ON A RANDOMIZED VERSION OF PARALLEL LOCAL SEARCH

Matthias Löwe

Fakultät für Mathematik
Universität Bielefeld
Postfach 100131
33501 Bielefeld
Germany

Abstract. We introduce a stochastic version of the parallel local search algorithm. Using techniques similar to those used to establish convergence of the simulated annealing algorithm we are able to give a cooling schedule for which the process converges and to analyze the speed of convergence by spectral gap methods.

Key words and phrases. Markov processes, Spectral gap, Optimization, Simulated Annealing.

1. Introduction

Let $G = (X, E)$ be a finite connected graph with vertex set X and edge set E. Let $U : X \to \mathbb{R}$ be a real valued function. A problem which often appears in applications is to determine the minimum value of U or, equivalently, the vertex where it is attained. This problem becomes extremely difficult when the vertex set X is very large as e. g. in the most typical examples, in the Traveling Salesman Problem, (where X consists of all closed tours through a given set of cities) or in the problem of restoring a disturbed image (where X is the set of all possible images, so $|X| \sim 10^{10^4}$ in this case). In these cases it is plain that an application of the so called global search algorithm, which calculates explicitly all values $U(x)$, is hopeless. On the other hand global search is one of the few deterministic strategies that definitely find the global minimum.

In contrast the versions of the local search algorithm often get trapped in a local minimum which usually is not the global one. To be more precise there are essentially two forms of local search, a sequential one and a second one which is basically parallel in concept. The first one, roughly speaking, works in the following way: Start in a point x_0; when having arrived at a vertex x_t at time t, choose a vertex $y \in \mathcal{N}(x_t)$ (where $\mathcal{N}(x)$ denotes the neighborhood of a vertex x in (X, E), and if $U(y) \le U(x_t)$ set $x_{t+1} = y$ and stay in x_t otherwise. In contrast to that the second algorithm when having arrived at a vertex x_t at time t evaluates $U(y)$ for all $y \in \mathcal{N}(x_t)$ (which is the parallel component of the algorithm) and then steps to one of those vertices in $\{x_t\} \cup \mathcal{N}(x_t)$ that minimize $U(\cdot)$.

Now for the first type of local search (the sequential one) there has been introduced a stochastic version by Kirkpatrick et.al. [K83], to overcome the difficulty of convergence to suboptimal points. This stochastic version is the well known Simulated Annealing (SA) Algorithm. The idea behind SA is the following: modify the local search scheme described above in such a way that it induces an irreducible and reversible, time inhomogeneous random walk on X with an invariant measure that is Gibbsian with energy U and at some inverse temperature β_t. While running this random walk we slowly cool down the process (i.e. we increase $\beta_t \to \infty$) such that the invariant measure finally becomes concentrated on the set of minima of $U(\cdot)$.

More precisely SA modifies the original random walk created by the first version of local search in such a way that each transition probability from x to y is multiplied by a factor $\exp(-\beta_t(U(x) - U(y)^+)$, where $x^+ = \max(x, 0)$.

Basically there are two questions intrinsic in the description of this algorithm, namely how can the cooling schedule β_t be chosen to ensure convergence of the algorithm to the set of global minimum of $U(\cdot)$, and in case of convergence, how fast does the process become concentrated on the set of minima of U?

The first convergence results of SA have been established by Mitra et al. [M85] and Gidas [G85]. Later Hajak [Ha88] was able to solve the problem of an asymptotically optimal cooling schedule by proposing a functional dependence of β_t from t of the form $\beta_t = \frac{1}{m} log(A + B_t)$, where A and B are positive constants and m is approximately the largest hill to be climbed to reach the global minimum. The right exponential order of the speed of convergence has been investigated by Catoni [C90] by using Freidlin–Wentzell techniques [FW84].

A new light was shed on the mechanism behind the convergence of SA by the work of Holley and Stroock [HS88] who investigated the speed of convergence by establishing and solving a differential inequality for the Radon–Nikodym density of the distribution of the process at time t with respect to the Gibbs measure at time t with the help of Sobolev inequalities. A refined analysis of these differential inequalities by Götze [G92] led to applicable bounds for the entrance time into typical subsets of the state space X.