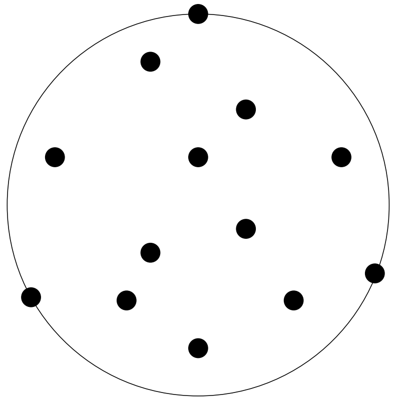
Randomizes Incremental Construction and Random Sampling

- Random Incremental Construction
  - Introduction
  - An example
- Random Sampling
  - Introduction
  - An example
- Linear Programming
  - Basic Facts
  - Random Incremental Construction Algorithm
  - Random Sampling Algorithm

Example: Smallest enclosing disks (SED)

Input: A set of points P in the plane.

Output: The disk with smallest radius containing P.



Random Incremental Construction

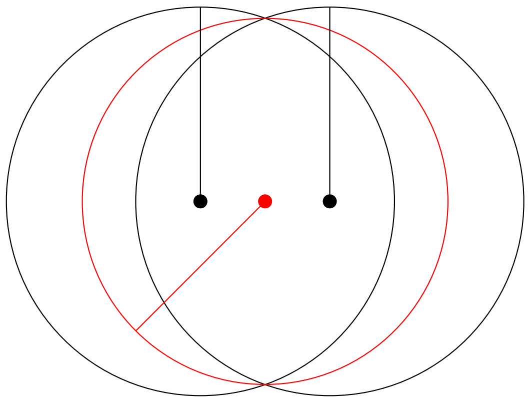
Input: A set of n objects.

Output: Something S depending on the objects.

Algorithm:

- Randomly permute the input set.
  Let $p_1, \ldots p_n$ be the elements indexed in the random order
- Iteratively compute S_i, the “something” for the set $\{p_1, \ldots p_i\}$

Smallest enclosing disk is unique



- Notation: $B(p, r)$ is ball around p with radius r.
- Assume $B(p, r)$ and $B(q, r)$ are smallest enclosing disks.
- $B((p+q)/2, \sqrt{r^2 - \|pq\|^2/4})$ contains $B(p, r) \cap B(q, r)$.

At most three points define SED

Claim: There is $S \subseteq P$ with $|S| \leq 3$, such that $SED(P) = SED(S)$.
Claim: If $SED(P \setminus \{p\}) \neq SED(P)$, then p lies on the boundary of $SED(P)$.

Proof: Later for a bit more general result.

The Algorithm

Input: Two sets of points P and S with $|S| \leq 3$.

Invariant: The $SED(P \cup S)$ has S in its boundary.

Output: $SED(P, S)$: SED for P with S on its boundary, provided such a disk exsists.

Algorithm: $SED(P, S)$:

1. if $|S| = 3$ return ball defined by S.
2. if $|P| = 2$ return $B((p+q)/2, \|pq\|/2)$, where $P = \{p, q\}$.
3. chose $p \in P$ at random
4. $B := SED(P \setminus \{p\}, S)$.
5. if $p \in B$ return B.
6. return $SED(P \setminus \{p\}, S \cup \{p\})$.

Idea of the Algorithm

Input: A set of point P.

Output: $SED(P)$: SED for P.

Algorithm: $SED(P)$:

1. if $|P| = 2$ return $B((p+q)/2, \|pq\|/2)$, where $P = \{p, q\}$.
2. chose $p \in P$ at random
3. $B := SED(P \setminus \{p\})$.
4. if $p \in B$ return B.
5. return $SED(P \setminus \{p\})$ with p on the boundary.

Analysis: Running Time

- $$\begin{aligned} T(n,d) &= 1 + T(n-1,d) + \\ &\quad Pr(p \notin SED(P \setminus \{p\}, S)) T(n-1, d+1) \end{aligned}$$
- $Pr(p \notin SED(P \setminus \{p\}, S)) \leq 3/n$
- Claim: $T(n, d) \in O(n)$
- Induction over d:
  - We know $T(n, 3) = O(1) \leq cn$.
    Assume $T(n, d+1) \leq cn$.
  - Claim by induction: $T(n, d) \leq (3c+1)n$.
  - $$\begin{aligned} T(n,d) &= 1 + T(n-1,d) + \tfrac{3}{n} c(n-1) \\ &\leq (1+3c)(n-1) + 1 + 3c. \end{aligned}$$

