Advanced School on

STOCHASTICS IN COMBINATORIAL OPTIMIZATION

CISM — Udine — Italy
September 22-25 1986

Edited by
G Andreatta
F Mason
P Serafini

World Scientific

Advanced School on

STOCHASTICS IN COMBINATORIAL OPTIMIZATION

CISM — Udine — Italy
September 22-25 1986

Edited by

G Andreatta
F Mason
P Serafini

World Scientific
Singapore • New Jersey • Hong Kong

Published by

World Scientific Publishing Co. Pte. Ltd.
P.O. Box 128, Farrer Road, Singapore 9128

U. S. A. office: World Scientific Publishing Co., Inc.
687 Hartwell Street, Teaneck NJ 07666, USA

Library of Congress Cataloging-in-Publication data is available.

STOCHASTICS IN COMBINATORIAL OPTIMIZATION

ISBN 9971-50-456-1

Printed in Singapore by Utopia Press.

PREFACE

This volume contains the proceedings of the advanced school on "Stochastics in Combinatorial Optimization" held at CISM, Udine, Italy from September 22 to 25, 1986. The planning of the school was motivated by the observation that there has been a growing interest in developing stochastic techniques to solve deterministic problems (e.g. "simulated annealing") and also by the fact that many real life problems are inherently stochastic leading to stochastic combinatorial optimization models (e.g. "stochastic vehicle routing").

Aim of the school was to present the state of the art for both aspects and to investigate possible common links. The interest in this subject is witnessed by the several lectures given by the speakers. Various points of view and results are reported in the papers collected in this volume.

A first group of papers deals with randomized methods for deterministic problems. S.B. Gelfand and S.K. Mitter discuss the properties of simulated annealing techniques. F. Maffioli first considers in general the effect of using randomness to get efficient algorithms and then presents particular randomized heuristics for NP-hard problems. Problems of establishing good stopping rules for stochastic algorithms are investigated by C.G. Bender, A.H.G. Rinnooy Kan and C. Vercellis.

Stochastic versions of combinatorial optimization problems are discussed in various papers. A.R. Odoni analyzes facility location problems in stochastic networks where travel times are random or queuing phenomena may arise. P. Mirchandani and H. Soroush address the problem of determining optimal paths, multicommodity flows and traffic equilibria in a stochastic environment and with nonlinear utility functions. A survey of models concerning the shortest path problem in stochastic networks is presented by G. Andreatta. Results concerning the computational complexity of the PERT, shortest path and maximum flow problems in different classes of stochastic networks are summarized by J. Kamburowski. The paper of P. Jaillet deals with recent results on the probabilistic traveling salesman problem (PTSP) and with a generalization of it to the probabilistic vehicle routing problem. Heuristic methods for the PTSP are

investigated by F. Rossi and I. Gavioli.

Other contributions do not exactly fit into either category. R. Hassin outlines a computing scheme for some network problems with random edge lengths that "almost surely" leads to computational improvements over existing algorithms. R. Szkatula and M. Libura investigate probabilistic properties of greedy-like algorithms for the knapsack problem and W. Woess outlines the properties of random walks on infinite graphs.

We would like to express our gratitude to those institutions who made possible the school : the Science Sector of Unesco, the Committee for Economics and the Committee for Mathematics of the Italian Research Council (CNR) for their financial contributions, and the International Center for Mechanical Sciences (CISM) for the organizational support. We would also like to thank the participants for the friendly atmosphere and the active interest shown during the school.

Giovanni Andreatta
Francesco Mason
Paolo Serafini

TABLE OF CONTENTS

SIMULATED ANNEALING

by

Saul B. Gelfand[1]
Sanjoy K. Mitter[2]

[1]Department of Electrical Engineering, Purdue University, Lafayette, Indiana

[2]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139

This research has been supported by the Air Force Office of Scientific Research under grant AFOSR-85-0227 and the Army Research Office under grants DAAG29-84-K-0005 and DAAL03-86-K-1071.

CHAPTER I
INTRODUCTION

Algorithms for finding a global extremum of a real-valued function may be classified into two groups: deterministic and random. The distinction here is of course that the random or Monte-Carlo algorithms make use of pseudo random variates whereas the deterministic algorithms do not. The earliest global optimization algorithms were of the deterministic type and were associated with evaluating the cost function at points on a grid. One drawback of these methods is that they typically require certain prior information about the cost function such as a Lipshitz constant. Most global optimization algorithms are of the random type and are related to the so-called multistart algorithm. In this approach, a local optimization algorithm is run from different starting points which are selected at random, usually from a uniform distribution on the domain of the cost function. See [5], [29] for a discussion of global optimization algorithms.

