

Multiway Cut, Pairwise Realizable Distributions, and Descending Thresholds

[Extended Abstract] *

Ankit Sharma
Carnegie Mellon University
Pittsburgh, PA
ankits@cs.cmu.edu

Jan Vondrák
IBM Almaden
San Jose, CA
jvondrak@us.ibm.com

ABSTRACT

We design new approximation algorithms for the Multiway Cut problem, improving the previously known factor of 1.32388 [Buchbinder et al., 2013].

We proceed in three steps. First, we analyze the rounding scheme of Buchbinder et al. [2013] and design a modification that improves the approximation to $\frac{3+\sqrt{5}}{4} \approx 1.309017$. We also present a tight example showing that this is the best approximation one can achieve with the type of cuts considered by Buchbinder et al. [2013]: (1) partitioning by exponential clocks, and (2) single-coordinate cuts with equal thresholds.

Then, we prove that this factor can be improved by introducing a new rounding scheme: (3) single-coordinate cuts with descending thresholds. By combining these three schemes, we design an algorithm that achieves a factor of $\frac{10+4\sqrt{3}}{13} \approx 1.30217$. This is the best approximation factor that we are able to verify by hand.

Finally, we show that by combining these three rounding schemes with the scheme of independent thresholds from Karger et al. [2004], the approximation factor can be further improved to 1.2965. This approximation factor has been verified only by computer.

Categories and Subject Descriptors

F.2 [Theory of Computation]: Analysis of Algorithms and Problem Complexity

General Terms

Approximation Algorithms

Keywords

Multiway Cut, Approximation Algorithm

*A full version of this paper is available at <http://arxiv.org/abs/1309.2729>.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

STOC '14 New York, NY USA

Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...\$15.00.

1. INTRODUCTION

The Multiway Cut problem is one of the classical graph optimization problems: Given a graph $G = (V, E)$ with edge weights $w : E \rightarrow \mathbb{R}_+$ and k terminals $t_1, t_2, \dots, t_k \in V$, we want to find a minimum-weight subset of edges $F \subseteq E$ such that no pair of terminals is connected in $(V, E \setminus F)$. Equivalently, we can search for a labeling of the vertices $\ell : V \rightarrow [k]$ so as to minimize the total weight of edges (v, w) such that $\ell(v) \neq \ell(w)$. This is a natural generalization of the Minimum s - t -Cut problem, which is the $k = 2$ case.

The study of Multiway Cut goes back to Dahlhaus et al. [1994] who proved that the problem is MAX SNP-hard for every $k \geq 3$, and gave a simple combinatorial algorithm using repeated applications of Min s - t -Cut that achieves a $(2 - 2/k)$ -approximation. A novel technique for Multiway Cut was introduced by Calinescu et al. [2000] who proposed a geometric relaxation for this problem. In this “CKR relaxation”, the graph is embedded into a simplex with each terminal at a distinct vertex, and finding the optimal embedding can be formulated as a linear program. A partitioning of the graph then corresponds to a partitioning of the simplex that separates all the vertices. Using this relaxation, Calinescu et al. [2000] designed a $(1.5 - 1/k)$ -approximation for Multiway Cut.

Cunningham and Tang [1999], and Karger et al. [2004] independently provided a $12/11$ -approximation for $k = 3$ and showed that this is the best approximation achievable using the CKR relaxation for $k = 3$, by presenting a matching integrality gap example. More generally, Karger et al. [2004] provided improved approximation algorithms for all values of k , with approximation factors tending to 1.3438 as $k \rightarrow \infty$. On the negative side, Freund and Karloff [2000] showed an integrality gap of $8/(7+1/(k-1))$ for each $k \geq 3$.

The importance of the CKR relaxation was bolstered further by the work of Manokaran et al. [2008] who proved that assuming the Unique Games Conjecture, it is NP-hard to achieve an approximation for Multiway Cut better than the integrality gap of the CKR relaxation (for any fixed k). This means that $12/11$ is indeed the best possible approximation for $k = 3$, and for every k the CKR relaxation provides the optimal approximation factor (assuming the UGC).

Recently, Buchbinder et al. [2013] made a new improvement on the algorithmic side and designed a 1.32388 approximation for Multiway Cut (for arbitrary k). They introduced an interesting new rounding scheme for the CKR relaxation that they called “partitioning using exponential clocks”. (In fact, they also showed that a threshold-based

rounding scheme from Kleinberg and Tardos [2002] could be used equivalently in place of the exponential clocks.) Combining this scheme with a modification of a thresholding scheme from Calinescu et al. [2000], they presented a very simple and elegant way to achieve a $4/3$ -approximation. Then they modified the rounding scheme further, to improve the approximation factor to 1.32388.

Our contribution. We build upon previous work and provide further improvements on the approximation factor for Multiway Cut. First, we study the rounding scheme of Buchbinder et al. [2013] and identify a modification of their scheme that leads to a factor of $\frac{3+\sqrt{5}}{4} \approx 1.309017$. (See Section 3.) We note that this number is equal to $\frac{1+\varphi}{2}$ where $\varphi = \frac{1+\sqrt{5}}{2}$ is the *golden ratio*. We also present a tight example showing that this is the best approximation factor that can be achieved by any combination of the techniques considered by Buchbinder et al. [2013]: the exponential clocks scheme, the Kleinberg-Tardos scheme, and thresholding schemes with equal thresholds for all terminals. (We provide more details in Appendix B.)

Secondly, we improve this approximation factor by introducing a new rounding scheme that we call *descending thresholds*. This scheme can be combined with the rounding schemes above in a way that achieves an approximation factor of $\frac{10+4\sqrt{3}}{13} \approx 1.30217$. The analysis of this algorithm is still quite simple and can be verified easily by hand (see Section 4). This factor is tight for any combination of partitioning using exponential clocks and thresholding schemes with “analysis based on two thresholds” (see Appendix C for details).

Finally, we show that this factor can be further improved by including another rounding scheme, the scheme of *independent thresholds* from Karger et al. [2004]. However, in this case we are not able to analyze the approximation factor manually anymore. With the help of an LP solver, we find a set of parameters that leads to an approximation factor of 1.2965 (see Section 5). The verification of this result reduces to finding the maximum of a certain function of 2 variables (involving polynomials and exponentials), which we have done by computer.¹

Pairwise realizable distributions. While searching for possible extensions of the rounding schemes with a random threshold for each terminal (see Section 4.1 for more details), we encountered the following question:

Given a joint distribution ρ of two random variables (X, Y) , can we design arbitrarily many random variables X_1, \dots, X_k such that $\forall i \neq j, (X_i, X_j)$ has the same distribution ρ ?

If this is the case, we call such a distribution ρ *pairwise realizable*. Not every distribution ρ is pairwise realizable (see Appendix A). Here we present the following answer (in discrete domains): ρ is pairwise realizable, if and only if ρ is a convex combination of symmetric product distributions, or in other words

$$\Pr_{(X,Y) \sim \rho} [X = a, Y = b] = \sum_s \alpha_s p_s(a) p_s(b)$$

where $\alpha_s \geq 0, \sum_s \alpha_s = 1$ and $\forall s; p_s(a) \geq 0, \sum_a p_s(a) = 1$. We provide a short proof in Appendix A; this result also follows from Trotter and Winkler [1998]. In particular, it

¹We have used *IBM ILOG CPLEX* for linear programming and *Wolfram Mathematica* for analytical manipulations.

is necessary that the matrix $P_{ab} = \Pr[X = a, Y = b]$ be *positive semidefinite*. We do not need this result directly for any of our algorithms, but this characterization is helpful in understanding what kinds of threshold distributions are worth considering. (See Section 4.1 and Appendix A for more details.)

Discussion. We have investigated several rounding schemes that might improve the approximation factor. While we have a good understanding of thresholding schemes that rely only on 2 relevant variables in the analysis (and we identify the best approximation factor in this setting - see Section C), the situation gets more complicated with the inclusion of additional variables as in the analysis of *independent thresholds*. Then the cut density is not a linear function of the underlying probability distributions anymore. Our approach in this case is a combination of intuition from tight examples and the use of an LP solver. The rounding scheme achieving a factor of 1.2965 that we present in this paper is the best one that we are able to describe in a simple form. Further (small) improvements might be achieved by finding more exhaustive descriptions of the probability distributions returned by the LP solver. However, we do not think that this would improve our understanding of the Multiway Cut problem.

It is interesting to note that as of now, all known approximation algorithms for Multiway Cut can be implemented using a sequence of label assignments based on a threshold condition ($x_{v,i} \geq \theta$). The only exception to our knowledge, the exponential clocks scheme of Buchbinder et al. [2013], can be replaced by the Kleinberg-Tardos algorithm, which uses a sequence of thresholds (with repeated use of variables). In fact Karger et al. [2004], citing computational experiments, speculated that the optimal approximation for Multiway Cut might be achievable using “sparcs”, which are sequences of k threshold cuts, one for each variable. Our scheme of descending thresholds is of this type. However, the exponential clocks scheme as well as the Kleinberg-Tardos scheme, one of which is still a necessary ingredient in our algorithm, are outside of this framework.

