## 10.3 Matroids and approximation

Given a family $\mathcal{F}$ of subsets of some finite set $X$, called the *ground-set*, and a weight function assigning each element $x \in X$ a non-negative real number $w(x)$, the *optimization problem* for $\mathcal{F}$ is to find a member $A \in \mathcal{F}$ whose weight $w(A) = \sum_{x\in A} w(x)$ is maximal. For example, given a graph $G = (V, E)$ with non-negative weights on edges, we might wish to find a matching (a set of vertex-disjoint edges) of maximal weight. In this case $X = E$ is the set of edges, and members of $\mathcal{F}$ are matchings. As it happens in many other situations, the resulting family is *hereditary*, that is, $A \in \mathcal{F}$ and $B \subseteq A$ implies $B \in \mathcal{F}$.

In general, some optimization problems are extremely hard—the so-called "NP-hard problems." In such situations one is satisfied with an "approximative" solution, namely, with a member $A \in \mathcal{F}$ whose weight is at least $1/k$ times the weight of an optimal solution, for some real constant $k \geq 1$.

One of the simplest algorithms to solve an optimization problem is the *greedy algorithm*. It first sorts the elements $x_1, x_2, \ldots, x_n$ of $X$ by weight, heaviest first. Then it starts with $A = \emptyset$ and in the $i$-th step adds the element $x_i$ to the current set $A$ if and only if the result still belongs to $\mathcal{F}$. A basic question is: for what families $\mathcal{F}$ can this trivial algorithm find a good enough solution?

Namely, say that a family $\mathcal{F}$ is *greedy $k$-approximative* if, for every weight function, the weight of the solution given by the greedy algorithm is at least $1/k$ times the weight of an optimal solution. Note that being greedy 1-approximative means that for such families the greedy algorithm always finds an optimal solution.

Given a real number $k \geq 1$, what families are greedy $k$-approximative?

In the case $k = 1$ (when greedy is optimal) a surprisingly tight answer was given by introducing a notion of "matroid." This notion was motivated by the following "exchange property" in linear spaces: If $A, B$ are two sets of linearly independent vectors, and if $|B| > |A|$, then there is a vector $b \in B \setminus A$ such that the set $A \cup \{b\}$ is linearly independent.

Now let $\mathcal{F}$ be a family of subsets of some finite set $X$; we call members of $\mathcal{F}$ *independent sets*. A *$k$-matroid* is a hereditary family $\mathcal{F}$ satisfying the following *$k$-exchange property*: For every two independent sets $A, B \in \mathcal{F}$, if $|B| > k|A|$ then there exists $b \in B \setminus A$ such that* $A + b$ is independent (belongs to $\mathcal{F}$). Matroids are $k$-matroids for $k = 1$.

Matroids have several equivalent definitions. One of them is in terms of maximum independent sets. Let $\mathcal{F}$ be a family of subsets of $X$ (whose members we again call independent sets), and $Y \subseteq X$. An independent set $A \in \mathcal{F}$ is a *maximum independent* subset of $Y$ (or a *basis* of $Y$ in $\mathcal{F}$) if $A \subseteq Y$ and $A + x \notin \mathcal{F}$ for all $x \in Y \setminus A$. A family is *$k$-balanced* if for every subset $Y \subseteq X$

* Here and in what follows, $A + b$ will stand for the set $A \cup \{b\}$.

and any two of its maximum independent subsets $A, B \subseteq Y$ we have that $|B| \leq k|A|$.

**Lemma 10.7.** *A hereditary family is $k$-balanced if and only if it is a $k$-matroid.*

*Proof.* ($\Leftarrow$) Let $Y \subseteq X$, and let $A, B \subseteq Y$ be two sets in $\mathcal{F}$ that are maximum independent subsets of $Y$. Suppose that $|B| > k|A|$. Then by the $k$-exchange property, we can add some element $b$ of $B \setminus A$ to $A$ and keep the result $A + b$ in $\mathcal{F}$. But since $A$ and $B$ are both subsets of $Y$, the set $A + b$ is also a subset of $Y$ and thus $A$ is not *maximum* independent in $Y$, a contradiction.

