Available online at www.sciencedirect.com
ScienceDirect
Discrete Applied Mathematics 155 (2007) 2039–2042
DISCRETE APPLIED MATHEMATICS
www.elsevier.com/locate/dam

Note

On the L_2-discrepancy

Jinsoo Oh, Byung-Ro Moon*

School of Computer Science and Engineering, Seoul National University, Seoul 151-744, South Korea

Received 21 July 2006; received in revised form 7 December 2006; accepted 13 April 2007
Available online 22 May 2007
Communicated by E. Boros

Abstract

For a family of subsets $\mathscr{S}$ of a finite set V, a coloring $\chi : V \to \{-1, 1\}$, and $S_j \in \mathscr{S}$, let $\chi(S_j) = \sum_{v \in S_j} \chi(v)$. We consider the problem to minimize $\sum_j \chi(S_j)^2$ and we call the problem L_2-discrepancy problem. We show that the problem is NP-complete, and we also provide an upper bound for the L_2-discrepancy.

Keywords: L_2-discrepancy; Hypergraph; NP-completeness

1. Introduction

The discrepancy problem is a well-known problem [4]. For a family of subsets $\mathscr{S}$ of a finite set V, a coloring $\chi : V \to \{-1, 1\}$, and $S_j \in \mathscr{S}$, let $\chi(S_j) = \sum_{v \in S_j} \chi(v)$. Define the discrepancy of $\mathscr{S}$ with respect to χ by

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

and the discrepancy of $\mathscr{S}$ by

$$\mathrm{disc}(\mathscr{S}) = \min_{\chi} \mathrm{disc}(\mathscr{S}, \chi).$$

The NP-completeness of the discrepancy problem was proved in [5]. A known upper bound of the problem is $5.32 \times \sqrt{|V|}$ when $|V| = |\mathscr{S}|$ [6]. In [3], it was proven that the discrepancy is at most $2d - 1$ where d is the maximum number of subsets containing a common element of V. There is a conjecture that there always exists a coloring in which the discrepancy is $O(\sqrt{d})$ [2].

In [1], the following L_2-discrepancy was defined. With the basic settings described above, the L_2-discrepancy of $\mathscr{S}$ with respect to χ is defined by

$$\mathrm{disc}_2(\mathscr{S}, \chi) = \sqrt{\frac{1}{|\mathscr{S}|} \sum_j |\chi(S_j)|^2}$$

* Corresponding author.
E-mail addresses: jsoh@soar.snu.ac.kr (J. Oh), moon@soar.snu.ac.kr (B.-R. Moon).

0166-218X/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.dam.2007.04.021

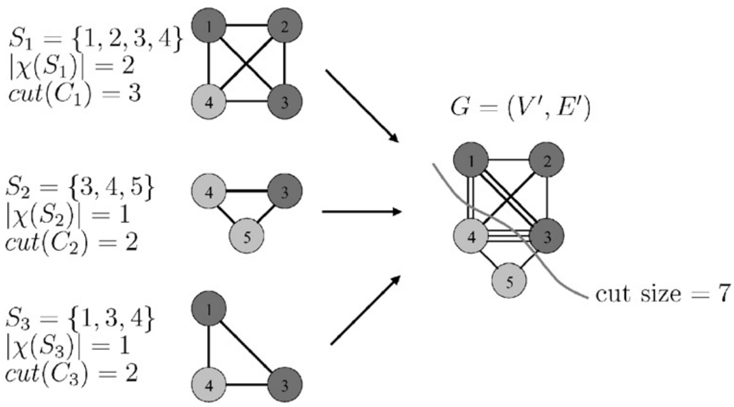

Fig. 1. Coloring example. When vertices 1, 2, 3 are colored −1 (dark gray) and vertices 4, 5 are colored +1 (light gray), $V_1' = \{1, 2, 3\}$ and $V_2' = \{4, 5\}$.

and the L_2-discrepancy of $\mathscr{S}$ by

$$\mathrm{disc}_2(\mathscr{S}) = \min_{\chi} \mathrm{disc}_2(\mathscr{S}, \chi).$$

In Section 2, we state the relationship between the L_2-discrepancy problem and the MAX-CUT problem. Based on the relationship in Section 2, we prove the NP-completeness of the L_2-discrepancy problem in Section 3 and provide an upper bound in Section 4.

2. Relation with MAX-CUT problem

Given an instance of L_2-discrepancy problem, we construct a multigraph $G = (V', E')$, as an instance of MAX-CUT problem, which is the edge (but not vertex)-disjoint union of the cliques $C_j = (S_j, \binom{S_j}{2})$ (Fig. 1).

