Best Cuts of a Set of Hyperrectangles

(Extended Abstract)

Viet Hai Nguyen* Thomas Roos* Peter Widmayer*

1 Introduction

Given a set of n d-dimensional (possibly intersecting) isothetic hyperrectangles, we consider the problem of separating these rectangles by means of a cutting isothetic hyperplane. If the cutting plane crosses a rectangle, this is cut into two non-overlapping hyperrectangles. We present optimal algorithms for computing several kinds of balanced cuts in $O(d\,n)$ time and space. Thereby, the balance function can be defined in different ways, leading also to different optimal partitions. We mainly consider two important balancing strategies and linear combinations. First, the balance of a partition is defined to be the maximum number of hyperrectangles on either side of the cutting hyperplane after the split. We call the corresponding optimal cut the so-called *best balanced cut*. In opposite to that, the second cutting objective is to minimize the *size* of partitions, i.e the total number of rectangles in the two subspaces, while preserving the balance of partitions, thereby minimizing the number of intersected rectangles. We call this simply an *optimal cut*. Our results include a generalization of recent results by [AmFr 92a] and [AmRoWi 93].

A solution to this problem has a number of applications, e.g., in the context of binary space partitions, partition trees, computer graphics and solid modeling, VLSI design, computer cartography and GIS, as well as geometric divide-and-conquer algorithms in many variations (see, e.g., [Be 75, Gu 84, NiHiSe 84, PaYa 89, Sa 90, PaYa 91, AmFr 92b, AmRoWi 93]).

In [AmFr 92a], it was shown that there always exists a cutting hyperplane separating the set S of n d-dimensional non-overlapping isothetic hyperrectangles into two sets each containing at most $\lceil \frac{2d-1}{2d} n \rceil$ hyperrectangles. Later, this result was extended to the more general case of overlapping rectangles by [AmRoWi 93]. It was shown that there always exists a cutting plane which creates two halfspaces with at most $\alpha = \lceil \frac{2d-1}{2d}(n-k) \rceil + k$ rectangles on each side, where k is the maximal number of rectangles in S that have a common point. (We call this k-overlapping). Here, the problem appears to be that the upper bound approaches n in the limit as d increases, thus becoming an overestimated upper bound (in general; not in the worst-case). The answer whether there is a better bound in other cases than the worst-case is still open.

A closer look at the problem shows that we can improve the previously established results to obtain a tighter upper bound by using a different measure than the overlapping factor k. We prove that there always exists a cutting plane which creates two halfspaces with at most $\lfloor \frac{n+k^*}{2} \rfloor$ rectangles on each side, where k^* is the "profile" of the given set S of rectangles. All proofs in this extended abstract are omitted.

*Departement Informatik, ETH Zentrum, CH-8092 Zürich, Switzerland

2 Preliminaries

At the beginning, we are given a set $S := \{R_1, \ldots, R_n\}$ of n (possibly overlapping) isothetic hyperrectangles[1]

$$R_i := [a_{i1}, b_{i1}] \times \ldots \times [a_{id}, b_{id}]$$

in d-dimensional real space $\mathbb{R}^d$. Any isothetic hyperplane C (which can be regarded as some degenerate isothetic rectangle) separates $\mathbb{R}^d$ into two closed half-spaces. Any rectangle R intersected by C is split into two non-overlapping parts corresponding to the respective sides of C, i.e. for any $c \in]a_i, b_i[$, the hyperplane $x_i = c$ splits the rectangle $R = [a_1, b_1] \times \ldots \times [a_d, b_d]$ into two parts: $R_{low} := [a_1, b_1] \times \ldots \times [a_i, c] \times \ldots \times [a_d, b_d]$ and $R_{high} := [a_1, b_1] \times \ldots \times [c, b_i] \times \ldots \times [a_d, b_d]$. So, any isothetic hyperplane C induces a partition of S in the sense that some rectangles lie entirely on some side of the hyperplane while others are intersected. Let $C^<$, $C^>$ denote the sets of rectangles lying entirely in the two halfspaces generated by the cut C and $C^=$ the set of rectangles intersected by C. Now, it is desirable to obtain a cutting plane such that

(i) the number of rectangles after the split is as small as possible, or equivalently, the number of intersected rectangles is minimized and

(ii) the difference of the number of rectangles on each side of the cutting plane is minimized.

