Asymptotic Properties of a Probabilistic Tabu Search Algorithm *

Yuri Kochetov, Novosibirsk

Summary: We present a new version of the Tabu Search algorithm with randomized neighborhood. It is shown that the algorithm generates a finite aperiodical Markov chain on a suitable set of outcomes. The Markov chain is irreducible if some conditions on the structure of the tabu list and randomized neighborhood hold. This property allows us to find an arbitrary feasible solution as well as an optimal solution. Computational results for hard benchmarks of the Uncapacitated Facility Location problem are discussed.

1 Introduction

The purpose of this paper is to present new theoretical and experimental results concerning Tabu Search algorithm (TS). We design a new version of the Probabilistic Tabu Search (PTS) and show that the algorithm generates a Markov chain on a suitable set of outcomes. This set is the product of the feasible domain and a set of tabu lists. The Markov chain is irreducible if some conditions on the structure of the tabu list and randomized neighborhood hold. This property allows us to find an arbitrary feasible solution and an optimal solution as well. We present experimental results for hard tests of the Uncapacitated Facility Location problem and compare deterministic and probabilistic versions of TS. We conclude that using randomized neighborhood reduces running time and increases the probability of finding an optimal solution.

2 Probabilistic Tabu Search

Consider the minimization problem of an objective function $f(x)$ on the boolean cube B^n:

$$\min\{f(x),\ x \in B^n\}.$$

In order to find the global optimum $f_{opt} = f(x_{opt})$ we apply PTS algorithm with the following features. Denote by $N(x)$ a neighborhood of the point x and assume that it contains all neighboring points $y \in B^n$ with Hamming distance $d(x, y) \leq 2$. A randomized neighborhood $N_p(x)$ with probabilistic threshold p, $0 < p < 1$, is a subset of $N(x)$. Each $y \in N(x)$ is included in $N_p(x)$ randomly with probability p and independently from other points.

Let us consider a finite sequence $\{x_t\}$, $1 \leq t \leq k$ with property $x_{t+1} \in N(x_t)$. An ordered set $\varphi = \{(i_k, j_k), (i_{k-1}, j_{k-1}), \ldots, (i_{k-l+1}, j_{k-l+1})\}$ is called a tabu list if vectors x_t and x_{t+1} differ by coordinates (i_t, j_t). The constant l is called the length

*This work was supported by the RFBR grant 99-01-00510.

of the tabu list. Note that i_t and j_t may be equal. By definition, we put $i_t = j_t = 0$ if $x_{t+1} = x_t$. Finally, let $N_p(x_t, \varphi)$ be a set of points $y \in N_p(x_t)$ which are not forbidden by the tabu list φ. Obviously, $N_p(x_t, \varphi)$ may be empty for nonempty set $N_p(x_t)$. Now algorithm PTS can be written as follows:

Algorithm PTS

1. Initialize $x_0 \in B^n$, $f^* := f(x_0)$, $\varphi := \emptyset$, $t := 0$.
2. While a stopping condition is not fulfilled do
   2.1. Generate neighborhood $N_p(x_t, \varphi)$.
   2.2. If $N_p(x_t, \varphi) = \emptyset$ then $x_{t+1} := x_t$,
   else find x_{t+1} such that $f(x_{t+1}) = \min\{f(y), y \in N_p(x_t, \varphi)\}$.
   2.3. If $f(x_{t+1}) < f^*$ then $f^* := f(x_{t+1})$.
   2.4. Update the tabu list φ and the counter $t := t + 1$.

The algorithm does not use any aspiration criteria, intensification or diversification rules [3]. This simple version of the Tabu Search method allows us to see a connection with Markov chains and find asymptotic properties.

3 Markov Chains

Suppose $l = 0$. In this case, the selection of the x_{t+1} depends on the current point x_t and does not depend on the previous points $x_s, s < t$. We may say that the sequence $\{x_t\}$ is a random walk on the boolean cube B^n or a Markov chain on the finite set of outcomes B^n. A Markov chain is called *irreducible* if for each pair of outcomes x, y there is a positive probability of reaching y from x in a finite number of steps [1]. The sequence $\{x_t\}$ is irreducible and we obtain $f^* = \min_{t \leq k} f(x_t) = f_{opt}$ for large k. Now we consider the case $l > 0$. Let Ω denotes the set of all pairs (x, φ), where $x \in B^n$, and φ is the tabu list. The sequence $\{(x_t, \varphi_t)\}$ is a finite Markov chain on Ω.

Theorem 1 Algorithm PTS generates an irreducible Markov chain.

