Optimization
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713646500

A semidefinite programming approach to the hypergraph minimum bisection problem

Changhui Choi[a]; Samuel Burer[b]
[a] Department of Mathematical and Statistical Sciences, University of Colorado Denver, Denver, CO 80217-3364, USA [b] Department of Management Sciences, University of Iowa, Iowa City, IA 52242-1000, USA

First published on: 23 April 2010

To cite this Article Choi, Changhui and Burer, Samuel(2011) 'A semidefinite programming approach to the hypergraph minimum bisection problem', Optimization, 60: 3, 413 — 427, First published on: 23 April 2010 (iFirst)
To link to this Article: DOI: 10.1080/02331930903233744
URL: http://dx.doi.org/10.1080/02331930903233744

Optimization
Vol. 60, No. 3, March 2011, 413–427

Taylor & Francis
Taylor & Francis Group

A semidefinite programming approach to the hypergraph minimum bisection problem

Changhui Choi[a]* and Samuel Burer[b]

[a]*Department of Mathematical and Statistical Sciences, University of Colorado Denver, Denver, CO 80217-3364, USA;* [b]*Department of Management Sciences, University of Iowa, Iowa City, IA 52242-1000, USA*

(*Received 17 October 2008; final version received 28 July 2009*)

The hypergraph minimum bisection (HMB) problem is the problem of partitioning the vertices of a hypergraph into two sets of equal size so that the total weight of hyperedges crossing the sets is minimized. HMB is an NP-hard problem that arises in numerous applications – for example, in digital circuit design. Although many heuristics have been proposed for HMB, there has been no known mathematical program that is equivalent to HMB. As a means of shedding light on HMB, we study the equivalent, complement problem of HMB (called CHMB), which attempts to maximize the total weight of non-crossing hyperedges. We formulate CHMB as a quadratically constrained quadratic program, considering its semidefinite programming relaxation and providing computational results on digital circuit partitioning benchmark problems. We also provide results towards an approximation guarantee for CHMB.

Keywords: hypergraph minimum bisection; semidefinite programming; digital circuit design

1. Introduction

Let $H=(V,E,w)$ be an edge-weighted hypergraph, where $V=\{1,\dots,n\}$ is the set of vertices; $E=\{E_1,\dots,E_m\}$ is a collection of m subsets of V denoting the hyperedges (each of which contains at least 2 vertices); and $w=(w_1,\dots,w_m)^T$ is a vector of non-negative weights, one for each hyperedge. We assume that n is an even number. The vertex j is said to be adjacent to E_i if $j\in E_i$, and two vertices j_1 and j_2 are adjacent if they are mutually adjacent to some E_i. We define $\tau_i:=|E_i|$ and $\tau:=\max_i\tau_i$.

For any $S\subseteq V$, let $S^c:=V\setminus S$ be the complement of S. We say that a hyperedge E_i is *cut* with respect to (S, S^c) if $E_i\cap S\neq\emptyset$ and $E_i\cap S^c\neq\emptyset$; otherwise, E_i is said to be *uncut*. The objective of the hypergraph minimum bisection (HMB) problem is to find (S,S^c) with $|S|=n/2$, which minimizes the total weight of cut hyperedges.

*Corresponding author. Email: changhui.choi@cudenver.edu

ISSN 0233–1934 print/ISSN 1029–4945 online

DOI: 10.1080/02331930903233744
http://www.informaworld.com

The objective can alternatively be expressed as maximizing the total weight of uncut hyperedges, which we specify by the function

$$w(S, S^c) := \sum_{\{i:E_i \text{ uncut}\}} w_i.$$

Formally, we refer to the problem of maximizing $w(S, S^c)$ as the *complement* of HMB, or CHMB. Note that these two problems are equivalent in that they share the same optimal solution set.

HMB is well known to be an NP-hard problem, even on ordinary graphs (i.e. with $\tau = 2$) [11]. Applications of HMB include task-scheduling, machine vision (or pattern recognition), design of digital circuits and database design [17,18].

The majority of studies on HMB have concerned the design of heuristic methods. A survey paper by Alpert and Kahng [2] contains an extensive list of such methods, including approaches based on simulated annealing, tabu-search and spectral partitioning, as well as move-based methods such as multilevel partitioning and clustering methods.

Much work has also been done on approximation results for minimum bisection on ordinary graphs, i.e. HMB with $\tau_i = \tau = 2$ for all i. Feige and Krauthgamer [8] prove an $O(\log^2 n)$ approximation ratio (see also [3]). Regarding the complement problem on ordinary graphs, Ye and Zhang [20] establish an approximation guarantee of 0.602 using a Goemans–Williamson-style algorithm [13] based on the semidefinite programming (SDP) relaxation.

For general instances of HMB, however, we are aware of only one study concerning the approximability of HMB. Berman and Karpinski [4] show that HMB on τ-uniform hypergraphs (i.e. $\tau_i = \tau$ for all i) is as hard to approximate, up to a factor $\frac{\tau}{3}$, as minimum bisection on ordinary graphs.

Please note that people work on approximation algorithms on the complement of minimum bisection on ordinary graphs because only the complement problem has an approximation guarantee. For the same reason, it makes sense to work with CHMB to achieve an approximation guarantee. Although HMB and CHMB are equivalent problems, they have different properties from the viewpoint of establishing an approximation guarantee.

We mention another common technique for working with HMB and CHMB. An instance of HMB can be translated into a minimum bisection instance on an ordinary graph by replacing each hyperedge with a clique. Care is taken when assigning weights to each edge of the clique so as to approximate the weight of a cut hyperedge by its corresponding cut clique, though a perfect translation is not possible [14]. For example, Choi and Ye [7] employed this technique as a part of heuristic for HMB, which was based on a Goemans–Williamson-style randomization algorithm applied to the SDP relaxation of a quadratically constrained quadratic programming (QCQP) formulation of minimum bisection on ordinary graphs.

