Facts, Conjectures, and Improvements for Simulated Annealing

Peter Salamon
Paolo Sibani
Richard Frost

siam
Monographs on Mathematical Modeling and Computation

Facts, Conjectures, and Improvements for Simulated Annealing

Peter Salamon

San Diego State University
San Diego, California

Paolo Sibani

University of Southern Denmark
Odense, Denmark

Richard Frost

San Diego State University
San Diego, California

siam
Society for Industrial and Applied Mathematics
Philadelphia

10 9 8 7 6 5 4 3 2 1

Library of Congress Cataloging-in-Publication Data

Salamon, Peter, 1950-
Facts, conjectures, and improvements for simulated annealing / Peter Salamon, Paolo Sibani, Richard Frost.
p. cm. — (SIAM monographs on mathematical modeling and computation)
ISBN 0-89871-508-3 (pbk.)
1. Simulated annealing (Mathematics) I. Sibani, Paolo, 1954- II. Frost, Richard, 1955- III. Title. IV. Series.

QA402 .S35 2002
519.3—dc21 2002029215

Preface

Simulated annealing is a simple and general algorithm for finding global minima. It operates by simulating the cooling of a (usually fictitious) physical system whose possible energies correspond to the values of the objective function being minimized. The analogy works because physical systems occupy only the states with the lowest energy as the temperature is lowered to absolute zero.

Simulated annealing has been developed by a wide and highly interdisciplinary community and used by an even wider one. As a consequence, its techniques and results are scattered through the literature and are not easily accessible for the computer scientist, physicist, or chemist who wants to become familiar with the field.

The present monograph is intended both as an introduction for the noninitiated and as a review for the more expert reader. We start by developing the subject from scratch. We then explain the methods and techniques indispensable for building state-of-the-art implementations. The physical background presented is meant to sharpen the reader's intuition for the field.

In our choices we have been somewhat biased toward the line of investigation developed in the last decade and a half of progress in one particular community: those who have worked on the theory of optimal thermodynamic processes and obtained most of their results by viewing simulated annealing as the cooling of a physical system. This point of view leads to some important improvements in algorithm design and these form the main topic discussed herein. While some proofs for these improvements are presented, most of the results require assumptions that may or may not be justified and should therefore be classified as conjectures. As such, they represent open problems in the area, and the tone of the discussions is in this spirit. Experimental data and heuristic arguments regarding various improvements are presented and discussed. Special care is taken to make the underlying assumptions explicit and to examine the likelihood of their validity.

The background assumed is minimal. The little physical background required is derived in Chapter 5; the reader with a background in physical science can probably skim this part. The mathematical background, mostly the theory of finite Markov chains, is developed in Chapter 6. The reader with an educational background in mathematics or computer science can probably skim section 6.1. The rest of Chapter 6 is required reading for all undergraduate backgrounds and brings together those optional topics from statistical physics and Markov chains that bear strongly on the reader's understanding of the effect that variants of the algorithms can have. Only readers with a graduate background in statistical physics can skim this part without compromising later understanding.

The book is divided into four parts, each with a distinct purpose and flavor. Minimal familiarity with a part can be gained by reading the brief synopsis at its beginning. Part I presents an overview of simulated annealing and how it fits into the family of algorithms for global optimization. This part is accessible to all readers and they should familiarize themselves with it before proceeding. Part II develops the theoretical aspects of complex systems that are indispensable for understanding simulated annealing and its refinements. We tried hard to keep this material to a minimum so as to make Part III of the text as accessible as possible to a wide audience. The further one has managed to work through Part II, the more the refinements in Part III will make sense. We have found it relatively easy to cover Parts I and II in their entirety with advanced undergraduate classes of mixed backgrounds (mathematics, computer science, physics, and engineering) in a one-semester course using selected topics from Part III as a supplement and source of projects. Part III is very different in nature from the previous parts. Theory is almost completely absent and each chapter focuses on an area where simulated annealing can be enhanced and describes the competing methods for improving performance. Each of the methods presented has been tested on some examples on which its creators claim improved performance. It is our belief that selecting the more promising algorithms will depend on the structure of the problem of interest, i.e., different enhancements improve performance on different problems. Some of the enhancements are outlined in sufficient detail to allow implementation; others require going back to the research articles cited for further details. In general, Part III requires significantly more sophistication on the part of the reader. Coding an enhancement based on our description requires thorough understanding. Part IV switches back to the style in Part II, albeit at a significantly more advanced level. It starts off roughly where Part II leaves off and continues our development of the physical understanding that one can get about global optimization problems. This is definitely graduate-level material and brings the reader to the research level in the field. While the theory developed here is not indispensable for utilizing the enhancements in Part III, it does shed further light on these methods and is, in our opinion, indispensable for the ultimate understanding that can bring this subject from the realm of heuristics to the realm of provably optimal algorithms.