It seems that the statement is too strong without any restrictions for the length of tabu list. If $l > |N(x)|$, then all points may be forbidden and $N_p(x, \varphi) = \emptyset$. For this case we get $x_{t+1} = x_t$ on the step 2.2. and $(i_t, j_t) = (0, 0)$. This means that the real length of tabu list is decreasing. The algorithm regulates the tabu list by itself. This element of self learning allows us to prove the theorem and get asymptotic properties of the algorithm.

Denote a randomized neighborhood of x by $N_r(x, \varphi)$ which contains exactly $r > 1$ unforbidden points from $N(x)$. The algorithm PTS with $N_r(x, \varphi)$ neighborhood generates a Markov chain as well. Unfortunately, we can not prove the irreducibility for this case and the analogous theorem may be false. In particular, for $r = |N(x)| - l$

we obtain the Deterministic Tabu Search algorithm (DTS). If the tabu list is too small, algorithm DTS finds a local optimum and has no opportunity to escape from it. Therefore we need a large tabu list. But it is easy to present a minimization problem for which algorithm DTS can not find the optimal solution for arbitrary large tabu lists.

4 Asymptotic Properties

The following properties of PTS are derived from Theorem 1.

Corollary 2 For an arbitrary initial point $x_0 \in B^n$ we have
1. $\lim_{t\to\infty} Pr\{f^* = f_{opt})\} = 1$;
2. there exist $b > 0$ and $1 > c > 0$ such that $Pr\{\min_{\tau\le t} f(x_\tau) \ne f_{opt}\} \le bc^t$;
3. The Markov chain $\{x_t, \varphi_t\}$ has a unique stationary distribution $\pi > 0$.

The first property guarantees that the algorithm will find an optimal solution with probability 1. We can not say that PTS converges to a global optimum as Simulated Annealing [1]. But for practical purposes this is unimportant. We can always select the best solution f^*. A variant of PTS that converges can be found in [2].

The second property guarantees a geometrical rate of convergence of the best value f^* to f_{opt}. To estimate the real convergence rate, we consider the Uncapacitated Facility Location problem [5]. It is strongly NP-hard problem with the following objective function:

$$f(x) = \sum_{i=1}^{n} c_i x_i + \sum_{j=1}^{m} \min_i\{g_{ij} | x_i = 1\},$$

where c_i, g_{ij} are real numbers. Experimental results are obtained for the instances from benchmarks library: http://math.nsc.ru/LBRT/k5/Kochetov/bench.html, case of hard matrices, $n = m = 100$. For computational convenience, we redefine the neighborhood $N(x) = N_1(x) \cup N_{sw}(x)$, where $N_1(x) = \{y \in B^n | d(x,y) \le 1\}$, and $N_{sw}(x) = \{y \in B^n | \exists^1 (i,j) : y_i = 1 - x_i, y_j = 1 - x_j, x_i \ne x_j\}$. All statements remain true for the new neighborhood as well.

In Figure 1 we present experimental results for algorithm PTS. We get 30 instances and apply the algorithm 30 times for each of them. The value $Fr(f^* \ne f_{opt})$ is the frequency that the best found solution is nonoptimal. The upper and lower curves in Figure 1 correspond to the threshold values $p = 0.1$ and $p = 0.5$ respectively. These curves are similar. It seems, one of them can be received from other one by a parallel shift. One should expect the curve to decrease when the threshold p increases. Surprisingly, this plausible reasoning is false. The curve in the middle corresponds to the maximal value of $p = 1$. This curve (DTS) is less sloped and dominates the others for large t.

