X2TE2109: Convex Optimization and Its Applications in Signal Processing

Handout 4: Some Applications of Linear Programming

Instructor: Anthony Man–Cho So March 22, 2022

1 Introduction

The theory of LP has found many applications in various disciplines. In this lecture, we will consider some of those applications and see how the machineries developed in previous lectures can be used to obtain some interesting results.

2 An Approximation Algorithm for Vertex Cover

The theory of LP has been employed very successfully in the design of approximation algorithms in recent years (see, e.g., [3]). We say that an algorithm $\mathcal{A}$ is an **α-approximation algorithm** for a minimization problem $\mathcal{P}$ if for every instance I of $\mathcal{P}$ it delivers a solution that is at most α times the optimum in polynomial time. Clearly, we have $\alpha \geq 1$, and the closer it is to 1, the better. In a similar fashion, one may define the notion of an α-approximation algorithm for a maximization problem. In this section we study the so-called *vertex cover* problem and see how the theory of LP can be used to obtain a 2-approximation algorithm for it.

To begin, consider a simple undirected graph $G = (V, E)$, where each vertex $v_i \in V$ has an associated cost $c_i \in \mathbb{R}_+$. A **vertex cover** of G is a subset $S \subset V$ such that for every edge $(v_i, v_j) \in E$, at least one of the endpoints belongs to S. We are interested in finding a vertex cover S of G of minimal cost.

Now, let $x_i \in \{0, 1\}$ be a binary variable indicating whether v_i belongs to the vertex cover S or not (i.e., $x_i = 1$ iff $v_i \in S$). Then, the minimum-cost vertex cover problem can be formulated as the following integer program:

$$\begin{array}{rcll} v^* & := & \text{minimize} & c^T x = \sum_{i=1}^{|V|} c_i x_i \\ & & \text{subject to} & x_i + x_j \geq 1 \qquad \text{for } (v_i, v_j) \in E, \\ & & & x \in \{0,1\}^{|V|}. \end{array} \tag{1}$$

Again, (1) is a non-convex optimization problem. Consider now the following LP relaxation of (1):

$$\begin{array}{rcll} v_r^* & := & \text{minimize} & c^T x \\ & & \text{subject to} & x_i + x_j \geq 1 \quad \text{for } (v_i, v_j) \in E, \\ & & & x \geq \mathbf{0}. \end{array} \tag{2}$$

Clearly, we have $v_r^* \leq v^*$. Suppose that x' is an optimal solution to (2). The question now is, can we convert x' into a solution x'' that is feasible for (1) such that $c^T x'' \leq \alpha v_r^*$ for some $\alpha > 0$? Note that if this is possible, then we would obtain an α-approximation algorithm for the minimum-cost vertex cover problem, since we have $c^T x'' \leq \alpha v_r^* \leq \alpha v^*$. As the following theorem shows, the answer to the above question is indeed yes:

Theorem 1 *(cf. [2]) Let $P \subset \mathbb{R}^{|V|}$ be the polyhedron defined by the following system:*

$$\begin{cases} x_i + x_j & \geq & 1 \quad \textit{for } (v_i, v_j) \in E, \\ x & \geq & \mathbf{0}. \end{cases}$$

Suppose that x is an extreme point of P. Then, we have $x_i \in \{0, 1/2, 1\}$ for $i = 1, \ldots, |V|$.

Proof Let $x \in P$, and consider the sets

$$\begin{aligned} U_{-1} &= \{i \in \{1, \ldots, |V|\} : x_i \in (0, 1/2)\}, \\ U_1 &= \{i \in \{1, \ldots, |V|\} : x_i \in (1/2, 1)\}. \end{aligned}$$

For $i = 1, \ldots, |V|$ and $k \in \{-1, 1\}$, define

$$y_i = \begin{cases} x_i + k\epsilon & \text{if } i \in U_k, \\ x_i & \text{otherwise} \end{cases} \quad ; \quad z_i = \begin{cases} x_i - k\epsilon & \text{if } i \in U_k, \\ x_i & \text{otherwise.} \end{cases}$$

By definition, we have $x = (y + z)/2$. If either U_{-1} or U_1 is non-empty, then we may choose $\epsilon > 0$ to be sufficiently small so that $y, z \in P$, and that x, y, z are all distinct. It follows that $U_k = \emptyset$ for $k \in \{-1, 1\}$ if x is an extreme point of P. $\square$

Corollary 1 *There exists a 2-approximation algorithm for the minimum-cost vertex cover problem.*

Proof We first solve the LP (2) and obtain an optimal extreme point solution x'. Now, by Theorem 1, all the entries of x' belong to $\{0, 1/2, 1\}$. Hence, the vector x'' defined by

$$x''_i = \begin{cases} x'_i & \text{if } x'_i = 0 \text{ or } 1, \\ 1 & \text{if } x'_i = 1/2 \end{cases} \qquad \text{for } i = 1, \ldots, |V|$$

is feasible for (1). Moreover, it has objective value $c^T x'' \leq 2c^T x' = 2v_r^* \leq 2v^*$. This completes the proof. $\square$