The text grew out of a number of courses taught at the Niels Bohr Institute, the Technical University of Denmark, the University of Southern Denmark (formerly Odense University), and San Diego State University. A version of this course also formed the material presented in a tutorial for the SIAM annual meeting during the summer of 1994 in San Diego.

Additional material related to the contents of this book can be found at `http://www.frostconcepts.com/`. Specific questions regarding errata and updates can be addressed to `salamon@math.sdsu.edu`.

Acknowledgments

Thanks to . . .

Our families for enduring many days and nights of absence.

Our colleagues and students for their many inspiring questions.

The publishers for their incredible patience with repeated delays in the writing process.

Jacob Mørch Pedersen, from whose Ph.D. dissertation we have borrowed numerous results.

Lene Mørch Pedersen for consenting to our use of her paper silhouette in Figure 13.1.

Gale Bamber of the San Diego Supercomputer Center for her graphic interpretation of energy landscapes displayed on the cover.

Tove Nyberg of SDU-Odense Universitet for drawing most of the figures.

Statens Naturvidenskabelige Forskningsråd for financial support.

Facts, Conjectures, and Improvements for Simulated Annealing

SIAM Monographs on Mathematical Modeling and Computation

About the Series

In 1997, SIAM began a new series on mathematical modeling and computation. Books in the series develop a focused topic from its genesis to the current state of the art; these books

- present modern mathematical developments with direct applications in science and engineering;
- describe mathematical issues arising in modern applications;
- develop mathematical models of topical physical, chemical, or biological systems;
- present new and efficient computational tools and techniques that have direct applications in science and engineering; and
- illustrate the continuing, integrated roles of mathematical, scientific, and computational investigation.

Although sophisticated ideas are presented, the writing style is popular rather than formal. Texts are intended to be read by audiences with little more than a bachelor's degree in mathematics or engineering. Thus, they are suitable for use in graduate mathematics, science, and engineering courses.

By design, the material is multidisciplinary. As such, we hope to foster cooperation and collaboration between mathematicians, computer scientists, engineers, and scientists. This is a difficult task because different terminology is used for the same concept in different disciplines. Nevertheless, we believe we have been successful and hope that you enjoy the texts in the series.

Joseph E. Flaherty

Peter Salamon, Paolo Sibani, and Richard Frost, *Facts, Conjectures, and Improvements for Simulated Annealing*

Lyn C. Thomas, David B. Edelman, and Jonathan N. Crook, *Credit Scoring and Its Applications*

Frank Natterer and Frank Wübbeling, *Mathematical Methods in Image Reconstruction*

Per Christian Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*

Michael Griebel, Thomas Dornseifer, and Tilman Neunhoeffer, *Numerical Simulation in Fluid Dynamics: A Practical Introduction*

Khosrow Chadan, David Colton, Lassi Päivärinta, and William Rundell, *An Introduction to Inverse Scattering and Inverse Spectral Problems*

Charles K. Chui, *Wavelets: A Mathematical Tool for Signal Analysis*

Contents

List of Figures

Part I

Overview

Part I of this book presents a stand-alone course on the very elementary aspects of annealing. The really impatient reader can go directly to Chapter 4, which presents the bare-bones annealing algorithm. On the other hand, the experienced reader may choose to skip Chapter 4.

Chapter 1 puts the annealing algorithm in context and compares and contrasts the annealing approach to algorithms based on biological heuristics. This chapter is strongly recommended for all readers. Chapter 2 presents six examples of simulated annealing problems. These problems are used throughout the text for illustration. The impatient reader can come back to these as needed. Chapter 3 consists of an elaborate table to be used as a dictionary for translating different terms used in the nomenclature of global optimization and its physical and biological counterparts. A glance at this table is recommended and it can be consulted as needed.

