Discrepancy Theory in Approximation Algorithms

Rajat Sen , Soumya Basu

May 8, 2015

1 Introduction

In this report we would like to motivate the use of discrepancy theory in algorithms. Discrepancy theory is a field in combinatorics that deals with coloring of elements of a set system. The important results in discrepancy theory dates back to a famous paper by Spencer [1], which is titled "six standards deviations suffice". But most of the results in this field were non-constructive. Only recently there has been some ground breaking results, and it is now possible to construct low discrepancy colorings in polynomial time [2, 3]. Quite surprisingly these constructuve algorithms can be useful in designing efficient approximation algorithms for seemingly unrelated problems like the classical bin-packing problem, the bin-packing with rejection problem and the train delivery problem[4]. In section 2 we introduce various aspects of discrepancy theory from an algorithmic perspective, then in section 3 we describe the entropy method which is the key tool that has been used in the bin-packing problem. In section 4, we give a brief sketch of the algorithm that gives a better approximation guarantee for the bin-packing with rejection problem. Finally, we list some future applications of this method in section 5.

2 Discrepancy Theory from an algorithmic perspective

The aim of this section is to give an overview of the applications of discrepancy theory in algorithms. Consider a set system (V, C), specified by a collection of elements $V = \{1, 2, ..., m\}$ and a collection of subsets $C = \{S_1, S_2, ..., S_n\}$ of V. Consider a coloring of V, given by $\chi : V \to \{1, -1\}$. The discrepancy of the set system under a coloring is a measure of how uniformly all the sets are colored under that coloring. The formal definition is as follows,

$$\mathrm{disc}(\chi, C) = \max_{S \in C} |\chi(S)|$$

where $|\chi(S)| = |\sum_{i \in S} \chi(i)|$. The discrepancy of the system is defined as $\mathrm{disc}(C) = \min_\chi \max_{S \in C} |\chi(S)|$, where the minimum is over all possible colorings. In many cases discrepancy is not the correct quantity to look at and it is known that is is NP hard to distinguish whether a set system with $n = O(m)$, has a discrepancy of 0 or $\Omega(\sqrt{m})$ [1]. The correct quantity to look at is hereditary discrepancy which is defined as follows,

$$\text{herdisc}(C) = \max_{V' \subseteq V} \text{disc}(C|_{V'})$$

where $C|_{V'}$ is the subset system C restricted to elements only in V'. It is known that herdisc(C) can be approximated to within a factor of $\log^{3/2} n$ efficiently [5]. Any set system (V, C) can be represented by an incidence matrix A, and in this case we will set herdisc(A) = herdisc(C). This notion can be generalized to a matrix $A \in \mathbb{R}^{n \times m}$. For any $V \subseteq \{1, 2, 3, ..., m\}$, let A^V denote the the matrix A restricted to columns belonging to V. We can define the discrepancy of A as $\text{disc}(A) = \min_{\chi \in \{-1,1\}^m} ||A\chi||_\infty$. The hereditary discrepancy of A is given by,

$$\text{herdisc}(A) = \max_{V \subseteq [m]} \text{disc}(A^V)$$

Next we introduce a theorem[6] due to Lovasz et al., which explains our interest in discrepancy from the point of view of approximation algorithms.

Theorem 1. *For any $x \in \mathbb{R}^m$ satisfying $Ax = b$, there is a $\tilde{x} \in \mathbb{Z}^m$ with $||\tilde{x} - x||_\infty < 1$, such that $||A(x - \tilde{x})||_\infty \leq \text{herdisc}(A)$.*

Proof. For $x = (x_1, x_2, ..., x_m)$, we consider the binary expansion of each x_i. That is we have $x_i = \lfloor x_i \rfloor + \sum_{j \geq 1} q_{ij} 2^{-j}$. Let $A^{(k)}$ be the sub-matrix of A restircted to the columns i for which $q_{ik} = 1$. By definition of hereditary discrepancy there exists a coloring of these columns, $\chi^{(k)}$, such that $||A\chi^{(k)}||_\infty \leq \text{herdisc}(A)$. Let us pad zeroes in the remaining elements of $\chi^{(k)}$ such that $\chi^{(k)} \in \mathbb{R}^m$. Consider the vector $x' = x + 2^{-k}\chi^{(k)}$. Now, the k^{th} bit of all x'_i is 0. However, $||Ax - Ax'||_\infty \leq 2^{-k}.\text{herdisc}(A)$. We can now iterate this process at bit positions $k - 1, k - 2, ..., 1$, to get $|Ax - Ax'| \leq (2^{-k} + 2^{-k+1} + ... + 2^{-1}).\text{herdisc}(A)$. Making k arbitrarily large implies the result. □

Moreover there is a randomized poly-time algorithm to find a coloring with a discrepancy at most $O((\log m \log n)^{1/2}\text{herdisc}(A)$ [7]. In order to get meaningful bounds we need upper bounds on the quantity herdisc(A). There are some good methods to upper bound herdisc(A)[5, 2], which would then translate to upper bounds for error in rounding LP solutions to integer values.