Formally, we define the *best sum cut* as a cut minimizing the following sum:

$$\Sigma_C = |C^<| + |C^>| + 2 * |C^=|$$

where $|C|$ denotes the cardinality of the set C. Note that, since for an arbitrary cut C we have $|C^<| + |C^>| + |C^=| = n$, minimizing Σ_C is equivalent to minimizing $|C^=|$.

In opposite to that, a *(best) balanced cut* is defined to be a cut which minimizes the difference between $C^<$ and $C^>$, i.e.

$$\Delta_C = \| \, |C^<| - |C^>| \, \|$$

Thereby, $\|x\|$ denotes the absolute value of x. In an arbitrary situation, where the end points of rectangles may coincide, these two conditions together will not always be satisfied simultaneously. So, we propose a compromise, that is to look for an *optimal cut* C which minimizes the weighted sum: $\alpha\Sigma_C + \beta\Delta_C$ for real positive parameters α and β. We shall discuss this optimal cut in more detail in Section 4.

The following definition is crucial. Let C_j be an arbitrary isothetic plane along the j-th coordinate direction. We denote by k_j^* the maximum number of rectangles intersected by any such cut C_j, i.e. $k_j^* = \max_{C_j} |C_j^=|$ and k^* the minimum over all k_j^*, i.e. $k^* = \min\{k_1^*, \ldots k_d^*\}$. The set of all endpoints of intervals forming the projections of the rectangles in S onto the j-th coordinate axis[2] create a set of strictly increasing coordinates $\{x_1, \ldots, x_m\}$ where $1 \leq m \leq 2n$ and $\forall i, j,\ x_i, x_j \in \{a_1, \ldots, a_n, b_1, \ldots, b_n\}$ and $i < j \Rightarrow x_i < x_j$. We abbreviate this set by $\{x_i\}_1^m$. Using the same notations as in [AmRoWi 93], we define two functions $left(x) = |C^<|$ and $right(x) = |C^>|$, where C is a cut at coordinate x. Note that our approach is more general than that of [AmFr 92a, AmRoWi 93].

[1] For the sake of simplicity, we often refer to hyperrectangles (hyper-planes) simply as rectangles (planes respectively), implicitely assuming that we always talk about scenes in d-dimensional real space, except when it is explicitly stated otherwise.

[2] In the following, we consider all coordinate directions separately.

3 The Best Balanced Cut

Lemma 1 Given a set S of n possibly overlapping segments on a real line with possibly coinciding end points. Then, we can always find an interval (m_1, m_2) such that any cut C at $x \in (m_1, m_2)$ satisfies either $\Delta_C \leq \frac{e_{m_1}}{2}$ or $\Delta_C \leq \frac{e_{m_2}}{2}$ where e_{m_1} and e_{m_2} are the numbers of coincident segment endpoints at m_1 and m_2, respectively.

This lemma justifies again our reasoning above that in case of disjoint endpoints, there always exists a sub-interval with $left(x) = right(x)$. Indeed, in this case we have $e_{m_1} = 1$. This implies $left(x) - right(x) \leq \frac{1}{2}$ or equivalently $left(x) - right(x) = 0$. Our first theorem now provides an upper bound for the best balanced cut.

Theorem 1 Given a set of n possibly overlapping isothetic rectangles in d-dimensional real space, there always exists a cutting plane which creates two halfspaces with at most $\alpha^* = \lfloor \frac{n+k^*}{2} \rfloor$ rectangles on each side.

