On Approximating the Depth and Related Problems

Boris Aronov

Polytechnic University, Brooklyn, NY

Sariel Har-Peled

UIUC, Urbana, IL

Motivation:

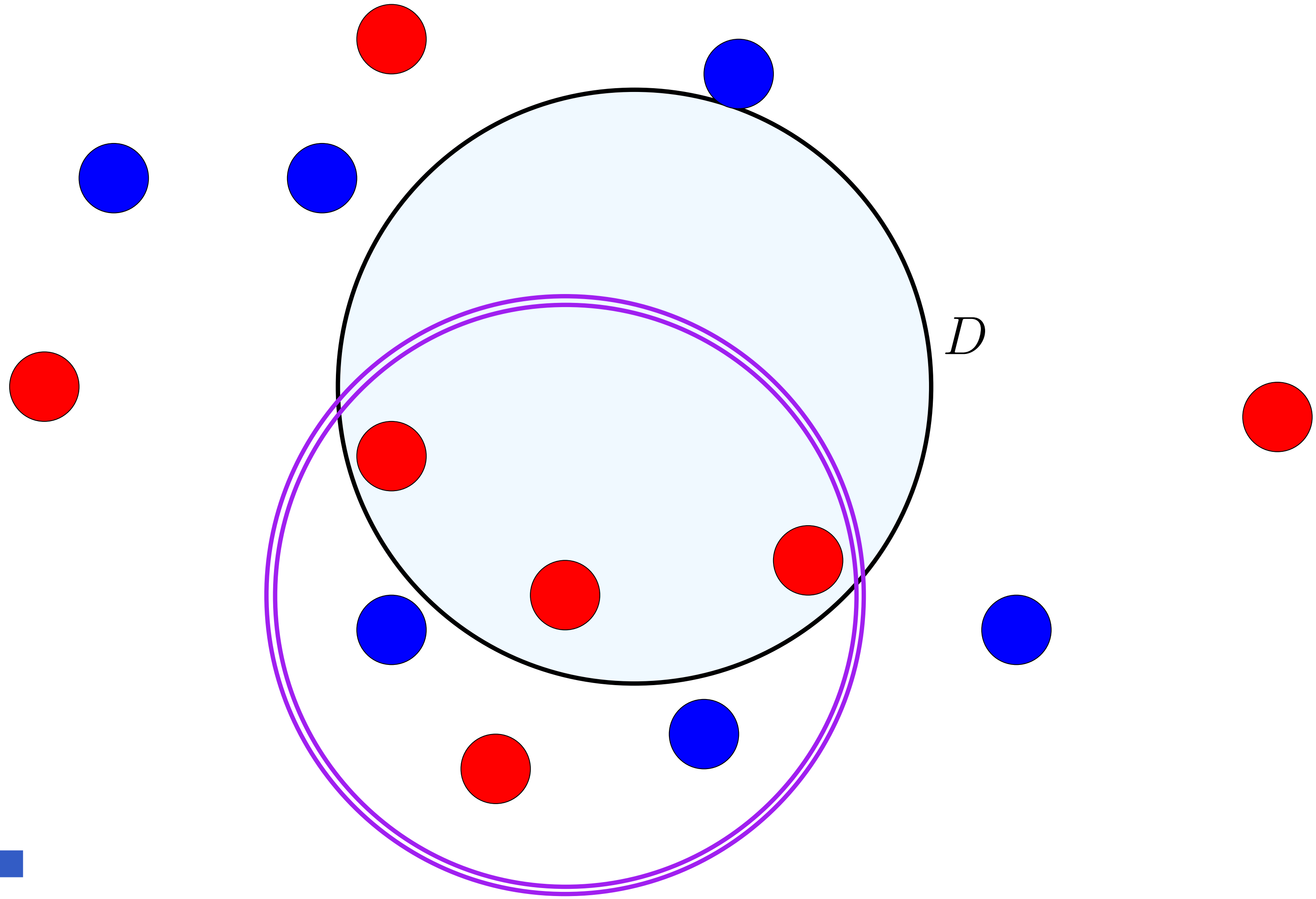
1: Operation Inflicting Freedom

Input:

- R - set red points in $\mathbb{R}^2$.
  (i.e., loc. of soldiers of the spindle of infamy)
- B - set blue points in $\mathbb{R}^2$.
  (i.e., loc. of peace loving and freedom spreading soldiers).