3 The Entropy Method

In this section we present the main ideas of [4] through a brief review of the rounding algorithm for any given LP. Consider an LP relaxation of some combinatorial problem given by

$$\min\{c^T x | Ax \geq b, x \geq \mathbf{0}\}.$$

where $A \in \mathbb{R}^{n \times m}$. The entropy rounding method aims to generate an integral rounding of the LP solution x^* to $\hat{x} \in \{0, 1\}^m$ ensuring $c^T\hat{x}/c^T x^*$ is not too large.

In order to present the rounding method we introduce the notion of random coloring, i.e., consider the random variable $A\chi$, where $\chi \in \{\pm 1\}^m$ is chosen uniformly at random. Using the common notation $H(\cdot)$ as the entropy of a random variable and the notation $[\cdot]$ as the nearest integer we have the following definition,

Theorem 2. *Let $A \in \mathbb{R}^{n \times m}$ be a matrix and $\Delta = \{\Delta_1, \ldots, \Delta_n\}$ be a vector with $\Delta_i > 0$. We define the Δ-approximate entropy of A as*

$$H_\Delta(A) = H_{\chi \in \{1,-1\}^m}\left(\left\{\left[\frac{A_i \chi}{2\Delta_i}\right]\right\}_{i=1,\ldots,n}\right) \leq \frac{m}{5}$$

Now that we have the necessary definitions , we are at a position to present the main theorem of [4].

Theorem 3. *Assume the following holds*

(1) $A \in \mathbb{R}^{n_A \times m}$, $\Delta = \{\Delta_1, \ldots, \Delta_n\} > \mathbf{0}$ *such that* $\forall J \in [m]$, $H_\Delta\left(A^J\right) \leq \frac{|J|}{10}$

(2) a matrix $B \in [-1,1]^{n_B \times m}$, *weights* $\mu_1, \ldots, \mu_{n_B} > 0$ *with* $\sum_i \mu_i \leq 1$

(3) a vector $x \in [0,1]^m$ *and an objective function* $c \in [-1,1]^m$.

Then there exists a vector $y \in \{0,1\}^m$ *with,*

- *Preserved expectation:* $\mathbb{E}\left(c^T y\right) = c^T x$, $\mathbb{E}(Ay) = Ax$, $\mathbb{E}(By) = Bx$.
- *Bounded difference:*
  - *(i)* $|c^T x - c^T y| \leq O(1)$
  - *(ii)* $|A_i x - A_i y| \leq \log(\min\{4n, 4m\}) .\Delta_i$, $\forall i \in [n_A], n := n_A + n_B$
  - *(iii)* $|B_i x - B_i y| \leq O\left(\sqrt{1/\mu_i}\right)$, $\forall i \in [n_B]$.

Data: Solution of the LP, x.
Result: Integral solution y, after rounding
Round x to the nearest larger or smaller integer multiple of $(\frac{1}{2})^K$;
while *x not integral* **do**
  Let $k \in \{1, 2, , \ldots, K\}$ be the index of the least bit in any entry of x ;
  $J = \{j \in [m] | x_j\text{'s } k\text{-th bit is } 1\}$;
  Choose the half coloring $\chi \in \{0, \pm 1\}^m$ with $\chi(j) = 0\ \forall j \in J$, $supp(\chi) \geq |J|/2$, $|A_i \chi| \leq \Delta_i$ and $|B_i \chi| \leq G^{-1}(\mu_i |J|/10)\sqrt{J}$, for all i;
  With probability $\frac{1}{2}$ flip all signs of χ;
  Update $x := x + (\frac{1}{2})^k \chi$;
end

Algorithm 1: Entropy Rounding Method

The proof of the above theorem relies on the existence of a proper half coloring in each iteration which is a direct consequence of Theorem 1. The construction of such a coloring in polynomial time is feasible due to the SDP algorithm given by Nikhil Bansal in [7]. The combination of these two key results in discrepancy theory enables us to effectively round the LP solution.