In this article, we develop and investigate a new approach for HMB, one which follows in the footsteps of the above studies. We first present in Section 2, a mathematical programming model for CHMB, which to our knowledge is the first such model for CHMB (and HMB by equivalence). The formulation is a QCQP and hence admits an SDP relaxation, which we explain. Then in Section 3 we investigate the power of the SDP relaxation from two points of view: (i) the tightness of bounds on the optimal value of CHMB and (ii) the quality of heuristic solutions for CHMB and HMB gotten by a Goemans–Williamson-style algorithm with post-processing by simple local search. Computational results are presented on real-world benchmark problems. In Section 4, we present results towards an approximation guarantee for CHMB based on analysing its SDP relaxation. Our analysis is based on some interesting geometric properties of SDP feasible solutions. Finally, some conclusions and ideas for future research directions are briefly discussed in Section 5.

1.1. *Notation and terminology*

We mention some notation and terminology that we will use throughout this article. For two vectors a and b of the same size, $\langle a, b\rangle := a^T b$ denotes their inner product. Also, $\|a\| := \sqrt{\langle a, a\rangle}$ is the 2-norm of a. The symbol $\bullet$ indicates the matrix inner product, i.e. for conformal matrices A and B, we define $A \bullet B := \operatorname{trace}(A^T B)$. For a symmetric matrix A, the notation $A \succeq 0$ means that A is positive semidefinite (i.e. all eigenvalues are non-negative). Diag(a) is a diagonal matrix made from the vector a. The notation $|\cdot|$ is used to denote absolute value when applied to a scalar and cardinality when applied to a set. I is the identity matrix, e_k is a column vector with one in the k-th position and zeros elsewhere and the all-ones vector is denoted e. The dimensions of I, e_k and e should be clear from context. For a random variable X and a random event x, $P[X=x]$ and $E[X]$ denote the probability and expected value, respectively. Moreover, we will often use the same notation to denote a random variable and a random sample from it; the particular usage will be made clear from its context.

2. A QCQP formulation of CHMB and its SDP relaxation

We are unaware of any exact mathematical models for CHMB (except for the case $\tau = 2$). In this section, we show how to formulate CHMB as a QCQP problem and relax it as an SDP. We consider the technique of SDP relaxation because it has provided high-quality bounds and solutions for similar graph partitioning problems [10,13,19,20]. (We will investigate bounds and solutions for CHMB based on its SDP relaxation in Section 3.)

We define the following variables: $x_j \in \{-1, 1\}$ for each vertex j and $y_i \in \{-1, 0, 1\}$ for each hyperedge E_i. The sets (S, S^c) corresponding to a bisection are defined by $S := \{j : x_j = 1\}$. So vertices j_1 and j_2 are on the same side of the bisection if and only if $x_{j_1} = x_{j_2}$. By convention, E_i will be uncut if and only if $y_i^2 = 1$. Consider the following QCQP:

$$\begin{aligned} w^* := \max\ & \textstyle\sum_{i=1}^m w_i y_i^2, \\ \text{s.t.}\ \ & x_j^2 = 1 \qquad \forall j = 1, \ldots, n, \\ & y_i^2 - y_i x_j = 0 \quad \forall j \in E_i \ \ \forall i = 1, \ldots, m, \\ & \textstyle\sum_{j=1}^n x_j = 0. \end{aligned} \tag{Q-CHMB}$$

The first two constraints guarantee $x_j \in \{-1, 1\}$ and $y_i \in \{-1, 0, 1\}$, and the third constraint guarantees $|S| = n/2$. The second constraint and the objective function capture whether a hyperedge is cut or not by the following observations: (i) if E_i is cut, that is, there exist $j_1, j_2 \in E_i$ satisfying $x_{j_1} \neq x_{j_2}$, then $y_i = 0$; and (ii) if E_i is uncut, that is, if $x_{j_1} = x_{j_2}$ for all $j_1, j_2 \in E_i$, then $y_i = x_{j_1}$ in order to maximize the objective. Let us give a simple example. Suppose $E_1 = \{1, 2, 3\}$. If all of x_1, x_2, x_3 do not have the same sign, then it must hold that $y_1 = 0$ in order to satisfy $y_1(y_1 - x_j) = 0$ where $j \in \{1, 2, 3\}$. On the other hand, if $x_1 = x_2 = x_3 = 1$, y_1 can either be 0 or 1. In this case, $y_1 = 1$ is chosen because we are maximizing $w_1 y_1^2$ with $w_1 \geq 0$. Also note that $y_1 = -1$ when $x_1 = x_2 = x_3 = -1$. We will show in the following proposition.

Proposition 2.1 *(Q-CHMB) is equivalent to CHMB.*

By formulating CHMB as a QCQP, we have made it possible to approximate CHMB by relaxing it as an SDP, which we now illustrate. The technique of SDP relaxation is considered because it is known to work well with similar graph partitioning problems [10,13,19,20].

First, we rewrite the original problem in matrix form using the rank-1 positive semidefinite matrix

$$\begin{pmatrix} X & Z^T \\ Z & Y \end{pmatrix} := \begin{pmatrix} xx^T & xy^T \\ yx^T & yy^T \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}^T,$$

where $x = (x_1, \ldots, x_n)^T$ and $y = (y_1, \ldots, y_m)^T$. For notational convenience, we also write

$$M := \begin{pmatrix} X & Z^T \\ Z & Y \end{pmatrix}. \tag{1}$$

This allows (Q-CHMB) to be expressed as

$$\begin{aligned} \max\ & \mathrm{Diag}(w) \bullet Y, \\ \text{s.t.}\ & e_j e_j^T \bullet X = 1 && \forall j = 1, \ldots, n, \\ & e_i e_i^T \bullet Y - e_i e_j^T \bullet Z = 0 && \forall j \in E_i \;\; \forall i = 1, \ldots, m, \\ & ee^T \bullet X = 0, \\ & \mathrm{Rank}(M) = 1. \end{aligned} \tag{M-CHMB}$$

Now we relax this last problem into an SDP by removing any appearance of x and y and enforcing positive semidefiniteness on M, keeping in mind that M and X, Y, Z are related according to (1)

$$\begin{aligned} w_+^* := \max\ & \mathrm{Diag}(w) \bullet Y, \\ \text{s.t.}\ & e_j e_j^T \bullet X = 1 && \forall j = 1, \ldots, n, \\ & e_i e_i^T \bullet Y - e_i e_j^T \bullet Z = 0 && \forall j \in E_i \;\; \forall i = 1, \ldots, m, \\ & ee^T \bullet X = 0, \\ & M \succeq 0. \end{aligned} \tag{S-CHMB}$$

By construction, we have that the optimal value of (S-CHMB) is not smaller than the optimal value of CHMB, i.e. $w_+^* \geq w^*$.