Q: Compute disk covering largest # of red points, avoiding all blue points.

2: Example



3: ”Dual Problem” - Deepest Point

- **Input:** $\mathcal{S}$ - Set of objects in $\mathbb{R}^d$
- **Q:** Find point covered by largest # objects of $\mathcal{S}$.
- **Minor Result:** For disks - **3sum**-hard.
  $\Longrightarrow$ Requires quadratic time.
- **Q:** Approximation?

4: From Shallow to Deep

- **depth thresholding:** Is $\mathfrak{MaxDepth}(\mathcal{S}) > k$?
- Assume done in $T(n,k)$ time.
  Example: $T(n,k) = O(nk)$.
- **Result:** Find point of depth
  $\geq (1-\varepsilon)\mathfrak{MaxDepth}(\mathcal{S})$.
  Running time: $O(T(n, \varepsilon^{-2}\log n) + n)$.
- Independent of value of $\mathfrak{MaxDepth}(\mathcal{S})$!
- **Idea:** Random sampling.
- **Known:** For large values of $\mathfrak{MaxDepth}(\mathcal{S})$.

Complement Problem

5: From Shallow to Shallower

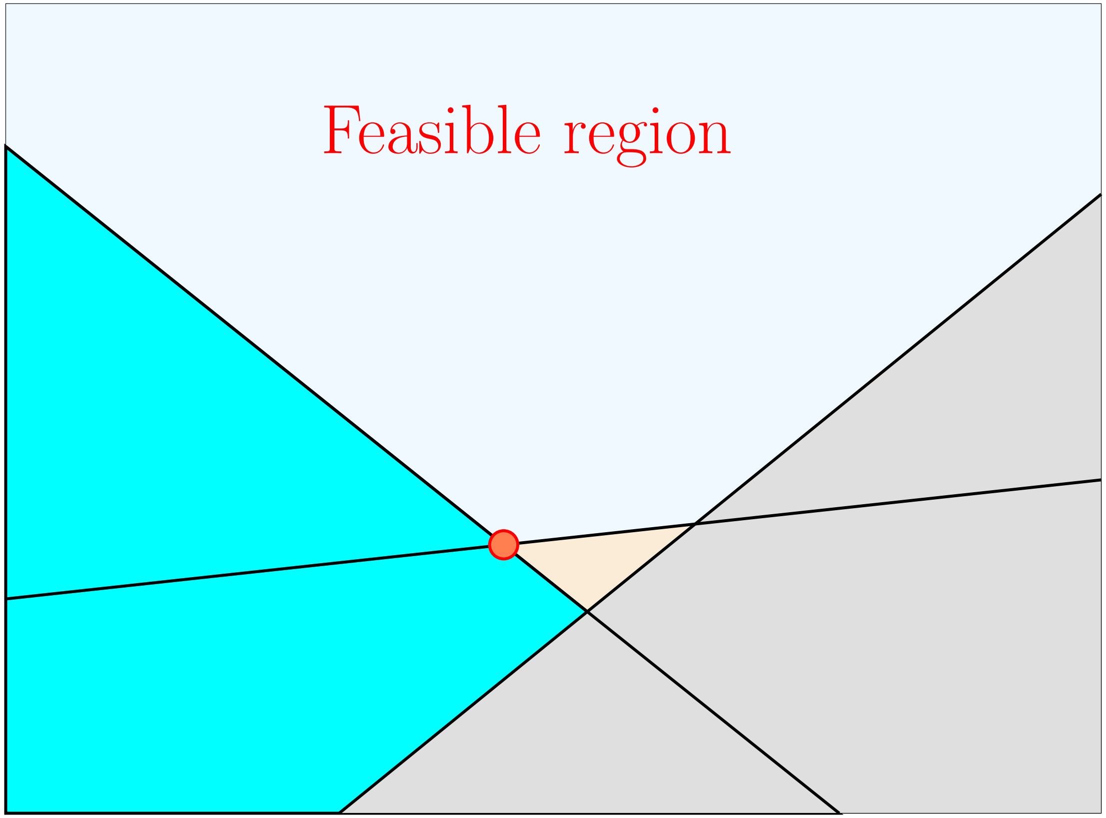
- **min-depth thresholding:** Is $\mathfrak{MinDepth}(\mathcal{S}) \leq k$?
- Assume done in $S(n, k)$ time.
- **Result:** Find point of depth $\leq (1+\varepsilon)\mathfrak{MinDepth}(\mathcal{S})$.
  Running time: $O(S(n, \varepsilon^{-2} \log n) + n)$.