2. THE CKR RELAXATION

Calinescu et al. [2000] proposed the following LP relaxation of the Multiway Cut problem, where V is the set of vertices, E is the set of edges with weights $w_{v,v'}$, and T denotes the set of terminals, $|T| = k$.

$$\min \frac{1}{2} \sum_{(v,v') \in E} w_{v,v'} \|\mathbf{x}_v - \mathbf{x}_{v'}\|_1 : \quad (1)$$

$$\forall v \in V, \|\mathbf{x}_v\|_1 = 1, \quad (2)$$

$$\forall t \in T, \mathbf{x}_t = \mathbf{1}_t, \quad (3)$$

$$\forall v \in V, \mathbf{x}_v \geq \mathbf{0}. \quad (4)$$

Here, $\mathbf{x}_v \in \mathbb{R}^k$ for each vertex $v \in V$, and $\mathbf{1}_t$ denotes the unit basis vector corresponding to terminal $t \in [k]$. The fractional solution can be viewed as an embedding of the graph in the unit simplex $\Delta = \{\mathbf{x} \in \mathbb{R}^k : \mathbf{x} \geq 0, \|\mathbf{x}\|_1 = 1\}$, with terminal at the vertices of Δ . Given a fractional solution, the objective of any rounding scheme is to assign each of the vertices to one of the terminals without increasing the objective value $\frac{1}{2} \sum_{(v,v') \in E} w_{v,v'} \|\mathbf{x}_v - \mathbf{x}_{v'}\|_1$ by much. For a rounding scheme to achieve an α approximation to the LP ($\alpha \geq 1$), it suffices to show that for every edge $(v, v') \in E$,

the probability that the edge is “cut” by the rounding scheme (that is v and v' are assigned to different terminals) is at most $\alpha \cdot \frac{1}{2} \|\mathbf{x}_v - \mathbf{x}_{v'}\|_1$. We call $\frac{1}{2} \|\mathbf{x}_v - \mathbf{x}_{v'}\|_1$ the *length* of the edge (v, v') .

Moreover, as has been shown by Calinescu et al. [2000], it suffices to consider the case where for each edge $(v, v') \in E$, the two end-points v and v' are mapped so that their corresponding vectors differ in only two coordinates. In other words, we can assume that $\mathbf{x}_v = (u_1, u_2, \dots, u_k)$ and $\mathbf{x}_{v'} = (u_1, u_2, \dots, u_i + \epsilon, \dots, u_j - \epsilon, \dots, u_k)$, where i and j are the two coordinates where the two vectors differ. Also, note that in this case, $\frac{1}{2} \|\mathbf{x}_v - \mathbf{x}_{v'}\|_1 = \epsilon$. The probability of cutting such an edge should be at most $\alpha\epsilon$. In fact, ϵ can be made arbitrarily small, by subdividing edges. Dividing the cut probability by the length of the edge and letting $\epsilon \rightarrow 0$, we obtain the notion of *cut density*.

DEFINITION 2.1. *A randomized rounding scheme is a probability distribution \mathcal{R} over labelings $\ell: \Delta \rightarrow [k]$. An edge of type (i, j) is an edge (v, v') where \mathbf{x}_v and $\mathbf{x}_{v'}$ differ only in coordinates i, j . For a randomized rounding scheme \mathcal{R} , the cut density for edges of type (i, j) at $\mathbf{x} \in \Delta$ is*

$$d_{ij}(\mathbf{x}) = \limsup_{\epsilon \rightarrow 0} \frac{\Pr_{\ell \sim \mathcal{R}}[\ell(\mathbf{x}) \neq \ell(\mathbf{x} + \epsilon \mathbf{1}_i - \epsilon \mathbf{1}_j)]}{\epsilon}.$$

As shown in Karger et al. [2004], to achieve an approximation factor of α for Multiway Cut it is sufficient to demonstrate a rounding scheme such that $d_{ij}(\mathbf{x}) \leq \alpha$ for all $i, j \in [k]$ and $\mathbf{x} \in \Delta$.

3. EXPONENTIAL CLOCKS & SINGLE THRESHOLD:

1.309017-APPROXIMATION

We begin with a rounding scheme based on the techniques of Buchbinder et al. [2013] that achieves a $(3 + \sqrt{5})/4 \approx 1.309017$ -approximation for the Multiway Cut problem. Buchbinder et al. [2013] use a combination of two rounding schemes, the “exponential clocks rounding scheme” and what we call the “single-threshold rounding scheme”. The two schemes are described in detail below.

Algorithm 1

With probability p , choose the Exponential Clocks Rounding Scheme (Algorithm 2).
 With probability $1 - p$, choose the Single Threshold Rounding Scheme (Algorithm 3).

Algorithm 2 Exponential Clocks Rounding Scheme

1: Choose independent random variables Z_i from the exponential distribution for $i = 1, \dots, k$.
 2: For each vertex v , assign v to $\arg \min_{i \in [k]} Z_i / x_{v,i}$.

Algorithm 3 Single Threshold Rounding Scheme

1: Choose a threshold $\theta \in (0, 1]$ with probability density $\phi(\theta)$.
 2: Choose a random permutation σ of the terminals.
 3: **for all** i in $[k - 1]$ **do**
 4: For every vertex $v = (v_1, \dots, v_k)$ such that v has not been assigned yet,
 5: assign v to terminal $\sigma(i)$ if $x_{v, \sigma(i)} \geq \theta$.
 6: **end for**
 7: Assign all remaining unassigned vertices to terminal $\sigma(k)$.

Algorithm 2 can be viewed as selecting a point \mathbf{z} uniformly in the simplex Δ (by taking $\mathbf{z} = \frac{(Z_1, Z_2, \dots, Z_k)}{Z_1 + Z_2 + \dots + Z_k}$) and then partitioning the simplex into k regions that meet at the point \mathbf{z} . Buchbinder et al. [2013] also observe that a rounding scheme from Kleinberg and Tardos [2002] can be used in place of Algorithm 2 and leads to the same cut density of edges. In Algorithm 3, a threshold θ is chosen from some distribution and the simplex is partitioned by going over a random permutation of the terminals and assigning all currently unassigned points \mathbf{x} such that $x_{\sigma(i)} \geq \theta$ to terminal $\sigma(i)$.

Let us recall the bounds on cut density for Algorithms 2 and 3 from Buchbinder et al. [2013]. By symmetry, we focus in the following on edges of type $(1, 2)$.

LEMMA 3.1 (BUCHBINDER ET AL. [2013]). *The cut density for edges of type $(1, 2)$ under Algorithm 2 is*

$$d_{12}(u_1, u_2, \dots) = 2 - u_1 - u_2.$$

The cut density for edges of type $(1, 2)$ under Algorithm 3, assuming $u_1 \leq u_2$, is

$$d_{12}(u_1, u_2, \dots) = \frac{1}{2} \phi(u_1) + \phi(u_2).$$

In order to balance Algorithm 2 and Algorithm 3, Buchbinder et al. [2013] define the probability distribution $\phi(u)$ as a certain power of u . Our rounding scheme differs in the choice of this probability distribution. (In fact we claim that we have identified the best possible distribution for this purpose — see Section B). We prove the following.

THEOREM 3.2. *Algorithm 1, with $p = \frac{5+3\sqrt{5}}{20} \approx 0.58541$ and probability density function (for Algorithm 3)*

$$\phi(u) = \begin{cases} a u & \text{for } 0 \leq u \leq b \\ \frac{a}{2} (u + b) & \text{for } b \leq u \leq 1 \end{cases} \quad (5)$$

where $a = \frac{4+2\sqrt{5}}{3}$ and $b = \sqrt{5} - 2$, achieves a $\frac{3+\sqrt{5}}{4} \approx 1.309017$ -approximation for Multiway Cut.

The intuition behind this construction is as follows. Assume that $u_1 < u_2$. Considering Lemma 3.1, we would ideally like to design ϕ so that $\frac{1}{2} \phi(u_1) + \phi(u_2) = c(u_1 + u_2) + d$ for some constants c, d , in order to combine it with the cut density of $2 - u_1 - u_2$ for Algorithm 2. Unfortunately, it is impossible to design ϕ in such a way: If such a function ϕ existed, it would have to satisfy $\frac{1}{2} \phi'(u_1) = \phi'(u_2)$ for every pair of values $u_1 < u_2$ which is not possible. Still, we can try to satisfy this property for many pairs, and our construction (Figure 1) achieves this for all pairs such that $u_1 < b < u_2$. This means that our analysis is going to be tight for all such (u_1, u_2) pairs.