Correctness

Assume P and R are point sets and there is a disk with R on its boundary containing P then:

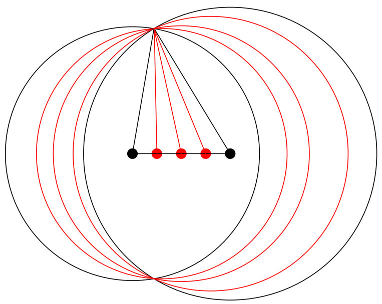
- $SED(P, R)$ is unique.
- If $p \notin SED(P \setminus \{p\}, R)$ then p is on the boundary of $SED(P, R)$, i.e. $SED(P, R) = SED(P \setminus \{p\}, R \cup \{p\})$.
- There is a set S of at most $\max(0, 3 - |R|)$ point in P such that $SED(P, R) = SED(S, R)$.

Proof of Lemma

- ($SED(P, R)$ is unique): $B_{1/2}$ has smaller radius and R on its boundary.
- (A necessary point is on the boundary): Use λ to transform $SED(P \setminus \{p\}, R)$ continuously to $SED(P, R)$.
  For some $0 < \lambda < 1$, p is on the boundary and the radius of this disk is smaller than the radius of $SED(P, R)$.
- (A total of three points suffice): Three points of the boundary define a disk. If four points are cocircular, you must argue more carefully.

Convex combination of two disks

For two disks $B(p_1, r_1)$ and $B(p_2, r_2)$ let B_λ be the smallest disk with center $(1 - \lambda)p_1 + \lambda p_2$ containing $B(p_1, r_1) \cap B(p_2, r_2)$.



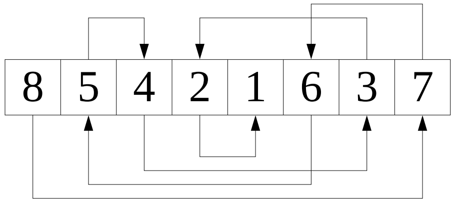
- For $0 < \lambda < 1$, the radius of B_λ is smaller than $\max(r_1, r_2)$.
- The intersection of the boundaries of $B(p_1, r_1)$ and $B(p_2, r_2)$ is on the boundary of B_λ.

Random Sampling or Randomized Divide-and-Conquer

General Problem:

- Given a ground set S and
  a function ϕ which maps all subsets $R \subseteq S$ to a value.
- Compute $\phi(S)$.
- We assume
  - ϕ can be evaluated for small subsets.
  - We can compute the *violators*, i.e.
    $V(R) = \{s \in S \mid \phi(R \cup \{s\}) \neq \phi(R)\}$.
  - The set of *extreme points*, i.e.
    $X(R) = \{s \in R \mid \phi(R \setminus \{s\}) \neq \phi(R)\}$, is small.

Searching in a sorted compact list



Input: n elements with distingt values.

Task: Test whether an element with value k is in the list.

Algorithm:

- Take a sample of size r.
- Find the smallest element with value at least k.
- Follow the links in the list until either k is found or an element smaller than k.

The Sampling Lemma

Let v_r be the expected number of violators for a random subset of cardinality r.

Let x_r be the expected number of extreme points for a random subset of cardinality r

$$v_r = x_{r+1}(n-r)/(r+1)$$

Example: Note that the number of extreme points is always 1.

Choosing a random sample of cardinality $\sqrt{n}$, the expected number of violators is $(n-1)/(\sqrt{n}+1) = O(\sqrt{n})$.

Analysis

In our general setting:

- $\phi(R) = \min\{s \in R \mid s \geq k\}$.
- $V(R) = \{s \in S \mid s \geq k \wedge s < \phi(R)\}$
- $X(R)$ the element which defines $\phi(R)$.

The expected running time is: $O(r) + O(E(V(R)))$.

Proof of Sampling Lemma

$$\begin{aligned}
\binom{n}{r} v_r &= \sum_{R \in \binom{S}{r}} \sum_{s \in S \setminus R} [s \text{ violates } R] \\
&= \sum_{R \in \binom{S}{r}} \sum_{s \in S \setminus R} [s \text{ is extreme in } R \cup \{s\}] \\
&= \sum_{Q \in \binom{S}{r+1}} \sum_{s \in Q} [s \text{ is extreme in } Q] \\
&= \binom{n}{r+1} x_{r+1}
\end{aligned}$$

Finally $\binom{n}{r+1} / \binom{n}{r} = (n-r)/(r+1)$.