

into the definition of the first autocorrelation coefficient,

$$r(s) \equiv \frac{E \{ \underline{X}(s_{n-1}) \underline{X}(s_n) \} - \mu^2(s)}{E \{ \underline{X}(s_n) \underline{X}(s_n) \} - \mu^2(s)},$$

we obtain the update formula for the efficient λ -schedule

$$s_+ = s + \lambda \frac{\rho_2(s)}{2\sigma^3(s)}. \quad (30)$$

This formula is valid if the two constraints (18) and (19) are satisfied. These constraints are usually satisfied for almost all the temperature range as we shall see in Section 4.1. Note that the efficient λ -schedule depends on the move generation strategy via $\rho_2(s)$.

We are left with the problem of characterizing the move generation strategy whose associated efficient λ -schedule results in the minimum final average energy among all efficient λ -schedules. Since minimizing the final average energy is equivalent to maximizing the decrement in average energy at every step and since, from (23) and (29), the decrement in average energy is

$$\bar{X}(s) - \bar{X}(s_+) = \lambda \frac{\rho_2(s_+)}{2\sigma(s_+)},$$

where $\sigma(s_+)$ does not depend on the move generation strategy, minimizing the final average energy is equivalent to maximizing $\rho_2(s_+)$ for all inverse temperatures s_+ . Consequently, in the sequel, we seek the move generation strategy that maximizes $\rho_2(s_+)$ at every step.

3. Models for simulated annealing

The choice of move generation strategy strongly influences the performance of simulated annealing. In order to analyze these strategies, we need a model that is simple, yet able to capture their essential properties. Since the goal of simulated annealing is to minimize the energy of the system, only those aspects that affect the energy need to be modeled. The conditional probability density function of the proposed energy given the current energy, modeled in this section, completely specifies the evolution of the energy and is adequate for that purpose.

For ease of presentation, we use X , X_p and X_+ , respectively, to symbolize $\underline{X}(s_{n-1})$, $\underline{X}(s_{n-1}) + \Delta X$ and $\underline{X}(s_n)$, where ΔX is the proposed energy increment; thus, X_p represents the proposed—but not necessarily accepted—energy which is different from X_+ , the energy at the next time step. Moreover, we use lower case for the values of these random variables.

3.1. Move generation model

3.1.1. Separable model

It is reasonable to assume that the frequency with which an energy is proposed depends on the number of states possessing that energy (specified by the energy density function): the more states with a given energy, the more likely that energy is to be proposed. Also, it is expected that the amplitude of the gap between the proposed energy and the current energy affects the frequency: moves that result in energy increments of same amplitude are proposed with nearly equal frequencies. Formulating both ideas into a separable function leads to the class of *separable models*:

$$f_p(x_p | x) \propto G(|x_p - x|)Q(x_p)$$

where $f_p(x_p | x)$ is the conditional probability density function of the *proposed energy*, x_p , given the current energy, x . The function $Q(x_p)$ models the effect of the energy density function, $P(x_p)$, on $f_p(x_p | x)$ while the function $G(|x_p - x|)$ models the effect of distance between x_p and x on $f_p(x_p | x)$. Rewriting this class of models to include the normalization factor, $K(x)$,

we have

$$f_p(x_p | x) = \frac{1}{K(x)} G(|x_p - x|) Q(x_p), \quad (31)$$

with

$$K(x) = \int_{-\infty}^{\infty} G(|y - x|) Q(y) dy = G(|x|) * Q(x) \quad (32)$$

where * denotes convolution.

Now we can compute $f(x_+ | x)$, the conditional probability density function of *the energy at the next time step*, needed for the computation of ρ_2 . The probability that a move with proposed energy x_p is accepted given the current energy x is determined by the Metropolis criterion:

$$\min(1, e^{-s(x_p - x)}).$$

Since the probability that such a move is proposed is $f_p(x_p | x)$, the probability that the move is proposed and accepted is

$$f_p(x_p | x) \cdot \min(1, e^{-s(x_p - x)}),$$

while the probability that the move is proposed and rejected is

$$f_p(x_p | x) \cdot [1 - \min(1, e^{-s(x_p - x)})].$$

Summing up the rejection probability of all possible moves,

$$\int_x^{\infty} f_p(y | x) (1 - e^{-s(y - x)}) dy,$$

and combining the probabilities of rejection and acceptance, we obtain

$$f(x_p | x) = \begin{cases} f_p(x_p | x) e^{-s(x_p - x)} & \text{if } x_p > x \\ f_p(x_p | x) + \delta(x_p - x) \int_x^{\infty} f_p(y | x) (1 - e^{-s(y - x)}) dy & \text{if } x_p \leq x \end{cases}$$

as the conditional probability density function of the energy at the next time step. The delta function, $\delta(x_p - x)$, ensures that the integral representing the rejection probability contributes to $f(x_p | x)$ only when the system remains at the same energy. Substituting expression (31) for $f_p(x_p | x)$ into the above equation and replacing x_p with x_+ lead to the final form of the conditional probability density function of the energy at the next time step:

$$f(x_+ | x) = \frac{1}{K(x)} \left[A_s(x_+ - x) Q(x_+) + \delta(x_+ - x) \int_x^{\infty} [G(|y - x|) - A_s(y - x)] Q(y) dy \right] \quad (33)$$

with

$$A_s(x_+ - x) = \begin{cases} G(|x_+ - x|)e^{-s(x_+ - x)} & \text{if } x_+ \geq x \\ G(|x_+ - x|) & \text{otherwise.} \end{cases} \quad (34)$$

The function $Q(x)$ is determined from the properties of the stationary process $\underline{X}(s)$. Since the process $\underline{X}(s)$ is stationary, for any two energies $x \neq y$, the transition probability from energy x to energy y must be equal to the transition probability from energy y to energy x . Thus,

$$f(y|x)P_s(x) = f(x|y)P_s(y)$$

where $P_s(x)$ is the probability that the system has energy x given by equation (9). Assuming, without loss of generality, that $y > x$ and substituting (33) into the above expression, we obtain

$$\frac{1}{K(x)} G(|y-x|)e^{-s(y-x)}Q(y)P_s(x) = \frac{1}{K(y)} G(|x-y|)Q(x)P_s(y).$$

hence, assuming $G(|y-x|) \neq 0$,

$$K(y)Q(y)e^{-sy}P_s(x) = K(x)Q(x)e^{-sx}P_s(y). \quad (35)$$

However, from equation (9)

$$P(y)e^{-sy}P_s(x) = P(x)e^{-sx}P_s(y).$$

Comparison of the last two expressions yields

$$Q(x) = \kappa \frac{P(x)}{K(x)} \quad (36)$$

where κ is a proportionality constant. We call $Q(x)$ the *implicit* energy density function and, accordingly, choose κ such that $\int_{-\infty}^{\infty} Q(x)dx = 1$.