3. Computational results for a digital circuit design problem

In this section, we investigate the use of the SDP relaxation (S-CHMB) for providing good bounds on the optimal w^* of (Q-CHMB) and also for generating quality solutions to CHMB (and HMB by equivalence). Our test cases arise from an application in digital circuit design.

3.1. *Digital circuit design and benchmark problems*

One of the most important applications of HMB (or CHMB) is in digital circuit design. A digital circuit is a cluster of electronic components with wires connecting multiple components simultaneously. A digital circuit can be identified with a hypergraph. Figure 1 is a simple example of a digital circuit. The circuit has six components (vertices) and four wires (hyperedges). For example, wire E_1 in Figure 1 connects components 1, 2 and 3 simultaneously.

Sometimes the size of a digital circuit is too big to fit in one 2-dimensional layer. In this case, partitioning it into multiple layers is considered while minimizing the wire connections between layers. This is the type of problem modelled by HMB (or CHMB) (although circuit design problems can be much more general). Minimizing the number of connections is important because:

(a) signal delays are reduced and
(b) the design is simpler.

In digital circuit design, this problem has been studied for more than two decades [9].

Since HMB is a special case of the digital circuit design problem just described, circuit design is an NP-hard combinatorial optimization problem. Hence, the majority of approaches for this problem have been heuristic methods [2]. In particular, METIS [15,16] is considered one of the best heuristics for solving HMB.

In the following subsections, we present computational results for the solution of some digital circuit benchmark problems [1] via CHMB and its SDP relaxation. Furthermore, we use METIS as a point of comparison. Table 1 shows the details of the benchmark problems, including name, number of vertices and number of hyperedges. For all experiments, the SDP solver used is SDPLR 1.02 [5,6] with

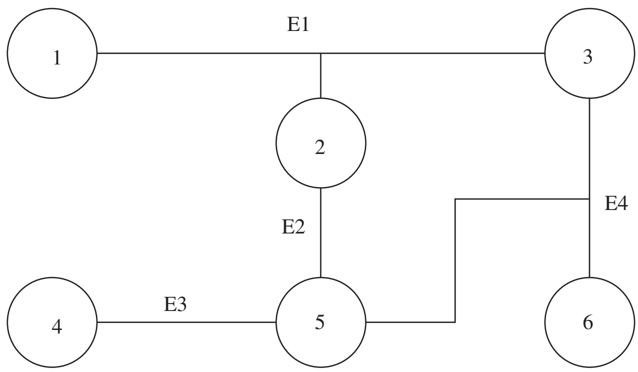

Figure 1. A simple digital circuit.

Table 1. Details of the digital circuit benchmark problems.

Name	n	m
balu	801	735
p1	833	902
bm1	882	903
t4	1515	1658
t3	1607	1618
t2	1663	1720
t6	1752	1541
struct	1952	1920
t5	2595	2750
19ks	2844	3282
p2	3014	3029
s9234	5866	5844
biomed	6514	5742
s13207	8772	8651
s15850	10,470	10,383

Table 2. Relative gap of the (S-CHMB) optimal value on benchmark problems.

Name	SDP bound	Value	Gap (%)
balu	711.6581	705	0.94
p1	870.2040	849	2.50
bm1	872.4735	852	2.40
t4	1625.3291	1607	1.14
t3	1582.3842	1558	1.57
t2	1672.0276	1631	2.52
t6	1596.0932	1478	7.99
struct	1898.3215	1886	0.65
t5	2705.0885	2678	1.01
19ks	3228.2448	3175	1.68
p2	2945.0347	2883	2.15
s9234	5830.8660	5801	0.51
biomed	5688.0000	5659	0.51
s13207	8636.2416	8595	0.48
s15850	10,371.6418	10325	0.45

default settings except for the infeasibility tolerance, which is set to 10^{-3} (instead of the default 10^{-5}). Note that we use older benchmark problems because newer benchmark problems are too large to be handled by SDPLR.

3.2. *Bound quality*

Since (S-CHMB) is a relaxation of (Q-CHMB), its optimal value w_+^* provides an upper bound on the optimal value w^* of (Q-CHMB). Table 2 investigates the

Table 3. Comparison of SDP, METIS and best-known solutions for the digital circuit benchmark problems with 45–55% balance tolerance.

Name	SDP	METIS	Best-known	SDP time
balu	27	27	27	159
p1	47	47	47	114
bm1	50	48	47	151
t4	48	54	48	1248
t3	56	58	57	1141
t2	91	90	87	956
t6	60	63	60	60
struct	33	33	33	775
t5	75	72	71	75
19ks	106	107	104	3705
p2	144	145	139	2289
s9234	45	40	40	9149
biomed	93	83	83	20,558
s13207	59	53	53	25,455
s15850	44	42	42	44,654

closeness of w_+^* and w^*. The bound w_+^* is calculated using SDPLR. In lieu of a global optimization algorithm to compute w^*, however, we calculate a heuristic solution of (CHMB) using METIS. For the benchmark problems, Table 2 shows the problem name, the SDP bound w_+^*, the heuristic value (denoted *val*), as well as the relative gap between these two numbers, which is got by the formula $100\% \times (w_+^* - \text{val})/\text{val}$.

