Algorithms	**Problem Set 4**
CS 6820 Fall 2009	**Due October 30, 2009**

Recall the general instructions for handing in homework:

- If possible, please typeset the homework (i.e. format your solutions as an electronic file using latex or Word with mathematical notation).
- Homework solutions done electronically can be handed in by directly uploading them to CMS. Please mail Ashwin (`ashwin85@cs.cornell.edu`) if you have any trouble with this.

(1) *(KT Exercise 11.4)* Consider the following *Hitting Set* problem. We are given a set $A = \{a_1, \ldots, a_n\}$ and a collection $B_1, B_2, \ldots, B_m$ of subsets of A. Also, each element $a_i \in A$ has a *weight* $w_i > 0$.

We say that a set $H \subseteq A$ is a *hitting set* if $H \cap B_i$ is not empty for each $i = 1, 2, \ldots, m$; in other words, if H "hits" each set B_i. Let $b = \max_i |B_i|$ denote the maximum size of any of the sets $B_1, B_2, \ldots, B_m$. Give a polynomial time approximation algorithm that finds a hitting set whose total weight is at most b times the minimum possible.

(2) *(KT Exercise 11.9)* Consider the following maximization version of the *three-dimensional matching* problem. We are given disjoint sets X, Y, and Z, together with a set $T \subseteq X \times Y \times Z$ of ordered triples; that is, each triple $t \in T$ has the form $t = (x_i, y_j, z_k)$ for $x_i \in X$, $y_j \in Y$, and $z_k \in Z$. You may assume $|X| = |Y| = |Z|$.

A subset $M \subset T$ is a *three-dimensional matching* if each element of $X \cup Y \cup Z$ is contained in at most one of the triples in M. The *maximum three-dimensional matching* problem is to find a three-dimensional matching M of maximum size. (The size of the matching is the number of triples it contains.)

Give a polynomial time algorithm that finds a three-dimensional matching of size at least $\frac{1}{3}$ times the maximum possible size.

(3) Many algorithmic problems involve working with a finite set of points on which some metric has been defined: we have a set P of n points, $P = \{p_1, p_2, \ldots, p_n\}$, and there is a distance $d(p_i, p_j)$ defined for each pair satisfying three properties:

- Non-negativity: $d(p_i, p_j) > 0$ when $p_i \neq p_j$, and $d(p_i, p_i) = 0$.
- Symmetry: $d(p_i, p_j) = d(p_j, p_i)$.
- Triangle inequality: $d(p_i, p_j) + d(p_j, p_k) \geq d(p_i, p_k)$.

Often, even when the points don't come from Euclidean space, it is useful to define a combinatorial notion of the "dimension" of the point set. One of the most useful such definitions is the *doubling dimension*, which is defined as follows.

First, for a point $p_i \in P$ and a distance r, we say that the ball of radius r centered at p_i, denoted $B_r(p_i)$, is the (finite) set of all points in P whose distance to p_i is at most r:

$$B_r(p_i) = \{p_j : d(p_i, p_j) \leq r\}.$$

Now, we say that P with distance function d has *doubling dimension* at most k if for every point $p_i \in P$ and every distance $r > 0$, the points of $B_r(p_i)$ can be covered by the union of at most 2^k balls of radius $r/2$ — that is, if there exist points $p_{i_1}, \ldots, p_{i_\ell}$ such that

$$B_r(p_i) \subseteq \bigcup_{j=1}^{\ell} B_{(r/2)}(p_{i_j})$$

where $\ell \leq 2^k$. The doubling dimension of P with distance function d is then the least k for which this holds.

Now, in a high-dimensional Euclidean space, it's possible to find large sets of points that are all approximately the same distance from one another. Here, we'll show that the same is true, in an approximate sense, for finite point sets using the more combinatorial notion of doubling dimension.

Specifically, let's say that a set of points $P' \subseteq P$ is *c-uniform* if for some distance r, we have

$$r \leq d(p_i, p_j) \leq cr$$

for all pairs of points $p_i, p_j \in P'$. That is, P' is c-uniform if all distances between points in P' are within a factor of c of each other.

Prove the following claim, which shows that large c-uniform sets are present in any point set with large doubling dimension:

> *Claim: Suppose that P with distance function d has doubling dimension greater than k. Then there exists a 4-uniform subset $P' \subseteq P$ such that P' contains more than 2^k elements.*

In other words, if P has large doubling dimension, then it must have a large subset of points that are all at approximately the same distance from each other.