3.1.2. Moments of the energy increment

We now derive an intermediate expression for ρ_2 using the properties of separable models. From the joint probability density function of the energy y at the next time step and of the current energy x ,

$$f(y, x) = f(y|x)P_s(x),$$

where $P_s(x)$ is the probability that the system has energy x , we compute the n^{th} order moment of the energy increment,

$$E\{\xi^n(s)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y-x)^n f(y, x) dy dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y-x)^n f(y|x)P_s(x) dy dx,$$

or equivalently, letting $z = y - x$,

$$E\{\xi^n(s)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z^n f(y|y-z)P_s(y-z) dy dz.$$

Substituting (33) into the above expression and noting that the integral associated with the delta function vanishes for $n > 0$, we obtain

$$E \{ \xi^0(s) \} = 1$$

and

$$E \{ \xi^n(s) \} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z^n \frac{A_s(z) Q(y)}{K(y-z)} P_s(y-z) dy dz \quad \text{for } n > 0.$$

The modified moments of the energy increment, ρ_n , defined as

$$\rho_n(s) \equiv \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z^n \frac{A_s(z) Q(y)}{K(y-z)} P_s(y-z) dy dz \quad \text{for } n \geq 0,$$

are identical to $E \{ \xi^n(s) \}$ except for $n = 0$; while $E \{ \xi^0(s) \} = 1$ is the probability of accepting or rejecting a move, ρ_0 , called the *acceptance ratio*, is the probability of accepting a move.

Replacing $P_s(y-z)$ with $P(y-z)e^{-s(y-z)} / Z(s)$ and $P(y-z) / K(y-z)$ with $Q(y-z) / \kappa$, we obtain

$$\rho_n(s) = \frac{1}{\kappa Z(s)} \int_{-\infty}^{\infty} z^n A_s(z) \int_{-\infty}^{\infty} Q(y) Q(y-z) e^{-s(y-z)} dy dz.$$

Defining $Q_s(x)$ as

$$Q_s(x) \equiv \frac{Q(x) e^{-sx}}{Z_Q(s)} \quad \text{with} \quad Z_Q(s) = \int_{-\infty}^{\infty} Q(x) e^{-sx} dx, \quad (37)$$

we rewrite ρ_n as

$$\rho_n(s) = \frac{Z_Q(s)}{\kappa Z(s)} \int_{-\infty}^{\infty} z^n A_s(z) H_s(z) dz \quad (38)$$

where

$$H_s(z) = \int_{-\infty}^{\infty} Q(y) Q_s(y-z) dy = Q(z) * Q_s(-z). \quad (39)$$

Finally, we eliminate the factor $Z_Q(s) / (\kappa Z(s))$ in (38) [12, Appendix 3.A] and arrive at

$$\rho_n(s) = \begin{cases} \frac{2 \int_0^{\infty} z^n G(|z|) H_s(-z) dz}{\int_{-\infty}^{\infty} G(|z|) H_s(-z) dz} & \text{if } n \text{ is even} \\ 0 & \text{if } n \text{ is odd.} \end{cases} \quad (40)$$

3.2. Locality model

3.2.1. Exponential locality model

The separable model for $f_p(x_p | x)$ is made up of two functions: the locality function G and the implicit energy density function Q . We shall model G in this section and Q in Section 3.3. We propose for the function G the *exponential locality model*:

$$G(|z|) \equiv e^{-\beta|z|}, \quad \beta > 0, \quad (41)$$

where $z = |x_p - x|$ and β depends on the move generation strategy. The parameter β models the effect of locality: the closer the proposed energy is to the current energy, the higher the probability that it is proposed.

Using this model, we rewrite the expression for the modified moments of the energy increment, ρ_n , into a form that is used in Section 3.3. We first decompose $G(|z|) = e^{-\beta|z|}$ into two parts,

$$g_+(z) = \begin{cases} e^{-\beta z} & \text{if } z \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad g_-(z) = \begin{cases} e^{\beta z} & \text{if } z < 0 \\ 0 & \text{otherwise} \end{cases},$$

with Fourier transforms

$$F\{g_+(z)\} = \frac{1}{\beta - j\omega} \quad \text{and} \quad F\{g_-(z)\} = \frac{1}{\beta + j\omega}.$$

Then, we define $L_+^{(0)}$ and $L_-^{(0)}$ as

$$L_+^{(0)}(s) \equiv \int_0^{\infty} G(|z|) H_s(-z) dz \quad \text{and} \quad L_-^{(0)}(s) \equiv \int_{-\infty}^0 G(|z|) H_s(-z) dz$$

and substitute (41) into these definitions to obtain

$$\begin{cases} L_+^{(0)}(\beta, s) = \int_0^{\infty} e^{-\beta z} H_s(-z) dz = \int_{-\infty}^{\infty} g_+(z) H_s(\beta, -z) dz \\ L_-^{(0)}(\beta, s) = \int_{-\infty}^0 e^{\beta z} H_s(-z) dz = \int_{-\infty}^{\infty} g_-(z) H_s(\beta, -z) dz \end{cases} \quad (42)$$

Applying Parseval's second theorem for real functions,

$$\int_{-\infty}^{\infty} F\{f(z)\} F^*\{g(z)\} d\omega = \int_{-\infty}^{\infty} f(z) g(z) dz$$

where F^* denotes the complex conjugate of F , and substituting the complex conjugate of the Fourier transform of equation (39) into the preceding expressions for $L_+^{(0)}$ and $L_-^{(0)}$, we get

$$\begin{cases} L_+^{(0)}(\beta, s) = \int_{-\infty}^{\infty} \frac{F\{Q(\beta, x)\} F^*\{Q_s(\beta, x)\}}{\beta - j\omega} d\omega \\ L_-^{(0)}(\beta, s) = \int_{-\infty}^{\infty} \frac{F\{Q(\beta, x)\} F^*\{Q_s(\beta, x)\}}{\beta + j\omega} d\omega \end{cases} \quad (43)$$

where the dependency of $Q(\beta, x)$ on β can be observed from (32) and (36). Finally, substituting

these expressions back into (40), we arrive at

$$\rho_n(\beta, s) = \begin{cases} \frac{2L_+^{(n)}(\beta, s)}{L_+^{(0)}(\beta, s) + L_-^{(0)}(\beta, s)} & \text{if } n \text{ is even} \\ 0 & \text{if } n \text{ is odd} \end{cases} \quad (44)$$

where $L_+^{(n)}$ is defined as

$$L_+^{(n)}(\beta, s) \equiv \int_0^\infty z^n e^{-\beta z} H_s(-z) dz = (-1)^n \left[\frac{\partial^n}{\partial \alpha^n} \int_0^\infty e^{-\alpha z} H_s(-z) dz \right]_{\alpha=\beta}$$

and can be rewritten in the form

$$L_+^{(n)}(\beta, s) = (-1)^n \left[\frac{\partial^n}{\partial \alpha^n} \int_{-\infty}^\infty \frac{F\{Q(\beta, x)\} F^*\{Q_s(\beta, x)\}}{\alpha - j\omega} d\omega \right]_{\alpha=\beta} \quad (45)$$

using (42) and (43).

