



Exercise X, Sublinear Algorithms for Big Data Analysis 2024-2025

These exercises are for your own benefit. Feel free to collaborate and share your answers with other students, and solve as many problems as you can. Problems marked (*) are more difficult, but also more rewarding. These problems have been taken from various sources on the Internet, too numerous to cite individually.

- 1 Give a single pass streaming algorithm that find a 2-approximation to the maximum matching in the input graph $G = (V, E)$ with n vertices using $O(n \log n)$ bits of space.
 - 2 Let $A \subseteq [n]$ be an arithmetic-progression-free set, i.e. a set such that for every triple $a, b, c \in A$ if $2b = a + c$, then $a = b = c$. Let $G = (V, E)$, $V = [3n]$, be defined as follows. For every $i \in [n]$ let $M_i = \{(a + i, a + 2i) : a \in A\}$. Let $E = \bigcup_{i \in [n]} M_i$. Show that M_i are induced, i.e. for every i it holds that the subgraph induced by the vertices matched by M_i contains only the edges of M_i .
 - 3 Pick integers d, s and let t be the smallest integer such that $(2d + 1)^t \geq n$. Let $A_{d,s}$ be the set of integers of the form $\sum_{i=0}^t a_i (2d + 1)^i$ with integers a_i satisfying
 1. For all i , $0 \leq a_i \leq d$
 2. $\sum_{i=0}^t a_i^2 = s$.
- 3a** For every d, s the set $A_{d,s}$ has no three-term arithmetic progressions.
- 3b** For $d = \lceil 2^{\sqrt{\log n}} \rceil$ and some choice of s , $|A_{d,s}| \geq n/2^{O(\sqrt{\log n})} = n^{1-o(1)}$.