Simulated annealing for graph coloring

Castellon, Joel and Liu, Chen.

May 21, 2016

1 Introduction

Exact optimization of integer valued functions can be challenging, this is because there is no simple criterion to verify optimality as in continuous optimization. The problem we study in this project is graph coloring. In this context the space of possible solutions increases exponentially with the input graph size. Then, we need a principled approach to be able to optimize in this huge search space while providing a measure of sub-optimality.

2 Problem statement

Let $G = (V, E)$ be a graph with $|V| = N$ and a set of colors $\{1, \ldots, q\}$. We define a state (a coloring) as $x = (x_v, v \in V) \in \{1, \ldots, q\}^V$. Then, we get a *proper coloring* when $x_v \neq x_w \forall (v, w) \in E$. Note, however, that such proper coloring might not even exist. Therefore, if we cast graph coloring as an optimization problem we need a target function to account for the sub-optimality of our solution.
Hence, our function will be the Hamiltonian energy function:

$$H(x) = \sum_{(v,w)\in E} 1_{x_v = x_w}$$

.
The Hamiltonian simply counts the number of spurious edges e.g. for which both endpoints have the same color. Now, the theoretical support for simulated annealing comes from the theory of Markov Chains. It is, as we will see, an instance of the Metropolis-Hastings algorithm [1].
Then, to minimize the Hamiltonian in the context of Markov Chains we need a target distribution in the stationary regime.For this purpose, we will use the *Gibbs-Boltzmann* probability distribution.

$$p_\beta(x) = \frac{e^{-\beta H(x)}}{Z_\beta}$$

Where $\beta = 1/T$ is the *inverse temperature* and $Z_\beta = \sum_{x\in\{1,\ldots,q\}^V} e^{-\beta H(x)}$ is the normalizing factor.
The main idea is that as β increases, the distribution $p_\beta(x)$ concentrates around the minima of $H(x)$. Namely, when $\beta \to \infty$ we sample from the distribution of optima.

$$p_\infty(x) = \frac{1_{x \ is \ a \ global \ minimum \ of \ H(x)}}{Z_\infty \ nbr. \ of \ global \ minima \ of \ H(x)}$$

To make it explicit, only colorings at which $H(x) = 0$ are non-zero when e has $-\infty$ as an exponent.
Then, we build the chain with the Metropolis-Hastings method.

$$p_{xy} = \begin{cases} \psi_{xy} a_{xy} \ when \ y \neq x \\ 1 - \sum_{z\neq x} \psi_{xz} a_{xz} \ when \ y = x \end{cases}$$

Here, ψ_{xy} are the transition probabilities in the *base chain*. Namely, we define ψ_{xy} as follows: given x we pick a vertex uniformly at random. We then switch the color c_x of such vertex to an uniformly chosen color in $\{1,\ldots,q\}-c_x$ to get y. All colorings y that can be obtained in the fore-mentioned rule define $N(x)$ The acceptance probabilities are defined as (π being the distribution in the stationary regime).

$$a_{ij} = min(1, \frac{\pi_y}{\pi_x} = \begin{cases} min(1, e^{-\beta(H(y)-H(x))}) \ when \ y \in N(X) \\ 0 \ otherwise \end{cases}$$

Finally, we emphasize the main advantage of this formulation is that we do not have to calculate the normalizing factor Z_β. In this application (graph coloring) represents a sum over an exponentially large set in the number of vertices ($q^{|}V|$). Moreover, without the Metropolis-Hastings method, Z_β would have to be computed every time we update β giving the algorithm super-exponential running time.