Thus far, the only constraint for β is $\beta > 0$. If we allow β to change by dynamically controlling move generation, theoretically, we can obtain a β that improves the performance of simulated annealing by maximizing ρ_2 . The necessary condition under which ρ_2 is maximized for a given s is

$$\frac{\partial \rho_2(\beta, s)}{\partial \beta} = 0. \quad (46)$$

In Section 4, we use this formula together with (43) to (45) and the energy density models to construct the final form of the efficient λ -schedule.

3.2.2. Test of the exponential locality model

To assess the quality of the exponential locality model, we compare the computed and measured conditional probability density function $f_p(x_p | x)$ for different values of x . The conditional probability density function $f_p(x_p | x)$ is computed from the expression

$$f_p(x_p | x) \propto e^{-\beta |x_p - x|} [P_s(x_p) e^{sx_p}]^{1/2}. \quad (47)$$

This expression is obtained by substituting (41) into (31),

$$f_p(x_p | x) \propto e^{-\beta |x_p - x|} Q(x_p),$$

replacing $Q(x_p)$ with $P(x_p)^{1/2}$ using (54) (derived in Section 3.3),

$$f_p(x_p | x) \propto e^{-\beta |x_p - x|} [P(x_p)]^{1/2},$$

and substituting $P_s(x_p) e^{sx_p}$ for $P(x_p)$ using (9). Since direct measurement of β is not possible, we estimate it at different temperatures from measured values of $\rho_2(s)$, the variance of the energy increment, using (65) (derived in Section 3.3),

$$\rho_2(s) \approx \frac{4(2\beta - s)}{\beta(2\beta + s)^2}.$$

The tests were performed on a 100-city traveling salesman problem and a graph partition problem with 500 vertices and average degree 20. The results for the graph partition problem are shown

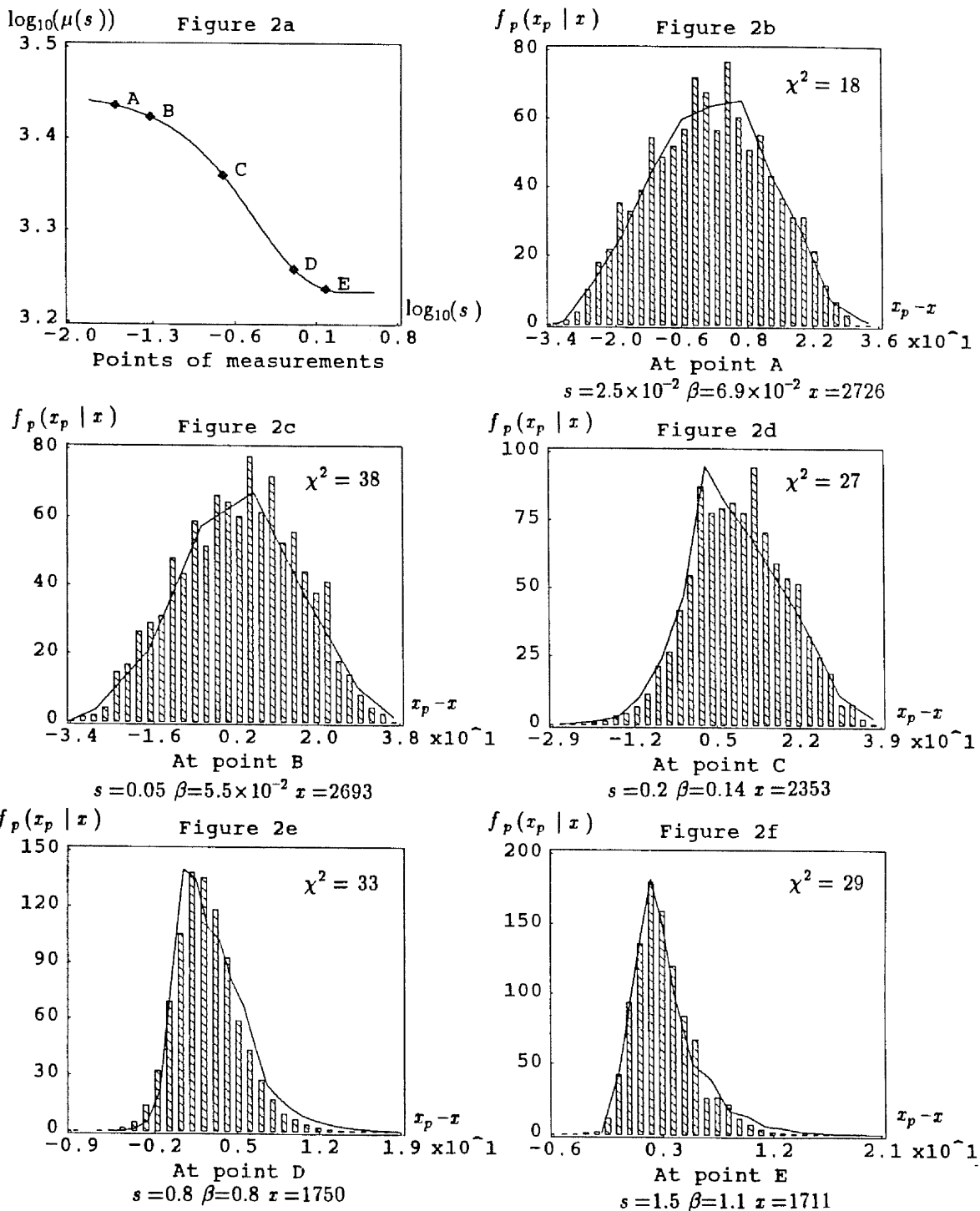


Figure 2: Test of the modeling of the conditional probability density function $f_p(x_p | x)$ for different values of x for a graph partition problem with 500 vertices and average degree 20. The inverse temperatures at which the tests were performed are indicated in Figure 2a while the measured $f_p(x_p | x)$ (histograms) and the computed $f_p(x_p | x)$ (curves) are shown in Figure 2b to Figure 2f.

in Fig. 2. The computed curves agree reasonably well with the measured data throughout the entire temperature range. This agreement suggests that the exponential locality model, although simple, captures the essence of move generation strategies. Note that these experiments also test the validity of the separable model of Section 3.1 and the energy density model of Section 3.3.

A difficulty arose when we attempted to measure $P_s(x_p)$ used in (47). Since the first autocorrelation coefficient, $r(s)$, is very close to 1 at low temperature, with typical values near 0.99, sequential measurements of the energy would yield highly correlated data, giving us little information about $P_s(x_p)$. So, instead of obtaining the data sequentially in one execution, we collected the data in 60 different executions with 1,000 data points in each execution.

3.3. Energy density model

3.3.1. Gamma density model

The implicit energy density function, $Q(x)$ which is part of the separable model, depends on the energy density function, $P(x)$. In order to find a closed form expression for ρ_2 we need to know $P(x)$. However, owing to the complexity of a typical optimization problem, we rarely know $P(x)$ and must, therefore, resort to modeling.