A couple of comments regarding the use of METIS for the heuristic values in Table 2 are in order. First, the METIS software unfortunately does not allow one to attain a true 50–50% partition of the nodes of the hypergraph. Instead, the closest one can guarantee is 49–51%; that is, METIS forces the user to permit a slight variation in the partition sizes. In this sense, the heuristic values in Table 2 may not be true lower bounds on w^*, but given the software limitations of METIS, we use them as under-approximations of w^*. Second, the reported heuristic value is actually the best obtained over 100 runs of METIS on each benchmark problem.

The tests show that (S-CHMB) provides very tight bounds on the benchmark problems since the relative gap is lower than a few percent on most of the problems. The average of the relative gaps is 1.97%. One outlier is t6, which has a relative gap of 7.99%.

3.3. *Heuristic solution quality*

In order to generate good solutions to CHMB and HMB, we combine two techniques to 'round' any SDP optimal solution of (S-CHMB) to feasible solutions for CHMB.

First, we use a modified version of the maximum cut approach of Goemans and Williamson [12]. Let (X, Y, Z) be an optimal solution of (S-CHMB). Since $X \succeq 0$, there exist polynomial-time computable vectors $\vec{x}_1, \ldots, \vec{x}_n$ such that $X_{jk} = \vec{x}_j^T \vec{x}_k$ for all pairs (j, k). This is the so-called Gram representation of X. Next, let v be a random sample from a multivariate-normal distribution $\mathcal{N}(0, I)$, where I is the identity matrix, and compute $v^T\vec{x}_j$ for all j. Set $\bar{x}_j = -1$ if $v^T\vec{x}_j$ is one of the $n/2$ smallest among $\{v^T\vec{x}_1, \ldots, v^T\vec{x}_n\}$ and $\bar{x}_j = 1$ otherwise. Finally, we assign vertices with the same $\bar{x}_j$ value to the same partition. This procedure guarantees a balanced partition because $\sum_{j=1}^{m} \bar{x}_j = 0$ and is similar to the technique that was proposed in [12], which assigns $\bar{x}_j$ according to the sign of $v^T\vec{x}_j$. It is common, in practice, to repeat this process many times to get the best $\bar{x}$ based on different random v (but the same $\vec{x}_j$), we repeat it $n+m$ times.

Then we apply one of the most basic greedy-search algorithms called *the FM method* [9] to refine the solutions that are generated by the above rounding technique.

In Table 3, we compare the rounded solutions of (S-CHMB) with METIS. Note that this table compares the values of HMB for a relaxed balance tolerance of 45–55%, which means S or S^c can have at most 55% of the vertices. We use 45–55% balance tolerance because it is commonly used in digital circuit partitioning to compare different methods [15,16].

The second column (SDP) of Table 3 is the number of cut hyperedges that was generated by solving (S-CHMB) and using the two-stage rounding technique described above. The third column (METIS) is the best results of running METIS 100 times with the default setting. The fourth column is the best-known solutions on these problems with 45–55% balance tolerance. The last column shows the time in seconds for SDPLR 1.02 to solve this problem.

Our approach found very good solutions for the benchmark problems and in some cases outperformed METIS with the default setting. Of course, the time for solving (S-CHMB) is a lot more than for METIS. For example, solving s15850 took SDPLR around 12 h but took less than 6 s for METIS.

4. Towards an approximation guarantee

As discussed in Section 1, the Goemans–Williamson procedure has been used to derive an approximation algorithm for CHMB when $\tau = 2$. Our motivation for this section is the following question: can the $\tau = 2$ approximation result be extended for general τ? Although we have not been able to resolve this question, in this section we report on what we feel is positive progress. The analysis relies on some interesting geometrical properties of the SDP (S-CHMB) established in Section 4.1.

4.1. *Some properties of the SDP relaxation*

Let M be a feasible solution of (S-CHMB). Because $M \succeq 0$ is a Gram matrix, there exist vectors $\vec{x}_1, \ldots, \vec{x}_n$ and $\vec{y}_1, \ldots, \vec{y}_m$ in $\Re^{n+m}$ such that

$$\begin{aligned} \langle \vec{x}_{j_1}, \vec{x}_{j_2} \rangle &= X_{j_1 j_2} & & j_1 = 1, \ldots, n,\ j_2 = 1, \ldots, n, \\ \langle \vec{x}_j, \vec{y}_i \rangle &= Z_{ij} & & j = 1, \ldots, n,\ i = 1, \ldots, m, \\ \langle \vec{y}_{i_1}, \vec{y}_{i_2} \rangle &= Y_{i_1 i_2} & & i_1 = 1, \ldots, m,\ i_2 = 1, \ldots, m. \end{aligned} \tag{2}$$

Throughout this section, we consider M, $\vec{x}_j$ and $\vec{y}_i$ to be related by the above equations.

The following proposition proves some basic geometric properties of $\{\vec{x}_j\}$ and $\{\vec{y}_i\}$.

Proposition 4.1 *Let M be a feasible solution of* (S-CHMB), *and let $\vec{x}_j$ $(j=1,\dots,n)$ and $\vec{y}_i$ $(i=1,\dots,m)$ be related to M according to* (2). *Then*:

(a) $\|\vec{x}_j\| = 1$ *for all j,*
(b) $\|\vec{y}_i\| \le 1$ *for all i and*
(c) *if $\|\vec{y}_i\| > 0$ for some i, then, for all $j \in E_i$, the angle between $\vec{x}_j$ and $\vec{y}_i$ is* $\arccos(\|\vec{y}_i\|)$.

