Balloon Darts: Fast Approximate Union Volume in High Dimensions with Line Samples

Scott A. Mitchell
Mohamed S. Ebeida
Muhammad A. Awad

SIAM GD/SPM
November 2013 talk only
Outline

• What is the problem – bunch of spheres overlapping, estimate the volume covered by one or more of them, the volume of their union.
  – Related problems – boxes, other booleans
• Standard power-cell in 2d, what goes wrong in high d
• Three variants, with sampling versions
  – Power Cells
  – Occlusion
  – Depth
• BF-Alg
  – First to implement it
• Neighbors
• Scaling studies
• 15 minutes, plus 5 for speaker transition
UnionVolume Problem Definition

• Collection of balls (spheres) that overlap arbitrarily.
  – Calculate the volume covered by one or more of them, 
    \(<\rightarrow\) calculate the volume of the union of spheres 
  – UnionVolume < sum of volumes
    • some volume is overcounted, in multiple balls

• Closely related problems
  – Collection of boxes, convex shapes
  – Intersection
Standard Approach in 2d

- Partition the shared-overlapped volume, assign piece to one sphere. Estimate each cell. Overlapped volume counted only once.
- Power cells
  - Partition into cells, subregions closer to the sphere than any other
    - Line through two points of sphere intersection separates cells
    - Define “closer” weighted by radius
- For each sphere
  - compute its power cell
    - intersect lines one point per pair
    - linear number of boundary points
  - subdivide power cell into sectors and triangles
    - Two types
    - Linear number of them
  - add up analytic area of the sectors and triangles
Why Hard in High Dimensions? d=30

- Partition the shared-overlapped volume, assign piece to one sphere. Estimate each cell. Overlapped volume counted only once.

- Power cells
  - Partition into cells, subregions closer to the sphere than any other
    - (d-1) hyperplane through (d-2) sphere of intersection separates cells
    - Define “closer” weighted by radius

- For each sphere
  - Compute its power cell
    - Intersection of hyperplanes for d+1 of them, there are k-choose-d (d-k)-dimensional faces
    - Combinatorial complexity explosion
  - Subdivide power cell into sectors and triangles
    - Many types combinatorial complexity explosion
    - Number of them exponential in d
  - Add up analytic area of the sectors and triangles
Why Hard in High Dimensions?

• Problem:
  – Computing faces of the boundary of the union or of cells is intractable, factorial in dimension

• Solution:
  – We don’t care about the boundary
  – Estimate the volume without constructing the boundary
Simple Point Estimation

• **Power cells**
  – Partition into cells, subregions closer to the sphere than any other

• For each ball
  – \( V \) compute its volume in isolation
  – sample points from the ball, \( S = \{\text{hits} \mid \text{misses}\} \)
    • if point in cell, its a hit. else miss.
  – \( V\text{-estimate} = \frac{\text{V hits}}{\text{hits + misses}} \)
  – Simple primitives:
    • generate point in sphere
    • compute weighted distances
    • hit if dist < other distances

\[
v\text{-est} = \frac{3}{3+2} = \frac{3v}{5}
\]

\[
v\text{-est} = \frac{4}{4+0} = v
\]
Simple Point Estimation

- **Occlusion Cells**
  - Order the sphere from 1-n
  - k is “above” k+1, owns the overlap volume

- **For each ball**
  - V compute its volume in isolation
  - sample points from the ball, S = {hits | misses}
    - if point in cell, it's a hit. else miss.
  - V-estimate = V hits / (hits + misses)
  - This part was the same! Only the cell definition was different.