4 Application : Bin-packing with rejection

The bin packing problem (BP) involves packing a set of items into the minimum number of bins of unit size. The bin packing with rejection (BPR) problem generalizes the above problem by allowing rejection of elements at some specific cost. Though both the problems are known to be NP-Hard [8] they admit Asymptotic FPTAS (AFPTAS). Denoting OPT as the optimal solution, the best known additive approximation to the BPR problem was $O\left(\frac{OPT}{\log(OPT)^{1-o(1)}}\right)$ [10]. Whereas, the $O\left(\log^2(OPT)\right)$ [9] approximation to BP problem was known for a long time.

In [4] the author proposes an AFPTAS with approximation $O\left(\log^2(OPT)\right)$ for the BPR problem through the use of discrepancy methods. More recently there has been a groundbreaking improvement as Rothvoss [11] designs an AFPTAS of approximation $O(\log(OPT)\log(\log(OPT)))$ for the bin packing with rejection.

In the BPR problem we have an input with items of size $s_1 \geq s_2 \geq \cdots \geq s_n$ and penalty $\{\pi_i\}_{i=1}^n$. The set of potential bin patterns is $\mathbf{B} = \{S | \sum_{i \in S} s_i \leq 1\}$, whereas the rejection pattern is given by $\mathbf{R} = \{\{i\} | i \in [n]\}$. The union of these sets give us the complete set system for the BPR problem, $\mathbf{S} = \mathbf{B} \cup \mathbf{R}$ with $|S| = m$. Each set $S \in B$ has a cost of $c_S := 1$ and rejection of item i has cost $c_i := \pi_i$, for all $i \in [n]$. Let $\mathbf{1}_S = \{0,1\}^n$ with $\mathbf{1}_{S,i} = \mathbb{1}(i \in S)$ denote the candidacy of elements in each set S. The natural LP relaxation to the BPR problem is

$$OPT_f = \{c^T x | \sum_{s \in \mathbf{S}} x_s \mathbf{1}_s = \mathbf{1}, x \geq \mathbf{0}\}. \tag{1}$$

The above LP (1), despite having an exponentially many variables, can be approximated in polynomial time to get a feasible solution x with cost $OPT_f + 1$ using ellipsoid method. Where OPT_f being the optimal solution of the LP satisfies $OPT_f \leq OPT$.

Define the pattern matrix as $P = \{\mathbf{1}_s\}_{s \in S}$. This matrix has no inherent structure as the variation in item sizes in each problem instance changes the potential bin patterns significantly. Instead we consider the cumulative pattern matrix A with rows $A_i = \sum_{i'=1}^{i} P_i$, which ensures all the columns in A are non-decreasing. For some constant $C > 0$, let $\Delta = \left(\left\{\frac{C}{s_i}\right\}_{i=1}^n\right)$. Then we can bound the Δ-approximate entropy of this matrix due to the following Lemma 4 as $H_\Delta(A) \leq m/10$.

Lemma 4 ([4]). *Let $A \in \mathbb{Z}_{\geq 0}^{m \times n}$ be any matrix with column j having non decreasing elements $\{0, \ldots, b_j\}$ for all $j \in [m]$; let $\sigma = \sum_{j \in [m]} b_j$ and $\beta = \max_{j \in [m]} b_j$.*

Then with any $\Delta > \mathbf{0}$ one has $H_\Delta(A) \leq O\left(\frac{\beta\sigma}{\Delta^2}\right)$.

The rounding method employed in the work uses two key observations. For each i, the items $1, \ldots, i$ are covered (either in some bin or rejected) by an integral solution $y \in \{0,1\}^m$ if the deficit $i - \sum_{S \in \mathbf{S}} y_S |S \cap \{1, \ldots, i\}|$ is not positive. Note that for a feasible solution x of LP (1), y has no deficit (covers all elements) if and only if $Ay - Ax \geq \mathbf{0}$. Also differentiate the items as 'large' items if $s_i \geq \epsilon$, and 'small' otherwise. We will introduce the value of ϵ later. We describe the rounding method for the solution to the LP (1) below,

- Firstly, discard all items i with $x_{\{i\}} > 1 - \epsilon$, which incurs an additional cost of ϵOPT_f. Therefore, in the solution, any remaining item j has at least ϵ fraction packed in bins.

- Let there be L large items remaining. We have $OPT_f \geq \epsilon^2 L$, because the optimal solution must cover an ϵ fraction of all the L large items.
- For each potential bin pattern B we group all the small items together, call it item $L+1$. Note this item can have different size for different set S. Let $B \in \mathbb{R}^{1\times m}$ be the vector of the cumulative size of small items in each set, so $B_S \leq 1$, $\forall s \in \mathbb{B}$. Consider the cumulative pattern matrix $A \in \mathbb{R}^{L\times m}$ for the L large items, the vector B and the parameter Δ, defined earlier. There exists a vector $y \in \{0,1\}^{L+1}$ such that, i) for the sets $\{1,\dots,i\}$ there is a deficit of $O(\log(L)/s_i)$ for all $i \in [L]$ and ii) the $(L+1)$-th item has $O(1)$ deficit [Theorem 3].
- The following subgroups of the large items give a finer partitioning based on size,

  $G^l = \{i | s_i \in [\left(\frac{1}{2}\right)^l, \left(\frac{1}{2}\right)^{l+1}]\}, \forall l = 0, \dots \log(1/\epsilon)$.