We model $P(x)$ as a gamma density function with parameters \bar{N} and \bar{b} ,

$$\frac{\bar{b}^{\bar{N}}}{\Gamma(\bar{N})} (x - x_0)^{\bar{N}-1} e^{-\bar{b}(x-x_0)}, \quad x > x_0,$$

where $\Gamma(\bar{N}) = \int_0^{\infty} x^{\bar{N}-1} e^{-x} dx$ is the gamma function. The main reasons for using this model are that it is simple, it leads to manageable analysis, and unlike the normal density function, it has a bounded left tail which corresponds to a finite global minimum (at $x = x_0$). The last reason is important, for all minimization problems have a finite global minimum that must be modeled in order to give accurate descriptions of energy density functions.

Without loss of generality, we assume that the global minimum is at $x_0 = 0$. The case $x_0 \neq 0$ can be treated similarly via a simple substitution of variable. From equation (36),

$$Q(\beta, x) = \kappa \frac{P(x)}{K(\beta, x)}, \quad (48)$$

and equation (32) with the exponential locality model,

$$K(\beta, x) = \int_{-\infty}^{\infty} e^{-\beta|y-x|} Q(\beta, y) dy, \quad (49)$$

we observe that if $Q(\beta, x)$ is specified, $P(x)$ can be obtained from $Q(\beta, x)$ by evaluating an integral; however, if $P(x)$ is specified, $Q(\beta, x)$ must be obtained from $P(x)$ by solving an integral equation. Therefore, we shall check that $P(x)$ is a gamma density function under the assumption that $Q(\beta, x)$ is a gamma density function.

Let $Q(\beta, x)$ be a gamma density function with parameters $N = N(\beta)$ and $b = b(\beta)$,

$$Q(\beta, x) = \frac{b^N}{\Gamma(N)} x^{N-1} e^{-bx}, \quad x > 0. \quad (50)$$

The mean and variance of $Q(\beta, x)$ are

$$\mu_Q = \frac{N}{b} \quad \text{and} \quad \sigma^2_Q = \frac{N}{b^2}, \quad (51)$$

respectively. Substituting (49) into (48), we obtain

$$P(x) = \frac{1}{\kappa} Q(\beta, x) \int_{-\infty}^{\infty} e^{-\beta|y-x|} Q(\beta, y) dy. \quad (52)$$

If the variance of $\beta e^{-\beta|y-x|}/2$ (the normalized form of $e^{-\beta|y-x|}$ in the above expression) is much smaller than the variance of $Q(\beta, y)$, specifically

$$\frac{2}{\beta^2} \ll \sigma^2_Q, \quad (53)$$

we can approximate $\beta e^{-\beta|y-x|}/2$ by the delta function $\delta(y-x)$ to get

$$P(x) \approx \frac{1}{\kappa} Q(\beta, x) \int_{-\infty}^{\infty} \delta(y-x) Q(\beta, y) dy = \frac{1}{\kappa} Q(\beta, x)^2 \quad (54)$$

Since (53) is easily satisfied and (54) shows that the dependency of $Q(\beta, x)$ on β is insignificant since $P(x)$ is independent of β , we use $Q(x)$ in place of $Q(\beta, x)$ hereafter! Substituting (50) into (54) yields

$$P(x) \propto \frac{b^{2N}}{\Gamma(N)^2} x^{2N-2} e^{-2bx}, \quad x > 0.$$

Including the normalization constant into the expression, we do arrive at

$$P(x) = \frac{\bar{b}^{\bar{N}}}{\Gamma(\bar{N})} x^{\bar{N}-1} e^{-\bar{b}x}, \quad x > 0,$$

a gamma density function with parameters $\bar{N} = 2N-1$ and $\bar{b} = 2b$, whose mean and variance are

$$\mu = \frac{\bar{N}}{\bar{b}} \quad \text{and} \quad \sigma^2 = \frac{\bar{N}}{\bar{b}^2}, \quad (55)$$

respectively. The variance of $Q(x)$ in (51) is about twice the variance of $P(x)$, hence we rewrite the condition in (53) as

$$\frac{1}{\beta^2} \ll \sigma^2. \quad (56)$$

3.3.2. Moments of the energy increment

Now we can find closed form expressions for ρ_0 , the acceptance ratio, and ρ_2 , the variance of the energy increment. We start with equation (44):

$$\rho_n(\beta, s) = \frac{2L_+^{(n)}(\beta, s)}{L_+^{(0)}(\beta, s) + L_-^{(0)}(\beta, s)} \quad \text{if } n \text{ is even} \quad (57)$$

in which

$$L_+^{(n)}(\beta, s) = (-1)^n \frac{\partial^n}{\partial \beta^n} \int_{-\infty}^{\infty} \frac{\phi(\omega) \phi_s^*(\omega)}{\beta - j\omega} d\omega \quad (58)$$

and

$$L_-^{(0)}(\beta, s) = \int_{-\infty}^{\infty} \frac{\phi(\omega) \phi_s^*(\omega)}{\beta + j\omega} d\omega \quad (59)$$

where $\phi(\omega)$ and $\phi_s^*(\omega)$ are the Fourier transforms of $Q(x)$ and $Q_s(-x)$, respectively. The expression for $Q_s(x)$,

$$Q_s(x) = \frac{1}{Z_Q(s)} \left[\frac{b^N}{\Gamma(N)} x^{N-1} e^{-cx} \right], \quad x > 0,$$

where $c = b + s$, is obtained by substituting (50) into (37). Since $Z_Q(s)$ is such that $\int_{-\infty}^{\infty} Q_s(x) dx = 1$, $Q_s(x)$ is a gamma density function with parameters N and c ,

$$Q_s(x) = \frac{c^N}{\Gamma(N)} x^{N-1} e^{-cx}, \quad x > 0. \quad (60)$$

Expressions for $L_+^{(0)}$ and $L_-^{(0)}$ are found in [12, Appendix 3.B]:

$$L_+^{(0)}(\beta, s) = \frac{\Gamma(2N)}{N \Gamma(N)^2} \left[\frac{bc}{(b+c)^2} \right]^N \frac{b+c}{2\beta+s} (1 + \Re_+^{(0)}), \quad \beta < 3b+s, \quad (61)$$

and

$$L_-^{(0)}(\beta, s) = \frac{\Gamma(2N)}{N \Gamma(N)^2} \left[\frac{bc}{(b+c)^2} \right]^N \frac{b+c}{2\beta-s} (1 + \Re_-^{(0)}), \quad \frac{s}{2} < \beta < 3b+2s, \quad (62)$$

with

$$\begin{aligned} |\Re_+^{(0)}| &\leq \left| \frac{2z_+(1-z_+)}{(N+1)(4z_+(1-z_+) - 1)} \right|, \\ |\Re_-^{(0)}| &\leq \left| \frac{2z_-(1-z_-)}{(N+1)(4z_-(1-z_-) - 1)} \right| \end{aligned}$$

where $z_+ = (b-\beta)/(b+c)$ and $z_- = (c-\beta)/(b+c)$. (The conditions on β ensure that $-1 < z_+$, $z_- < 1/2$.) The expression for $L_+^{(2)}$ [12, Appendix 3.D] is

$$L_+^{(2)}(\beta, s) = \frac{\Gamma(2N)}{N \Gamma(N)^2} \left[\frac{bc}{(b+c)^2} \right]^N \frac{8(b+c)}{(2\beta+s)^3} (1 + \Re_+^{(2)}), \quad \beta < 3b+s, \quad (63)$$