Proof Part (a) follows from the constraint $e_j e_j^T \bullet X = X_{jj} = 1$ and $\langle \vec{x}_j, \vec{x}_j \rangle = \|\vec{x}_j\|^2 = X_{jj} = 1$. Now we show (b) and (c). For $j \in E_i$, let θ_{ij} denote the angle between $\vec{x}_j$ and $\vec{y}_i$. Then we have

$$0 = e_i e_i^T \bullet Y - e_i e_j \bullet Z = \|\vec{y}_i\|^2 - \langle \vec{y}_i, \vec{x}_j \rangle = \|\vec{y}_i\|^2 - \|\vec{y}_i\| \|\vec{x}_j\| \cos\theta_{ij},$$

which, if $\|\vec{y}_i\| > 0$, implies $\|\vec{y}_i\| = \cos\theta_{ij}$. ■

Proposition 4.1 is critical because, for non-zero $\vec{y}_i$, it shows that all $\vec{x}_j$ with $j \in E_i$ make the same acute angle with $\vec{y}_i$. This also provides an interesting geometric insight. For example, say that $E_1 = \{1, 2, 3, 4\}$ and $\|\vec{y}_1\| > 0$. Then Figure 2 represents how the $\vec{x}_j$ $(j=1,\dots,4)$ are positioned relative to $\vec{y}_1$. One can see that $\vec{x}_j$ appears on the boundary of a disc, which contains $\vec{y}_1$ as its centre.

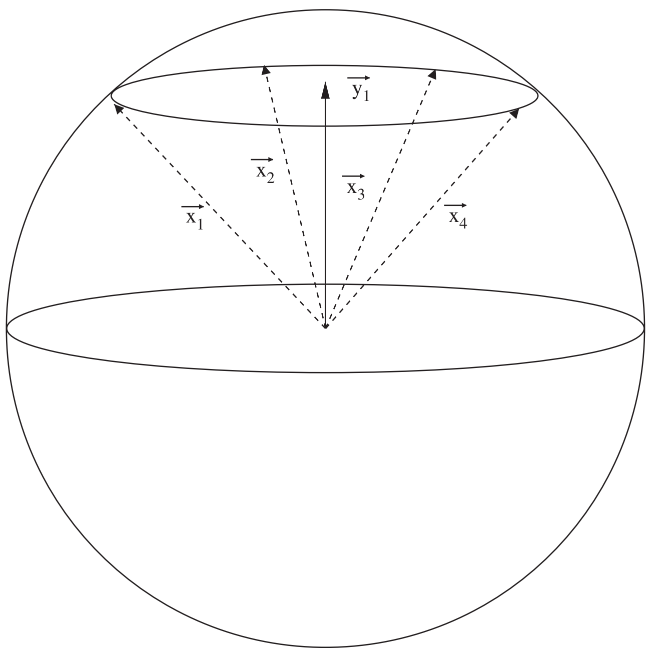

Figure 2. Geometric structure of a feasible solution of (S-CHMB) with respect to a single hyperedge with four vertices.

We formalize the above idea as follows. Given $u \in \Re^{n+m}$ such that $0 < \|u\| \leq 1$, define

$$\mathcal{D}(u) := \{p \in \Re^{m+n} : \|p\| \leq 1,\ (p-u)^T u = 0\}. \tag{3}$$

$\mathcal{D}(u)$ can be written equivalently as

$$\left\{p \in \Re^{m+n} : \|p-u\| \leq \sqrt{1-\|u\|^2},\ (p-u)^T u = 0\right\},$$

which emphasizes the role that u plays as the centre of the disc. Now we can give a geometric interpretation of the SDP relaxation of CHMB. Since the radius of $\mathcal{D}(\vec{y}_i)$ gets smaller as $\|\vec{y}_i\|$ increases, maximizing $w_i\|\vec{y}_i\|^2$ tends to gather $\vec{x}_j$'s close to each other for $j \in E_i$. The following corollary of Proposition 4.1 is straightforward.

Corollary 4.2 *Let $i \in \{1, \ldots, m\}$ and suppose $\|\vec{y}_i\| > 0$. Then $\vec{x}_j \in \mathcal{D}(\vec{y}_i)$ for all $j \in E_i$.*

The next lemma provides necessary and sufficient conditions for a given hyperplane $\mathcal{H}(h) := \{x : x^T h = 0\}$ to intersect a disc $\mathcal{D}(y)$. This lemma will be used later in Section 4.2.1.

Lemma 4.3 *Let $u, h \in \Re^{n+m}$ with $0 < \|y\| \leq 1$ and $\|h\| = 1$ be given, and consider the disc $\mathcal{D}(y)$ defined by (3). Then $h^T p \geq 0$ for all $p \in \mathcal{D}(y)$ if and only if $\cos\theta \geq \sqrt{1-\|y\|^2}$, where θ is the angle between y and h.*

Proof By a rigid motion of $\Re^{n+m}$, we may reduce to the case $y = (\lambda, 0, \ldots, 0)^T$, where $0 < \lambda \leq 1$. Then

$$D(y) = \{p \in \Re^{n+m} : p_1 = \lambda,\ p_2^2 + \cdots + p_{n+m}^2 \leq 1 - \lambda^2\},$$

and the statement of the lemma becomes: $h^T p \geq 0$ for all $p \in \mathcal{D}(y)$ if and only if $h_1 \geq \sqrt{1-\lambda^2}$.