($\Rightarrow$) We will show that if $\mathcal{F}$ does *not* satisfy the $k$-exchange property, then it is not $k$-balanced. Let $A$ and $B$ be two independent sets such that $|B| > k|A|$ but no element of $B \setminus A$ can be added to $A$ to get a result in $\mathcal{F}$. We let $Y$ be $A \cup B$. Now $A$ is a maximum independent set in $Y$, since we cannot add any of the other elements of $Y$ to it. The set $B$ may not be a maximum independent set in $Y$, but if it isn't there is some subset $B'$ of $Y$ that contains it and is maximum independent in $Y$. Since this set is at least as big as $B$, it is strictly bigger than $k|A|$ and we have a violation of the $k$-balancedness property. $\square$

For $k = 1$, the ($\Leftarrow$) direction of the following theorem was proved by Rado (1942), and the ($\Rightarrow$) direction by Edmonds (1971).

**Theorem 10.8.** *A hereditary family is greedy $k$-approximative if and only if it is a $k$-matroid.*

*Proof.* ($\Leftarrow$) Let $\mathcal{F}$ be a $k$-matroid over some ground-set $X$. Fix an arbitrary weight function, and order the elements of the ground-set $X$ according to their weight, $w(x_1) \geq w(x_2) \geq \ldots \geq w(x_n)$. Let $A$ be the solution given by the greedy algorithm, and $B$ an optimal solution. Our goal is to show that $w(B)/w(A) \leq k$.

Let $Y_i := \{x_1, \ldots, x_i\}$ be the set of the first $i$ elements considered by the greedy algorithm. The main property of the greedy algorithm is given by the following simple claim.

**Claim 10.9.** For every $i$, the set $A \cap Y_i$ is a maximum independent subset of $Y_i$.

*Proof.* Suppose that the independent set $A \cap Y_i$ is not a maximum independent subset of $Y_i$. Then there must exist an element $x_j \in Y_i \setminus A$ (an element *not* chosen by the algorithm) such that the set $A \cap Y_i + x_j$ is independent. But then $A \cap Y_{j-1} + x_j$ (as a subset of an independent set) is also independent, and should have been chosen by the algorithm, a contradiction. $\square$

Now let $A_i := A \cap Y_i$. Since $A_i \setminus A_{i-1}$ is either empty or is equal to $\{x_i\}$,

$$w(A) = w(x_1)|A_1| + \sum_{i=2}^{n} w(x_i)(|A_i| - |A_{i-1}|)$$
$$= \sum_{i=1}^{n-1}(w(x_i) - w(x_{i+1}))|A_i| + w(x_n)|A_n| \,.$$

Similarly, letting $B_i := B \cap Y_i$, we get

$$w(B) = \sum_{i=1}^{n-1}(w(x_i) - w(x_{i+1}))|B_i| + w(x_n)|B_n| \,.$$

Using the inequality $(a + b)/(x + y) \le \max\{a/x, b/y\}$ we obtain that $w(B)/w(A)$ does not exceed $|B_i|/|A_i|$ for some $i$. By Claim 10.9, the set $A_i$ is a maximum independent subset of $Y_i$. Since $B_i$ is also a (not necessarily maximum) independent subset of $Y_i$, the $k$-balancedness property implies that $|B_i| \le k|A_i|$. Hence, $w(B)/w(A) \le |B_i|/|A_i| \le k$, as desired.

($\Rightarrow$) We will prove that if our family $\mathcal{F}$ fails to satisfy the $k$-exchange property, then there is some weight function on which the greedy algorithm fails to approximate an optimal solution by a factor of $1/k$.