with

$$\Re_+^{(2)} = R_1 - \frac{5}{2(1-2z_+)^2(N+1)} (1 + R_2) + \frac{3}{(1-2z_+)^4(N+1)(N+2)} (1 + R_3),$$

where

$$\begin{aligned} |R_1| &\leq \left| \frac{2z_+(1-z_+)}{(N+1)(4z_+(1-z_+) - 1)} \right|, \\ |R_2| &\leq \left| \frac{12z_+(1-z_+)}{(N+2)(4z_+(1-z_+) - 1)} \right|, \\ |R_3| &\leq \left| \frac{30z_+(1-z_+)}{(N+3)(4z_+(1-z_+) - 1)} \right|, \end{aligned}$$

and $z_+ = (b - \beta) / (b + c)$. Substituting the expressions for $L_+^{(0)}$, $L_-^{(0)}$ and $L_+^{(2)}$ into (57), we get an expression for the acceptance ratio,

$$\rho_0(\beta, s) = \frac{2L_+^{(0)}(\beta, s)}{L_+^{(0)}(\beta, s) + L_-^{(0)}(\beta, s)} = \frac{2\beta - s}{2\beta}(1 + \mathfrak{R}_0), \quad (64)$$

and an expression for the variance of the energy increment,

$$\rho_2(\beta, s) = \frac{2L_+^{(2)}(\beta, s)}{L_+^{(0)}(\beta, s) + L_-^{(0)}(\beta, s)} = \frac{4(2\beta - s)}{\beta(2\beta + s)^2}(1 + \mathfrak{R}_2), \quad (65)$$

subject to the condition $s/2 < \beta < 3b + s$, where the bounds for \mathfrak{R}_0 and \mathfrak{R}_2 can be found from the bounds for $L_+^{(0)}$, $L_-^{(0)}$ and $L_+^{(2)}$. Since β is not a directly measurable parameter, it is advantageous to express ρ_2 in terms of ρ_0 , which may be measured directly. Eliminating β from (65) using (64), we finally arrive at the closed form expression for ρ_2 :

$$\rho_2(s) \approx \frac{8\rho_0(1 - \rho_0)^2}{s^2(2 - \rho_0)^2}, \quad \frac{s}{2} < \beta < 3b + s. \quad (66)$$

4. Efficient annealing schedule - final form

4.1. Schedule and constraints

To obtain the condition under which ρ_2 is maximized, we use the necessary condition in (46),

$$\frac{\partial \rho_2(\beta, s)}{\partial \beta} = 0.$$

Applying this formula to (65) and noting that the second derivative of ρ_2 with respect to β is always negative, we obtain an expression for the *desired* β that maximizes ρ_2 :

$$\beta_{des} \approx 0.89s. \quad (67)$$

Since the annealing schedule should operate with β close to β_{des} , the condition, $s/2 < \beta < 3b + s$, for the validity of (66) is clearly satisfied. Substituting β_{des} into (64) and (65), we obtain the *desired* acceptance ratio,

$$\rho_{0,des}(s) \approx 0.44, \quad (68)$$

and the *desired* variance of the energy increment,

$$\rho_{2,des}(s) \approx \frac{0.45}{s^2}. \quad (69)$$

Upon substitution of (69) into the efficient λ -schedule (30), we would arrive at the optimized (with respect to move generation) efficient λ -schedule

$$s_+ = s + \lambda \frac{0.45}{s^2 \sigma^3(s)} .$$

However, in general, we cannot use the optimized version for the entire temperature range because our means of controlling β so that β is close to β_{des} are limited and imprecise. Therefore, instead of using the optimized formula which would depart from the efficient λ -schedule associated with the actual move generation strategy, we try to keep β close to β_{des} and employ the annealing schedule

$$s_+ = s + \lambda \frac{4\rho_0(1 - \rho_0)^2}{s^2(2 - \rho_0)^2 \sigma^3(s)} , \quad (70)$$

obtained by substituting (66) into (30).

Time is ripe for the investigation of the constraints for the efficient λ -schedule. Since the efficient λ -schedule should operate close to its optimized version, investigations of the constraints for which β assumes its desired value, β_{des} , are appropriate.

We first show that constraint (19) from Section 2.3. is satisfied. Since $P(x)$ is a gamma density function with parameters \bar{N} and \bar{b} , $P_\zeta(x)$ is also a gamma density function, with parameters \bar{N} and $\bar{b} + \zeta$. Thus, the variance of $P_\zeta(x)$ is

$$\sigma^2(\zeta) = \frac{\bar{N}}{(\bar{b} + \zeta)^2} , \quad (71)$$

and

$$\left| \frac{\partial \sigma(\zeta)}{\partial \zeta} \right| = \frac{\sqrt{\bar{N}}}{(\bar{b} + \zeta)^2} .$$

Substituting these two expressions into (19), we obtain the inequality $\sqrt{\bar{N}} \gg \lambda$. Since λ is less than 1 this inequality will be satisfied if $\bar{N} \gg 1$. Using (55) this condition may be expressed as $\mu^2(0)/\sigma^2(0) \gg 1$ or, accounting for the generally non-zero minimum energy x_0 ,

$$\frac{(\mu(0) - \mu(\infty))^2}{\sigma^2(0)} \gg 1 . \quad (72)$$

The investigation of (18), the other constraint from Section 2.3, requires results from Section 3.3 and hence is postponed till after the constraints from Section 3.3 have been analyzed. Three restrictions were used for the derivation and simplification of ρ_2 in Section 3.3. First, in order to ensure that the expression (66) for ρ_2 is a good approximation (assuming the models are good), the expressions for the bounds on the error terms in (62) to (64) should be small. The bounds for $\mathfrak{R}_\mp^{(0)}$ and $\mathfrak{R}_\mp^{(2)}$ increase as $z_+ \rightarrow 1/2$ while the bound for $\mathfrak{R}_-^{(0)}$ increases as $z_- \rightarrow 1/2$. Since z_- is larger than z_+ , a condition on s that ensures that z_- is sufficiently far away from $1/2$ such that $\mathfrak{R}_-^{(0)}$ is less than a few percent also ensures that $\mathfrak{R}_\mp^{(0)}$ and $\mathfrak{R}_\mp^{(2)}$ are small and hence that (66) is a good approximation. Such a restriction is derived in [12, Appendix 3.C]:

$$s \geq \frac{7}{\sigma(0)}. \quad (73)$$

Second, we show that the condition in (56),

$$\frac{1}{\beta^2} \ll \sigma^2(0),$$

needed for the validity of (54), is satisfied if (73) is satisfied. Substituting β_{des} into (56), we get

$$\frac{1.26}{s^2} \ll \sigma^2(0)$$

which, upon substitution of the restriction in (73), yields

$$\frac{1.26\sigma^2(0)}{7^2} \ll \sigma^2(0)$$

which is obviously satisfied.

Third, the bounds on the approximation errors $\mathfrak{R}_+^{(0)}$, $\mathfrak{R}_-^{(0)}$, R_1 , R_2 and R_3 are in fact exact up to an error term of the order of N^{-1} , hence condition (73) is accurate if

$$2N - 1 = \bar{N} \gg 1,$$

which is identical to condition (72).