Chapter 1

The Place of Simulated Annealing in the Arsenal of Global Optimization

This text presents one general approach to the problem of global optimization. The basic problem is to find the lowest possible value of a function[1] whose graph looks typically like a many-dimensional version of Fig. 1.1. Six examples of such problems are described in Chapter 2.

Calculus provides a powerful tool for characterizing and locating *local* minima. The problem of *global* minima has remained much more elusive. In the special case of a convex function,[2] the familiar conditions from calculus are necessary and sufficient conditions with which to establish the optimality of a point. Lacking convexity, one is left with little recourse except to enumerate all possibilities and examine each one (sometimes referred to as a "grand tour"). This is not a viable option for the type of problems we are interested in, which typically have far too many candidates for exhaustive enumeration.[3] Lacking a better strategy, enumeration until the point of exhaustion does give a viable strategy for finding better than average solutions even if the very best is unreachable. Computers have had a large impact here since they have extended the point of exhaustion by many orders of magnitude. In fact, in the 1950s the computer was heralded as a wonderful tool for global optimization because it made long random search runs possible [Con80]. A method often advocated was to evaluate the function at a few million points and select the best value found. A variant of this scheme is a good strategy in low dimensions or in problems with relatively few minima: use many random points as initial states and feed them to a local minimizer.[4]

Random search was not the last word in global optimization, but it did bring to the field the important tool of a random number generator. Methods that incorporate the use of random number generators are known as Monte Carlo methods. Typically such methods

[1] The problem of finding the largest possible value of a function f is equivalent to finding the lowest possible value of $-f$.

[2] A function is convex iff a line segment connecting two points on the graph of the function is always above the graph of the function [Lue84].

[3] Berry et al. [BBK+96], however, carried out such a program for chemical structure problems in clusters of moderate size (see Example B in Chapter 2).

[4] A more sophisticated variant uses simulated annealing combined with a local minimizer [DW96, MRX92]. This technique is called basin hopping and is discussed in Chapter 10.

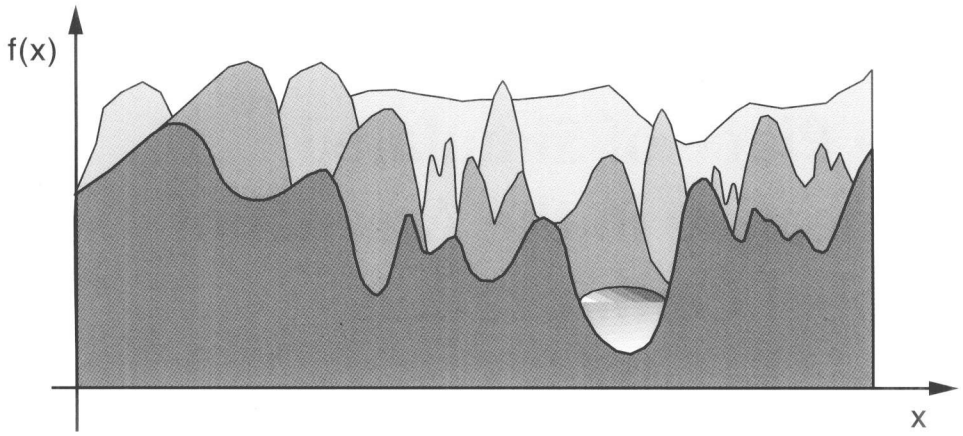

Figure 1.1. *A landscape with many local minima.*

are superior to deterministic algorithms for problems in high dimensions (for example, in numerical evaluation of integrals). Clever implementations of Monte Carlo techniques amount to refinements of random search known as importance sampling [BH97]. Such sampling techniques can greatly increase the rate of convergence of Monte Carlo algorithms by sampling preferentially according to some distribution. Choosing the best distribution for rapid convergence is often an art [BH97]. We will find that this idea of *importance sampling* is very useful for comparing simulated annealing with other Monte Carlo–based global optimization procedures. The differences between the techniques are nicely characterized by differences in their criteria of importance during the sampling.