Recently, motivated by hard combinatorial optimization problems such as arise in computer design and operations research, Kirkpatrick et. al. [19] and independently Cerny [3] have proposed a different kind of random algorithm called *simulated annealing*. The annealing algorithm is based on an analogy between large scale optimization problems and statistical mechanics. For our purposes this analogy consists simply of viewing the cost function as an energy function defined on a finite state space of an imaginary physical system. The annealing algorithm is then seen as a variation on a Monte-Carlo algorithm developed by Metropolis et. al. [25] for making statistical mechanics calculations, which we now describe. It is well-known that the states of a physical system in thermal equilibrium obey a Gibbs distribution $\propto \exp[-U(\cdot)/T]$, where $U(\cdot)$ is an energy function and T is the temperature. The Metropolis algorithm was developed for obtaining samples from such a Gibbs distribution and for computing estimates of functionals averaged over the Gibbs distribution. The Metropolis algorithm proceeds as follows:

> Given a state i of the system, select a candidate state j in a random manner corresponding to a small perturbation of the system, and

compute the change in energy $\Delta U = U(j) - U(i)$. If $\Delta U \leq 0$ accept state j as the new state for the next iteration of the algorithm. If $\Delta U > 0$ accept state j with probability $\exp[-\Delta U/T]$; otherwise the algorithm starts at state i for the next iteration.

The annealing algorithm consists of identifying the cost function to be minimized with the energy function $U(\cdot)$ and taking the temperature T as a function of time and slowly lowering it to zero. Suppose that the distribution of a candidate state is independent of past states given the current state. Then it is clear that the Metropolis algorithm simulates the sample paths of a Markov chain, and it can be shown that if the candidate states are selected in a suitable manner then this chain infact has a Gibbs distribution $\propto \exp[-U(i)/T]$ as its (unique) equilibrium distribution (see Chapter 2 for details). Furthermore as the temperature T is decreased to zero the Gibbs distribution concentrates more and more on the lower energy states. The motivation behind the annealing algorithm is that if $T \rightarrow 0$ slowly enough such that the system is never far away from equilibrium, then presumably there is convergence (in some probabilistic sense) to the global minima of $U(\cdot)$.

The annealing algorithm stands in contrast to heuristic methods for combinatorial optimization which are based on iterative improvement, allowing only decreases in the cost function at each iteration. Iterative improvement algorithms in statistical mechanics terms correspond to rapidly quenching a system from a high to a very low temperature. Such quenching can result in the system getting trapped in a so-called metastable state, and analogously the iterative improvement algorithm getting trapped in a strictly local minimum of the cost function. On the other hand, the annealing algorithm corresponds to slowly cooling a system. Such cooling should result in the system spending most of its time among low energy states and analogously the annealing algorithm finding a global or nearly global minimum of the cost function.

The annealing algorithm as described above is suitable for combinatorial optimization. Motivated by optimization problems with continuous variables which arise in image processing problems, Geman and independently Grenander [13] have proposed a diffusion-type algorithm called the *Langevin algorithm* (as coined by Gidas [11]). Consider the diffusion solution of the Langevin equation

$$dx(t) = -\nabla U(x(t))dt + \sqrt{2T}\, dw(t)$$

where $U(\cdot)$ is now a smooth function on r-dimensional Euclidean space (again called energy), T is a positive constant (again called temperature), and $w(\cdot)$ is

a standard r-dimensional Wiener process. The Langevin equation describes the motion of a particle in a viscous fluid. The Langevin algorithm consists of identifying the cost function to be minimized with the energy function $U(\cdot)$ and taking the temperature T as a function of time and slowly lowering it to zero. Now it is well known that under suitable conditions on $U(\cdot)$ the diffusion solution of the Langevin equation has a Gibbs density $\propto \exp[-U(\cdot)/T]$ as its (unique) equilibrium density, and as the temperature T is decreased to zero this density becomes more and more concentrated on the lower energy states. Like the annealing algorithm, the motivation behind the Langevin algorithm is that if $T \rightarrow 0$ slowly enough such that the system is never far away from equilibrium, then presumably there is convergence (in some probabilistic sense) to the global minima of $U(\cdot)$.

The annealing algorithm has been applied with varying success to a wide range of problems including circuit placement and wire routing for VLSI chip design [19], image reconstruction [8], and assorted hard combinatorial problems which arise in operations research [3], [12], [18], [19]. There has also been intense theoretical interest in both the annealing algorithm [8], [10], [11], [14], [15], [26], [31] and the Langevin algorithm [4], [9], [11], [15], [21].