Now we can show that the constraint in (18) is superseded by (73). Assuming (72) and (73) are satisfied, we can substitute $\rho_{2,des}$ into (29) to get

$$1 - r(s_+) = \frac{0.23(\bar{b} + s_+)^2}{\bar{N}s_+^2}. \quad (74)$$

Substituting this formula and (71) into (18) gives

$$\frac{\sqrt{\bar{N}}}{\bar{b} + s_+} \geq \lambda \frac{s_+^2 \bar{N}}{0.23(\bar{b} + s_+)^2} \left| \frac{\partial r(s_+)}{\partial s_+} \right|. \quad (75)$$

Since, from (74),

$$\left| \frac{\partial r(s_+)}{\partial s_+} \right| = \frac{0.45\bar{b}(\bar{b} + s_+)}{\bar{N}s_+^3},$$

(75) is equivalent to

$$\frac{\sqrt{\bar{N}}}{\bar{b} + s_+} \geq \lambda \frac{s_+^2 \bar{N}}{0.23(\bar{b} + s_+)^2} \frac{0.45\bar{b}(\bar{b} + s_+)}{\bar{N}s_+^3}.$$

which simplifies to

$$s_+ \geq \frac{2\lambda\bar{b}}{\sqrt{\bar{N}}},$$

and can be rewritten as

$$s_+ \geq \frac{2\lambda}{\sigma(0)}.$$

by making use of (55). Since $s_+ \geq s$ and $\lambda \leq 1$, this constraint is not as stringent as (73) which

was used to derive it. Therefore this constraint is superseded by (73).

Finally, combining the constraints from Section 2.3 and Section 3.3 and including them into the efficient λ -schedule, we arrive at

$$s_+ = s + \lambda \frac{4\rho_0(1-\rho_0)^2}{s^2(2-\rho_0)^2\sigma^3(s)}, \quad s \geq \frac{7}{\sigma(0)} \quad \text{and} \quad \frac{(\mu(0)-\mu(\infty))^2}{\sigma^2(0)} \gg 1. \quad (76)$$

Recall that the above expressions for the constraints were derived assuming $\beta = \beta_{des}$ and thus might not apply if β were to differ significantly from β_{des} .

4.2. Sensitivity analysis

Now we present arguments indicating that the results obtained under the gamma density model assumption should remain essentially valid for the energy density functions encountered in practice, i.e., that relations (66) and (76) are not very sensitive to the energy function. The gamma density model, while typically a good model of an energy density function at a few standard deviations around its mean (hence for high temperatures), represents the *left tail* much less satisfactorily. However an accurate representation of the left tail is important at low temperatures. Since the mathematical form of the gamma model is what allowed us to carry out the derivations of the efficient λ -schedule in Section 3.3, we seek approximations of the actual energy density functions whose form is a natural extension of the gamma model. Since the classical orthogonal polynomials associated to a gamma weight function are Laguerre polynomials, our approximations will be projections on basis functions constructed out of these polynomials. The Laguerre polynomials, $\mathcal{L}_n^{(\alpha)}(x)$, are defined recursively as

$$\begin{aligned} \mathcal{L}_{-1}^{(\alpha)}(x) &= 0, \quad \mathcal{L}_0^{(\alpha)}(x) = 1, \\ (n+1) \mathcal{L}_{n+1}^{(\alpha)}(x) - [2n+\alpha+1-x] \mathcal{L}_n^{(\alpha)}(x) + (n+\alpha) \mathcal{L}_{n-1}^{(\alpha)}(x) &= 0, \quad n = 0, 1, \dots, \end{aligned}$$

and are of the form

$$\mathcal{L}_n^{(\alpha)}(x) = \sum_{i=0}^n l_i^{(n)} x^i. \quad (77)$$

Without loss of generality we can assume in this analysis, as we did in Section 3.3, that all functions are defined on the half-line $[0, \infty)$, i.e., the ground states have energy $x_0 = 0$. The functions

$$\psi_n(N, b; x) = \left[\frac{\Gamma(n+1)}{\Gamma(2N+n-1)} \right]^{1/2} (2b)^{N-1/2} x^{N-1} e^{-bx} \mathcal{L}_n^{(2N-2)}(2bx)$$

$$n = 0, 1, \dots, \quad b > 0, \quad N > 1/2,$$

form an orthogonal Cartesian basis on $[0, \infty)$. Given an implicit energy function $Q(x)$ and given N, b and M , the approximation of the form

$$q_M(N, b; x) = \sum_{i=0}^M a_i \psi_i(N, b; x) \quad (78)$$

which minimizes the approximation error

$$D(Q, q_M(N, b)) \equiv \int_0^{\infty} [Q(x) - q_M(N, b; x)]^2 dx.$$

has coefficients a_i equal to $\int_0^{\infty} Q(x) \psi_i(N, b; x) dx$.

If $Q(x)$ is a gamma density function and the parameters N and b of the basis are identical to those of Q , a single term (hence $M = 0$) is sufficient to provide a perfect approximation. Moreover, by changing the parameters of the basis with the temperature and using N and $c = b + s$, one still needs a single term to approximate $Q_s(x) = Q(x)e^{-sx}/Z_Q(s)$ perfectly. This observation leads us, for arbitrary density functions, to allow the parameters of the basis to change with s . Our goal is to approximate $Q_s(x)$ as well as possible using only a few terms, therefore at each temperature we should select the parameters that minimize M while giving a sufficiently small approximation error. Typically, if the parameters are properly selected at each temperature, $M = 0$ leads to a satisfactory approximation of $Q_s(x)$; yet, since $Q(x)$ is not a gamma function, the *evolution* of the basis parameters with s departs significantly from what would happen for a gamma function.

The expressions for $\rho_0(\beta, s)$ and $\rho_2(\beta, s)$ are found by first computing the expressions for $L_{\mp}^{(0)}$, $L_{\pm}^{(0)}$, and $L_{\mp}^{(2)}$. We start with the approximation for $Q_s(x)$

$$q_{s, M}(x) = \sum_{i=0}^M f_i \psi_i(N, c; x), \quad (79)$$

where we assume that N and c are selected so that M and the approximation error are small, and approximate $Q(x)$ by

$$\sum_{i=0}^M d_i \psi_i(N, c; x) e^{sx}$$

with $d_i = f_i Z_Q(s)$. As s changes, the parameters N , $b \equiv c - s$ and M may vary. If M is much smaller than N , the expressions for $L_{\mp}^{(0)}$ and $L_{\pm}^{(0)}$ are well approximated by [12, Appendix 3.G]

$$L_{\mp}^{(0)}(\beta, s) \approx U(N, M, b, s) \frac{b + c}{2\beta + s}, \quad \beta < 3b + s, M \ll N,$$

and

$$L_{\pm}^{(0)}(\beta, s) \approx U(N, M, b, s) \frac{b + c}{2\beta - s}, \quad \frac{s}{2} < \beta < 3b + 2s, M \ll N,$$

where

$$U(N, M, b, s) = \sum_{n=0}^M g_n \sum_{j=0}^n l_j^{(n)}(2c)^j \sum_{m=0}^M h_m \sum_{i=0}^m l_i^{(m)}(2c)^i \frac{\Gamma(2N + i + j)}{(N + i)(b + c)^{2N + i + j}},$$

with

$$g_n = f_n (2c)^{N-1/2} \left[\frac{\Gamma(n+1)}{\Gamma(2N+n-1)} \right]^{1/2} \quad \text{and} \quad h_m = d_m (2c)^{N-1/2} \left[\frac{\Gamma(m+1)}{\Gamma(2N+m-1)} \right]^{1/2}.$$