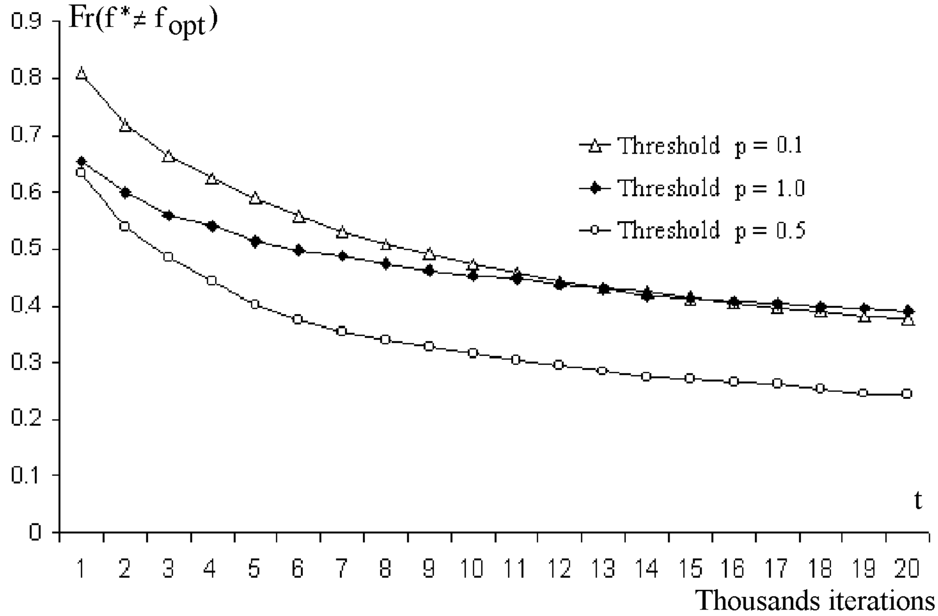

Figure 1: The convergence rate of algorithm PTS for $l = 50$

The third property guarantees that PTS generates an ergodic Markov chain with nonzero limiting distribution $\pi(x, \varphi) > 0$. So we can find the global optimum from an arbitrary initial point. It is interesting to know how the probability $Pr\{f^* \neq f_{opt})$ depends on the choice of the initial point. For this purpose we take as x_0 the most remote point $1-x_{opt}$ and the solution x_{gr} obtained by a randomized greedy algorithm [4]. For our benchmarks the average value $d(x_{opt}, x_{gr}) \approx 16$ and $d(x_{opt}, 1 - x_{opt}) = 100$. As a rule the optimal solution is unique and $d(x_{opt}, 0) \approx 12$. In Figure 2 the results of such comparisons are shown. For the first thousand iterations we can notice a slight preference of the greedy solutions. As the number of iterations increases, this preference disappears.

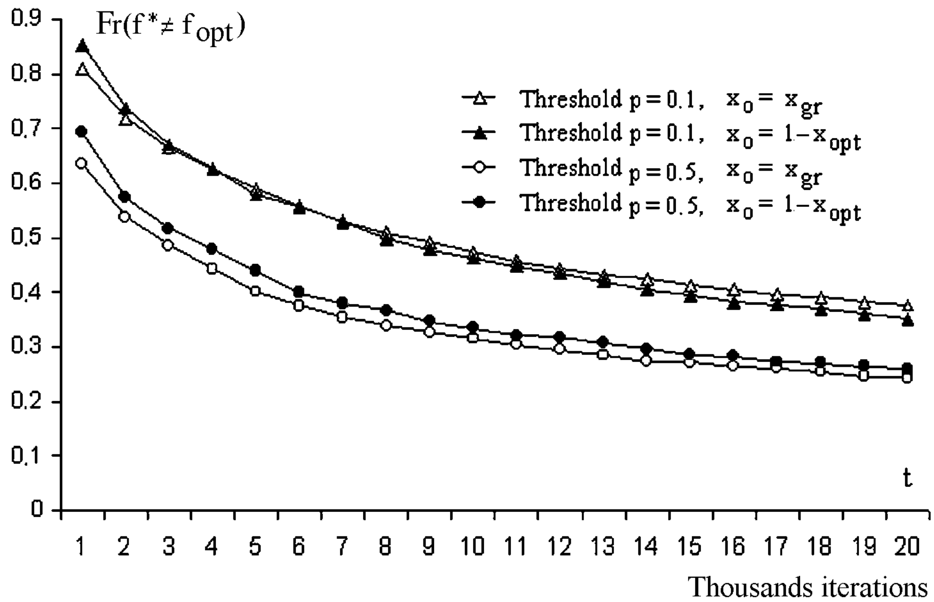

Figure 2: Performance of algorithm PTS for the different initial points

5 Stopping Rules

There are many stopping rules for Markov chains [6]. Two of them are the most popular in the field of the local search methods: 1) stopping after a prescribed number of steps; 2) stopping if the best solution f^* so far does not change during a prescribed number of steps. These rules are simple and convenient but can not help us to understand the current state of the search process. Now we present a new stopping rule exploits the history of the search process more carefully.

Denote by $H(x,y)$ the expected number of steps to reach y from x. Suppose that at the t-th step we are at the point x_t. By Theorem 1 we will return to x_t.

Corollary 3 For each $x \in B^n$ we have $\pi(x) = \sum_{\varphi} \pi(x,\varphi) = 1/H(x,x)$.