We consider the optimization problem $\min\{h^T p : p \in \mathcal{D}(y)\}$. Because this is linear optimization over a convex set, the optimal solution will occur on the boundary of $D(y)$, and so it suffices to consider

$$\min\{h^T p : p_1 = \lambda,\ p_2^2 + \cdots + p_{n+m}^2 = 1 - \lambda^2\}. \tag{4}$$

A stationary point $\bar{p}$ of (4) is characterized by the existence of scalars a and b such that the gradient of the Lagrangian function

$$L_{a,b}(p) = h^T p + a(\lambda - p_1) + b(1 - \lambda^2 - p_2^2 - \cdots - p_{n+m}^2)$$

vanishes at $\bar{p}$, i.e.

$$\nabla L_{a,b}(\bar{p}) = h - ae_1 - 2b(0, \bar{p}_2, \ldots, \bar{p}_{n+m})^T = 0 \Longleftrightarrow h = (a, 2b\bar{p}_2, \ldots, 2b\bar{p}_n).$$

From this equality, we have $a = h_1$, and since $\|h\| = 1$ and $\bar{p}_2^2 + \cdots + \bar{p}_{n+m}^2 = 1 - \lambda^2$, b is determined as follows:

$$h_1^2 + \sum_{j=2}^{n+m}(2b\bar{p}_j)^2 = 1 \Longleftrightarrow b^2 = \frac{1-h_1^2}{4(1-\lambda^2)} \Longleftrightarrow b = \pm\frac{\sqrt{1-h_1^2}}{2\sqrt{1-\lambda^2}}.$$

Moreover, the objective value equals

$$h^T\bar{p} = h_1\lambda + \sum_{j=2}^{n+m} h_j\bar{p}_j = h_1\lambda + \sum_{j=2}^{n+m}(2b\bar{p}_j)\bar{p}_j = h_1\lambda + 2b\sum_{j=2}^{n+m}\bar{p}_j^2 = h_1\lambda + 2b(1-\lambda^2).$$

Hence, at any stationary point $\bar{p}$, there are only two possible values for $h^T\bar{p}$ based on the two possibilities for b. Because $1-\lambda^2 \geq 0$, the overall minimum value of (4) is thus

$$h_1\lambda - 2\left(\frac{\sqrt{1-h_1^2}}{2\sqrt{1-\lambda^2}}\right)(1-\lambda^2) = h_1\lambda - \sqrt{1-h_1^2}\sqrt{1-\lambda^2}.$$

What conditions guarantee that this minimum value is non-negative? Because $\lambda > 0$, a necessary condition is $h_1 \geq 0$. Another necessary condition is as follows:

$$h_1\lambda - \sqrt{1-h_1^2}\sqrt{1-\lambda^2} \geq 0 \Longrightarrow h_1^2 \geq 1-\lambda^2.$$

These two necessary conditions can be combined as $h_1 \geq \sqrt{1-\lambda^2}$. It is now not difficult to see that the condition $h_1 \geq \sqrt{1-\lambda^2}$ is also sufficient. ■

4.2. *A Goemans–Williamson procedure and analysis*

In Section 3, we used a variant of Goemans and Williamson's rounding technique for its simplicity. In this section, we apply the original version of Goemans and Williamson's rounding technique [13] for a more careful discussion and analysis.

Let (M, X, Y, Z) be an optimal solution of (S-CHMB) with Gram representation (2), and let h be a random vector generated uniformly over the surface of the unit sphere in $\Re^{n+m}$. Then define a random cut $\hat{x}$ in the hypergraph according to

$$\hat{x}_j := \text{sign}\langle \vec{x}_j, h\rangle \quad \forall j = 1, \ldots, n,$$

where 'sign' returns either -1 or $+1$ based on the sign of its argument. Note that, based on the random variable h, $\hat{x}$ is itself a random variable. This is the basic hyperplane rounding technique of Goemans and Williamson, but it may not produce a balanced cut in the hypergraph. To ensure this, one can, for example, greedily shift vertices from one side to the other.

This procedure is precisely the one used by Ye and Zhang [20] for their approximation analysis of CHMB when $\tau = 2$. Although their analysis contained several ingredients, one key component was estimating the probability that the i-th edge E_i is uncut by $\hat{x}$ (even before the greedy balancing procedure).

Although we do not provide the details here in the interest of space, we have been able to show that, when τ is arbitrary, if one can bound away from zero the probability that E_i is uncut, then Ye and Zhang's analysis goes through with little change, thus allowing an approximation algorithm for the general CHMB.

Unfortunately, we have not been able to prove this last ingredient, i.e. bounding the uncut probability away from 0. Nevertheless, we can provide a non-trivial lower bound on the probability as we describe next. (Note that this lower bound tends to 0 as $n + m \to \infty$, see Section 4.3).

4.2.1. *Lower bound*

A helpful geometrical interpretation of generating $\hat{x}$ and its associated cut (before greedy balancing) can be seen by defining the following two random half-spaces (the first closed, the second open):

$$\mathcal{H}^+(h) := \{v \in \Re^{n+m} : v^T h \geq 0\} \quad \text{and} \quad \mathcal{H}^-(h) := \{v \in \Re^{n+m} : v^T h < 0\}.$$

Then we have $\hat{x}_j = 1$ for all $j \in \mathcal{H}^+(h)$ and $\hat{x}_j = -1$ for all $j \in \mathcal{H}^-(h)$. Moreover, $\hat{x}_{j_1} = \cdots = \hat{x}_{j_\ell}$ (i.e. E_i is uncut) if and only if all $\vec{x}_{j_1}, \ldots, \vec{x}_{j_\ell}$ lie completely in one of the two half-spaces where $E_i = \{j_1, j_2, \ldots, j_l\}$.

Related to this geometrical interpretation, we introduce the following terminology: for any set $Q \subseteq \Re^{n+m}$, Q is said to be *uncut* by the hyperplane

$$\mathcal{H}(h) := \{v \in \Re^{n+m} : v^T h = 0\}$$

if $Q \subseteq \mathcal{H}^+(h)$ or $Q \subseteq \mathcal{H}^-(h)$; otherwise, Q is said to be *cut* by $\mathcal{H}(h)$. So E_i is uncut by the randomized procedure if and only if $\{\vec{x}_{j_1}, \ldots, \vec{x}_{j_\ell}\}$ is uncut by $\mathcal{H}(h)$. We say that E_i is cut or uncut by $\mathcal{H}(h)$ when referring to the cut or uncut status of $\{\vec{x}_{j_1}, \ldots, \vec{x}_{j_\ell}\}$.