The following lemma relates the k-overlapping factor of a rectangular scene with the definition of the "profile" k^* of the scene.

Lemma 2 Given a set of n k-overlapping rectangles in d-dimensional real space, then the following inequality holds:

$$k \leq k^* \leq \frac{(d-1)n + k}{d}$$

Using this, we prove that the upper bound α^* is (in general) better than the upper bound α, i.e. we have

$$\alpha^* \leq \alpha$$

Indeed, it is easy to produce examples showing that in many cases α^* can be considerably smaller than α.

The proof of Theorem 1 gives rise to an $O(d\,n)$ algorithm to compute the best balanced cut C or, more precisely, an interval which contains all best balanced cuts. (Note that there may be infinitely many cuts which satisfy the balance condition). Afterwards, what we have to do now is to find a sub-interval I where Δ_I is minimal. Under the general position assumption, the proof of theorem 1 implies that $\Delta_I = 0$ and we obtain a very simple algorithm to compute the best balanced cut in $O(d\,n)$ time.

We now outline a lemma assuring the correctness of the algorithm. For this, we call $[m_1, m_2]$ the *median interval* of a set of n segments if m_1 and m_2 are the n-th and n+1-th elements of the set of disjoint segment end points, respectively. With that, we obtain the following surprising lemma.

Lemma 3 Given a set of n possibly overlapping segments on a real line, let $[m_1, m_2]$ be the median interval of this set. For any cut C at $x \in [m_1, m_2]$ and any cut C' at $x \notin [m_1, m_2]$, we have $\Delta_C \leq \Delta_{C'}$.

With that, we have the following straightforward algorithm to compute a best balanced cut. In a first step, we compute for any dimension j the median interval $[m_{j,1}, m_{j,2}]$ of the projections of the rectangles in S onto the j-th coordinate axis. Afterwards, we compute Σ_{C_j} where C_j is a cut at some $x_j \in [m_{j,1}, m_{j,2}]$. Finally, we choose the best cut among those C_j's over all dimensions $j = 1, \ldots, d$, i.e. the one with minimum Σ_{C_j}.

Theorem 2 Given a set of n possibly overlapping isothetic rectangles in d-dimensional real space, a best balanced cut can be computed in optimal $O(d\,n)$ time and space.

A more careful investigation of the problem suggests that the balanced cut is not always the optimal one. In the following section we shall discuss this optimal cut in more detail.

4 The Optimal Cut

In this section, we consider the question of satisfying the balance and the minimal sum conditions, as defined in Section 2, at the same time. Generally, such an absolutely optimal solution does not exist. In other words, one can not always obtain an optimal cut C_O such that for all other cuts C, we have $\Delta_{C_O} \leq \Delta_C$ and $\Sigma_{C_O} \leq \Sigma_C$. But we can make a compromise between these two conflicting conditions by considering a weighted sum of both balancing functions. Formally, we define the *optimal cut* as the one minimizing the following weighted sum

$$\Sigma^* = \alpha \Sigma + \beta \Delta$$

Thereby, α and β denote real positive parameters. Their relation represents the priority we put on the balance condition or the minimal sum condition. Particularly, $\alpha < \beta$ implies that the balance condition is more preferable and $\alpha > \beta$ means that a best sum cut has higher priority than the best balance cut.

Actually, up to now we have considered the first of the above cases by trying to obtain a cut which best satisfies the balance condition. In the following, we shall consider the cut which minimizes Σ^* and supports the minimal sum condition, i.e. with $\alpha > \beta$. The idea is to minimize $\min(|C^<|, |C^>|)$ while maintaining $\max(|C^<|, |C^>|)$ obtained by the balanced cut C. In effect, this is equivalent to reducing $|C^=|$ as much as possible.