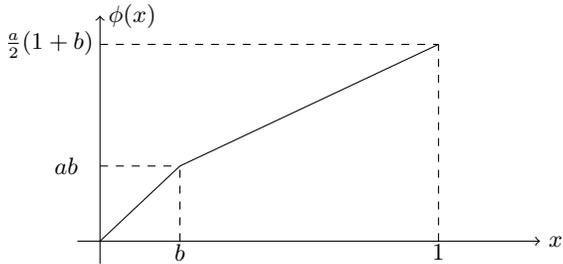


Figure 1: The probability density function $\phi(x)$.

Now we calculate the optimal values of a and b . A constraint on a and b is that since $\int_0^1 \phi(u)du = 1$, we should have $\frac{1}{2}ab^2 + \frac{a}{2}(\frac{1}{2}(1+b) + b)(1-b) = \frac{1}{4}a(-b^2 + 2b + 1) = 1$ ($a \geq 0$ and $b \in [0, 1]$). The following lemma gives the total cut density under Algorithm 1.

LEMMA 3.3. *The cut density for an edge of type $(1, 2)$ at (u_1, u_2, \dots) , $u_1 \leq u_2$ under Algorithm 1 is at most*

$$p \cdot (2 - u_1 - u_2) + (1 - p) \cdot \frac{a}{2}(u_1 + u_2 + b).$$

Proof: Observe that for any $a, b \geq 0$, the density function $\phi(u)$ can be written equivalently as $\phi(u) = \min\{au, \frac{a}{2}(u + b)\}$. Plugging this expression into Lemma 3.1, the cut density under Algorithm 3 is at most

$$\frac{1}{2}\phi(u_1) + \phi(u_2) \leq \frac{1}{2}au_1 + \frac{a}{2}(u_2 + b) = \frac{a}{2}(u_1 + u_2 + b).$$

Since we take Algorithm 2 with probability p and Algorithm 3 with probability $1 - p$, the lemma follows. \square

Hence, we can upper-bound the cut density under Algorithm 1 by $p(2 - u_1 - u_2) + (1 - p)\frac{a}{2}(u_1 + u_2 + b)$. To eliminate the dependence on $u_1 + u_2$, we set $p = (1 - p)\frac{a}{2}$, which means $p = a/(2 + a)$. This makes the bound equal to $2p + (1 - p)\frac{a}{2}b = 2p + pb = (2 + b)a/(2 + a)$. Hence, the final expression that we would like to minimize is $(2 + b)a/(2 + a)$ subject to the constraint $\frac{1}{4}a(-b^2 + 2b + 1) = 1$. The minimum is achieved at $a = \frac{2}{3}(2 + \sqrt{5})$ and $b = \sqrt{5} - 2$, where the bound on cut density is $(2 + b)a/(2 + a) = \frac{1}{4}(3 + \sqrt{5})$. The probability p in Algorithm 1 is $p = a/(2 + a) = \frac{1}{20}(5 + 3\sqrt{5})$. This proves Theorem 3.2.

4. DESCENDING THRESHOLDS:

1.30217-APPROXIMATION

As we show in Appendix B, the Exponential Clocks Rounding Scheme combined with the Single Threshold Rounding Scheme (under any threshold distribution) achieves exactly the factor of $\frac{3 + \sqrt{5}}{4}$ and not better. In this section, we present an improved $\frac{10 + 4\sqrt{3}}{13} \approx 1.30217$ -approximation for Multiway Cut. This is achieved by combining the techniques of Buchbinder et al. [2013] with a new rounding scheme that we call *descending thresholds*. Before we describe our algorithm, let us discuss the ideas and considerations that led us to this rounding scheme.

4.1 Pairwise realizable distributions

It is natural to consider different thresholds for different terminals, but the space of possibilities (all joint probability distributions of $(\theta_1, \theta_2, \dots, \theta_k)$) seems too vast to explore.

However, let us make the following observation: for an edge of type (i, j) , only the thresholds θ_i, θ_j corresponding to terminals i, j can potentially cut this edge. Therefore, the cut density for edges of type (i, j) is primarily determined by the distribution of the thresholds θ_i, θ_j . Assuming that threshold θ_i is applied before threshold θ_j and the joint distribution of θ_i, θ_j is given by a density function $\pi(\theta_i, \theta_j)$, we can write the following bound on the cut density of an edge of type i, j located at \mathbf{u} :

$$\int_0^1 \pi(u_i, u)du + \int_{u_i}^1 \pi(u, u_j)du.$$

Note the asymmetry here: the edge is always cut in coordinate u_i if θ_i cuts it, but it is cut in coordinate u_j only if θ_j cuts it and the edge was not captured by terminal i before.

Eventually, we apply a sequence of cuts to a permutation of terminals, either an independently random one, or one correlated with the values of the thresholds in some way. Given such a rounding scheme, we can compute a bound on the cut density as above for edges of all types. Since the bound is linear as a function of π , one can formulate a linear program that searches for the best distribution π . As shown in Karger et al. [2004], the symmetry of the terminals implies that whatever rounding scheme we have, we can apply it after a random re-labeling of the terminals. Therefore, we can assume that overall, every pair of thresholds (θ_i, θ_j) has the same joint distribution, which is a symmetrization of the distribution π above: $\rho(\theta_i, \theta_j) = \frac{1}{2}(\pi(\theta_i, \theta_j) + \pi(\theta_j, \theta_i))$. A basic question that we encountered here is: What distributions ρ are actually realizable for all pairs of terminals at the same time? In Appendix A, we provide the following characterization, at least in a discrete setting: A distribution ρ can be realized for all pairs of terminals simultaneously, if and only if ρ is a convex combination of symmetric product distributions, in other words ρ is in the form

$$P_{ab} = \Pr[X = a, Y = b] = \sum_s \alpha_s p_s(a) p_s(b)$$

where $\alpha_s, p_s(a) \geq 0$ and $\sum_s \alpha_s = 1, \sum_a p_s(a) = 1$ (see Theorem A.2). In particular, this implies that the matrix $P_{ab} = \Pr[X = a, Y = b]$ must be *positive semidefinite*.

Going back to our original motivation, this characterization gives us a hint as to what kinds of distributions over thresholds are worth considering. It is sufficient to consider a combination of rounding schemes, where the thresholds in each scheme are chosen independently from a certain distribution, and then applied in a certain order (which is possibly correlated with their values; *this is an aspect independent of the notion of pairwise realizability*). Unfortunately, we do not know how to search efficiently over the space of all such distributions. As far as we know, the condition of being pairwise-realizable is not equivalent to P_{ab} being positive semidefinite. However, when we computed the best distribution π^* to be combined with the Exponential Clocks Rounding Scheme, without the restriction of being pairwise-realizable, the symmetrization of this distribution $\rho^* = \frac{1}{2}(\pi^* + \pi^{*T})$ just happened to be in the pairwise realizable form. Furthermore, we identified this optimal solution π^* as a convex combination of two rounding schemes, the Single Threshold Rounding Scheme under a certain distribution ϕ , and a “Descending Thresholds Rounding Scheme” under a certain distribution ψ — we describe this scheme in the following section.

4.2 Descending thresholds

Based on the discussion above and our computational experiments, we propose the following new rounding scheme.

Algorithm 4 Descending Thresholds Rounding Scheme

- 1: For each $i \in [k]$, choose independently a threshold $\theta_i \in (0, 1]$ with probability density $\psi(\theta)$.
 - 2: Let σ be a permutation of $[k]$ such that $\theta_{\sigma(1)} \geq \theta_{\sigma(2)} \geq \dots \geq \theta_{\sigma(k)}$.
 - 3: **for all** i in $[k-1]$ **do**
 - 4: For every vertex $v \in V$ that has not been assigned yet,
 - 5: if $x_{v, \sigma(i)} \geq \theta_{\sigma(i)}$ then assign v to terminal $\sigma(i)$.
 - 6: **end for**
 - 7: Assign all remaining unassigned vertices to terminal $\sigma(k)$.
-

This scheme is in fact quite easy to analyze. The cut density for an edge of given location is given by the following lemma.