We are now ready to prove the lower bound, whose proof is based on the geometry of the SDP relaxation – in particular, the discs $\mathcal{D}(\vec{y}_i)$ discussed in Section 4.1. Although this lower bound is not as strong as we would wish (in fact, it goes to 0 as $n+m \to \infty$ as shown in Section 4.3), we remain hopeful that this perspective of integrating the geometry of the SDP into the probability analysis may yet prove fruitful.

Proposition 4.4 *Consider the Goemans–Williamson randomized procedure for generating a cut with SDP solution* (M, X, Y, Z), *Gram matrix representation* (2) *and random vector h. Then, for all* $i = 1, \ldots, m$,

$$P[E_i \textit{ is uncut}] \geq 2P\left[\cos\theta_i \geq \sqrt{1 - Y_{ii}}\right],$$

where θ_i *is the angle between* $\vec{y}_i$ *and h.*

Proof If $Y_{ii} = 0$, then the inequality follows easily. So we may assume that $Y_{ii} > 0$. Let $\vec{y}_i \in \Re^{n+m}$ be related to Y according to (2); in particular, $\|\vec{y}_i\| = Y_{ii}^{1/2} > 0$. Corollary 4.2 thus implies $\{\vec{x}_{j_1}, \ldots, \vec{x}_{j_\ell}\} \subset \mathcal{D}(\vec{y}_i)$. Hence, the event that $\{\vec{x}_{j_1}, \ldots, \vec{x}_{j_\ell}\}$ is uncut contains the event that $\mathcal{D}(\vec{y}_i)$ is uncut. So

$$\begin{aligned} P[E_i \text{ is uncut}] = P[\hat{x}_{j_1} = \cdots = \hat{x}_{j_\ell}] &= P\big[\{\vec{x}_{j_1}, \ldots, \vec{x}_{j_\ell}\} \text{ is uncut by } \mathcal{H}(h)\big] \\ &\geq P\big[\mathcal{D}(\vec{y}_i) \text{ is uncut by } \mathcal{H}(h)\big] \\ &= P\big[\mathcal{D}(\vec{y}_i) \subset \mathcal{H}^+(h) \text{ or } \mathcal{D}(\vec{y}_i) \subset \mathcal{H}^-(h)\big] \\ &= 2P\big[\mathcal{D}(\vec{y}_i) \subset \mathcal{H}^+(h)\big], \end{aligned}$$

where the last equality follows by symmetry. Lemma 4.3 then yields the desired result. ∎

4.3. *Limit of the lower bound*

We now show that the lower bound established in Proposition 4.4 necessarily tends to 0 with order $\mathcal{O}((n+m)^{-2})$.

The following standard proposition gives a concrete way to generate h in the Goemans–Williamson procedure of Section 4.2.

Lemma 4.5 *Suppose $g \in \Re^{n+m}$ follows the multivariate normal distribution $\mathcal{N}(0, I)$. Then $h := g/\|g\|$ is uniformly distributed on the surface of the unit sphere in $\mathcal{R}^{n+m}$.*

This lemma allows us to estimate the quantity $P\left[\cos\theta_i \geq \sqrt{1 - Y_{ii}}\right]$ in Proposition 4.4 as $n+m$ gets larger and larger.

Proposition 4.6 *Consider the context of Proposition 4.4 for various, increasing values of $n+m$. It holds that*

$$\lim_{n+m\to\infty} P\left[\cos\theta_i \geq \sqrt{1 - Y_{ii}}\right] = 0$$

irrespective of Y_{ii} and $\vec{y}_i$.

Proof If $Y_{ii}=0$, then the proposition is clear. Assuming that $Y_{ii} > 0$, we first use Lemma 4.5 to derive a more concrete expression for $P\left[\cos\theta_i \geq \sqrt{1 - Y_{ii}}\right]$.

By a rigid motion of $\Re^{n+m}$, we may assume without loss of generality that $\vec{y}_i$ points in the direction e_1. Then, since $Y_{ii} = \|\vec{y}_i\|^2$ by (2), we have, in fact, $\vec{y}_i = \sqrt{Y_{ii}}e_1$. Then $\cos\theta_i = h^T\vec{y}_i/\|h\|\|\vec{y}_i\| = h_1 = g_1/\|g\|$ by Lemma 4.5. Thus,

$$\begin{aligned}
P\left[\cos\theta_i \geq \sqrt{1 - Y_{ii}}\right] &= P\left[\frac{g_1^2}{\|g\|^2} \geq 1 - Y_{ii}\right] \\
&= P\left[g_1^2 \geq \left(g_1^2 + g_2^2 + \cdots + g_{m+n}^2\right)(1 - Y_{ii})\right] \\
&= P\left[Y_{ii}g_1^2 \geq \left(g_2^2 + \cdots + g_{m+n}^2\right)(1 - Y_{ii})\right] \\
&= P\left[\frac{g_1^2}{g_2^2 + \cdots + g_{m+n}^2} \geq \frac{1 - Y_{ii}}{Y_{ii}}\right] \\
&= P\left[(n+m-1)\cdot\frac{g_1^2}{g_2^2 + \cdots + g_{m+n}^2} \geq (n+m-1)\cdot\frac{1 - Y_{ii}}{Y_{ii}}\right] \\
&= P\left[V \geq (n+m-1)\cdot\frac{1 - Y_{ii}}{Y_{ii}}\right],
\end{aligned}$$

where

$$V := (n+m-1)\cdot\frac{g_1^2}{g_2^2 + \cdots + g_{m+n}^2}$$

follows the F distribution having degrees of freedom $(1, n+m-1)$, expected value $E[V] = (n+m-1)/(n+m-3)$ and variance $\mathrm{Var}(V) = (2(n+m-1)^2(n+m-2))/((n+m-3)^2(n+m-5))$.