CHAPTER II
FINITE STATE ANNEALING TYPE ALGORITHMS

2.1 Introduction to the Annealing Algorithm

In Chapter 1 we briefly described the annealing algorithm and discussed the heuristic motivation based on the connection that Kirkpatrick [19] has suggested between statistical mechanics and large-scale optimization problems. Mathematically, the annealing algorithm consists of simulating a nonstationary finite-state Markov chain whose state space is the domain of the cost function (called energy) to be minimized. In this Section we shall discuss in detail the annealing algorithm and describe some of the considerable literature which has been devoted to its analysis.

We first give some standard finite state space Markov chain notation (c.f. [6], [7]). Let Σ be a finite set. $P = [p_{ij}]_{i,j\in\Sigma}$ is a stochastic matrix on Σ if $p_{ij} \geq 0$ for all $i,j\in\Sigma$ and

$$\sum_{j\in\Sigma} p_{ij} = 1 \qquad \forall\, i\in\Sigma\,.$$

$\{P^{(k,k+1)}\} = \{[p_{ij}^{(k,k+1)}]\}$ are the 1-step transition matrices for a Markov chain $\{\xi_k\}$ with state space Σ if for every $k\in\mathbb{N}$ $P^{(k,k+1)}$ is a stochastic matrix on Σ and

$$P\{\xi_{k+1} = j\,|\xi_k = i\} = p_{ij}^{(k,k+1)} \quad (\text{if } P\{\xi_k = i\} > 0) \tag{2.1}$$

for all $i,j\in\Sigma$. Conversely, given a sequence $\{P^{(k,k+1)}\} = \{[p_{ij}^{(k,k+d)}]\}$ of stochastic matrices on Σ we can construct on a suitable probability space (Λ,F,P) a Markov chain $\{\xi_k\}$ with state space Σ which satisfies (2.1). For each $d\in\mathbb{N}$ let

$$P^{(k,k+d)} = P^{(k,k+1)} \cdot \ldots \cdot P^{(k+d-1,k+d)}\,.$$

$P^{(k,k+d)} = [p_{ij}^{(k,k+d)}]$ is a stochastic matrix on Σ and

$$P\{\xi_{k+d} = j\,|\xi_k = i\} = p_{ij}^{(k,k+d)} \quad (\text{if } P\{\xi_k = i\} > 0)$$

for all $i,j\in\Sigma$. It will be convenient to have a fixed version of the conditional probability of ξ_{k+d} given ξ_k which we define by

$$P\{\xi_{k+d}\in A|\xi_k = i\} = \sum_{j\in A} p_{ij}^{(k,k+d)}$$

for all $i\in\Sigma$ and $A\subset\Sigma$.

We now define the annealing algorithm. Let $U(\cdot)$ be a nonnegative function on Σ, called the *energy function*. The goal is to find a point in Σ which minimizes or nearly minimizes $U(\cdot)$. Let $\{T_k\}$ be a sequence of positive numbers, called the *temperature schedule*. Let $Q = [q_{ij}]$ be a stochastic matrix on Σ. Now let $\{\xi_k\}$ be the Markov chain with state space Σ and 1-step transition matrices $\{P^{(k,k+1)}\} = \{[p_{ij}^{(k,k+1)}]\}$ given by

$$p_{ij}^{(k,k+1)} = \begin{cases} q_{ij} \exp\left[-\dfrac{U(j) - U(i)}{T_k}\right] & \text{if } U(j) > U(i) \\ q_{ij} & \text{if } U(j) \le U(i),\ j\neq i \\ 1 - \sum\limits_{\ell\neq i} p_{i\ell}^{(k,k+1)} & \text{if } j = i \end{cases} \tag{2.2}$$

for all $i,j\in\Sigma$. $\{\xi_k\}$ shall be called the *annealing chain*. For each $d\in\mathbb{N}$ let $Q^d = [q_{ij}^{(d)}]$. Recall that Q is *irreducible* if for every $i,j\in\Sigma$ there exists a $d\in\mathbb{N}$ such that $q_{ij}^{(d)} > 0$. Also, Q is *symmetric* if $q_{ij} = q_{ji}$ for all $i,j\in\Sigma$. In the special case where Q is irreducible and symmetric and $T_k = T$, a positive constant, $\{\xi_k\}$ is the stationary Markov chain introduced by Metropolis et. al. [25] for computing statistics of a physical system in thermal equilibrium at temperature T. It was Kirkpatrick et. al. [19] and Cerny [3] who suggested that the Metropolis scheme could be used for minimizing $U(\cdot)$ by letting $T = T_k \longrightarrow 0$. We shall call the algorithm which simulates the sample paths of $\{\xi_k\}$ with $T_k\longrightarrow 0$ the *annealing algorithm*.