3 Blind Separation of Non-Negative Sources

In various image processing applications, a problem of fundamental interest is that of separating non-negative source signals in a blind fashion. For simplicity, consider the following linear mixture model:

$$x[\ell] = As[\ell] \quad \text{for } \ell = 1, \ldots, L, \tag{3}$$

where $s[\ell] \in \mathbb{R}_+^n$ is the ℓ-th source vector, $x[\ell] \in \mathbb{R}^m$ is the ℓ-th observation vector, and $A \in \mathbb{R}^{m \times n}$ is a mixing matrix describing the input-output relationship. Our goal here is to extract the source vectors $s[1], \ldots, s[L] \in \mathbb{R}^n$ from the observation vectors $x[1], \ldots, x[L] \in \mathbb{R}^m$ without knowledge of the mixing matrix $A \in \mathbb{R}^{m \times n}$. Note that such a task is not well-defined. For instance, if the pair $\left(\{s[\ell]\}_{\ell=1}^L, A\right)$ satisfies (3), then so does the pair $\left(\{s[\ell]/c\}_{\ell=1}^L, cA\right)$ for any constant $c > 0$. Thus, it is necessary to impose additional assumptions on the model (3). Towards that end, let us first rewrite (3) as

$$x^i = \sum_{j=1}^n a_{ij} s^j \quad \text{for } i = 1, \ldots, m, \tag{4}$$

where $x^i = (x^i[1], \ldots, x^i[L]) \in \mathbb{R}^L$ is the i-th observed signal and $s^j = (s^j[1], \ldots, s^j[L]) \in \mathbb{R}_+^L$ is the signal from the j-th source. We shall make the following assumptions:

(a) The mixing matrix $A \in \mathbb{R}^{m \times n}$ has full column rank (in particular, $m \geq n$) and satisfies

$$\sum_{j=1}^{n} a_{ij} = 1 \quad \text{for } i = 1, \ldots, m.$$

Moreover, the number of observations L satisfies $L \gg m$.

(b) Each source signal is *locally dominant*; i.e., for each source $j \in \{1, \ldots, n\}$, there exists an (unknown) undex $\ell(j) \in \{1, \ldots, L\}$ such that $s^j[\ell(j)] > 0$ and $s^k[\ell(j)] = 0$ for all $k \neq j$.

The above assumptions are satisfied in a wide variety of settings; see, e.g., [1] for a more detailed discussion. Now, observe that from Assumption (a), we have $x^i \in \text{aff}(\{s^1, \ldots, s^n\})$ for $i = 1, \ldots, m$. In fact, more can be said:

Proposition 1 *Under Assumption (a), we have* $\text{aff}(\{x^1, \ldots, x^m\}) = \text{aff}(\{s^1, \ldots, s^n\})$.

Proof Suppose that $x \in \text{aff}(\{x^1, \ldots, x^m\})$. Then, there exist $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$ such that

$$x = \sum_{i=1}^{m} \alpha_i x^i \quad \text{and} \quad \sum_{i=1}^{m} \alpha_i = 1.$$

Substituting this into (4) yields

$$x = \sum_{i=1}^{m} \sum_{j=1}^{n} \alpha_i a_{ij} s^j = \sum_{j=1}^{n} \beta_j s^j,$$

where $\beta_j = \sum_{i=1}^{m} \alpha_i a_{ij}$, for $j = 1, \ldots, n$. Using Assumption (a), we have

$$\sum_{j=1}^{n} \beta_j = \sum_{i=1}^{m} \alpha_i \left(\sum_{j=1}^{n} a_{ij} \right) = 1.$$

It follows that $x \in \text{aff}(\{s^1, \ldots, s^n\})$.

Conversely, suppose that $x \in \text{aff}(\{s^1, \ldots, s^n\})$. Then, there exist $\beta_1, \ldots, \beta_n \in \mathbb{R}$ such that

$$x = \sum_{j=1}^{n} \beta_j s^j \quad \text{and} \quad \sum_{j=1}^{n} \beta_j = 1.$$

Consider now the following system of linear equations in $\alpha = (\alpha_1, \ldots, \alpha_m) \in \mathbb{R}^m$:

$$\beta_j = \sum_{i=1}^{m} \alpha_i a_{ij} \quad \text{for } j = 1, \ldots, n. \tag{5}$$

By Assumption (a), we have $m \geq n$, which implies that the above system is solvable. Upon summing (5) over $j = 1, \ldots, n$, we have

$$1 = \sum_{j=1}^{n} \beta_j = \sum_{i=1}^{m} \alpha_i \left(\sum_{j=1}^{n} a_{ij} \right) = \sum_{i=1}^{m} \alpha_i.$$

