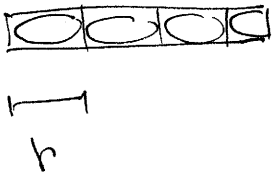
BMG variant

Idea to search from succesful iteration in BM:

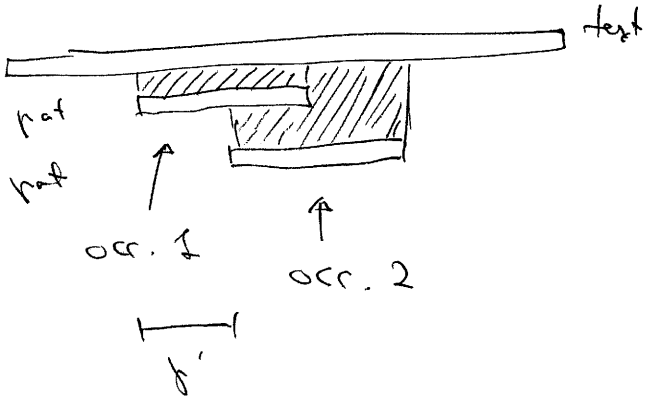
Let p = period(pattern).

So pat = 

p

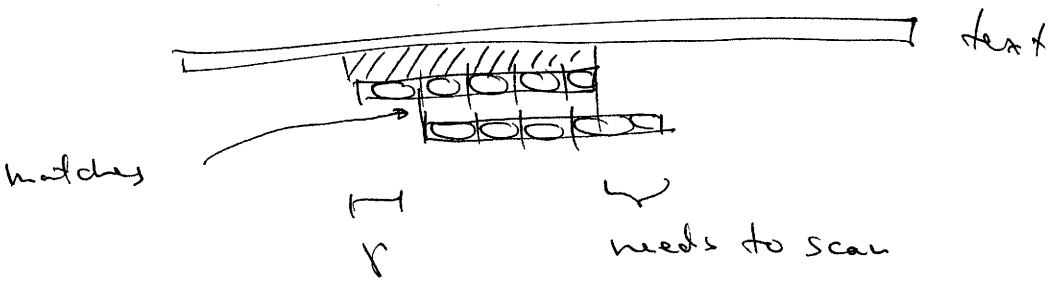
(and p shortest so) **)

Hence, after a match, we can safely (by *) and **)) skip p forward.



We see that p' is a period of pattern. *)

And will match a prefix of size $m-p$ after:



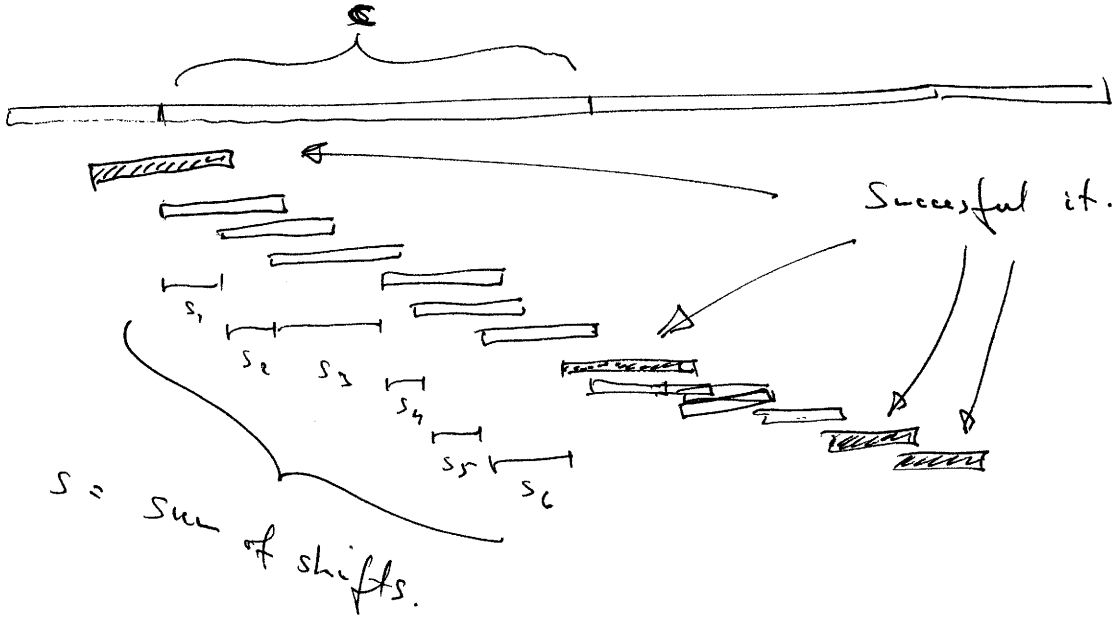
INVARIANT:

In BMG alg.: mem = known matching prefix $\in \{0, m-p\}$

```
p = period(pat)
i = 0
mem = 0
while i ≤ n - m
    j = m
    while j > mem AND pat[j] == text[i+j]
        j = j - 1
    if j = mem                    (succesful it.)
        report occ. at i
        mem = m - p
        i = i + p
    else                          (unsuccesful it.)
        mem = 0
        i = i + BM_Shift(j)
```

Correctness follows from safeness of shift p for succesful iterations (and correctness of mem = m-p after shift), and safeness of BM_Shift for unsuccesful iterations.