6: Caring. The hard part.

- Can find deep/shallow point quickly.
- **Q:** Who cares?
- **A:** Linear programming with violations!
- Connection to learning, etc.

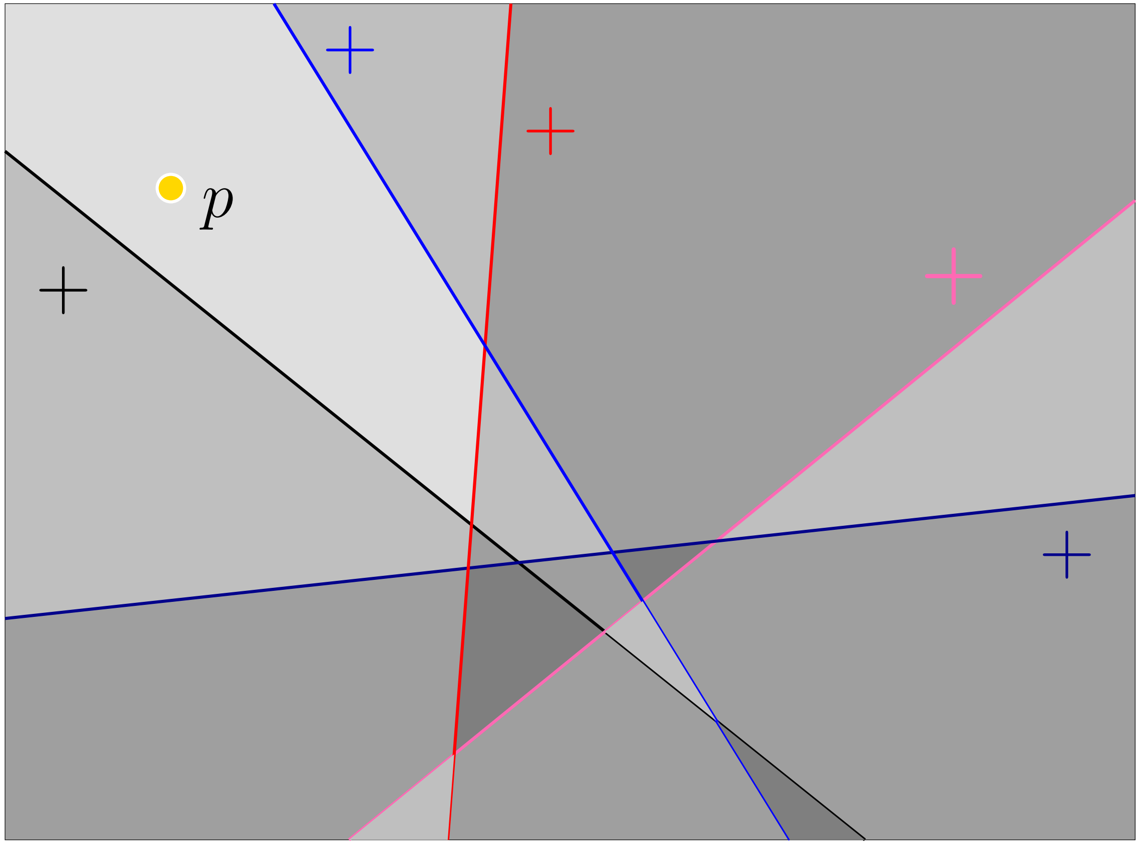
7: Linear Programming

- **Input:** H set of linear inequalities in $\mathbb{R}^d$.
- **Find:** Feasible point in $\mathbb{R}^d$