LEMMA 4.1. *The cut density under Algorithm 4 for edges of type (1, 2) at (u_1, u_2, \dots, u_k) such that $u_1 \leq u_2$ is at most*

$$\left(1 - \int_{u_1}^{u_2} \psi(u) du\right) \psi(u_1) + \psi(u_2).$$

Proof: The edge can be cut in coordinate u_1 only if θ_2 is not between u_1 and u_2 : if $u_1 < \theta_2 < u_2$, threshold θ_2 would be considered before θ_1 and the entire edge would be assigned to terminal 2. This accounts for the first term. The edge can be also cut in coordinate u_2 , which accounts for the second term. We neglect the possibility of the entire edge being assigned to another terminal, which can only decrease the probability of being cut. \square

This lemma provides another way to explain why descending thresholds might be advantageous. The Exponential Clocks Rounding Scheme cuts edges with probability density $2 - u_1 - u_2$, so our goal in the remaining schemes is to achieve a cut density bounded by a linear function of $u_1 + u_2$. The Single Threshold Rounding Scheme achieves this with some slack, because of the asymmetry in its analysis: it cuts edges with $u_1 = u_2$ less often than edges where u_1 and u_2 are far apart. The Descending Thresholds Rounding Scheme alleviates this problem, because its cut density is lower for edges where u_1 and u_2 are far apart due to the $\int_{u_1}^{u_2} \psi(u) du$ term.

4.3 The 1.30217-approximate rounding scheme

It remains to describe how we combine the Descending Thresholds Rounding Scheme with the Exponential Clocks and the Single Threshold Rounding Scheme, in particular how we choose the probability distributions ϕ and ψ . Based on computational evidence, we determined that ϕ is piecewise linear with a breakpoint at $b \in (0, 1)$, and the cut density due to the Single Threshold Rounding Scheme is $\frac{1}{2}\phi(u_1) + \phi(u_2)$, piecewise linear in u_1 and u_2 . Ideally we want a function in the form $a(u_1 + u_2)$, but as we saw in Section 3 that is hard to achieve. This is where the scheme of Descending Thresholds comes in: We choose the probability distribution ψ for decreasing thresholds uniform in the interval $[0, b]$, so that the cut density due to this scheme,

$(1 - \int_{u_1}^{u_2} \psi(u) du)\psi(u_1) + \psi(u_2)$, is again piecewise linear in u_1 and u_2 . Then we are able to balance the parameters so that the total cut density is a constant throughout most of the simplex. Our algorithm will be in the following form.

Algorithm 5

- With probability p_1 , choose the Exponential Clocks Rounding Scheme (Algorithm 2).
 - With probability p_2 , choose the Single Threshold Rounding Scheme (Algorithm 3), where the threshold is chosen with the following probability density:
 - For $0 \leq u \leq b$, $\phi(u) = a u$
 - For $b < u \leq 1$, $\phi(u) = c u + d$
 - With probability p_3 , choose the Descending Thresholds Rounding Scheme (Algorithm 4), where the thresholds are chosen with the following probability density:
 - For $0 \leq u \leq b$, $\psi(u) = \frac{1}{b}$
 - For $b < u \leq 1$, $\psi(u) = 0$
-

Using Lemma 3.1 and Lemma 4.1, the cut density for edges of type (1, 2) located at (u_1, u_2, \dots) is equal to $q = q_1 + q_2 + q_3$, where:

Case 1: $0 \leq u_1 \leq u_2 \leq b$.

- Exponential Clocks: $q_1 = p_1 (2 - u_1 - u_2)$.
- Single Threshold: $q_2 = p_2 (\frac{a}{2} u_1 + a u_2)$.
- Descending Thresholds: $q_3 = p_3 ((1 - \frac{u_2 - u_1}{b})\frac{1}{b} + \frac{1}{b})$.

$$q = 2p_1 + p_3 \frac{2}{b} + \left(-p_1 + p_2 \frac{a}{2} + p_3 \frac{1}{b^2}\right) u_1 + \left(-p_1 + p_2 a - p_3 \frac{1}{b^2}\right) u_2 .$$

Case 2: $0 \leq u_1 \leq b < u_2 \leq 1$.

- Exponential Clocks: $q_1 = p_1 (2 - u_1 - u_2)$.
- Single Threshold: $q_2 = p_2 (\frac{a}{2} u_1 + c u_2 + d)$.
- Descending Thresholds: $q_3 = p_3 \frac{u_1}{b^2}$.

$$q = 2p_1 + p_2 d + \left(-p_1 + p_2 \frac{a}{2} + p_3 \frac{1}{b^2}\right) u_1 + (-p_1 + p_2 c) u_2 .$$

Case 3: $b < u_1 \leq u_2 \leq 1$.

- Exponential Clocks: $q_1 = p_1 (2 - u_1 - u_2)$.
- Single Threshold: $q_2 = p_2 (\frac{c}{2} u_1 + c u_2 + \frac{3}{2} d)$.
- Descending Thresholds: $q_3 = 0$.

$$q = 2p_1 + \frac{3}{2} p_2 d + \left(-p_1 + p_2 \frac{c}{2}\right) u_1 + (-p_1 + p_2 c) u_2 .$$

Optimizing the parameters. Let us denote $\tilde{a} = p_2 a$, $\tilde{c} = p_2 c$ and $\tilde{d} = p_2 d$. Given the cut density formulas above, we would like to minimize z subject to the constraints

$$\forall 0 \leq u_1 \leq u_2 \leq b,$$

$$z \geq 2p_1 + \frac{2}{b} p_3 + (-p_1 + \frac{1}{2} \tilde{a} + \frac{1}{b^2} p_3) u_1 + (-p_1 + \tilde{a} - \frac{1}{b^2} p_3) u_2 \quad (6)$$

$$\forall 0 \leq u_1 \leq b \leq u_2 \leq 1,$$

$$z \geq 2p_1 + \tilde{d} + (-p_1 + \frac{1}{2} \tilde{a} + \frac{1}{b^2} p_3) u_1 + (-p_1 + \tilde{c}) u_2 \quad (7)$$

$$\forall b \leq u_1 \leq u_2 \leq 1,$$

$$z \geq 2p_1 + \frac{3}{2} \tilde{d} + (-p_1 + \frac{1}{2} \tilde{c}) u_1 + (-p_1 + \tilde{c}) u_2 \quad (8)$$

$$p_1 + \frac{1}{2} \tilde{a} b^2 + \frac{1}{2} \tilde{c} (1 - b^2) + (1 - b) \tilde{d} + p_3 = 1 \quad (9)$$

$$0 \leq b \leq 1 \quad (10)$$

$$p_1, p_3, \tilde{a}, \tilde{c}, \tilde{d} \geq 0 \quad (11)$$

Equation (9) follows from combining the probability normalization conditions $p_1 + p_2 + p_3 = 1$ and $\int_0^1 \phi(u) du = \frac{1}{2}ab^2 + \frac{1}{2}c(1-b^2) + (1-b)d = 1$. In order to eliminate the variables u_1, u_2 , we impose the following conditions: $p_1 = \tilde{c}$, $p_3 = \frac{b^2}{3}\tilde{c}$, $\tilde{a} = \frac{4}{3}\tilde{c}$, and $\tilde{d} = \frac{2}{3}b\tilde{c}$. This replaces all the constraints on z by $z \geq (2 + \frac{2}{3}b)\tilde{c}$. (In constraint (8), the right-hand side becomes $2\tilde{c} + b\tilde{c} - \frac{1}{2}\tilde{c}u_1$ which is dominated by $2\tilde{c} + \frac{1}{2}b\tilde{c}$ for $u_1 \geq b$.) Constraint (9) becomes $\tilde{c}(\frac{3}{2} + \frac{2}{3}b - \frac{1}{6}b^2) = 1$.

Therefore, we want to minimize $(2 + \frac{2}{3}b)\tilde{c}$ subject to $\tilde{c}(\frac{3}{2} + \frac{2}{3}b - \frac{1}{6}b^2) = 1$, which can be done by hand. The optimal solution is $b = 2\sqrt{3} - 3$ (≈ 0.464102) and $\tilde{c} = (6 + 5\sqrt{3})/26$ (≈ 0.563856), which gives $z = (10 + 4\sqrt{3})/13 \leq 1.30217$. The value of the other parameters can be derived from the equations above: $p_1 = (6 + 5\sqrt{3})/26$ (≈ 0.563856), $p_2 = (19 - 8\sqrt{3})/13$ (≈ 0.395661), $p_3 = (11\sqrt{3} - 18)/26$ (≈ 0.040483), $\tilde{a} = (12 + 10\sqrt{3})/39$ (≈ 0.75181) and $\tilde{d} = (4 - \sqrt{3})/13$ (≈ 0.17446). It can be verified that the cut density in all cases is bounded by $z = (10 + 4\sqrt{3})/13$. We summarize in the following theorem.

