18.997 Topics in Combinatorial Optimization	April 29th, 2004
Lecture 21	
Lecturer: Michel X. Goemans	*Scribe: Mohammad Mahdian*

1 The Lovasz splitting-off lemma

Lovasz's splitting-off lemma states the following.

Theorem 1 *Let $G = (V \cup \{s\}, E)$ be a graph such that*

$$\forall \emptyset \neq U \subsetneq V : \quad d(U) \geq k, \tag{1}$$

where $d(U)$ denotes the number of edges between U and $\bar{U}$, and $k \geq 2$. Also, assume that $d(s)$ (the degree of the vertex s) is even. Then for every $(s,t) \in E$, there exists $(s,u) \in E$ such that the graph $G' = (V \cup s, E \setminus \{(s,t),(s,u)\} \cup \{(t,u)\})$ also satisfies the condition (1).

Proof: Let S denote the set of neighbors of s in G (i.e., $S = \{u \in V : (s,u) \in E\}$). Fix a $t \in S$. We would like to show that there exists a $u \in S$ such that condition (1) holds for the graph $G' = (V \cup s, E \setminus \{(s,t),(s,u)\} \cup \{(t,u)\})$. For the sake of contradiction, assume this does not hold. This means that for every $u \in S$, there exists a set U, $\emptyset \neq U \subsetneq V$, such that $d(U) \leq k+1$ and $u, t \in U$ (See Figure 1). In other words, the collection of all sets U with $d(U) \leq k+1$ and $t \in U$ covers S. Let $\mathcal{C}$ be a collection of maximal sets U with $d(U) \leq k+1$ and $t \in U$ that covers S.

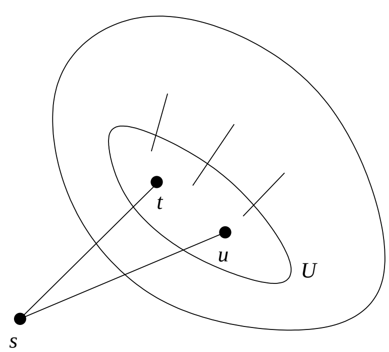

Figure 1: A set U with $d(U) \leq k+1$ and $u \in U$

For every $U \in \mathcal{C}$, we have $d(U) \leq k+1$ and $d(U \cup \{s\}) \geq k$ (the latter inequality holds because $d(U \cup \{s\}) = d(V \setminus U) \geq k$ by (1)). Therefore,

$$1 \geq d(U) - d(U \cup \{s\}) = d(s,U) - d(s, V \setminus U),$$

and so $d(s, V \setminus U) + 1 \geq d(s,U)$. On the other hand, $d(s, V \setminus U) + d(s,U)$ is equal to the degree of s, which is an even number. Thus, $d(s, V \setminus U)$ and $d(s,U)$ have the same parity. Therefore, $d(s, V \setminus U) \geq d(s,U)$. In other words, $d(s,U) \leq \frac{1}{2} d(s)$. This, together with the fact that $t \in U$ for

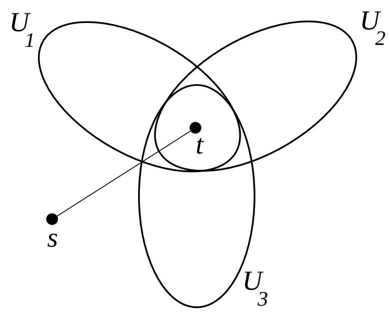

Figure 2: Three sets U_1, U_2, U_3 satisfying properties (2)

every $U \in \mathcal{C}$, shows that two of the sets $U \in \mathcal{C}$ are not enough to cover S (i.e., $U_i \cup U_j \neq S$ for every $U_i, U_j \in \mathcal{C}$). Therefore, $\mathcal{C}$ must contain at least three sets U_1, U_2, U_3 such that

$$\begin{aligned} t &\in U_1 \cap U_2 \cap U_3 \\ U_1 \setminus (U_2 \cup U_3) &\neq \emptyset \\ U_2 \setminus (U_1 \cup U_3) &\neq \emptyset \\ U_3 \setminus (U_1 \cup U_2) &\neq \emptyset. \end{aligned} \tag{2}$$

See Figure 2. We now use the following inequality which is a consequence of the *three-way submodularity* of the function d.

$$\begin{aligned} d(U_1) + d(U_2) + d(U_3) \quad \geq \quad & d(U_1 \cap U_2 \cap U_3) \\ & + d(U_1 \setminus (U_2 \cup U_3)) \\ & + d(U_2 \setminus (U_1 \cup U_3)) \\ & + d(U_3 \setminus (U_1 \cup U_2)). \end{aligned} \tag{3}$$