The expression for $L_{\mp}^{(2)}$ is obtained, according to (45), by differentiating $L_{\mp}^{(0)}$ twice with respect to β

$$L_{\mp}^{(2)}(\beta, s) \approx U(N, M, b, s) \frac{8(b+c)}{(2\beta+s)^3}, \quad M \ll N. \quad (80)$$

Note that the result holds even if N and b vary with s . The difference with Section 3.3 is that we cannot derive analytic expressions for the error bounds and the approximation errors are clearly larger than for a gamma density function. Substituting the expressions for $L_{\mp}^{(0)}$, $L_{\pm}^{(0)}$ and $L_{\mp}^{(2)}$ into (57), we get, for the acceptance ratio,

$$\rho_0(\beta, s) = \frac{2L_{\mp}^{(0)}(\beta, s)}{L_{\mp}^{(0)}(\beta, s) + L_{\pm}^{(0)}(\beta, s)} \approx \frac{2\beta - s}{2\beta} \quad (81)$$

and, for the variance of the energy increment,

$$\rho_2(\beta, s) = \frac{2L_{\mp}^{(2)}(\beta, s)}{L_{\mp}^{(0)}(\beta, s) + L_{\pm}^{(0)}(\beta, s)} \approx \frac{4(2\beta - s)}{\beta(2\beta + s)^2} \quad (82)$$

subject to the conditions $s/2 < \beta < 3b + s$ and $M \ll N$. These expressions are identical to the corresponding expressions obtained for the gamma density model.

Since the expressions (81) and (82) for ρ_0 and ρ_2 are the same as those obtained for the gamma density model we obtain the same desired value for the acceptance ratio $\rho_{0,des}(s) \approx 0.44$ and the same expression as in (76) for the efficient λ -schedule. Furthermore, it can be argued [12] that these results are valid under the same constraints as in (76). While more precise statements could be obtained for specific density functions by evaluating the approximation errors numerically, the above analysis does indicate that the results obtained under the gamma density model can be applied to real problems and remain close to optimal.

To test for the accuracy of the approximation (78) for $Q_s(x)$, we must construct the histogram of $Q_s(x)$ from the histogram of $P_s(x)$, the quantity that can be measured directly. To this purpose we employ the relation

$$Q_{s/2}(x) = \left[\frac{\kappa Z(s)}{Z_Q(s/2)^2} P_s(x) \right]^{1/2},$$

obtained by substituting the expression $P(x) = Q(x)^2/\kappa$ into the definition of $P_s(x)$ to get

$$P_s(x) = \frac{1}{\kappa Z(s)} Q(x)^2 e^{-sx}.$$

and then noting that

$$Q(x)^2 e^{-sx} = Z_Q(s/2)^2 Q_{s/2}(x)^2.$$

Since we need to measure $P_s(x)$, the problem of highly correlated data associated with the testing of the exponential locality model would occur. Therefore, we need to combine data from multiple executions, as described in Section 3.2.2, in order to obtain an accurate description of $Q_s(x)$. We performed the tests on a 100-city travel salesman problem and a graph partition problem. The results for the traveling salesman problem are shown in Fig. 3. We used single term approximations, $M = 0$, to approximate $Q_s(x)$ throughout the whole temperature range. Note that the parameters N and $b = c - s$ are different at different inverse temperatures. (The results for the graph partition problem are similar [12].)

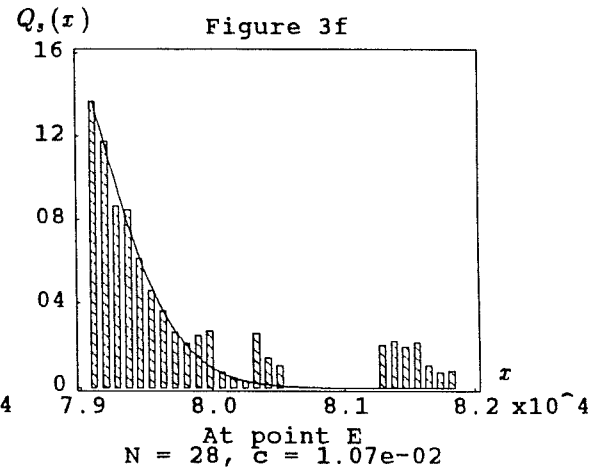
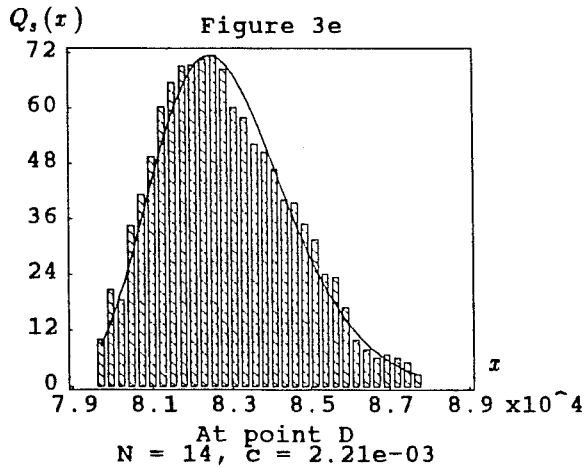
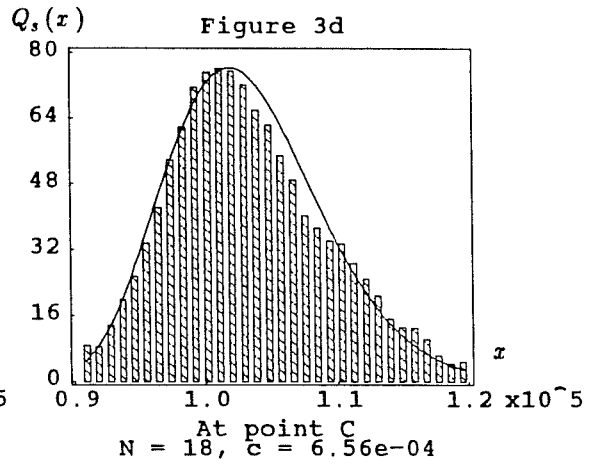
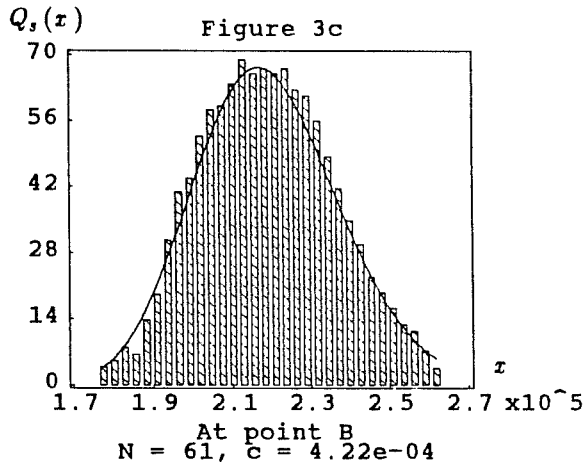
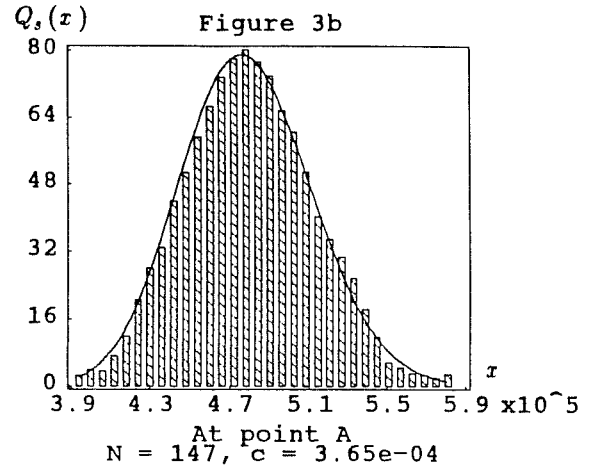
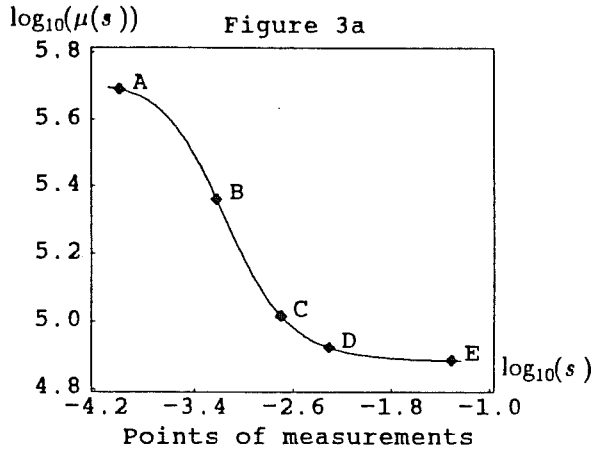