THEOREM 4.2. *Algorithm 5 with parameters $p_1 = (6 + 5\sqrt{3})/26$, $p_2 = (19 - 8\sqrt{3})/13$, $p_3 = (11\sqrt{3} - 18)/13$, $\tilde{a} = (12 + 10\sqrt{3})/39$, $b = 2\sqrt{3} - 3$, $\tilde{c} = (6 + 5\sqrt{3})/26$ and $\tilde{d} = (4 - \sqrt{3})/13$ gives a $(10 + 4\sqrt{3})/13 \simeq 1.30217$ -approximation for the Multiway Cut problem.*

5. INDEPENDENT THRESHOLDS: GETTING BELOW 1.3

As we show in Appendix C, the approximation factor of $\frac{10+4\sqrt{3}}{13} \approx 1.30217$ is the best one we can achieve if we combine Exponential Clocks with any sequence of k single-variable cuts, as long as the analysis for edges of type (i, j) works only with variables u_i, u_j . An obvious question is what happens if we consider the role of coordinates other than u_i, u_j . In particular, an edge of type (i, j) can be “captured” by another terminal ℓ , in the sense that both of its endpoints are labeled ℓ before coordinates u_i, u_j are even considered. Karger et al. [2004] rely on this argument to improve their analysis; they use a rounding scheme in the following form.

Algorithm 6 Independent Thresholds Rounding Scheme

- 1: For each $i \in [k]$, choose independently a threshold $\theta_i \in (0, 1]$ with probability density $\xi(\theta)$.
 - 2: Let σ be a uniformly random permutation of $[k]$.
 - 3: **for all** i in $[k - 1]$ **do**
 - 4: For every vertex $v \in V$ that has not been assigned yet,
 - 5: if $x_{v, \sigma(i)} \geq \theta_{\sigma(i)}$ then assign v to terminal $\sigma(i)$.
 - 6: **end for**
 - 7: Assign all remaining unassigned vertices to terminal $\sigma(k)$.
-

In this section, we pursue further improvements to the approximation ratio, using the Independent Thresholds Rounding Scheme. Unfortunately, the inclusion of further coordinates in the analysis leads to more involved expressions which are non-linear in the underlying distributions. In contrast to the results above, we are no longer able to find the

best parameters and verify the approximation ratio by hand. Our algorithm in this section has been found with the help of an LP solver; we provide more details on our computational experiments below. Let us remark that the LP solution involves a discretized description of a probability distribution ϕ , which makes it difficult to even present a description of the algorithm in a concise form. However, we have been able to find an (approximate) closed-form expression for ϕ and describe the algorithm in a compact form (see below). This incurs a small loss in the approximation factor. In this section, we do not claim that our approximation is optimal for any particular class of rounding schemes; our aim is to demonstrate that the approximation factor can be pushed below 1.3.

Why the gains are small. Before we get into the details of our final algorithm and its analysis, let us comment on why it seems difficult to achieve very substantial gains at this point. The gain from considering coordinates other than u_i, u_j comes from the fact that a threshold θ_ℓ might capture an edge of type (i, j) if the respective coordinate u_ℓ is above θ_ℓ . However, given u_i and u_j , all the other coordinates could be very small, namely $u_\ell = (1 - u_i - u_j)/(k - 2)$. For the rounding schemes that we considered above (Single Threshold and Descending Thresholds), this means that we do not get any improvement, because for any $u_i, u_j > 0$, we can have $u_\ell < u_i, u_j$ for all $\ell \neq i, j$; then, it never happens that terminal ℓ captures the edge before i or j . In order to exploit this argument, we need to use another rounding scheme, such as Independent Thresholds. (Other schemes are possible but computational experiments show that the Independent Thresholds Rounding Scheme is the most effective one out of those we were able to analyze.) Moreover, the probability density $\xi(u)$ should be relatively high for $u \rightarrow 0$, to make the probability $\Pr[\theta_\ell \leq u_\ell]$ significant. On the other hand, this is somewhat contrary to the goal of balancing cut densities with the Exponential Clocks Rounding Scheme where the cut density is $2 - u_i - u_j$, i.e. maximized for edges with u_i, u_j close to 0. In other words, we cannot use Independent Thresholds with a very high density near 0 because this would have significant impact on the cut density of edges with u_i, u_j close to 0. This is our interpretation of why the benefit of the Independent Thresholds Rounding Scheme on top of the other techniques is rather limited.

5.1 Our 1.2965-approximation algorithm

Based on computational experiments, we propose a rounding scheme in the following form, with a probability density ϕ , parameter $b \in [0, 1]$ and probabilities $p_1 + p_2 + p_3 + p_4 = 1$ to be determined.

Algorithm 7

- With probability p_1 , choose the Exponential Clocks Rounding Scheme (Algorithm 2).
 - With probability p_2 , choose the Single Threshold Rounding Scheme (Algorithm 3), where the threshold is chosen with probability density ϕ .
 - With probability p_3 , choose the Descending Thresholds Rounding Scheme (Algorithm 4), where the thresholds are chosen uniformly in $[0, b]$.
 - With probability p_4 , choose the Independent Thresholds Rounding Scheme (Algorithm 6), where the thresholds are chosen uniformly in $[0, b]$.
-

We defer the proofs of the results in this section to the full version of the paper. For the following result, we appeal to [Karger et al. \[2004\]](#) for the analysis of the Independent Thresholds Rounding Scheme.

LEMMA 5.1. *Given a point $(u_1, u_2, \dots, u_k) \in \Delta$ and the parameter b of Algorithm 6, let $a = \frac{1-u_1-u_2}{b}$. If $a > 0$, the cut density for an edge of type (1, 2) located at (u_1, u_2, \dots, u_k) under the Independent Thresholds Rounding Scheme with parameter b is at most*

- $\frac{2(1-e^{-a})}{ab} - \frac{(u_1+u_2)(1-(1+a)e^{-a})}{a^2b^2}$, if all the coordinates u_1, \dots, u_k are in $[0, b]$.
- $\frac{a+e^{-a}-1}{a^2b}$, if $u_1 \in [0, b], u_2 \in (b, 1]$ and $u_\ell \in [0, b]$ for all $\ell \geq 3$.
- $\frac{1}{b} - \frac{u_1+u_2}{6b^2}$, if $u_1, u_2 \in [0, b]$ and $u_\ell \in (b, 1]$ for some $\ell \geq 3$.
- $\frac{1}{3b}$, if $u_1 \in [0, b], u_2 \in (b, 1]$ and $u_\ell \in (b, 1]$ for some $\ell \geq 3$.
- 0, if $u_1, u_2 \in (b, 1]$.

For $a = 0$, the cut density is given by the limit of the expressions above as $a \rightarrow 0$.

We also refine the analysis of the Single Threshold and Descending Thresholds Rounding Schemes, depending on the value of the remaining coordinates.

LEMMA 5.2. *For an edge of type (1, 2) located at (u_1, u_2, \dots, u_k) , the cut density under the Single Threshold Rounding Scheme is at most*

- $\frac{1}{2}\phi(u_1) + \phi(u_2)$, if $u_\ell \leq u_1 \leq u_2$ for all $\ell \geq 3$.
- $\frac{1}{3}\phi(u_1) + \phi(u_2)$, if $u_1 < u_\ell \leq u_2$ for some $\ell \geq 3$.
- $\frac{1}{3}\phi(u_1) + \frac{1}{2}\phi(u_2)$, if $u_1 \leq u_2 < u_\ell$ for some $\ell \geq 3$.

LEMMA 5.3. *For an edge of type (1, 2) located at (u_1, u_2, \dots, u_k) , the cut density under the Descending Thresholds Rounding Scheme is at most*

- $(1 - \int_{u_1}^{u_2} \psi(u)du)\psi(u_1) + \psi(u_2)$, if $u_\ell \leq u_1 \leq u_2$ for all $\ell \geq 3$.
- $(1 - \int_{u_1}^{u_2} \psi(u)du)(1 - \int_{u_1}^{u_\ell} \psi(u)du)\psi(u_1) + \psi(u_2)$, if $u_1 < u_\ell \leq u_2$ for some $\ell \geq 3$.
- $(1 - \int_{u_1}^{u_2} \psi(u)du)(1 - \int_{u_1}^{u_\ell} \psi(u)du)\psi(u_1) + (1 - \int_{u_2}^{u_\ell} \psi(u)du)\psi(u_2)$, if $u_1 \leq u_2 < u_\ell$ for some $\ell \geq 3$.

Given these lemmas, we formulate the expressions for the total cut density under Algorithm 7.

