SIMULATED ANNEALING

The random search procedure called simulated annealing is in some ways like Markov chain Monte Carlo but different since now we're searching for an absolute maximum or minimum, such as a maximum likelihood estimate or M-estimate respectively. Suppose we're searching for the minimum of f (or equivalently, the maximum of $-f$). One keeps in memory the smallest value of f found so far and its location, and perhaps some other values if they are not too much larger, and their locations (the "simanneal" R code only keeps the "best," i.e. smallest, value and its location). One continues a random search, making proposed moves if they decrease f, but only with some probability if they don't, as in Metropolis–Hastings. Also, the search returns from time to time to places where the smallest value (or few values) so far had been found.

A distinguishing feature of simulated annealing is the "temperature" T which decreases as one goes along. Let T_k be the temperature at the kth stage. Moves that increase f are accepted with larger probability when T_k is large than when it is small. One specific rule, implemented in "simanneal," is as follows. Suppose after k steps we are at a point $x = x_k$ and by the proposal distribution, a transition to y is proposed. If $f(y) \le f(x)$, the move is accepted and we set $x_{k+1} = y$. If $f(y) > f(x)$, the move is accepted only with probability

$$\exp\{(f(x) - f(y))/T_k\}, \tag{1}$$

specifically if $U_k \le \exp\{(f(x) - f(y))/T_k\}$, where U_k is a $U[0,1]$ variable independent of those used earlier. One can see that if T_k is large (relative to $f(y) - f(x)$), as it may be early in the process, the acceptance probability will be close to 1; whereas if T_k is small, as it will become late in the process, the acceptance probability will be low.

Specifically, at what rate should T_k decrease? In the code "simanneal" the process is done for just n steps, with $T_k = 1 - \frac{k}{n+1}$ for $k = 1, ..., n$. Bertsimas and Tsitsiklis (1993), equation (1), give the same acceptance probability as in (1) above. They consider discrete state spaces. In their equation (4), they give

$$T_k = d/(\log k) \tag{2}$$

Date: 18.465, April 17, 2015.

for a constant $d > 0$, which converges to 0 much more slowly as k increases. They mention a theorem of Hajek (1988), according to which the simulated annealing Markov chain finds the absolute minimum of f (assuming it exists and is unique) with probability converging to 1 as k becomes large for $d \geq d^*$ for a given $d^* > 0$ but not if $d < d^*$.

If in the continuous state space $\mathbf{R}^d$ one uses a normal proposal density $N(x, \sigma^2 I)$ as in Metropolis–Hastings, then one may want to adjust σ from time to time, increasing it if the acceptance rate for moves is too high, or decreasing it if the acceptance rate is too low. Results found previous to a change of σ would not need to be discarded. Rather, they would give us a smallest value of f found thus far and a location from which we could continue.

Annealing. The Wikipedia article "Annealing (metallurgy)" (accessed 2012?) says annealing is a heat treatment, for example of steel. One is trying to improve the properties of a body of steel in which there are imperfections in the crystal structure or "dislocations" which produce "stress" and increase the "Gibbs free energy." One wants to reduce and to the extent possible minimize this free energy. The energy is some function of the state the body of metal is in. The process involves heating the metal quite hot ("glowing"). The heat provides energy to break some of the chemical bonds involved in the imperfections and allow the metal to reach a state with less free energy. The metal is then cooled. In the case of steel in particular, the cooling needs to be slow. It is slow as opposed to abrupt in the "simanneal" code, but much slower still with the logarithmic cooling rate (2).

Notes. The acceptance probability in (1) appears on p. 93 (the third page) of Aarts et al. (1997), who say it is the "original simulated annealing version" published by Kirkpatrick, Gelatt and Vecchi (1983) and Černý (1995). Other sources I've seen agree that the Kirkpatrick et al. paper was the first on simulated annealing, with some mentioning the paper by Metropolis et al. on Markov chain Monte Carlo as a predecessor. In fact Kirkpatrick et al. p. 672, column 3, give a form of (1) in the text and attribute it to Metropolis et al.

Kirkpatrick, Gelatt, and Vecchi were all working for IBM research at the time of their article. Dmitris Bertsimas and John Tsitsiklis, at the time of their paper and now, have been on the MIT faculty, in Sloan and EECS respectively, also associated with LIDS (Laboratory for Information and Decision Sciences) and the OR (Operations Research) Center.

The Wikipedia article "Simulated annealing," accessed April 15, 2015, describes it in rather general terms. It says that Černý (1985) independently rediscovered the method. Kirkpatrick et al. had treated the traveling salesman problem also as an example in their paper.

REFERENCES

Aarts, E. H. L., Korst, J. H. M., and van Laarhoven, P. J. M., "Simulated annealing," Chapter 4 of *Local Search in Combinatorial Optimization*, Eds. E. Aarts and J. K. Lenstra, Wiley, 1997, pp. 91–120.

Bertsimas, D., and Tsitsiklis, J. (1993). Simulated annealing. *Statistical Science* **8**, 10–15.

*Černý, V. (1985). A thermodynamic approach to the traveling salesman problem: An efficient simulation. *J. Optimization Theory Appl.* **45**, 41–51.

*Hajek, B. (1988). Cooling schedules for optimal annealing. *Math. Oper. Res.* **13**, 311–329

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983), Optimization by simulated annealing, *Science* **220**, 671–680.

* — I saw these references mentioned in secondary sources, not in the original.