Lemma 4 Given a set of n segments on a real line. Let (m_1, m_2) be the median interval and b_l (a_r, resp.) the first right (left, resp.) end point on the left (on the right, resp.) of (m_1, m_2). Then, the cut C at either b_l or a_r is optimal. That is, for any cut C', not at b_l nor at a_r, we have

$$\Sigma^*_{C'} \geq \Sigma^*_C$$

This provides an algorithm to compute the optimal cut. At first, for each dimension j, we compute the median m of the median interval $(m_{j,1}, m_{j,2})$ of the set of endpoints of the intervals. Afterwards, we find the nearest right end point b_l to the left of m and the nearest left end point a_r to the right of m. Then, we select the optimal cut C_j from C_{b_l} and C_{a_r}. Finally, we choose the optimal cut C out of all C_j's. Note that if b_l and a_r coincide with m_1 and m_2 in the optimal dimension, the balance cuts may be optimal too. The following theorem is now straightforward.

Theorem 3 Given a set of n possibly overlapping isothetic rectangles in d-dimensional real space, an optimal cut can be found in $O(d\,n)$ time and space.

We can also determine the whole sub-interval of optimal cuts. After computing b_l (a_r, resp.), we can find the nearest left end point a_l (right end point b_r, resp.) to the right of b_l (to the left of a_r, resp.) but not beyond m_1 (m_2, resp.). Of course, this is not necessary if already $b_l = m_1$ and $a_r = m_2$. One of the two sub-intervals $[b_l, a_l]$ and $[b_r, a_r]$, is then the optimal sub-interval. It is obvious that the optimal cut also satisfies Theorem 1.

5 Conclusion and Generalizations

In fact, our approach can be generalized to more general classes of objects. In this case, our algorithm and all investigations of the best achievable balance apply to the *bounding boxes* of the objects, as well. All we have to demand to preserve our runtime and space bounds is that the boundaries of the bounding boxes of the objects can be computed in $O(d)$ time, each. In contrast, in our general case, no corresponding results are known if we drop the restriction that the cutting hyperplanes be isothetic.

References

[AmFr 92a] F. d'Amore and P.G. Franciosa, *Separating Sets of Hyperrectangles*, to appear in Intern. Journal of Comp. Geometry & Appl., 1992

[AmFr 92b] F. d'Amore and P.G. Franciosa, *On the Optimal Binary Plane Partition for Sets of Isothetic Rectangles*, to appear in Proc. 4th Canadian Conference on Computational Geometry, St. John's, Newfoundland, Canada, August 1992

[AmRoWi 93] F. d'Amore, T. Roos and P. Widmayer, *An optimal algorithm for computing a best cut of a set of hyperrectangles*, Proc. International Conference on Computer Graphics ICCG'93, February 1993, Bombay, India

[Be 75] J.L. Bentley, *Multidimensional Binary Search Trees used for Associative Searching*, Comm. ACM, Vol. 18, No. 9, Sept. 1975, pp 509 – 517

[Gu 84] A. Guttman, *R-trees: A Dynamic Index Structure for Spatial Searching*, Proc. ACM SIGMOD, 1984

[NiHiSe 84] J. Nievergelt, H. Hinterberger and K.C. Sevcik, *The Grid File: An Adaptable, Symmetric Multikey File Structure*, ACM Trans. on Database Systems, Vol. 9, No. 1, March 1984, pp 38 – 71

[PaYa 89] M.S. Paterson and F.F. Yao, *Binary Space Partitions with Applications to Hidden Surface Removal and Solid Modelling*, Proc. 5th Annual ACM Symp. on Comp. Geometry, Saarbrücken, 1989, pp 23 – 32

[PaYa 91] M.S. Paterson and F.F. Yao, *Optimal Binary Space Partitions for Orthogonal Objects*, Proc. 1st ACM-SIAM Symp. on Discrete Algorithms, San Francisco, 1991, pp 100 – 106

[Sa 90] H. Samet, *The Design and Analysis of Spatial Data Structures*, Addison-Wesley, Reading, 1990