COROLLARY 5.4. *Let $a(u_1, u_2) = (1-u_1-u_2)/b$. The cut density under Algorithm 7 for an edge of type (1, 2) located at $\mathbf{u} = (u_1, u_2, \dots)$ is $d_{12}(\mathbf{u})$ where*

- **Case I.** If $u_1 \leq u_2 \leq b$ and $u_\ell \leq b$ for all $\ell \geq 3$,
- $$d_{12}(\mathbf{u}) \leq p_1(2 - u_1 - u_2) + p_2 \left(\frac{1}{2}\phi(u_1) + \phi(u_2) \right) + p_3 \left(2 - \frac{1}{b}(u_2 - u_1) \right) \frac{1}{b} + p_4 \left(\frac{2(1 - e^{-a(u_1, u_2)})}{a(u_1, u_2)b} - \frac{(u_1 + u_2)(1 - (1 + a(u_1, u_2))e^{-a(u_1, u_2)})}{(a(u_1, u_2)b)^2} \right).$$

- **Case II.** If $u_1 \leq u_2 \leq b$ and $u_\ell > b$ for some $\ell \geq 3$,

$$d_{12}(\mathbf{u}) \leq p_1(2 - u_1 - u_2) + p_2 \left(\frac{1}{3}\phi(u_1) + \frac{1}{2}\phi(u_2) \right) + p_3 \left(\left(1 - \frac{1}{b}(u_2 - u_1) \right) u_1 + u_2 \right) \frac{1}{b^2} + p_4 \left(1 - \frac{1}{6b}(u_1 + u_2) \right) \frac{1}{b}.$$

- **Case III.** If $u_1 \leq b < u_2$ and $u_\ell \leq b$ for all $\ell \geq 3$,

$$d_{12}(\mathbf{u}) \leq p_1(2 - u_1 - u_2) + p_2 \left(\frac{1}{2}\phi(u_1) + \phi(u_2) \right) + p_3 \frac{u_1}{b^2} + p_4 \frac{a(u_1, u_2) + e^{-a(u_1, u_2)} - 1}{(a(u_1, u_2))^2 b}.$$

- **Case IV.** If $u_1 \leq b < u_2$ and $u_\ell > b$ for some $\ell \geq 3$,

$$d_{12}(\mathbf{u}) \leq p_1(2 - u_1 - u_2) + p_2 \left(\frac{1}{3}\phi(u_1) + \phi(u_2) \right) + p_3 \frac{u_1^2}{b^3} + p_4 \frac{1}{3b}.$$

- **Case V.** If $b < u_1 \leq u_2$,

$$d_{12}(\mathbf{u}) \leq p_1(2 - u_1 - u_2) + p_2 \left(\frac{1}{2}\phi(u_1) + \phi(u_2) \right).$$

For $a(u_1, u_2) = 0$, the expressions should be interpreted as limits when $a(u_1, u_2) \rightarrow 0$.

We remark that Case IV and Case V are applicable only when $b < 1/2$ (otherwise we cannot have two variables above b). These cases were necessary for us to explore different choices of b , but eventually we identified $b = 6/11$ as the optimal choice. (Curiously, [Karger et al. \[2004\]](#) make the same choice, even though they use only two out of the four rounding schemes present here. We do not quite understand this coincidence.) Hence, Case IV and Case V do not arise in our final verification.

Formulation as an LP. Given the analysis of Algorithm 7 above, it remains to choose the parameters b, p_1, p_2, p_3, p_4 and the probability distribution $\phi(u)$ so that $d_{12}(\mathbf{u})$ can be upper-bounded by a constant as small as possible. It is easy to see for a fixed b , this can be formulated as a linear program (after discretization). The only seemingly non-linear part of the problem is the product $p_2\phi(u)$, but this can be easily folded into a single variable $\tilde{\phi}(u) = p_2\phi(u)$. Finally, we have the normalization constraint $p_1 + p_2 + p_3 + p_4 = p_1 + \int_0^1 \tilde{\phi}(u)du + p_3 + p_4 = 1$. We minimize an upper-bound on $d_{12}(\mathbf{u})$ over all values $0 \leq u_1 \leq u_2 \leq 1$, under a suitable discretization.

We obtained a solution which involves a discretized description of a probability distribution ϕ . In order to obtain a concise description, we approximated this probability distribution by a piecewise-polynomial density function which we describe here. Our solution is as follows.

- $b = 6/11$
- $p_1 = 0.31052$
- $p_2 = 0.305782$
- $p_3 = 0.015338$
- $p_4 = 0.36836$
- $\tilde{\phi}(u) = p_2\phi(u)$
 $= 0.14957u - 0.0478u^2 + 0.45u^3$ for $0 \leq u \leq 0.23$
 $= -0.00484 + 0.1995u - 0.1067u^2 + 0.158u^3$

$$\begin{aligned}
& \text{for } 0.23 < u \leq 6/11 \\
& = 0.47639 + 0.21685u - 0.02388u^2 - 0.021u^3 \\
& \text{for } 6/11 < u \leq 0.61 \\
& = 0.47368 + 0.2816u - 0.18365u^2 + 0.079u^3 \\
& \text{for } 0.61 < u \leq 0.77 \\
& = 0.32195 + 0.75u - 0.6476u^2 + 0.2239u^3 \\
& \text{for } 0.77 < u \leq 1.
\end{aligned}$$

Computational verification confirms that with these parameters, the cut density is upper-bounded in all cases by 1.296445, for $0 \leq u_1 \leq u_2 \leq 1$ such that $u_1 + u_2 \leq 1$ and u_1, u_2 are multiples of $\delta = 1/2^{16}$. Finally, we use the following (standard) argument to bound the error arising from the discretization.

LEMMA 5.5. *For any function $f : [x_0, x_0 + \delta] \times [y_0, y_0 + \delta] \rightarrow \mathbb{R}$ such that $\frac{\partial^2 f}{\partial x^2}, \frac{\partial^2 f}{\partial y^2} \geq -d$,*

$$\begin{aligned}
& \max_{x_0 \leq x \leq x_0 + \delta, y_0 \leq y \leq y_0 + \delta} f(x, y) \\
& \leq \max\{f(x_0, y_0), f(x_0 + \delta, y_0), f(x_0, y_0 + \delta), f(x_0 + \delta, y_0 + \delta)\} \\
& \quad + \frac{1}{4}d\delta^2.
\end{aligned}$$

We apply this lemma to the function $d_{12}(\mathbf{u})$ from Corollary 5.4. It can be verified that for all $\mathbf{u} \in [0, 1]^2$, $\frac{\partial^2 d_{12}}{\partial u_1^2}, \frac{\partial^2 d_{12}}{\partial u_2^2} \geq -16$. (The second derivatives are mostly easy to evaluate, except for the expressions for the Independent Thresholds Rounding Scheme — appearing with a multiplier of p_4 ; we have verified these by *Mathematica*.) By Lemma 5.5, the discretization error can be bounded by $\frac{1}{4}d\delta^2 = 4 \cdot 2^{-32} = 2^{-30} < 10^{-9}$. This proves the following theorem.

THEOREM 5.6. *Algorithm 6 with parameters as above provides a 1.2965-approximation for the Multiway Cut problem.*

Acknowledgment.

We would like to thank T. S. Jayram for bringing the references to exchangeable sequences (de Finetti [1937], Diaconis [1977], Diaconis and Freedman [1980]) to our attention.

References

N. Buchbinder, J. Naor, and R. Schwartz. Simplex partitioning via exponential clocks and the multiway cut problem. *ACM STOC*, 2013.

G. Calinescu, H. J. Karloff, and Y. Rabani. An improved approximation for multiway cut. *Journal of Computer and System Sciences*, 60 (3):564–574, 2000.

W. H. Cunningham and L. Tang. Optimal 3-terminal cuts and linear programming. *IPCO*, 1999.

E. Dahlhaus, D. S. Johnson, C. H. Papadimitriou, P. D. Seymour, and M. Yannakakis. The complexity of multi-terminal cuts. *SIAM Journal on Computing*, 23:864–894, 1994.

B. de Finetti. La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henry Poincaré*, 7(1):1–68, 1937.

P. Diaconis. Finite forms of de Finetti's theorem on exchangeability. *Synthese*, 36:271–281, 1977.

P. Diaconis and D. Freedman. Finite exchangeable sequences. *Annals of Probability*, 8(4):639–859, 1980.

A. Freund and H. J. Karloff. A lower bound of $8/(7 + 1/(k-1))$ on the integrality ratio of the Calinescu-Karloff-Rabani relaxation for multiway cut. *Information Processing Letters*, 75(1–2):43–50, 2000.

D. R. Karger, P. N. Klein, C. Stein, M. Thorup, and N. E. Young. Rounding algorithms for a geometric embedding of minimum multiway cut. *Mathematics of Operations Research*, 29(3):436–461, 2004.

J.F.C. Kingman. Uses of exchangeability. *Annals of Probability*, 6(2):183–197, 1978.

J. Kleinberg and E. Tardos. Approximation algorithms for classification problems with pairwise relationships: metric labeling and Markov random fields. *Journal of ACM*, 49(5):616–639, 2002.

R. Manokaran, J. S. Naor, P. Raghavendra, and R. Schwartz. SDP gaps and UGC hardness for multiway cut, 0-extension and metric labeling. *STOC*, pages 11–20, 2008.

W. Trotter and P. Winkler. Ramsey theory and sequences of random variables. *Combinatorics, Probability and Computing*, 7:221–238, 1998.