Complexity



We earlier (for BM alg.) analysed a seq. of unsuccesfull iteration. Its cost was covered by $c + 3s$

Here, we have several such seq., separated by one (or more) successful it.

For two unsuccesful seq., their s's cannot overlap (seen as areas of text), neighboring

~~(they add to $\leq n$ in total)~~,

and their c's can only overlap by $< m$.

Each succesful it. costs at most m.

Hence, the entire alg. costs $O(n + r \cdot m)$ when there are r occurences. (r succesful it.).

If $r \geq m/2$, all occ. are $m/2$ apart
(r is smallest dist. possible between ~~adjacent~~ occ.,
by def. [see page ①].)

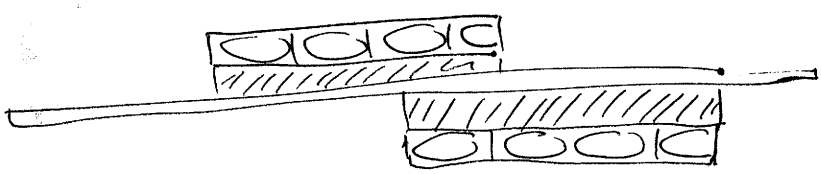
So $r \leq n/(m/2) = 2 \cdot \frac{n}{m}$. So $n + r \cdot n$
is $\leq n + 2n = 3n$

If $r \leq m/2$:

pat = 

periodic (by minimality of r).

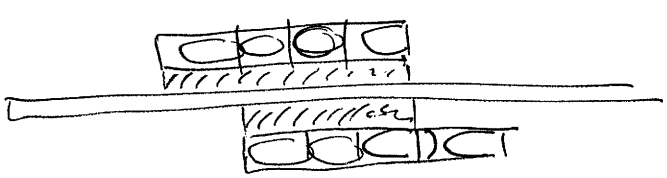
Assume later match has overlap $\geq r$:



conflict with prop. on cyclic shifts
if not "aligned" (is not a shift of multiple
of r).

Look at test after match:

Shift r

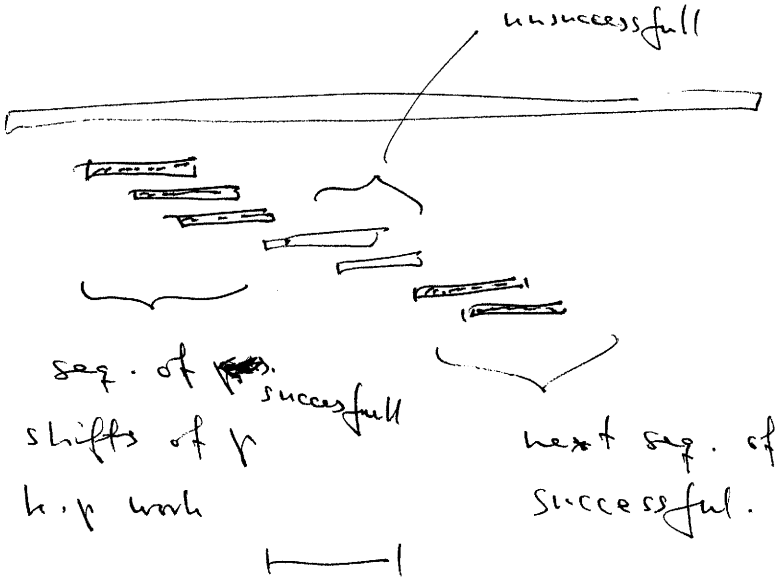


test here
(r chars)

If pos., repeat
(shift r, test r chars)

If neg., next
occ. is earliest
after a total shift
of $n - r \geq m/2$

So :



Same argument type as before now applies.

(At most $m/(m/2)$ seq. of succesful).

Period p can be calculated as $B_{pat}^{(m)}$ in $O(m)$ time [via KMP preprocessing].

□.

BMG is

Hence, $\forall$ an $O(n)$ BM-based alg. for reporting all occ.