![Diagram showing occlusion cells with spheres and points indicating hits and misses.]

\[ \text{v-est} = \frac{v}{2/(2+3)} = \frac{2v}{5} \]

Hint, order the balls so the big ones are not occluded and volume is analytically exact.
Simple Point Estimation

- **Depth Cells**
  - Overlapped regions owned equally

- For each ball
  - V compute its volume in isolation
  - sample points from the ball
    - depth = how many balls it is in
  - V-estimate =
    \[ V \left( \frac{1}{\text{#points}} \right) \sum_p \left( \frac{1}{\text{depth}} \right) \]
- Simple primitives, about as much work as occluded samples

\[ \text{depth} = \]
Line Sample Estimation

• For any of the three methods: power cells, occlusion cells, depth cells

• For each ball
  – V compute its volume in isolation
  – sample radial lines
  – get segments of the line in the cell
    • weight by distance from center = swept volume
  – V-estimate = V average weighted swept volume

\[
\begin{align*}
2-d: & \quad dv = r \, d\theta \\
\text{high-d:} & \quad dv = r^{d-1} \, d\theta
\end{align*}
\]
Bringman Friedrich Algorithm

• State of the art UnionVolume estimate in theory
  – Estimates frequency that random ball point is in another ball

• Repeat
  – Pick a random ball (uniform by volume)
    • p: pick a point from the ball
  – Repeat
    • B: pick another random ball
      (uniform by index, could be same ball)
    • If B contains p, break to outer loop
    • If iteration threshold, quit

• V = sum of volumes
• V-estimate = V outer-loop-iterations / inner-loop-iterations
• Iteration threshold is linear in the number of balls
to get an estimate with epsilon relative accuracy with \( \frac{3}{4} \) probability
• We were the first to implement it (says BF)
  – constants matter 😊
Bringman Friedrich Algorithm

- Iteration threshold is linear in the number of balls to estimate with epsilon relative accuracy with $\frac{3}{4}$ probability
  - Iteration Threshold $O(N / \epsilon^2)$
  - Runtime $O(N d / \epsilon^2)$

BF-ApproxUnion relative error ($\epsilon$) and required # samples per ball.

<table>
<thead>
<tr>
<th>$\epsilon$</th>
<th>$S/n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>40</td>
</tr>
<tr>
<td>.75</td>
<td>50</td>
</tr>
<tr>
<td>0.5</td>
<td>100</td>
</tr>
<tr>
<td>0.1</td>
<td>$1.8 \times 10^3$</td>
</tr>
<tr>
<td>0.01</td>
<td>$1.7 \times 10^5$</td>
</tr>
<tr>
<td>0.001</td>
<td>$1.7 \times 10^7$</td>
</tr>
</tbody>
</table>

Here the oracles are assumed to be perfectly accurate; otherwise more samples are needed.
Neighbors

• Are two balls close enough to overlap?
  – If known, reduces time to check if point, or segment, is in another ball
  – Neighbors in brute force $N^2$ time in high $d$
  – This $N^2$ may be less than $N/\epsilon^2$, number of samples needed in BF
  – All algs have $1/\epsilon^2$ trend, could be $>> N$
  – Trend
    suppose neighbors worth it for less than 1000 balls and $\epsilon = 0.1$
    1,000 balls $\leftrightarrow$ 0.1 $\epsilon$
    100,000 balls $\leftrightarrow$ 0.01 $\epsilon$
    10,000,000 balls $\leftrightarrow$ 0.001 $\epsilon$
  – and the constants matter 😊

$N^2$ isn’t a big deal compared to the other factors in high dimensions
Experimental Results: low-d

2-d Time to Achieve a Standard Deviation by Volume Estimation Method

3-d Time to Achieve a Standard Deviation by Volume Estimation Method

Implemented all the algorithms described above
Experimental Results: scaling by $d$, 2-20

Noisy trends. Need more replicates.

100 random balls
radius = distance to blue circle
BF is fast per sample, but each sample gives very little information

Line darts achieve a given level of accuracy faster
Conclusion

• Pre-finding overlapping balls is worth it if
  – High fidelity = relative error 0.01 or less
  – and << N balls overlap

• Times: (worst) BF-alg > point darts > line darts (best)

• Occluded cells with line samples is fast, and simpler than the alternatives
  – Need more experiments over more ball distribution types
    • Poisson-disks, or highly overlapping disks?

• In progress, extensions
  – more samples for larger balls & some overlaps analytically
  – selective sample locations & higher-dimensional samples?