8: LP with Violations

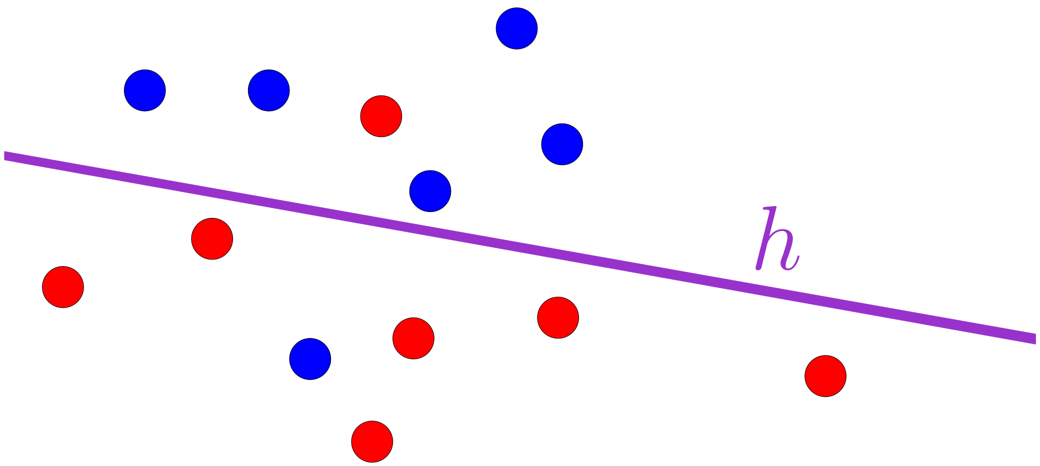
- **Input:** H set of linear inequalities in $\mathbb{R}^d$.
- **Find:** Point $p \in \mathbb{R}^d$ - min. # of violated constraints.



LP with Violations

9: Connection to learning

- P - positive points, N - Negative points
- Find hyperplane separating pos from neg.
- Might not be feasible.
- Find hyperplane min. # of misclassified points.
- $\Rightarrow$ LP with violations in the dual.



10: LP with Violations - Hardness.

- **NP-Hard**
- **[Amaldi and Kann (1995)]**
  NP-Hard to approximate within n^{ε}.
  d is large!

11: LP with Violations - low dim

- **[Matoušek (1995)]**
  LP with k violations in $\mathbb{R}^d$.
  Running time: $O_d\left(nk^{d+1}\right)$.
- **[Chan (2002)]**: "Small" speedup for $d = 2, 3$.
- Slow if, say, $k = \sqrt{n}$.
- k_{opt}: min # of constraints violated
- **Result:** Sol. violating $\leq (1+\varepsilon)k_{\text{opt}}$ constraints.
  Running time: $O\left(n\left(\varepsilon^{-2}\log n\right)^{d+1}\right)$.

Result

12: Disk covering largest # of red points

- R - set of red points
- B - set of blue points
- n points overall.
- **Result:** Find a disk covering $\geq (1-\varepsilon)k_{\text{opt}}$ red points while avoiding all blue points.
- Running time: $O\left(\frac{n}{\varepsilon^2}\log^2 n\right)$.

