A probabilistic analysis for the Feedback Vertex Set problem

Karthekeyan Chandrasekaran
Mentor: Santosh Vempala

1 Introduction

Randomization is a powerful tool that has provided simple solutions to several algorithmic questions giving rise to the broad area of randomized algorithms. Through my research, I intend to investigate a classic NP-hard problem to come up with algorithms to solve it efficiently when the inputs are arising from a distribution.

A randomized algorithm uses a source of randomness to make decisions during its execution. The output is "good" on all inputs for most settings of the randomness. Consequently, the expected number of computations done by such algorithms is often less than the number of computations done by their deterministic counterparts.

Probabilistic analysis is essentially the science of considering inputs as arising according to a probability distribution. A strong motivation for probabilistic analysis of algorithms comes from NP-hard problems. A typical issue arising with NP-hard problems is that, while most instances of the problem are easy to solve by an algorithm, a few input instances remain hard. However, if input instances come from certain probability distributions, it can often be proven that the hard input instances occur with very low probability. Consequently, one can show that the algorithm solves most of the instances with good probability. This illustrates the benefit of randomizing the input instances.

2 Feedback Vertex Set Problem (FVS)

The *Feedback Vertex Set* (FVS) problem is a special case of the Hitting Set problem. Formally, the optimization version of the Hitting Set problem presents a set T with elements as subsets of a universal set X; the objective is to find a subset $S \subseteq X$ of elements of minimum cardinality so that every subset in T contains at least one element from S. One can find an easy factor d-approximation algorithm for the hitting set problem if the size of each subset in T is restricted to be at most d. The number of subsets to consider in T by an algorithm is potentially as high as $\Omega(2^{|X|})$. I would like to find an algorithm that can judiciously pick a polynomial-sized choice of the subsets T' to hit so that, once a hitting set S' is picked for T', with high probability S' is also the hitting set for T. While this is unlikely to work for all inputs (due to the NP-hardness of the problem), I would like to demonstrate broad classes of inputs for which it does.

An important special case of the hitting set problem where the number of subsets to hit is exponential in the cardinality of the universal set X, is the *Minimum Feedback Vertex Set* (MFVS) problem. The MFVS problem is the following: given a graph $G(V, E)$, find a subset S of vertices of smallest cardinality so that every cycle in the graph contains at least one vertex from S. If the graph has n vertices, then the number of possible cycles in the graph could be exponential in n. On the other hand, given any subset of vertices, one can quickly determine whether that subset is a FVS or find a cycle that is not hit by the subset. I would like to come up with a randomized algorithm that considers only a polynomial number of cycles in order to find a FVS whose cardinality is as close to that of the MFVS as possible. I will consider both the directed and undirected versions of this problem.

The motivation for the existence of such an algorithm comes from random graphs $G(n, p)$. A graph from the family $G(n, p)$ consists of n vertices in which every edge is present with probability

p. It is clear that the probability of a large cycle containing a specific vertex is much lesser than the probability of smaller cycles containing the same vertex. Hence, intuitively, one would expect a feedback vertex set that hits all cycles of small size, to hit all (or almost all) cycles of larger sizes as well with good probability. I would like to prove this for random graphs. Following this, I plan to identify the choice of picking cycles and vertices from the cycles to find a FVS of size as small as possible.

A relaxed version of the Hitting Set problem is the planted hitting set problem. Here, a subset S' of k elements from X are selected. For every subset of T, an element chosen at random from S' is added to it. The objective is to recover a hitting set S of size at most k. In the graph-theoretical notion of feedback vertex set, one considers a random graph $G(V,E)$ from the family $G(n,p)$ and picks a subset of vertices S' of cardinality k. Now, we are presented with a subgraph $G'(V,E')$ of G on the same vertex set V such that, for every connected component in $G \setminus S'$, there exists exactly one spanning tree over the vertices present in that component in $G' \setminus S'$. The objective is to find a FVS of cardinality at most k in G'. An algorithm proposed along the above lines for the planted FVS problem could also act as a tool to understand the more general idea of identifying planted hitting sets by considering only a polynomially many number of subsets.

The decision version of the feedback vertex set is NP-complete on both directed graphs as well as undirected graphs [6]. For undirected graphs, the MFVS can be approximated upto a factor of 2 [1, 2]. It has also been proved that the approximation ratio cannot be improved unless the approximation factor of Vertex Cover can be improved beyond 2 [1]. It remains open if one can approximate the MFVS in directed graphs upto a constant factor. For arbitrary directed graphs Seymour [11] gives an algorithm which approximates the MFVS to within $O(\log k \log\log k)$ factor, where k is the size of the MFVS. The MFVS problem is known to be solvable in polynomial time for various other families of graphs including cubic graphs [7, 12], permutation graphs [8] and interval and comparability graphs [9].