We now use the one-tailed Chebyshev inequality to prove that $P[V \geq (n+m-1)\kappa]$ goes to zero for any $\kappa \geq 0$ (in our case, $\kappa = (1 - Y_{ii})/Y_{ii}$). We have the following upper bound:

$$\begin{aligned}
P[V \geq (n+m-1)\kappa] &= P[V - E(V) \geq (n+m-1)\kappa - E(V)] \\
&\leq \frac{\mathrm{Var}(V)}{\mathrm{Var}(V) + ((n+m-1)\kappa - E(V))^2}.
\end{aligned}$$

As $n+m$ goes to infinity, $E(V)$ and $\mathrm{Var}(V)$ converge to 1 and 2, respectively, and so the above probability goes to 0 with order $\mathcal{O}((n+m)^{-2})$. ■

Although Proposition 4.6 makes sense from a theoretical point of view, it is inconsistent with our computational results in which our method seems to work well on large-scale circuit partitioning problems.

5. Conclusion and remarks

One of the major contributions of this article is the presentation of a mathematical model for CHMB. While heuristic methods for HMB flourish, there has been little advance in mathematical approaches and theoretical results for HMB. This is mainly due to the fact that it was not clear how to pose HMB as an equivalent mathematical problem. We believe that having a mathematical model is a good starting point for studying the hypergraph minimum-bisection problem. For most of the heuristics, it might be difficult to come up with modifications that can make them perform better on general problems. On the other hand, it is quite possible that some improvements in our method may lead to an overall performance improvement. Tightening SDP relaxations for QCQPs is one of the most intensively studied topics in optimization and we are hoping to achieve a tighter relaxation of our model in the near future. The gap between METIS and our method is slim. We hope to improve our method and close the gap or even overcome it in the near future.

Although we succeeded in constructing a mathematical model for CHMB and achieving competitive computational results, our research on the theoretical aspects of our method is incomplete. Based on our computational results, we were hoping to achieve an approximation guarantee on the hypergraph minimum-bisection problem. However, a modified application of well-known techniques for similar problems does not seem to achieve this goal. This theoretical result is somewhat disappointing since we observe that our method works well computationally on relatively large problems. It is one of our future challenges.

Finally, another possible future improvement can be made in the software SDPLR. While SDPLR has the advantage of being able to solve very large-scale SDP problems, we realized that it can be improved even further in terms of running time, which is critical in solving the large-scale problems such as the digital circuit minimum-bisection problem.

Acknowledgement

Research by S. Burer was supported in part by NSF Grants CCR-0203426 and CCF-0545514.

References

[1] C.J. Alpert, *The ISPD98 circuit benchmark suite*, Proceedings of the 1998 International Symposium on Physical Design, 1998, pp. 80–85.

[2] C.J. Alpert and A.B. Kahng, *Recent directions in netlist partitioning: A survey*, PhD thesis, Computer Science Department, UCLA, Los Angeles, 1995.

[3] S. Arora, S. Rao, and U. Vazirani, *Expander flows, geometric embeddings and graph partitioning*, Proceedings of the 36th Annual ACM Symposium on Theory of Computing, 2004, pp. 222–231.

[4] P. Berman and M. Karpinski, *Approximability of hypergraph minimum bisection*, Technical Report, 2003. Available at http://www.maths.ox.ac.uk/rand-apx/, rAND-APX Thematic Network.

[5] S. Burer and C. Choi, *Computational enhancements in low-rank semidefinite programming*, Optim. Methods Software 21 (2006), pp. 493–512.

[6] S. Burer and R. Monteiro, *A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization*, Math. Program. (Series B) 95 (2003), pp. 329–357.

[7] C. Choi and Y. Ye, *Application of semidefinite programming to circuit partitioning*, in *Approximation and Complexity in Numerical Optimization: Continuous and Discrete Problems*, P. Pardalos, ed., Kluwer Academic Press, Dordrecht, 2000, pp. 130–136.

[8] U. Feige and R. Krauthgamer, *A polylogarithmic approximation of the minimum bisection*, SIAM J. Comput. 31 (2002), pp. 1090–1118.

[9] C.M. Fiduccia and R.M. Mattheyses, *A linear time heuristic for improving network partitions*, Proceedings of the 19th IEEE Design Automation Conference, 1982, pp. 175–181.

[10] A. Frieze and M. Jerrum, *Improved approximation algorithms for max k-cut and max bisection*, Algorithmica 18 (1997), pp. 61–77.

[11] M.R. Garey, D.S. Johnson, and L. Stockmeyer, *Some simplified NP-complete graph problems*, Theor. Comput. Sci. 1 (1976), pp. 237–267.

[12] M. Goemans and D. Williamson, *878-Approximation algorithms for MAX CUT and MAX 2SAT*, Proceedings of the Symposium of Theoretical Computer Science, 1994, pp. 422–431.

[13] M. Goemans and D. Williamson, *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*, J. ACM 42 (1995), pp. 1115–1145.

[14] E. Ihler, D. Wagner, and F. Wagner, *Modeling hypergraphs by graphs with the same mincut properties*, Inform. Process. Lett. 45 (1993), pp. 171–175.

[15] G. Karypis, R. Aggarwal, V. Kumar, and S. Shekhar, *Multilevel hypergraph partitioning: Applications in VLSI domain*, IEEE Trans. VLSI Syst. 7 (1999), pp. 69–79.

[16] G. Karypis and V. Kumar, *Metis: A software package for partitioning unstructured graphs, partitioning meshes, and computing fill-reducing orderings of sparse matrices, version 4.0*, Technical Report, Department of Computer Science, University of Minnesota, Minneapolis, MN 55455, USA, 1998.

[17] G. Karypis and V. Kumar, *Multilevel k-way hypergraph partitioning*, Proceedings of the 36th ACM Design Automation Conference, 1999, pp. 343–348.

[18] Y.G. Saab, *A fast network bisection algorithm*, IEEE Trans. Comput. 44 (1995), pp. 903–913.

[19] Y. Ye, *A .699-Approximation algorithm for max-bisection*, Math. Program. 90 (2001), pp. 101–111.

[20] Y. Ye and J. Zhang, *Approximation of dense-n/2-subgraph and the complement of minbisection*, J. Global Optim. 25 (2003), pp. 55–73.