The heuristic motivation behind the annealing algorithm was discussed (briefly) in Chapter 1. Here we give the motivation in more mathematical terms. Suppose that Q is irreducible and symmetric, and let $\{\xi_k^T\}$ be the stationary chain with 1-step (stationary) transition matrix $P^T = [p_{ij}^T]$ given by the r.h.s of (2.2) with $T_k = T$, a positive constant. Then it can be shown that P^T has an invariant Gibbs vector $\Pi^T = [\pi_i^T]$ (a row vector), i.e.,

$$\Pi^T = \Pi^T P^T$$

where

$$\pi_i^T = \frac{\exp\,[-\,U(i)/T]}{\sum\limits_{j\in\Sigma} \exp\,[-\,U(j)/T]} \qquad \forall\, i\in\Sigma\,.$$

This follows from the detailed reversibility

$$\pi_i^T p_{ij}^T = \pi_j^T p_{ji}^T \qquad \forall\, i,j\in\Sigma\ .$$

Furthermore, Q irreducible and symmetric implies that $\{\xi_k^T\}$ is an irreducible† (and aperiodic) chain and by the Markov Convergence Theorem [6, p. 177]

$$\lim_{k\to\infty} P\{\xi_k^T = i\} = \pi_i^T \qquad \forall\, i\in\Sigma\,. \tag{2.3}$$

Let S be the set of global minima of $U(\cdot)$, i.e.

$$S = \{i\in\Sigma:\ U(i) \le U(j)\ \ \forall\, j\in\Sigma\}\,.$$

Now

$$\lim_{T\to 0} \pi_i^T = \pi_i^* \qquad \forall\, i\in\Sigma \tag{2.4}$$

where $\Pi^* = [\pi_i^*]$ is a probability vector with support in S. In view of (2.3) and (2.4) the idea behind the annealing algorithm is that by choosing $T = T_k \to 0$ slowly enough hopefully

$$P\{\xi_k = i\} \approx \pi_i^{T_k} \qquad (k \text{ large}) \tag{2.5}$$

and then perhaps

$$\lim_{k\to\infty} P\{\xi_k = i\} = \pi_i^* \qquad \forall\, i\in\Sigma \tag{2.6}$$

and consequently ξ_k converges in probability to S.

In Chapter 1 we roughly described the procedure by which the sample paths of the annealing chain are simulated. It is seen that the Q matrix governs the small perturbations in the system configurations which are then accepted or rejected probabilistically depending on the corresponding energy changes and the temperature. More precisely, the annealing chain may be simulated as follows. Suppose $\xi_k = i$. Then generate a Σ-valued random variable η with $P\{\eta = j\} = q_{ij}$. Suppose $\eta = j$. Then set

†A stationary chain is irreducible if its 1-step (stationary) transition matrix is irreducible.

$$\xi_{k+1} = \begin{cases} j & \text{if } U(j) \leq U(i) \\ j & \text{if } U(j) > U(i) \text{ with probability } \exp\left[-\dfrac{U(j) - U(i)}{T_k}\right] \\ i & \text{else} \end{cases}$$

There are two in depth numerical studies of simulated annealing of which we are aware. Johnson et. al. [18] applied the annealing algorithm to four well-studied problems in combinational optimization: graph partitioning, number partitioning, graph coloring, and the travelling salesman problem. They compare the annealing algorithm with the best of the traditional algorithms for each problem. They found that although annealing is able to produce quite good solutions on three of the four problems, only on one of the four (graph partitioning) does it outperform the best of its rivals. Golden and Skiscim [12] have tested the annealing algorithm on routing and location problems, specifically the travelling salesman problem and the p-median problem. They conclude that there are more efficient and effective heuristics for these problems.

We shall now outline the convergence results on the annealing algorithm which are known to us. We refer the reader to the specific papers for full details.

Geman and Geman [8] were the first to obtain a convergence result for the annealing algorithm. The consider a version of the annealing algorithm which they call the *Gibbs sampler*. They show that for temperature schedules of the form

$$T_k = \frac{c}{\log k} \qquad (k \text{ large})$$

that if c is sufficiently large then (2.6) is obtained.

Gidas [10] also considers the convergence of the annealing algorithm and similar algorithms based on Markov chain sampling methods related to the Metropolis method.