3 My Background

A simple example of a randomized algorithm is in the area of Markov Chain Monte Carlo methods to generate samples from desired distributions. Given a nonnegative, integrable function, the objective is to generate points distributed according to this function. A standard technique here is to construct a Markov Chain whose stationary distribution is the desired distribution. In two independent results, I showed that the ball walk with the Metropolis filter can be used to sample points beyond the setting of convex sets and logconcave functions. In [4], we prove that the class of s-concave functions for $s \geq -1/(n-1)$ is efficiently sampleable, and this represents the limit of convexity-based isoperimetry. In [3], we presented the first well-known class of nonconvex bodies for which uniform random samples can be generated efficiently, namely star-shaped sets. This result required introducing a new tool for proving isoperimetric inequalities for nonconvex sets. Both results involve an intricate understanding of the local properties of the ball walk and the dependency between the distributions obtained after taking one ball-walk step from two different points.

Another project I have worked on is Efficient Deterministic Algorithms for the Lovász Local Lemma. In joint work with Navin Goyal and Bernhard Haeupler [5], I recently found an efficient derandomization of the Moser and Tardos algorithm [10] to find a point in the probability space which makes none of the events happen, given the conditions of the Lovász Local Lemma. Our result is easily stated for the k-CNF problem: for a k-CNF formula with m clauses such that every clause shares variables with at most $2^{k/(1+\epsilon)}/e$ other clauses for some $\epsilon \in (0,1)$, we give an algorithm that finds a satisfying assignment in time $\tilde{O}(m^{2(1+1/\epsilon)})$. This improves upon the deterministic algorithms of Moser and of Moser-Tardos with running time $m^{\Omega(k^2)}$ which is superpolynomial for $k = \omega(1)$, and upon other previous algorithms which work only for $d \leq 2^{k/16}/e$.

References

[1] V. Bafna, P. Berman, and T. Fujito. A 2-approximation algorithm for the undirected feedback vertex set problem. *SIAM J. Discret. Math.*, 12(3):289–297, 1999.

[2] R. Bar-Yehuda and D. Geiger. Approximation algorithms for the feedback vertex set problem with applications to constraint satisfaction and bayesian inference. *SIAM J. Comput.*, 27(4):942–959, 1998.

[3] K. Chandrasekaran, D. Dadush, and S. Vempala. Thin Partitions: Isoperimetric Inequalities and Sampling Algorithms for some Nonconvex Families. *To appear in Proceedings of ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2010.

[4] K. Chandrasekaran, A. Deshpande, and S. Vempala. Sampling s-Concave Functions: The Limit of Convexity Based Isoperimetry. *Proceedings: 13th Intl. Workshop on Randomization and Computation - RANDOM 2009*, 5687:420–433, 2009.

[5] K. Chandrasekaran, N. Goyal, and B. Haeupler. Deterministic Algorithms for the Lovász Local Lemma. *To appear in Proceedings of ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2010.

[6] R. Karp. Reducibility among combinatorial problems. *Proc. Sympos. IBM Thomas J. Watson Research Center*, pages 85–103, 1972.

[7] D. M. Li and Y. P. Liu. A polynomial algorithm for finding the minimum feedback vertex set of a 3-regular simple graph. *Acta Math Sci (Engl Ed)*, 19(4):375–381, 1999.

[8] Y. D. Liang. On the feedback vertex set problem in permutation graphs. *Inf. Process. Lett.*, 52(3):123–129, 1994.

[9] Y. D. Liang and M.-S. Chang. Minimum feedback vertex sets in cocomparability graphs and convex bipartite graphs. *Acta Informatica*, 34(5):337–346, 1997.

[10] R. A. Moser and G. Tardos. A constructive proof of the general Lovász Local Lemma. *CoRR*, abs/0903.0544, 2009.

[11] P. D. Seymour. Packing directed circuits fractionally. *Combinatorica*, 15(2):281–288, 1995.

[12] S. Ueno, Y. Kajitani, and S. Gotoh. On the nonseparating independent set problem and feedback set problem for graphs with no vertex degree exceeding three. *Discrete Math.*, 72(1-3):355–360, 1988.