It is straightforward to check all cases for an edge e and show that in each case, e is counted at least as many times on the left-hand side as it is counted on the right-hand side. This proves the above inequality. In fact, there is at least one edge st that is counted three times on the left-hand side, but only once on the right-hand side. Therefore, we can strengthen inequality (3) by adding a $+2$ to its right-hand side. Since every term on the left-hand side of (3) is at most $k+1$ (by the definition of U_i's) and every term on the right-hand side is at least k (by assumption (1) on the graph G and properties (2)), the above inequality implies:

$$3k + 3 \geq 4k + 2 \Rightarrow k \leq 1.$$

This gives us a contradiction since K was assumed to be at least 2. $\square$

2 Submodular function minimization

In the rest of this lecture, we sketch an algorithm for submodular function minimization. This is from Chapter 45 of Lex Schrijver's book.

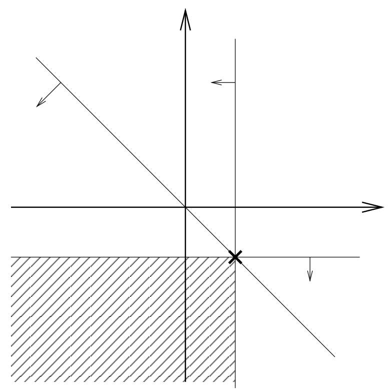

Figure 3: The extended polymatroid EP_f

Problem Statement. Given an oracle for a function $f : 2^S \mapsto \mathbb{Z}$, find a set $U \subseteq S$ that minimizes $f(U)$ over all subsets of S. We assume, without loss of generality, that $f(\emptyset) = 0$, otherwise we can minimize the function $f(U) - f(\emptyset)$ instead of f.

This problem has many applications. As an example, consider the matroid intersection problem that we discussed in previous lectures. We showed that the convex hull of the of the intersection of two matroids is the set of all vectors x such that for every $U \subseteq S$, $x(U) \le r_1(U)$ and $x(U) \le r_2(U)$, where r_1 and r_2 are rank functions of the matroids. Therefore, we can optimize over the intersection of two matroids by solving a linear program with the above constraints. It is not obvious how to solve this linear program, since it is of exponential size. However, we can get polynomial-time separation by minimizing $r_i(U) - x(U)$ over all $U \subseteq S$, and checking if the minimum is non-negative (for $i = 1, 2$). This can be done using an algorithm that solves the submodular function minimization, since $r_1 - x$ is a submodular function.

Notice that it is not obvious that minimizing a submodular function given by an oracle is possible in polynomial time. Clearly, without the assumption of submodularity, it is not possible to find the minimum of the function before calling the oracle on all 2^n points on which the function is defined. However, in the rest of this lecture we will sketch an algorithm due to Lex Schrijver that solves this problem for submodular functions in polynomial time.

We start by defining two polyhedra related to a submodular function f. The first polyhedron is called the *extended polymatroid* associated with f, and is defined as follows:

$$EP_f = \{x \in \mathbb{R}^S : \ x(U) \le f(U), \ \forall\, U \subseteq S\}.$$

Notice that this definition does not require $x \ge 0$. As an example, if $S = \{1, 2\}$ and f is defined by $f(\emptyset) = 0, f(\{1\}) = 1, f(\{2\}) = -1, f(\{1, 2\}) = 0$, then the extended polymatroid EP_f is the shaded area in Figure 3.

Prior to the algorithm of Schrijver, there was a polynomial-time (but not strongly polynomial-time) algorithm for submodular function minimization based on the ellipsoid algorithm and the polyhedron EP_f.

We define the second polyhedron, which is called the *base polyhedron*, as follows:

$$B_f = \{x \in \mathbb{R}^S : \ x(S) = f(S), \ x(U) \le f(U), \ \forall\, U \subset S\}.$$

For example, for the function in the previous example, the polyhedron B_f consists of one point that is marked by a cross in Figure 3.

Our goal is the following.