3 Our strategy

Simulated annealing is very sensitive to algorithm parameters such as initial guess of temperature (e.g. where to start decreasing from) and the cooling schedule (e.g. period, in number of iterations, for the update of T). We follow the intuitive principle that having high temperature (T) encourages exploration, and can be useful in the early stages of the algorithm so we don't converge to local minima. Evidently, as we decrease the temperature to zero we should converge to the distribution of global minima.
With the previous concept in mind, we devise a simple criterion to obtain a better initial estimate of the temperature than random guessing. For this purpose, we count the number of "descent steps",namely, when $dE \leq 0$ where $dE = H(y) - H(x)$ when we make the transition from state x to y. Hence, our method starts with an arbitrary initial T for one cooling period. Then, we increase the temperature until we hit a *descent threshold* (descent steps/nbr.of iterations, a typical value is 0.9). Finally, with the temperature found on the previous step we decrease T until we hit 0 (approximately).

4 Experiments

For the evaluation of our strategy and implementation we used the $G(n,p)$ random graph model [2]. The main parameters of our experiments are: the number of allowed colors q, the average degree in $G(n,p)$ noted as c and the initial guess in temperature T. The following results are representative of many re-runs.

4.1 Energy trace

First, we plot the trace of energy values until convergence in Fig. 1. We observe an initial spike and jittering in the regime when we increase T, then the energy curve decreases smoothly as we decrease T. We see that for a $G(n,p)$ instance of 1000 vertices the algorithm performs quite well finding proper 30-colorings for $c = 10, 25$ (coloring is guaranteed when $q > max_degree$). However, for $c = 50$ even if a proper coloring might not exist we still get good approximations consistently. In Fig. 1 we get $E_{min} = 5$ for $c = 50$ which means we only have 5 spurious edges over 25000 present edges on average.

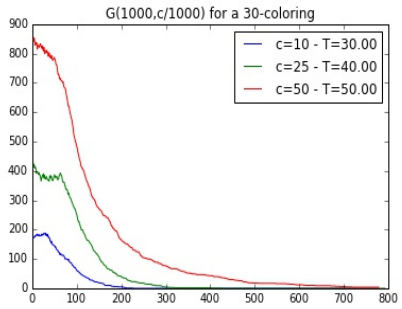

Figure 1: H(x) as a function of time. We vary the mean degree to test difficulty of the problem.

4.2 Coloring by mean degree

Finally, we plot the final Energy found (min Energy) as a function of mean degree instances for $G(n, p)$ (n=100 fixed) in Fig.2. The following curves are by the number of colors allowed. The trend of Fig. 2 was expected. First, the curves are increasing with c because larger degrees increase the probability of neighbors having the same color, thus require higher q. On the other hand, the curve for the minimal q ($q = 3$) is strictly above all others, this is because satisfying a lower q is always harder.

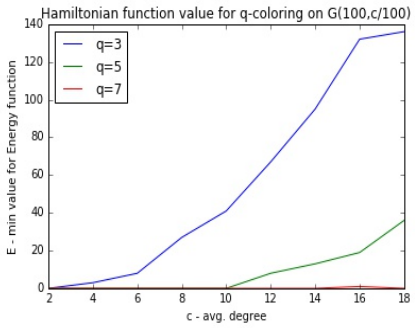

Figure 2: H(x) value found at convergence (Min. Energy) as a function of the mean degree.

5 Conclusion

We used simulated annealing together with the hamiltonian energy function. At a high level, simulated annealing incorporates randomness in local decisions to encourage exploration and therefore avoiding convergence to local optima. This is a principled approach as it is supported by the Markov-Chain theory. Nonetheless, tuning of the algorithm's parameters proves to be critical to the quality of the obtained solution, and it is often specific to the problem instance. Then, there is no silver bullet approach for the selection of such parameters. However, by incorporating a simple criterion for the diagnosis of performance (descent threshold) we facilitate greatly the tuning task.

References

[1] Siddhartha Chib and Edward Greenberg. *Understanding the Metropolis Hastings Algorithm.* 1995. American Statistical Journal.

[2] Paul Erdos and Alfred Renyi. *On the evolution of Random Graphs.* 1960. Publication of the Mathematical Institute of The Hungarian Academy of Sciences.