  By removing the largest $\log(L)$ items from each subgroup l and packing them in $O(\log(L)\log\left(\frac{1}{\epsilon}\right))$ new bins we compensate for the deficit for the large items.
- The remaining deficit in small items is overcome by adding $O(1)$ extra bins. The drawback for combining the small items is that $(L+1)$-th element in each bin, maybe packed fractionally. Therefore, we can cover the rejections and the large items integrally but small items fractionally with $\delta = O(\epsilon OPT_f) + O(1) + O(\log(L)\log\left(\frac{1}{\epsilon}\right))$ extra space.
- Finally, we only need $\epsilon(OPT_f + \delta)$ extra bins to construct an integral packing for the smaller items from the fractional packing. This result in a complete integral solution with the total excess cost of $\delta(1+\epsilon) + \epsilon OPT_f$.
- The careful choice of $\epsilon = \log(OPT_f)/OPT_f$ achieves the bound of $O(\log^2(OPT_f))$.

Therefore, the application of discrepancy theory gives a $O\left(\log^2(OPT)\right)$ approximation to the BPR problem.

5 Possible future applications

In the previous sections we gave a brief survey on how to use discrepancy theory in the design of rounding techniques for an LP. Our objective is to internalize the process of entropy rounding and apply it on some existing or new LP relaxation of combinatorial optimization problem. We are interested in the problem of packing cycles in graphs [12]. Suppose we have a graph given by $G = (V,E)$ and let $\mathcal{C}$ denote the set of all cycles in the graph, which can be exponential in number. We would like to find the largest number of edge disjoint cycles in the graph. The problem has a natural ILP formulation given by,

$$\text{OPT} = \{\max \sum_{C\in\mathcal{C}} x_C | \sum_{C\ni e} x_C \leq 1 \text{ for all } e \in E, x_C \in \{0,1\}, \forall C \in \mathcal{C}\}$$

The corresponding LP relaxation can be solved using the ellipsoid method. The structure of this problem is very similar to that of the bin-packing problem. Therefore, we believe that the application of the entropy method to this problem will yield meaningful results.

References

[1] J. Spencer, "Six standard deviations suffice," *Transactions of the American Mathematical Society*, vol. 289, no. 2, pp. 679–706, 1985.

[2] N. Bansal, "Algorithmic aspects of combinatorial discrepancy," in *A Panorama of Discrepancy Theory*. Springer, 2014, pp. 425–457.

[3] S. Lovett and R. Meka, "Constructive discrepancy minimization by walking on the edges," in *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on*. IEEE, 2012, pp. 61–67.

[4] T. Rothvoß, "The entropy rounding method in approximation algorithms," in *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*. SIAM, 2012, pp. 356–372.

[5] A. Nikolov and K. Talwar, "Approximating hereditary discrepancy via small width ellipsoids," *arXiv preprint arXiv:1311.6204*, 2013.

[6] L. Lovász, J. Spencer, and K. Vesztergombi, "Discrepancy of set-systems and matrices," *European Journal of Combinatorics*, vol. 7, no. 2, pp. 151–160, 1986.

[7] N. Bansal, "Constructive algorithms for discrepancy minimization," in *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*. IEEE, 2010, pp. 3–10.

[8] E. G. Coffman Jr, M. R. Garey, and D. S. Johnson, "Approximation algorithms for bin-packingan updated survey," in *Algorithm design for computer system design*. Springer, 1984, pp. 49–106.

[9] N. Karmarkar and R. M. Karp, "An efficient approximation scheme for the one-dimensional bin-packing problem," in *Foundations of Computer Science, 1982. SFCS'08. 23rd Annual Symposium on*. IEEE, 1982, pp. 312–320.

[10] L. Epstein and A. Levin, "Afptas results for common variants of bin packing: A new method for handling the small items," *SIAM Journal on Optimization*, vol. 20, no. 6, pp. 3121–3145, 2010.

[11] T. Rothvoss, "Approximating bin packing within o(log opt * log log opt) bins," 2013.

[12] A. Caprara, A. Panconesi, and R. Rizzi, "Packing cycles in undirected graphs," *Journal of Algorithms*, vol. 48, no. 1, pp. 239–256, 2003.