

# Simulated Annealing, Random Restarts

For homework 2, on the implementation of random restart hill climbing and simulated annealing to solve the SAT problem, there is a persistent question I find very surprising. Specifically, people seem to think that simulated annealing should always perform better than random restart hill climbing, and that if it is not they are perhaps doing something wrong.

While you should worry about aspects of your implementation, and you should make an effort to have your methods be as good as they can be, please don't assign too much emphasis on making sure that simulated annealing outperforms random restart hill climbing. This is not what the question is about.

Know that there is no guarantee that simulated annealing is unambiguously superior to random restarts. It is **very** possible for random restart hillclimbing to outperform simulated annealing with reasonable design decisions. (It is **also** possible for simulated annealing to outperform random restarts with bad design decisions.) We're far more interested in you demonstrating your sanity with respect to your design decisions for both implementations than we are on the final numbers – I expect, come Thursday, to award full credit to implementations where random restart hill climbing appears superior "by the numbers." Though we are interested in the performance of your method, it will likely not be a grade determiner, except if it alerts us that something is wrong or not.

To answer the larger question though, should simulated annealing outperform random restart hill climbing? Often, yes, but not always.

I had a **very** simple temperature schedule (a constant) in my reference implementation, and, **with the right temperature**, managed to get solutions reasonably quickly which were, on average, better than random restarts. You'll note my use of the phrase "on average." The process is very high variance. By "high variance" I mean that in one individual trial, it was very unpredictable which of the two would be better: across 100 trials, simulated annealing outperformed random restarts only ~70% of the time. (Even a blind dog finds a bone once in a while, or whatever the saying is.) So, unless you run a similar test, don't be sure one of your methods does better than the other.

On the other hand, who knows? Maybe you're doing something "pretty smart" for your restart procedure so that, on the whole, you **are** better off with random restarts? It wouldn't be hard to come up with something better than what I did.

Obviously, all of this is not to say that inferior performance of methods could not also be the result of poor decisions or mistakes in your implementation.