This shows that $x \in \text{aff}(\{x^1, \dots, x^m\})$, as desired. $\square$

The upshot of Proposition 1 is that although we do not know the source vectors $s^1, \dots, s^n \in \mathbb{R}^L_+$, their affine hull is completely determined by the observed vectors $x^1, \dots, x^m \in \mathbb{R}^L$. Such an observation can help us in recovering the source vectors $s^1, \dots, s^n \in \mathbb{R}^L_+$. Indeed, consider the polyhedron

$$\mathcal{P} = \text{aff}(\{s^1, \dots, s^n\}) \cap \mathbb{R}^L_+. \tag{6}$$

Since $s^i \in \mathbb{R}^L_+$, we have $s^i \in \mathcal{P}$ for $i = 1, \dots, n$. As the following result shows, the source vectors can be recovered by considering the vertices of $\mathcal{P}$:

Proposition 2 *Under Assumption (b), we have $\mathcal{P} = \text{conv}(\{s^1, \dots, s^n\})$. Moreover, the vertices of $\mathcal{P}$ are $\{s^1, \dots, s^n\}$.*

Proof Suppose that $x \in \mathcal{P}$. Then, there exist $\beta_1, \dots, \beta_n \in \mathbb{R}$ such that

$$\mathbf{0} \le x = \sum_{j=1}^{n} \beta_j s^j \quad \text{and} \quad \sum_{j=1}^{n} \beta_j = 1.$$

For each $j \in \{1, \dots, n\}$, we have $0 \le x[\ell(j)] = \beta_j s^j[\ell(j)]$ by Assumption (b), which implies that $\beta_j \ge 0$. Thus, we have $x \in \text{conv}(\{s^1, \dots, s^n\})$.

Conversely, suppose that $x \in \text{conv}(\{s^1, \dots, s^n\})$. Since $s^1, \dots, s^n \in \mathbb{R}^L_+$, it is clear that $x \in \mathcal{P}$.

Lastly, for a fixed $k \in \{1, \dots, n\}$, suppose that $s^k = \theta x^1 + (1-\theta)x^2$, where $x^1, x^2 \in \mathcal{P}$ and $\theta \in (0, 1)$. Then, there exist $\alpha^1_1, \dots, \alpha^1_n \in \mathbb{R}_+$ and $\alpha^2_1, \dots, \alpha^2_n \in \mathbb{R}_+$ such that

$$s^k = \sum_{j=1}^{n} \left(\theta\alpha^1_j + (1-\theta)\alpha^2_j\right) s^j \quad \text{and} \quad \sum_{j=1}^{n} \alpha^1_j = \sum_{j=1}^{n} \alpha^2_j = 1.$$

Now, by Assumption (b), we have

$$s^k[\ell(k)] = \left(\theta\alpha^1_k + (1-\theta)\alpha^2_k\right) s^k[\ell(k)],$$

which implies that $\theta\alpha^1_k + (1-\theta)\alpha^2_k = 1$. This is possible if and only if $\alpha^1_k = \alpha^2_k = 1$, or equivalently, $x^1 = x^2 = s^k$. It follows that s^k is a vertex of $\mathcal{P}$, as desired. $\square$

To obtain a representation of $\mathcal{P}$ that is more amenable to computation, we first observe that $\dim\left(\text{aff}(\{s^1, \dots, s^n\})\right) = n - 1$. Thus, by Proposition 1 and Assumption (b), there are $n-1$ linearly independent vectors in the collection $\{x^i - x^1\}_{i=2}^m$. Now, let $v^1, \dots, v^{L-n+1} \in \mathbb{R}^L$ be a basis of $\text{span}(\{x^2 - x^1, \dots, x^m - x^1\})^\perp$, which can be computed by the Gram-Schmidt orthogonalization procedure. Then, we have

$$\text{aff}(\{x^1, \dots, x^m\}) = \left\{x \in \mathbb{R}^L : (v^i)^T x = (v^i)^T x^1 \quad \text{for } i = 1, \dots, L-n+1\right\},$$

which, together with (6), implies that

$$\mathcal{P} = \left\{x \in \mathbb{R}^L_+ : (v^i)^T x = (v^i)^T x^1 \quad \text{for } i = 1, \dots, L-n+1\right\}.$$

References

[1] T.-H. Chan, W.-K. Ma, C.-Y. Chi, and Y. Wang. A Convex Analysis Framework for Blind Separation of Non–Negative Sources. *IEEE Transactions on Signal Processing*, 56(10):5120–5134, 2008.

[2] G. L. Nemhauser and L. E. Trotter, Jr. Properties of Vertex Packing and Independence System Polyhedra. *Mathematical Programming*, 6:48–61, 1974.

[3] V. V. Vazirani. *Approximation Algorithms*. Springer–Verlag, Berlin/Heidelberg, 2001.