We next discuss the work of Mitra et. al. [26]. The idea behind their work is similar to that of Geman and Geman and also Gidas in that they show that for temperature schedules which vary slowly enough the annealing chain reaches "quasiequilibrium", i.e., something like (2.5) holds. In order to state Mitra et. al.'s result we will need the following notation. Let

$$N(i) = \{j \in \Sigma : \ q_{ij} > 0\} \qquad \forall\, i \in \Sigma \ .$$

Let S_M be the set of states that are local maxima of $U(\cdot)$, i.e.,

$$S_M = \{i \in \Sigma : \ U(i) \geq U(j) \quad \forall\, j \in N(i)\} \ .$$

Let

$$r = \min_{i \in \Sigma \setminus S_M} \max_{j \in \Sigma} d(i,j)$$

where $d(i,j)$ is the minimum number of steps to get from state i to state j. Finally, let

$$L = \max_{i \in \Sigma} \max_{j \in N(i)} |U(j) - U(i)| \ .$$

Here is Mitra et. al.'s result:

Theorem 2.1 (Mitra et. al. [26]) Assume Q is irreducible and symmetric†. Let $T_k \downarrow 0$ and

$$\sum_{k=1}^{\infty} \exp\left(-\frac{r\,L}{T_{kr-1}}\right) = \infty \ . \tag{2.7}$$

Then

$$\lim_{k \to \infty} P\{\xi_k = i\} = \pi_i^* \qquad \forall\, i \in \Sigma \ . \tag{2.8}$$

Remarks

(1) If $T_k = c/\log k$ then (2.7) holds iff $c \geq r\,L$.

(2) An estimate of the rate of convergence in (2.8) is obtained for annealing schedules of the form $T_k = c/\log k$ for $c \geq r\,L$. Let

$$w = \min_{i \in \Sigma} \min_{j \in N(i)} q_{ij} \ ,$$

$$\gamma = \min_{i \in \Sigma \setminus S} U(i) - \min_{j \in S} U(j) \ .$$

It is shown that

$$P\{\xi_k = i\} = \pi_i^* + O\left(\frac{1}{k^{\min\{\alpha,\beta\}}}\right) \qquad \text{as} \quad k \to \infty \tag{2.9}$$

where

†or just $q_{ij} > 0$ iff $q_{ji} > 0$ for all $i,j \in \Sigma$

$$\alpha = \frac{w^r}{r^{r\,L/c}} \quad , \quad \beta = \frac{\gamma}{c} \quad .$$

Since α and β are increasing and decreasing respectively with increasing c, it is suggested that $c \geq r\,L$ be chosen to maximize $\min\{\alpha,\beta\}$.

We next discuss the work of Hajek [14]. The idea behind his work is that for temperature schedules which vary slowly enough, the annealing chain escapes from local minima of $U(\cdot)$ at essentially the same rate as for a constant temperature. In order to state Hajek's result we will need the following notation. We shall say that given states i and j, i can *reach* j if there exists a sequence of states $i = i_o,\ldots,i_p = j$ such that $q_{i_n,i_{n+1}} \geq 0$ for all $n = 0,\ldots,p-1$; if $U(i_n) \leq E$ (a nonnegative number) for all $n = 0,\ldots,p$ then we shall say that i can *reach* j *at height* E. We shall say that the annealing chain is *strongly irreducible* if i can reach j for all $i,j\in\Sigma$. Clearly, strong irreducibility is equivalent to Q irreducible, but we introduce strong irreducibility to conform with Hajek's notation. We shall also say that the annealing chain is *weakly reversible* if for every $E > 0$, i can reach j at energy E iff j can reach i at energy E, for all $i,j\in\Sigma$. Let S_m be the states that are local minima of $U(\cdot)$, i.e.,

$$S_m = \{i\in\Sigma : \ U(i) \leq U(j) \qquad \forall\, j\in N(i)\} \quad .$$

For each $i\in S_m\backslash S$ let $\Delta(i)$ be the smallest number E such that i can reach some $j\in\Sigma$ with $U(j) < U(i)$ at height $U(i) + E$. $\Delta(i)$ is the "depth" of the local (but not global) minimum i. Let

$$\Delta^* = \max_{i\in S_m\backslash S} \Delta(i) \ . \tag{2.10}$$

Here is Hajek's result:

Theorem 2.2 (Hajek [14]) Assume that the annealing chain is strongly irreducible and weakly reversible. Let $T_k\downarrow 0$. Then

$$\lim_{k\to\infty} P\{\xi_k\in S\} = 1 \tag{2.11}$$

iff

$$\sum_{k=1}^{\infty} \exp\left(-\frac{\Delta^*}{T_k}\right) = \infty \ . \tag{2.12}$$