Range Searching

13: Relation between emptiness and counting

- **Range searching:** P - set of n points in $\mathbb{R}^d$.
  $\mathfrak{R}$ - set of ranges.
  Q: Preprocess P s.t. given $r \in \mathfrak{R}$ compute:
  - **emptiness:** Is $r \cap P$ empty?
  - **counting:** $|r \cap P|$?
  - **reporting:** $r \cap P$?
- Gap in complexity between counting and emptiness.
- **Result:** Approx. range count. as hard as emptiness.

Range Searching

14: Approximate counting in three dimensions

- Ranges – half-spaces in $\mathbb{R}^3$
- Using near linear space
- **emptiness queries:** $O(\log n)$
- **Counting queries:** $\widetilde{\Theta}(\sqrt{n})$
- **Result:** $(1+\varepsilon)$-approx. counting queries
  Query time: $O(\mathrm{poly}(1/\varepsilon, \log n))$.

15: Techniques used

“The thing that hath been, it is that which shall be; and that which is done is that which shall be done: and there is no new thing under the sun.” — **Ecclesiastes 1:9**

- Use random sampling to estimate a quantity.
- Classical technique in randomized algorithms.
- Technique probably known to Ecclesiastes.

16: Finding deepest point.

- $\mathcal{S}$ set of n objects in the plane.
- p - point of depth k.
- Pick $r \in \mathcal{S}$ to random sample R, with probability: $O\left(\frac{1}{k} \cdot \frac{1}{\varepsilon^2} \log n\right)$
- μ = Depth of p in $\mathcal{A}(R)$
- μ “good” estimate to depth of p in $\mathcal{A}(\mathcal{S})$. (By Chernoff inequality.)
- However p is “shallow” at $\mathcal{A}(R)$.
- $\mathrm{E}[\mathrm{depth}_R(p)] = O(\varepsilon^{-2} \log n)$.

17: Technicalities.

- $\mathcal{S}$ set of n objects in the plane.
- k : guess for depth.
- Pick $r \in \mathcal{S}$ to random sample, with "right" probability.
- If correct, deepest point in $\mathcal{A}(R)$ is of small depth μ.
- If $\mathfrak{MaxDepth}(\mathcal{S}) >> k$ then $\mathfrak{MaxDepth}(R) >> \mu$.
- Check $\mathfrak{MaxDepth}(R) >> \mu$ by depth thresholding.
- Binary search for right depth...

18: Conclusions

- Solving depth problems via depth thresholding.
- Applications:
  - LP with violations approximately in near linear time.
  - shape fitting with outliers.
  - Finding a disk with maximum number of red points.
  - More...???
- Approximate counting equivalent to emptiness.
- Compute deepest point in arr quickly.

19: Conclusions - open problems

- Compute best shape covering red points for other shape families (squares, ellipses, etc).
- Finding disk covering max red points - 3sum-hard?
- More reductions?
  Range searching with minimizations
  $\Longrightarrow$ approx. counting.
- **Approximately**, how many angels can stand on the head of Buffon's needle?

References

Amaldi, E. and Kann, V. (1995). The complexity and approximability of finding maximum feasible subsystems of linear relations. *Theoretical Computer Science*, 147(1–2):181–210.

Chan, T. M. (2000). Random sampling, halfspace range reporting, and construction of $(\leq k)$-levels in three dimensions. *SIAM J. Comput.*, 30:561–575.

Chan, T. M. (2002). Low-dimensional linear programming with violations. In *Proc. 43th Annu. IEEE Sympos. Found. Comput. Sci.*, pages 570–579.

Har-Peled, S. and Indyk, P. (2000). When crossings count - approximating the minimum spanning tree. In *Proc. 16th Annu. ACM Sympos. Comput. Geom.*, pages 166–175.

Indyk, P. (2000). *High-Dimensional Computational Geometry*. PhD thesis, Stanford.

Indyk, P. and Motwani, R. (1998). Approximate nearest neighbors: Towards removing the curse of dimensionality. In *Proc. 30th Annu. ACM Sympos. Theory Comput.*, pages 604–613.

Matoušek, J. (1995). On geometric optimization with few violated constraints. *Discrete Comput. Geom.*, 14:365–384.