During the last 15 years, Monte Carlo methods have taken over the field of global optimization in the form of heuristic approaches based on analogies [Hol75, Aarts89, Otten89, Gold89, HRS98, Koza99]. The algorithms set up a correspondence between the optimization problem and a physical or biological system. These algorithms simulate the behavior of the system to find states with good values of the objective function. While myriad variations on both themes exist, the main choices for the correspondence are either the slow cooling of a physical system to find low-energy states or the evolution of a biological population to find high-fitness states. The first is called simulated annealing and is the basis of physical heuristics. The second is called either genetic algorithms or evolutionary programming and is the basis of biological heuristics. In the context of importance sampling, the major difference between the two schemes can be understood as a degree of greediness or a degree of urgency.

An optimization algorithm is called *greedy* if it chooses the best alternative at each step without regard to long-term benefits. The standard example is the method of steepest descent [Lue84]. The relevance for optimization heuristics is as follows. All heuristics in the family under discussion search the state space by a sequence of random choices of states. They all sample preferentially "near" states with good values of the objective. This is in keeping with the dictates of importance sampling. Biological algorithms tend to add a bias in favor of those states with good values of the objective that were found quickly. That

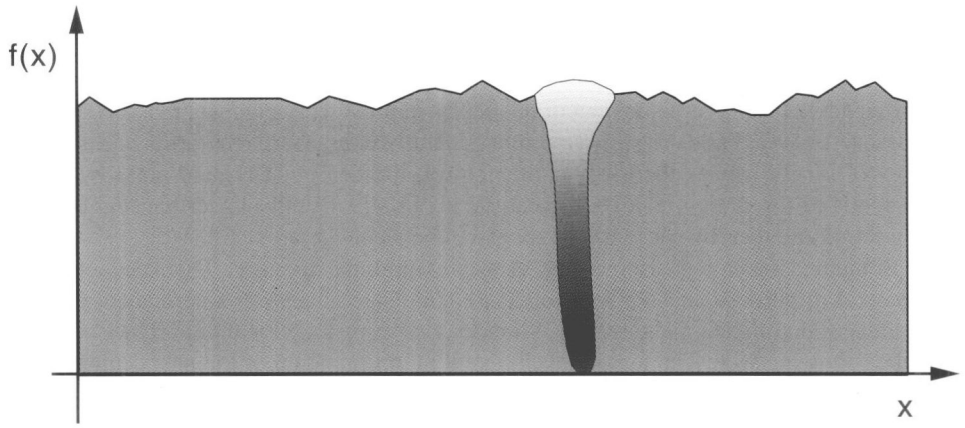

Figure 1.2. *A golf-hole landscape.*

is, they preferentially "breed" those members of the population (states) with high fitness. The physical heuristics on the other hand search to the same extent near all states at a given value of the objective. To readers with a background in biochemistry, this distinction may appear familiar as the one of thermodynamic versus kinetic control of a reaction. In the language of importance sampling, the issue is what criterion constitutes importance. The physical heuristic places importance only on the quality of the objective and not on how quickly it is reached. The biological heuristics place importance also on how quickly the high-fitness state is reached. In this sense, the biological algorithms are more greedy.

How the performances of various heuristic methods compare with each other as well as with more traditional local methods such as quasi-Newton or steepest descents depends on the structure of the problem. For example, if we are dealing with a convex function in n real variables, then calculus-based tools (quasi-Newton, conjugate gradient, etc.) are the best choices. At the opposite extreme, the values of the objective function are randomly assigned to states, with zero correlation between nearness of the points and values of the objective. In this case, sampling near states with good values of the objective does not help, and we might as well use random search. Similarly, no algorithm can beat random search on the so-called golf-hole problem, a variant of which is shown in Fig. 1.2. Such problems are characterized by the fact that there is no information regarding the global optimum at points outside a small neighborhood of this optimum. Between these extremes it is not known which structures lead to which purportedly optimal choices of algorithms beyond the usually exaggerated claims of the inventors. The results depend heavily on who runs the experiments and what examples they choose.[5] One of the themes of the present book is the importance of finding a classification scheme for hard optimization problems. In lieu of firm principles, we are left only with the guidance afforded by the old metatheorem valid for all applications of mathematics: The more we exploit the structure of the problem, the better.

One conclusion based on this metatheorem is that further study of the structure of a problem enables one to design better, more specific algorithms. An example of how

[5] See, however, the experiments described by Johnson et al. [JAMS89, JAMS91] which achieve a higher level of thoroughness.