Goal. Find a set $U \subseteq S$ and a vector $x \in B_f$ such that

$$x(v) \leq 0 \qquad \forall v \in U \tag{4}$$
$$x(v) \geq 0 \qquad \forall v \notin U \tag{5}$$
$$x(U) = f(U) \tag{6}$$
$$\tag{7}$$

Claim 2 *If we can find a set U and vector $x \in B_f$ satisfying properties (4)–(6), then U is the set that minimizes $f(U)$.*

Proof: This is because for every set $W \subset S$,

$$f(U) = x(U) \leq x(W) \leq f(W),$$

where the first equality follows from property (6), the second inequality follows from (4) and (5), and the third inequality is a consequence of $x \in B_f$. $\square$

It is not clear how one can *prove* that a vector x belongs to B_f, since B_f is defined by exponentially many inequalities. We do this by expressing x as a convex combination of elements that are "obviously" in B_f. Such elements are defined below. In fact, these elements are extreme points (and the only extreme points) of B_f, but we do not need this fact in our proof.

Choose a total order $\prec$ on S. For every $v \in S$, we define $v_\prec = \{w \in S : \ w \prec v\}$. The vector $b^\prec \in \mathbb{R}^S$ is defined by

$$b^\prec(v) = f(v_\prec \cup \{v\}) - f(v_\prec).$$

Claim 3 *For every total order $\prec$ on S, $b^\prec \in B_f$.*

Proof: By the definition of $b^\prec$, $b^\prec(S)$ is a telescopic sum that is equal to $f(S) - f(\emptyset) = f(S)$. Now, we prove that for every $U \subseteq S$, $b^\prec(U) \leq f(U)$. We can prove this by induction on the size of U. If $|U| = 0$, then the statement is trivial. Otherwise, let v be the maximal element of U (with respect to $\prec$), and apply the induction hypothesis on $U \setminus \{v\}$. This gives us $b^\prec(U \setminus \{v\}) \leq f(U \setminus \{v\})$. Since v is the maximal element of U, we have $U \subseteq v_\prec \cup \{v\}$. Therefore, by the submodularity of f, we have $b^\prec(v) = f(v_\prec \cup \{v\}) - f(v_\prec) \leq f(U) - f(U \setminus \{v\})$. By adding this inequality with the previous inequality we obtain $b^\prec(U) = b^\prec(U \setminus \{v\}) + b^\prec(v) \leq f(U)$. $\square$

We will find a vector x satisfying properties (4)–(6) and express the vector x as a convex combination of $b^\prec$'s, thereby showing that $x \in B_f$. We do this by starting from an arbitrary x that can be written as a convex combination of $b^\prec$'s, and modify x (along with the expression that gives x as a convex combination of $b^\prec$'s) until there exists a U such that (x, U) satisfies the desired properties.

Suppose we have a vector x that can be written as $x = \sum_{i=1}^{k} \lambda_i b^{\prec_i}$, where $\lambda_i > 0$ for all i. Since $B_f \subseteq \mathbb{R}^S = \mathbb{R}^n$ and all points in B_f must satisfy an equality, the dimension of B_f is at most $n-1$, and hence x can be expressed as a convex combination of n extreme points. So, we assume $k \leq n$.

For every $\prec$ and every $v \in S$, $b^{\prec}(v_{\prec}) = f(v_{\prec})$. We call the set $v_{\prec}$ a *prefix* (also known as a *lower ideal*) of $\prec$. Therefore, if $U \subset S$ is a prefix of $\prec_i$ for every $1 \le i \le k$, then $x(U) = \sum_{i=1}^{k} \lambda_i b^{\prec_i}(U) = \sum_{i=1}^{k} \lambda_i f(U) = f(U)$. Thus, if we can find a set U that is a prefix of $\prec_i$ for every $1 \le i \le k$ and satisfies $x(v) \le 0$ for $v \in U$ and $x(v) \ge 0$ for $v \notin U$, then we are done. This motivates the following definition: Let $D = (S, A)$ be a directed graph on the set of vertices S with the arc set $A = \{(u,v) : u \prec_i v \text{ for some } 1 \le i \le k\}$. By this definition, a set U is a prefix of every $\prec_i$ if and only if $\delta^{in}(U) = \emptyset$ in D.

Now, let $\mathcal{P} = \{v : \ x(v) > 0\}$ and $\mathcal{N} = \{v : \ x(v) < 0\}$. We consider two cases:

Case 1. D has no directed path from $\mathcal{P}$ to $\mathcal{N}$. In this case, let U be the set of vertices v such that there is a path from v to some vertex of $\mathcal{N}$. Therefore, U contains $\mathcal{N}$ but nothing from $\mathcal{P}$, and is a prefix for every $\prec_i$. Therefore, (x, U) satisfy the properties (4)–(6), and we are done.