APPENDIX

A. PAIRWISE REALIZABLE DISTRIBUTIONS

DEFINITION A.1. *A probability distribution ρ over $\mathcal{D} \times \mathcal{D}$ is pairwise realizable, if for every integer k , there exist k random variables X_1, X_2, \dots, X_k such that for each pair $i \neq j$, the joint distribution of (X_i, X_j) is ρ .*

In particular, we require that the distribution of (X_i, X_j) is the same as the distribution of (X_j, X_i) , so ρ should be symmetric. But even for symmetric distributions, this is not always possible. For example, we cannot design 3 random variables X_1, X_2, X_3 such that for each pair, we have $(X_i = 0, X_j = 1)$ or $(X_i = 1, X_j = 0)$ with probability $1/2$. Here, we present a necessary and sufficient condition for a distribution to be pairwise realizable. To avoid technical complications, we restrict ourselves to discrete domains \mathcal{D} .

THEOREM A.2. *A probability distribution ρ over a discrete domain $\mathcal{D} \times \mathcal{D}$ is pairwise realizable if and only if ρ is a convex combination of symmetric product distributions:*

$$\rho(a, b) = \sum_{s=1}^r \alpha_s p_s(a) p_s(b),$$

where $p_s : \mathcal{D} \rightarrow [0, 1]$, $\sum_{a \in \mathcal{D}} p_s(a) = 1$, $\alpha_s \in [0, 1]$, and $\sum_{s=1}^r \alpha_s = 1$.

Exchangeable sequences. Let us point out a connection with the concept of *exchangeable sequences* [de Finetti, 1937, Diaconis, 1977]. An infinite sequence of random variables X_1, X_2, X_3, \dots is exchangeable if for every n and a permutation π on $[n]$, (X_1, X_2, \dots, X_n) and $(X_{\pi(1)}, X_{\pi(2)}, \dots, X_{\pi(n)})$ have the same distribution. De Finetti's Theorem states that the distribution of every exchangeable sequence is a convex combination of distributions of sequences where X_1, X_2, \dots are identical and independent [de Finetti, 1937]. Our result can also be viewed as a statement about infinite

sequences of random variables: If X_1, X_2, \dots is an infinite sequence such that every *pair* has the same joint distribution, then this distribution is a convex combination of distributions of two identical and independent random variables. It has been pointed out to us by David Aldous that this statement can be also derived from the theory of exchangeability using either compactness arguments or Ramsey theory [Kingman, 1978, Trotter and Winkler, 1998]; here we give a simple proof using Farkas' lemma.

Proof: [Theorem A.2] The easy direction first: If $\rho(a, b) = \sum_{s=1}^r \alpha_s p_s(a) p_s(b)$ as above, we can generate arbitrarily many random variables as follows: With probability α_s , we generate k independent random variables X_1, X_2, \dots, X_k , each with probability distribution p_s . It is easy to see that for each $i \neq j$, the joint distribution of (X_i, X_j) is ρ .

The reverse direction relies on the convex separation theorem. Define \mathcal{S} to be the convex hull of symmetric product distributions on $\mathcal{D} \times \mathcal{D}$, which is exactly the set of all distributions ρ satisfying the conclusion of the theorem:

$$\mathcal{S} = \left\{ \rho : \mathcal{D} \times \mathcal{D} \rightarrow [0, 1] \mid \rho(a, b) = \sum_{s=1}^r \alpha_s p_s(a) p_s(b), \right. \\ \left. \sum_{a \in \mathcal{D}} p_s(a) = 1, p_s(a) \geq 0, \sum_{s=1}^r \alpha_s = 1, \alpha_s \geq 0 \right\}.$$

We note that \mathcal{S} is a convex compact set in $\mathbb{R}^{\mathcal{D} \times \mathcal{D}}$. Hence, if $\tilde{\rho} : \mathcal{D} \times \mathcal{D} \rightarrow [0, 1]$ is a distribution outside of \mathcal{S} , then by the convex separation theorem there is $y : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$ and $\lambda_1 < \lambda_2$ such that

- $\sum_{a, b \in \mathcal{D}} y(a, b) \tilde{\rho}(a, b) = \lambda_1,$
- $\forall \rho \in \mathcal{S}; \sum_{a, b \in \mathcal{D}} y(a, b) \rho(a, b) \geq \lambda_2.$

In particular, for every distribution $p : \mathcal{D} \rightarrow [0, 1], \sum_{a \in \mathcal{D}} p(a) = 1$, we have

$$\sum_{a, b \in \mathcal{D}} y(a, b) p(a) p(b) \geq \lambda_2. \quad (12)$$

Assume for a contradiction that for any k , there are random variables X_1, X_2, \dots, X_k such that $\forall i \neq j$ the distribution of (X_i, X_j) is $\tilde{\rho}$. In particular, we choose $k > \frac{1}{\lambda_2 - \lambda_1} (\max_{a \in \mathcal{D}} y(a, a) - \lambda_1)$. (The right-hand side is at least 1 by equation (12).) We consider the following quantity:

$$\sum_{a_1, a_2, \dots, a_k \in \mathcal{D}} \Pr[X_1 = a_1, X_2 = a_2, \dots, X_k = a_k] \sum_{i \neq j} y(a_i, a_j). \quad (13)$$

First, fix any choice of $a_1, a_2, \dots, a_k \in \mathcal{D}$ and consider the sum $\sum_{i \neq j} y(a_i, a_j)$. Define $p(a) = \frac{1}{k} |\{i \in [k] : a_i = a\}|$. Then we have

$$\sum_{i \neq j} y(a_i, a_j) = \sum_{i, j=1}^k y(a_i, a_j) - \sum_{i=1}^k y(a_i, a_i) \\ = k^2 \sum_{a, b \in \mathcal{D}} p(a) p(b) y(a, b) - k \sum_{a \in \mathcal{D}} p(a) y(a, a).$$

By the properties of $y(a, b)$, we get

$$\sum_{i \neq j} y(a_i, a_j) \geq k^2 \lambda_2 - k \sum_{a \in \mathcal{D}} p(a) y(a, a) \\ \geq k^2 \lambda_2 - k \max_{a \in \mathcal{D}} y(a, a).$$

Now we use the fact that $k > \frac{1}{\lambda_2 - \lambda_1} (\max_{a \in \mathcal{D}} y(a, a) - \lambda_1)$, and therefore $k \lambda_2 > k \lambda_1 + (\max_{a \in \mathcal{D}} y(a, a) - \lambda_1) = (k - 1) \lambda_1 + \max_{a \in \mathcal{D}} y(a, a)$. We obtain

$$\sum_{i \neq j} y(a_i, a_j) > k(k - 1) \lambda_1.$$

To summarize, expression (13) which is a convex combination of such terms, is lower-bounded strictly by

$$\sum_{a_1, a_2, \dots, a_k \in \mathcal{D}} \Pr[X_1 = a_1, X_2 = a_2, \dots, X_k = a_k] \sum_{i \neq j} y(a_i, a_j) \\ > k(k - 1) \lambda_1. \quad (14)$$

On the other hand, by switching the sums, we have

$$\sum_{a_1, a_2, \dots, a_k \in \mathcal{D}} \Pr[X_1 = a_1, X_2 = a_2, \dots, X_k = a_k] \sum_{i \neq j} y(a_i, a_j) \\ = \sum_{i \neq j} \sum_{a_i, a_j \in \mathcal{D}} y(a_i, a_j) \sum_{(a_\ell \in \mathcal{D} : \ell \neq i, j)} \Pr[X_1 = a_1, X_2 = a_2, \dots, X_k = a_k] \\ = \sum_{i \neq j} \sum_{a_i, a_j \in \mathcal{D}} y(a_i, a_j) \Pr[X_i = a_i, X_j = a_j] \\ = \sum_{i \neq j} \sum_{a_i, a_j \in \mathcal{D}} y(a_i, a_j) \tilde{\rho}(a_i, a_j) = k(k - 1) \lambda_1$$

again by the properties of $y(a, b)$. This is a contradiction with (14). \square

B. A TIGHT EXAMPLE FOR $\frac{3+\sqrt{5}}{4}$

Here we show the following lower bound.

THEOREM B.1. *No combination of the Single Threshold Rounding Scheme (with any distribution) and the Exponential Clocks Rounding Scheme can provide an approximation factor better than $(3 + \sqrt{5})/4$.*

We show this by considering a game between the algorithm and an adversary. The strategy space of the adversary is to pick an edge in the simplex. The strategy space of the algorithm is to pick a partition of the simplex into k parts of the following type:

1. Exponential Clocks: a random partition generated by Algorithm 2, or
2. Single Threshold: a partition generated by Algorithm 3 for any fixed value of θ

and any probability distribution over the strategies above. The game is a zero-sum game where if the endpoints of the edge picked by the adversary belong to two different parts of the partition picked by the algorithm, then the algorithm pays to the adversary a cost of 1 divided by the length of the edge (where recall that the length of the edge is defined to be $\frac{1}{2} \times$ the L_1 -distance between its endpoints). If the edge belongs to one part of the partition, there is 0 payoff to both players.