Suppose there are two sets $A$ and $B$ in $\mathcal{F}$, with $|B| > k|A|$, such that no element of $B \setminus A$ can be added to $A$ while keeping the result in $\mathcal{F}$. Let $m = |A|$. Take any two positive numbers $a$ and $b$ such that $0 < a - b \le 1/k$. Define the weight function as follows: elements in $A$ have weight $m + a$, elements in $B \setminus A$ have weight $m + b$, and other elements have weight 0. Then the greedy algorithm tries elements of weight $m + a$ first, gets all $m$ of them, but then is stuck because no element of weight $m + b$ fits; hence, the total score of the greedy algorithm is $m(m + a)$. But the optimum is at least the total weight $(m + b)|B| \ge (m + b)(km + 1)$ of elements in $B$. Thus, the greedy algorithm can $(1/k)$-approximate this particular optimization problem only if $(m + b)(km + 1) \le km(m + a)$, or equivalently, if $k(a - b) \ge 1 + b/m$. But this is impossible because $a - b \le 1/k$ and $b > 0$. $\square$

When trying to show that a given family is a $k$-matroid, the following somewhat easier to verify property, suggested by Mestre (2006), is often useful. We say that a family $\mathcal{F}$ is $k$-*extendible* if for every sets $A \subset B \in \mathcal{F}$ and for every element $x \notin B$ the following holds: If the set $A + x$ is independent then the set $B + x$ can be made independent by removing from $B$ at most $k$ elements not in $A$, that is, there exists $Y \subseteq B \setminus A$ such that $|Y| \le k$ and the set $B \setminus Y + x$ is independent.

**Lemma 10.10.** *Every $k$-extendible hereditary family is a $k$-matroid.*

*Proof.* Given two independent sets $A$ and $B$ with $|B| > k|A|$, we need to find an element $z \in B \setminus A$ such that the set $A + z$ is independent. If $A \subset B$ then we are done since all subsets of $B$ are independent. Suppose now that $A \not\subseteq B$. The idea is to pick an element $x \in A \setminus B$ and apply the $k$-extendibility property to the sets $C := A \cap B$ and $D := B$ to find a subset $Y \subseteq D \setminus C = B \setminus A$ with at most $k$ elements such that the set $B' = B \setminus Y + x$ is independent. If $A$ is still not a subset of $B'$, then repeat the same procedure. Since, due to the condition $Y \subseteq B \setminus A$, at any step none of the already added elements of $A$ are removed, after at most $|A \setminus B|$ steps we will obtain an independent set $B'$ such that $A \subseteq B'$. From $|B| > k|A|$, we have that $|B \setminus A| > k|A \setminus B|$. Since in each step at most $k$ elements of $B$ are removed, at least one element $z \in B \setminus A$ must remain in $B'$, that is, $A$ is a *proper* subset of $B'$. But then the set $A + z$ is independent, because $B'$ is such, and we are done. □

In the case of matroids ($k = 1$) we also have the converse.

**Lemma 10.11.** *Every matroid is* 1*-extendible.*

*Proof.* Let $\mathcal{F}$ be a matroid. Given sets $A \subset B \in \mathcal{F}$ and an element $x \notin B$ such that the set $A + x$ is independent, we need to find an element $y \in B \setminus A$ such that $B - y + x$ is independent. If necessary, we can repeatedly apply the matroid property to add elements of $B \setminus A$ to $A$ until we get a subset $A'$ such that $A \subseteq A' \subset B$, $A' + x \in \mathcal{F}$ and $|A' + x| = |B|$. Since $x \notin B$, this implies that $B \setminus A'$ consists of just one element $y$. But then $B - y + x = A' + x$ belongs to $\mathcal{F}$, as desired. □

It can be shown (see Exercise 10.12) that for $k \geq 2$ the converse of Lemma 10.10 does not hold, that is, not every $k$-matroid is $k$-extendible. Still, together with Theorem 10.8, Lemma 10.10 gives us a handy tool to show that some unrelated optimization problems can be approximated quite well by using the trivial greedy algorithm.

*Example 10.12* (Maximum weight $f$-matching). Given a graph $G = (V, E)$ with non-negative weights on edges and degree constraints $f : V \to \mathbb{N}$ for vertices, an $f$-*matching* is a set of edges $M$ such that for all $v \in V$ the number $\deg_M(v)$ of edges in $M$ incident to $v$ is at most $f(v)$. The corresponding optimization problem is to find an $f$-matching of maximal weight.