Case 2. There is a directed path from $\mathcal{P}$ to $\mathcal{N}$. In this case, we change either x or the way x is expressed as a convex combination of $b^{\prec}$'s. Pick s and t on the path from $\mathcal{P}$ to $\mathcal{N}$ such that $t \in \mathcal{N}$, $s \notin \mathcal{N}$, and there is an arc from s to t in D (details of the selection rule is omitted). We would like to change x or its representation to kill the path from $\mathcal{P}$ to $\mathcal{N}$. We can do this either by removing t from $\mathcal{N}$ (i.e., increasing x_t), or by removing the arc (s,t) from D. The arc (s,t) is present because $s \prec_i t$ for some i. We focus on one such i, and will try to get s closer to t in $\prec_i$.

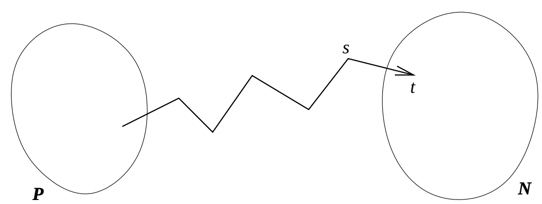

Figure 4: Path from $\mathcal{P}$ to $\mathcal{N}$ and the vertices s and t

Let χ^t denote the unit vector along along the coordinate t (and similarly for χ^s). We show that for some $\delta \ge 0$, $x + \delta(\chi^t - \chi^s) \in B_f$, and moreover, we can write $x + \delta(\chi^t - \chi^s)$ as a convex combination in which s is *closer* to t. More precisely, we use the following lemma as a subroutine.

Lemma 4 *Given $\prec$, s, and t, express the vector $b^{\prec} + \delta(\chi^t - \chi^s)$ for* some $\delta \ge 0$ *as a convex combination of $b^{\prec_{s,u}}$ for $u \in (s,t]_{\prec} = \{u : \ s \prec u \preceq t\}$, where $\prec_{s,u}$ is the total order that is obtained from $\prec$ by moving u before s.*

Proof: We assume that $b^{\prec} = 0$. This assumption is without loss of generality, because we can replace $f(U)$ by $f(U) - b^{\prec}(U)$ and apply the argument on this new function. By the submodularity of f, we have

$$b^{\prec_{s,u}}(v) = \begin{cases} b^{\prec}(v) = 0 & \text{if } v \prec s \text{ or } v \succ u \\ \le 0 & \text{if } s \preceq v \prec u \\ \ge 0 & \text{if } v = u. \end{cases}$$

The following table shows the pattern of non-negative and non-positive entries in $b^{\prec_{s,u}}$'s, for every $u \in (s,u]_{\prec}$. In this table, $-$ denotes a non-positive entry, $+$ denotes a non-negative entry, and 0 denotes an entry that is zero.

$$
\begin{array}{c|cccccccc}
u & s & & & & & & & t \\
\hline
s & & & & & & & & \\
 & 0 & - & + & 0 & & & & \\
 & 0 & - & - & + & 0 & & & \\
 & 0 & - & - & - & + & 0 & & 0 \\
 & & \vdots & & & & \ddots & & \\
t & 0 & - & - & - & \dots & & - & +
\end{array}
$$

If the non-negative entry $(+)$ on one row is zero (i.e., $b^{\prec_{s,u}}(u) = 0$ for some u), then all other entries of that row must also be zero, since the sum of the entries in each row must be the same as the sum of the entries in $b^{\prec}$ (since they are both equal to $f(S)$), which is zero. Therefore, in this case, we can take $\delta = 0$ and use $b^{\prec_{s,u}}$ as the desired convex combination. The other case is when $b^{\prec_{s,u}}(u) > 0$ for every $u \in (s, t]_{\prec}$. In this case, we can start from the last row of the table (i.e., the vector $b^{\prec_{s,t}}$), and for every row of the table, from the row before t to the row after s, add a multiple of the row to the current vector so that the $+$ entry cancels out the corresponding entry in the current vector. At the end, we will obtain a vector that has only one negative entry at position s and one positive entry at position t. Furthermore, since the sum of all entries should be zero, the absolute value of these two entries are equal. This means that for some $\delta > 0$, we can write $\delta(\chi^t - \chi^s)$ as a convex combination of $b^{\prec_{s,t}}$'s. $\square$

We iterate the procedure in the above lemma. Intuitively, every time we get s closer to t in at least one of $\prec_i$'s, without changing other $\prec_i$'s. We might also increase x_t. Therefore, after a finite number of iterations, we will either remove the arc (s, t) from D, or will remove t from $\mathcal{N}$. For how s and t are chosen and the analysis of the running time of this procedure, see Lex Schrijver's book.