

Optimization by simulated annealing

Posted by [dcoetzee](#) on March 19, 2009

Citation: S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi. Optimization by Simulated Annealing. Science, May 1983, Vol. 220, no. 4598, pp. 671-680. ([PDF](#)).

Abstract: There is a deep and useful connection between statistical mechanics (the behavior of systems with many degrees of freedom in thermal equilibrium at a finite temperature) and multivariate or combinatorial optimization (finding the minimum of a given function depending on many parameters). A detailed analogy with annealing in solids provides a framework for optimization of the properties of very large and complex systems. This connection to statistical mechanics exposes new information and provides an unfamiliar perspective on traditional optimization problems and methods.

Discussion: This 1983 paper introduced the heuristic optimization technique of [simulated annealing](#), inspired by physical simulation algorithms in [statistical mechanics](#), and applied it to problems of hardware design and the [traveling salesman problem](#).

[Annealing](#) is a centuries-old technique in metallurgy, once practiced by blacksmiths and now widely used in industry. It aims to regularize the crystal microstructure of the material, removing defects, which softens it and makes it easier to work. The [Gibbs free energy](#) of the material, a measure of the potential energy contained in its structure, is effectively minimized by this procedure.

To perform annealing, a metal is heated to its annealing temperature, slightly above its freezing point, and kept at that temperature for a while. Then it is allowed to slowly cool until it recrystallizes. The system can be seen as a massively parallel solver for minimizing the potential energy: at high temperature, molecules travel rapidly around the substance in search of a global minimum, but will fail to converge on a local minimum. As the substance cools, they can no longer travel as far and they begin to seek and converge on the nearest local energy minimum.

From a more algorithmic point of view, this can be modelled as a random walk on a search graph. Each node represents a solution or state and has a certain fitness value, and the nodes adjacent to each node represent other similar solutions (in the sense that their fitness value is not expected to be dramatically different). At each step the algorithm selects an adjacent node at random. It will transition to the new state if it is a better solution; it may also transition to the new state if it is a worse solution, with a probability determined by the global “temperature” parameter. This behavior allows the search to escape local minimums. As the temperature drops, this becomes less likely and the search drops into a nearby local minimum.

A simple example of this is the travelling salesman problem: given a set of cities and distances between each pair of cities, find a minimum-length tour that passes through each city once and returns to the starting point. Finding the minimum-length tour is an NP-complete problem and so not likely to admit an efficient solution any time soon, but simulated annealing can effectively find a solution close to the minimum. In this case a node in the search graph represents a particular tour of the cities, and the neighbors of a node are found by reversing two adjacent cities on the path. Solution quality is determined by total tour length.

Prior to simulated annealing, the most prominent heuristic methods were based on random-restart hill climbing, where we choose a number of random starting states and then make small changes to the solution as long as those small changes improve the fitness. From the physical perspective, this is the equivalent of taking, say, a set of 100 swords, heating them each to the annealing point, and then rapidly quenching them in water and picking the one that turns out softest. It's ineffective at locating global minimums in large problems with many parameters.

The specific parameters used in simulated annealing, for example determining the probability with which a transition to a higher-energy state should be made, were borrowed from statistical

mechanics and in particular the [Metropolis–Hastings algorithm](#) for simulation of microscopic systems.

Most of the applications demonstrated in the original simulated annealing paper were not simple classical ones like TSP but rather practical hardware design problems, largely because the authors had background in this area. This proceeds in stages: first the circuits are partitioned into groups small enough to fit on chips, then the locations of the chips are selected (placement), then routing of wires between the chips is done. The main parameter being minimized is the number of signals crossing between each group of circuits.

Unfortunately, simulated annealing is also a notoriously difficult technique to analyze and requires expertise to set up – its performance depends on how the probabilities are calculated, the rate at which the temperature is lowered, the allowed transitions between states, and so on, each of which generally must be empirically determined for a specific problem.

It's also more applicable to some types of problems than others: for example, it tends to do well for problems for which many near-optimal solutions exist and are a relatively small number of steps away from the starting random solution. Physical substances tend to have this property: it's extremely unlikely that they would enter a perfect lowest energy configuration, but with high probability annealing will bring them to a configuration close to the lowest possible energy.

Simulated annealing also tends to perform poorly for problems where the fitness function has many “deep wells,” very low local minimums that are difficult to escape from even at high temperature. Substances like glass exhibit this property in real life: even when cooled slowly, they will tend to fail to arrange themselves in a regular crystal structure, preferring the amorphous structure of glass.

Since the original work a number of improvements have been proposed to the method. [Adaptive simulated annealing](#) (Ingber, 1989) attempts to avoid the design problem of choosing parameters like the cooling schedule by automatically choosing these as the algorithm progresses, based on how the search is behaving. The methods of [quantum annealing](#) (Apolloni et al, 1989) and [stochastic tunneling](#) (Wenzel and Hamacher, 1999) seek to escape deep wells by occasionally “jumping” across nearby energy barriers. Other [stochastic optimization](#) techniques, such as genetic algorithms and ant colony optimization, generate new solutions in ways other than local modification.

An interesting question to ask as we enter the age of parallel computing is: if we reconsider the original motivating physical problem for simulated annealing, a group of particles independently seeking minima and responding to the global temperature, could a parallel method be deduced from this whereby a large number of small circuits or processors each simulate the behavior of a single component as the global temperature decreases? I haven't yet seen any work in this area.

The author releases all rights to all content herein and grants this work into the public domain, with the exception of works owned by others such as abstracts, quotations, and WordPress theme content.