To find the relation between L_2-discrepancy and the maxcut size of G, we observe the relation between the cut size of clique C_j and $|\chi(S_j)|$ of S_j. Assume $|C_j| = k$ and $|\chi(S_j)| = l$. That is, the elements of S_j are partitioned into S_j^1 and S_j^2 which are colored by −1 and +1, respectively, and $|S_j^1| = \frac{k+l}{2}$ and $|S_j^2| = \frac{k-l}{2}$ (or $|S_j^1| = \frac{k-l}{2}$ and $|S_j^2| = \frac{k+l}{2}$). Then, the cut size $cut(C_j) = \frac{k^2-l^2}{4}$. Therefore,

$$cut(C_j) = \frac{1}{4}(|C_j|^2 - |\chi(S_j)|^2).$$

So,

$$\sum_j cut(C_j) = \tfrac{1}{4}\left(\sum_j |C_j|^2 - \sum_j |\chi(S_j)|^2\right). \tag{1}$$

Note that $\sum_j cut(C_j)$ is the cut size of G and $\sum_j |C_j|^2$ is constant. Maximizing the cut size of graph G thus becomes minimizing $\sum_j |\chi(S_j)|^2$.

3. NP-completeness

The decision problem of L_2-discrepancy is to determine if there exists a coloring χ such that $\sum_j |\chi(S_j)|^2 \leqslant D$. The decision problem of MAX-CUT is to determine if there exists a partition of cut size $\geqslant K$.

Theorem 1. *L_2-discrepancy is NP-complete.*

Proof. When the coloring χ is given, it is easy to check $\sum_j |\chi(S_j)|^2 \leqslant D$ in polynomial time. Thus, L_2-discrepancy is in NP.

Next, we reduce MAX-CUT to L_2-discrepancy. Given an instance $G=(V, E)$ of MAX-CUT, $\mathscr{S} := E$. Note that we can construct the $G=(V, E)$ using $\mathscr{S}$, reversely. Finally, D is set to $\sum_j |C_j|^2 - 4K$. As shown in (1), max cut size is greater than or equal to K if and only if $\sum_j |\chi(S_j)|^2$ is less than or equal to D. □

4. Upper Bound

Theorem 2. *There is a coloring χ such that* $\sqrt{\frac{1}{|\mathscr{S}|}\sum_j |\chi(S_j)|^2} \leqslant \sqrt{|V|}$.

Proof. From (1),

$$\sum_j |\chi(S_j)|^2 = \sum_j |C_j| - 4\left(\sum_j cut(C_j) - \frac{1}{2}\frac{\sum_j |C_j|(|C_j|-1)}{2}\right).$$

Note that $\frac{\sum_j |C_j|(|C_j|-1)}{2}$ is the total weight W of the graph $G=(V', E')$. Consider the well-known greedy algorithm for the MAX-CUT problem described as follows: Provided two empty bins B, $\bar{B}$ and an arbitrary ordering of the vertices, we initially put the first vertex in B and then put each subsequent vertex in one of the two bins so that the resulting cut size is maximized. This algorithm guarantees the cut size greater than or equal to $\frac{W}{2}$. χ can be obtained from the cut constructed, so that

$$\sum_j |\chi(S_j)|^2 \leqslant \sum_j |C_j|.$$

Trivially,

$$\sum_j |C_j| \leqslant |\mathscr{S}||V|. \tag{2}$$

Therefore,

$$\sqrt{\frac{1}{|\mathscr{S}|}\sum_j |\chi(S_j)|^2} \leqslant \sqrt{|V|}. \quad \square$$

With a proper assumption, we draw the following corollary.

Corollary 1. *Assume that $|V| \leqslant |\mathscr{S}|$. Then, there is a coloring χ such that* $\sqrt{\frac{1}{|\mathscr{S}|}\sum_j |\chi(S_j)|^2} \leqslant \sqrt{d}$ *where d is the maximum number of subsets containing a common element of V.*

Proof. By the assumption, the right-hand side of (2) can be changed as follows:

$$\sum_j |C_j| \leqslant d|V| \leqslant d|\mathscr{S}|.$$

So,

$$\sqrt{\frac{1}{|\mathscr{S}|}\sum_j |\chi(S_j)|^2} \leqslant \sqrt{d}. \quad \square$$

References

[1] N. Alon, B. Doerr, T. Luczak, T. Schoen, On the discrepancy of combinatorial rectangles, Random Structures Algorithms 21 (2002) 205–215.
[2] N. Alon, J.H. Spencer, The Probabilistic Method, Wiley, New York, 2000 pp. 199–213.

[3] J. Beck, T. Fiala, Integer-making theorems, Discrete Appl. Math. 3 (1981) 1–8.
[4] J. Beck, V.T. Sós, Discrepancy theory, in: R. Graham, M. Grötschel, L. Lovász (Eds.), Handbook of Combinatorics, Elsevier, Amsterdam, 1995, pp. 1405–1446.
[5] T.C. Biedl, E. Čenek, T.M. Chan, E.D. Demaine, M.L. Demaine, R. Fleischer, M.-W. Wang, Balanced k-colorings, Discrete Math. 254 (2002) 19–32.
[6] J.H. Spencer, Six standard deviations suffice, Trans. Amer. Math. Soc. 289 (1985) 679–706.