The value of $\pi(x)$ equals the probability to be in the point x on the step t for large t. Obviously, $\pi(x)$ depends on the parameters p and l of algorithm PTS. Figure 3 shows average values of $1/\pi(x_{opt}) = H(x_{opt}, x_{opt})$ as a function of p. We consider three cases of tabu list: $l = 10, 50, 100$. For large values of the threshold $p \geq 0.9$, the least value of $H(x_{opt}, x_{opt})$ is achieved on the large tabu list, $l = 100$. For small p, best results are obtained with small tabu lists $l = 10, 50$. The pair $p = 0.6, l = 50$ seems to be best values of the parameters, $H(x_{opt}, x_{opt}) \approx 500$, $\pi(x_{opt}) \approx 0.002$. For comparison, the $H(x_{gr}, x_{gr})$ is greater than 10^6 for the same values of p and l.

Corollary 3 provides the following stopping rule. We run algorithm PTS and analyze the history of the search process. If PTS returns to x^* *very often* then the algorithm is stopped and restarted with a new initial point. As we see, the intensification strategy is used automatically by the stopping rule.

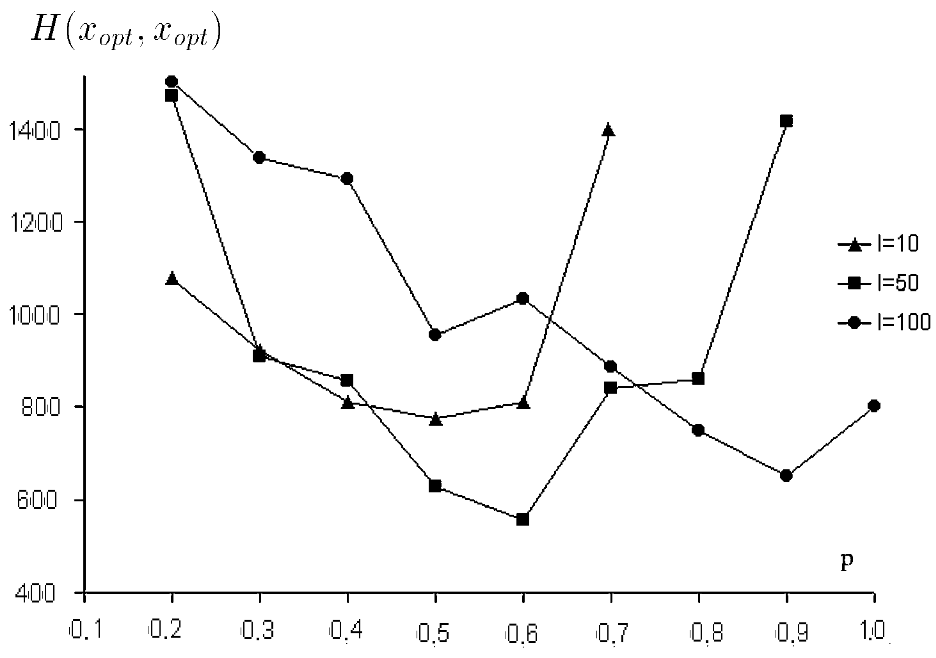

Figure 3: $H(x_{opt}, x_{opt})$ as a function of the threshold p

6 Further Research

Simulated Annealing (SA), Probabilistic Tabu Search and Genetic Algorithms (GA) are the most powerful heuristics for NP-hard optimization problems. They are closely related to finite Markov chains and find the global optimum with probability 1. But Uniform Random Walk obtains these properties as well. The meta-heuristics have something more. We may call it purposefulness. GA achieves it with the help of the selection device. PTS attains it selecting the best solution in the randomized neighborhood. SA obtains this property by the acceptance probability. It would be interesting to design a Markov chain with the most advanced guide-line strategy.

References

[1] Aarts, E. H. L.; Korst, J. H. M.; Laarhoven, P. J. M. (1997): Simulated Annealing. Aarts, E. H. L.; Lenstra, J. K. (EDS.) *Local Search in Combinatorial Optimization,* Wiley, Chichester, 91-120.

[2] Faigle, U.; Kern, W. (1992): Some Convergence Results for Probabilistic Tabu Search, *ORSA Journal on Computing 4:1*, 32-37.

[3] Glover, F.; Laguna, M. (1997): Tabu Search, Dortrecht: Kluwer Academic Publishers.

[4] Goncharov, E.; Kochetov, Yu. (1999): Behavior of Randomized Greedy Algorithms for the Multi-Stage Uncapacitated Facility Location Problem, *Discrete Analysis and Operations Research 6:1*, 12-32 (in Russian).

[5] Krarup, J.; Pruzan, P. (1983): The Simple Plant Location Problem: Survey and Synthesis, *European Journal of Operational Research 12:1*, 36-81.

[6] Lovász, L.; Winkler, P. (1995) Efficient Stopping Rules for Markov Chains, *Proc. 27th ACM Symp.Theory of Comp.* 76-82.