Clearly, if there is an algorithm in the strategy space that achieves a cut density bounded by α , then this implies a strategy for the algorithm player that pays at most α in expectation against any adversary. We present a strategy for the adversary such that the algorithm has to pay at least $\frac{3+\sqrt{5}}{4} - O(\frac{1}{k})$ in expectation. I.e., a cut density better than $\frac{3+\sqrt{5}}{4} - O(\frac{1}{k})$ cannot be achieved by this type of algorithm.

B.1 Probability distribution over edges

Here we define a strategy for the adversary. For each *ordered* pair of terminals (i, j) ($i, j \in [k]$, $i \neq j$), we have three sets of edges: A_{ij} , B_{ij} and C_{ij} . The edges in each of the sets A_{ij} , B_{ij} and C_{ij} have the property that their endpoints differ only in the coordinates i and j . Hence, the edges $(\mathbf{v}, \mathbf{v}')$ in these sets are of the form $\mathbf{v} = (u_1, \dots, u_k)$ and $\mathbf{v}' = (u_1, \dots, u_i - \epsilon, \dots, u_j + \epsilon, \dots, u_k)$. In all the three sets, we shall have $\forall k \neq i, j, u_k = (1 - u_i - u_j)/(k - 2)$. Hence, each edge is defined by the values of u_i and u_j .

- A_{ij} consists of all edges such that $u_i = x$ for some $x \in [3b, 1]$ and $u_j = 0$.
- B_{ij} consists of all edges such that $u_i + 2u_j = 3b$ with $b \leq u_i \leq 3b$ (and hence, $0 \leq u_j \leq b$).
- C_{ij} consists of a single edge defined by $u_i = 1$ and $u_j = 0$.

This completes the description of the sets A_{ij} , B_{ij} and C_{ij} except for setting the values of the parameters b and ϵ . The three sets are depicted in Figure 2. We shall set $b = \sqrt{5} - 2$ (as in Section 3) and $\epsilon = (1 - 2b)/(k - 2)$.

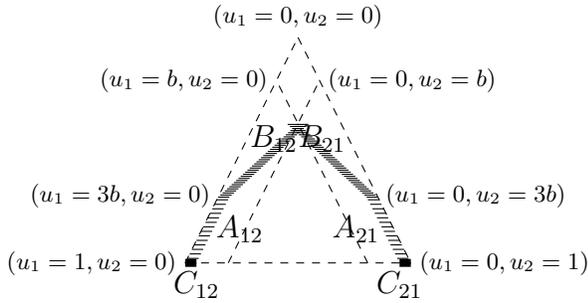


Figure 2: Tight example: edges of type (1, 2).

We now define the probability distribution over $\bigcup_{(i,j)} (A_{ij} \cup B_{ij} \cup C_{ij})$ that defines the strategy of the adversary. Starting with C_{ij} , the single edge is chosen with probability $\frac{2(1-2b)}{(1-b)(k-2)} \frac{1}{k(k-1)}$. Hence, the total probability mass over the edges in $\bigcup_{(i,j)} C_{ij}$ is $\frac{2(1-2b)}{(1-b)(k-2)}$.

The adversary has a uniform distribution over edges in $\bigcup_{(i,j)} A_{ij}$. Similarly, the adversary has a uniform distribution over edges in $\bigcup_{(i,j)} B_{ij}$. The total probability mass of the edges in $\bigcup_{(i,j)} A_{ij}$ is $\frac{1-3b}{1-b} (1 - \frac{1}{k-2})$, and the total probability mass of the edges in $\bigcup_{(i,j)} B_{ij}$ is $(\frac{2b}{1-b} - \frac{1}{k-2})$. Note that there are $k(k-1)$ ordered pairs (i, j) , hence for a particular ordered pair (i, j) , the sets A_{ij} and B_{ij} carry a total probability mass of $\frac{(1-3b)}{(1-b)} (1 - \frac{1}{k-2}) \frac{1}{k(k-1)}$ and $(\frac{2b}{1-b} - \frac{1}{k-2}) \frac{1}{k(k-1)}$ respectively.

We defer the analysis of the max-min value of the game to the full version of the paper.

C. A TIGHT EXAMPLE FOR $\frac{10+4\sqrt{3}}{13}$

THEOREM C.1. *For any combination of the Exponential Clocks Rounding Scheme and schemes with k threshold cuts, one for each terminal (and an arbitrary joint distribution), as long as the analysis of threshold cuts considers only the thresholds θ_i, θ_j for edges of type (i, j) (and not the possibility of being captured by another terminal), it cannot achieve a factor better than $\frac{10+4\sqrt{3}}{13}$.*

As in Section B, we define a probability distribution of edges in the simplex such that any partition strategy of the following type has to pay at least $\frac{10+4\sqrt{3}}{13}$ in expectation:

1. Exponential Clocks: a random partition generated by Algorithm 2, or
2. Simple Thresholds: any sequence of k threshold cuts, one for each terminal (such as the Single Threshold or Descending Thresholds Rounding Scheme).

Here, we modify the game to reflect the fact that we are considering analysis that depends only on the two coordinates u_i, u_j for edges of type (i, j) : if the partition player uses thresholds θ_i, θ_j in these coordinates, and say θ_i is applied first, he pays whenever the edge is cut in coordinate u_i or the edge is cut in coordinate u_j and *not captured* in coordinate u_i (i.e. other coordinates are not considered for the purposes of determining the payment).

C.1 Probability distribution over edges

Let us define the distribution over edges of type (1, 2) (Figure 3). Edges of type (i, j) are distributed analogously. We have $\alpha = \frac{-3+4\sqrt{3}}{13}$, $\gamma = \frac{19-8\sqrt{3}}{26}$ and $b = 2\sqrt{3} - 3$. The location of each edge satisfies $u_3 = u_4 = \dots = u_k = (1 - u_1 - u_2)/(k - 2)$, i.e. it is determined by u_1 and u_2 .

- Region R_A : Uniform density in the region defined by $u_1 \in [0, b]$ and $u_2 \in [0, b]$ with a total probability mass of α .
- Region R_{B1} : Uniform density in the region defined by $u_1 - \frac{1-2b}{b}u_2 \geq b$ on the simplex with a total probability mass of $\alpha/4$.
- Region R_{B2} : Uniform density in the region defined by $u_2 - \frac{1-2b}{b}u_1 \geq b$ on the simplex with a total probability mass of $\alpha/4$.
- Region R_{C1} : Density proportional to $(u_1 - (1 - b))$ in the region defined by $u_1 \in [1 - b, 1]$ and $u_1 + u_2 = 1$, with a total probability mass of $\frac{1}{1-b}(\alpha/4)$.
- Region R_{C2} : Density proportional to $(u_2 - (1 - b))$ in the region defined by $u_2 \in [1 - b, 1]$ and $u_1 + u_2 = 1$, with a total probability mass of $\frac{1}{1-b}(\alpha/4)$.
- Region R_{D1} : Density proportional to $(1 - b - u_1)$ in the region defined by $u_1 - \frac{1-2b}{b}u_2 = b$ and $u_1 \in [b, 1 - b]$, with a total probability mass of $\frac{1-2b}{1-b}(\alpha/4)$.
- Region R_{D2} : Density proportional to $(1 - b - u_2)$ in the region defined by $u_2 - \frac{1-2b}{b}u_1 = b$ and $u_1 \in [b, 1 - b]$, with a total probability mass of $\frac{1-2b}{1-b}(\alpha/4)$.
- Region R_{E1} : Uniform density in the region defined by $u_1 \in [b, 1]$ and $u_2 = 0$, with a total probability mass of γ .
- Region R_{E2} : Uniform density in the region defined by $u_2 \in [b, 1]$ and $u_1 = 0$, with a total probability mass of γ .

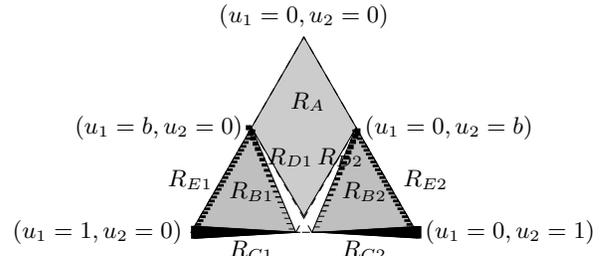


Figure 3: Tight example: edges of type (1, 2) have been shown.

We defer the analysis to the full version of the paper.