In this case we have a family $\mathcal{F}$ whose ground-set is the set $X = E$ of edges of $G$ and $f$-matchings are independent sets (members of $\mathcal{F}$). Note that $\mathcal{F}$ is already not a matroid when $f(v) = 1$ for all $v \in V$: if $A = \{a, b\}$ and $B = \{\{c, a\}, \{b, d\}\}$ are two matchings, then $|B| > |A|$ but no edge of $B$ can be added to $A$. We claim that this family is 2-extendible, and hence, is a 2-matroid.

To show this, let $A + x$ and $B$ be any two $f$-matchings, where $A \subset B$ and $x = \{u, v\}$ is an edge not in $B$. If $B + x$ is an $f$-matching, we are done. If not, then $\deg_B(u) = f(u)$ or $\deg_B(v) = f(v)$ (or both). But we know that $\deg_A(u) < f(u)$ and $\deg_A(v) < f(v)$, for otherwise $A + x$ would not be an $f$-matching. Thus, we can remove at most two edges of $B$ not in $A$ so that the resulting graph plus the edge $x$ forms a $f$-matching.

*Example 10.13* (Maximum weight traveling salesman problem). We are given a complete directed graph with non-negative weights on edges, and we must find a maximum weight Hamiltonian cycle, that is, a cycle that visits every vertex exactly once. This problem is very hard: it is a so-called "NP-hard" problem. On the other hand, using Theorem 10.8 and Lemma 10.10 we can show that the greedy algorithm can find a Hamiltonian cycle whose weight is at least one third of the maximum possible weight of a Hamiltonian cycle.

The ground-set $X$ of our family $\mathcal{F}$ in this case consists of the directed edges of the complete graph. A set is independent if its edges form a collection of vertex-disjoint paths or a Hamiltonian cycle. It is enough to show that $\mathcal{F}$ is 3-extendible.

To show this, let $A+x$ and $B$ be any two members of $\mathcal{F}$, where $A \subset B$ and $x = (u, v)$ is an edge not in $B$. First remove from $B$ the edges (if any) out of $u$ and into $v$. There can be at most two such edges, and neither of them can belong to $A$ since otherwise $A+(u, v)$ would not belong to $\mathcal{F}$. If we add $(u, v)$ to $B$ then every vertex has in-degree and out-degree at most one. Hence, the only reason why the resulting set may not belong to $\mathcal{F}$ is that there may be a non-Hamiltonian cycle which uses $(u, v)$. But then there must be an edge in the cycle, not in $A$, that we can remove to break it: if all edges, except for $(u, v)$, of the cycle belong to $A$, then $A + (u, v)$ contains a non-Hamiltonian cycle and could not belong to $\mathcal{F}$. Therefore we need to remove at most three edges in total.

## 10.4 The Kruskal–Katona theorem

A *neighbor* of a binary vector $v$ is a vector which can be obtained from $v$ by flipping one of its 1-entries to 0. A *shadow* of a set $A \subseteq \{0, 1\}^n$ of vectors is the set $\partial(A)$ of all its neighbors. A set $A$ is *$k$-regular* if every vector in $A$ contains exactly $k$ 1-entries. Note that in this case $\partial(A)$ is $(k-1)$-regular.

A basic question concerning shadows is the following one: What can one say about $|\partial(A)|$ in terms of the total number $|A|$ of vectors in a $k$-regular set $A$?

In general one cannot improve on the trivial upper bound $|\partial(A)| \leq k|A|$. But what about *lower* bounds? The question is non-trivial because one and the same vector with $k-1$ ones may be a neighbor of up to $n-k+1$ vectors in $A$. Easy counting shows that

$$|\partial(A)| \geq \frac{k}{n-k+1}|A| = \frac{|A|}{\binom{n}{k}}\binom{n}{k-1}.$$

This can be shown by estimating the number $N$ of pairs $(u, v)$ of vectors such that $v \in A$ and $u$ is a neighbor of $v$. Since every $v \in A$ has exactly $k$ neighbors, we have that $N = k|A|$. On the other hand, every vector $u$