Figure 3: Test results for $Q_s(x)$ at different values of s for a 100-city traveling salesman problem. The inverse temperatures at which the tests were performed are indicated in Figure 3a while the measured $Q_s(x)$ (histograms) and the fitted gamma density functions (curves) with parameters N and c are shown in Figure 3b to Figure 3f.

5. Conclusions

The increment in inverse temperature in the efficient λ -schedule can be decomposed into three factors:

$$s_{n+1} = s_n + \left[\frac{\lambda}{\sigma(s_n)} \right] \left[\frac{1}{s_n^2 \sigma^2(s_n)} \right] \left[\frac{4\rho_0(1-\rho_0)^2}{(2-\rho_0)^2} \right].$$

The first factor, which depends on the quasi-equilibrium criterion, is the same as the factors in Aarts's and Huang's annealing schedules. The second factor relates the specific heat, which is the rate of change of the equilibrium average energy with respect to the temperature,

$$\frac{\partial \mu(s)}{\partial(1/s)} = s^2 \sigma^2(s),$$

to the temperature decrement. The presence of this factor is in agreement with the common belief that the larger is the specific heat, the slower should be the cooling. The third factor relates the acceptance ratio and, hence, the move generation strategy to the temperature decrement. As mentioned earlier, to maximize this factor, the move generation strategy should be controlled in order to arrive at a desired acceptance ratio of 44%. This is similar to the suggestion made by Binder [2, p. 11] on Monte Carlo methods in statistical physics that the magnitude of the proposed energy change should be chosen such that the acceptance ratio is close to 50%. Thus, not only the dynamics of the system is accounted for by our annealing schedule, but also the move generation strategy is actively controlled.

In [11], we discuss practical aspects of applying the efficient λ -schedule, including parameter estimation and move generation control, and assess the performance of the efficient λ -schedule by comparing it with the annealing schedules surveyed in the introduction. Furthermore, we evaluate the performance of simulated annealing as a general combinatorial optimization method by comparing it with other problem specific heuristics on some classical combinatorial optimization problems.

References

- [1] E. Aarts, and P. van Laarhoven, "Statistical Cooling Algorithm: A General Approach to Combinatorial Optimization Problems," *Philips J. of Research*, Vol. 40, No. 4, 193-226, 1985.
- [2] K. Binder, *Monte Carlo Methods in Statistical Physics*, 2nd Edition, Springer-Verlag, 1986.
- [3] L. Cooper and M. Cooper, *Introduction to Dynamic Programming*, Pergamon Press, 1981.
- [4] A. El Gamal, L. Hemachandra, I. Shperling, and V. Wei, "Using Simulated Annealing to Design Good Codes," *IEEE Trans. Information Theory*, Vol. 33, No. 1, 116-123, 1987.
- [5] S. Geman, and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 6, 721-741, 1984.
- [6] B. Hajek, "A Tutorial Survey of Theory and Applications of Simulated Annealing," *Proc. of the 24th IEEE CDC*, 755-760, 1985.
- [7] M. Huang, F. Romeo, and A. Sangiovanni-Vincentelli, "An Efficient General Cooling Schedule for Simulated Annealing," *Proc. of the IEEE ICCAD*, 381-384, 1986.

- [8] N. Van Kampen, *Stochastic Processes in Physics and Chemistry*, North-Holland, 1981.
- [9] S. Kirkpatrick, C. Gelatt Jr., and M. Vecchi, "Optimization by Simulated Annealing," *IBM Research Report, RC 9355*, 1982.
- [10] P. van Laarhoven, and E. Aarts, *Simulated Annealing: Theory and Applications*, D. Reidel Publishing Company, 1987.
- [11] J. Lam, and J.-M. Delosme, "An Efficient Simulated Annealing Schedule: Implementation and Evaluation," Technical Report 8817, Department of Electrical Engineering, Yale University, 1988.
- [12] J. Lam, "An Efficient Simulated Annealing Schedule," Ph.D. Dissertation, Department of Computer Science, Yale University, 1988.
- [13] M. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, "Equation of State Calculations by Fast Computing Machines," *J. of Chemical Physics*, Vol. 21, 1087-1092, 1953.
- [14] D. Mitra, F. Romeo, and A. Sangiovanni-Vincentelli, "Convergence and Finite-time Behavior of Simulated Annealing," *Proc. of the 24th IEEE CDC*, 761-767, 1985.
- [15] C. Sechen, and A. Sangiovanni-Vincentelli, "The TimberWolf Placement and Routing Package," *IEEE J. of Solid-State Circuits*, Vol. 20, No. 2, 510-522, 1985.
- [16] S. White, "Concepts of Scale in Simulated Annealing," *Proc. of the IEEE ICCD*, 646-651, 1984.