6.1 Fast approximation schemes for packing and covering LP's

In this lecture we show how to remove the dependence on the width in the multiplicative weights update/randomized fictitious play method, in the case when the entries of the matrix are non-negative. Our presentation follows mostly the framework of Koufogiannakis and Young [KY07]. In the next lecture will show how to derive the result of Garg and Könemann [GK98] for multicommodity flows from this framework.

6.1.1 Packing and covering LP's

A packing-covering pair of LP's looks like the following

$$\max\{c^T x : \; A^T x \leq b, \; x \geq 0\} = \min\{b^T y : \; Ay \geq c, \; y \geq 0\}, \tag{6.1}$$

where A is a *non-negative* $m \times n$ matrix, b and c are *non-negative* vectors.

It will simplify matters to assume that $c = e$ and $b = e$ are the vectors of all ones. This assumption, which we will make in the rest of this lecture, entails no loss of generality if we look for relative approximation errors, made precise in the following definition.

Definition 6.1. *Let $\varepsilon > 0$ be a constant. An ε-approximation for (6.1) is a primal-dual feasible pair (x, y) such that $b^T y \leq (1 + \epsilon)c^T x$.*

In this lecture we consider the following problem.

ϵ-APPROXIMATION OF PACKING-COVERING LP'S

Input: A matrix $A \in \mathbb{R}_+^{m\times n}$, vectors $b \in \mathbb{R}_+^m$, $c \in \mathbb{R}_+^n$, and a desired accuracy $\varepsilon \in (0, 1)$

Output: A primal-dual feasible pair (x, y)

Objective: (x, y) is an ε-approximation

Exercise 1. *Show that for the purpose of obtaining ε-approximations for a packing-covering pair of LP's, one can assume, without loss of generality, that the objective functions and the right-hand sides are the vectors of all ones.*

A *fully polynomial-time approximation scheme* for (6.1) is one that finds an ε-approximation in time polynomial in n, m and $\frac{1}{\varepsilon}$. We proceed to present such a scheme. To get an intuition first, it is instructive to recall the *Lagrangian relaxation* of the two LP's:

$$\max_{A^T x \leq e,\ x \geq 0} e^T x = \max_{x \geq 0} \min_{y \geq 0} (e^T x + \sum_{j=1}^{n} y_j (1 - x^T A^j)) \tag{6.2}$$

$$\min_{Ay \geq e,\ y \geq 0} e^T y = \min_{y \geq 0} \max_{x \geq 0} (e^T y + \sum_{i=1}^{m} x_i (1 - A_i y)) \tag{6.3}$$

where, as usual, A_i an A^j denote the ith row and the jth column of A, respectively. Thus, we can think of x_i and y_j as *penalties* that we pay for violating constrains i in the dual and j in the primal, respectively. To put this into a procedure that gradually constructs a primal-dual feasible pair, we start by initializing $X(0) = 0$ and $Y(0) = 0$; we think of $X_i(T)$ (respectively, $Y_j(t)$) as the total penalty we pay for violating primal constraint i (respectively, dual constraint j) upto time t. At time t, the values of $X(t)$ and $Y(t)$ will be updated to reflect the current violation of the constraints. The intuition we get from (6.2) is that, if $X(t-1)^T A^j$ is "large", then $Y_j(t)$ should be "large". Similarly, we intuitively get from (6.3) that if $A_i Y(t-1)$ is large then $X_i(t)$ should be "small". This seems, in a sense, similar to fictitious play, which is not surprising, if we recall the fact that a matrix game is equivalent to a pair of packing-covering LP's (c.f Exercise 2 of Lecture 1).

Thus, we may try to apply randomized fictitious play in the same way we did for matrix games. More precisely, at each time t, we increase the penalties $X_i(t)$ and $Y_j(t)$ with probabilities proportional to $p_i(t) = e^{-\frac{\varepsilon A_i Y(t-1)}{2}}$ and $q_j(t) = e^{\frac{\varepsilon X(t-1)^T A^j}{2}}$, respectively. If we update $X_i(t)$ and $Y_j(t)$ by the same value $\delta(t)$ (at time time t), we guarantee at the end of the procedure that the primal and dual objectives are the same (since we assume that both b and c are the vectors of all ones). However, we have to scale $X(t)$ and $Y(t)$ to guarantee feasibility. It is clear that the scaling factors for $X(t)$ and $Y(t)$, should be respectively,

$$M(t) = \max_{j \in [n]} \{X(t)^T A^j\} \quad \text{and} \quad m(t) = \min_{i \in [m]} \{A_i Y(t)\}. \tag{6.4}$$

As the case for matrix games, we can show that after enough time, namely $t \geq \frac{6\rho^2 \ln(2nm)}{\varepsilon^2}$ where $\rho = \max_{i,j} a_{ij}$, we have (with high probability)

$$\frac{X(t)^T A^j}{t} \leq \frac{A_i Y(t)}{t} + \varepsilon \quad \text{for all } i, j,$$

or in other words, $M(t) \leq m(t) + \varepsilon t$. Furthermore, if we assume that the update values are uniform, $\delta(t') = 1$ for all t', we get at the end that $e^T X(t) = e^T Y(t) = t$, and thus the scaled objective values satisfy $\frac{e^T y(t)}{e^T x(t)} = \frac{M(t)}{m(t)} \leq 1 + \varepsilon e^T y(t)$, where $x(t) = \frac{X(t)}{M(t)}$ and $y(t) = \frac{Y(t)}{m(t)}$ are the final solutions.

There are two problems with this approach. First, the approximation guarantee is $1 + \varepsilon$ only if $e^T y(t) \leq 1$. Second, and more critical, the running time depends on ρ which is unavoidable unless we try somehow to utilize the non-negativity of the matrix A.

The following procedure, due to Koufogiannakis and Young [KY07], which builds on ideas from [GK95] and [GK98] fixes both problems. The basic new ingredients to the standard approach are:

- The use of a nonuniform update value $\delta(t)$, which essentially replaces the explicit scaling step by ρ; this would guarantee that the change vectors $A\Delta Y(t)$ and $\Delta X(t)^T A$ are bounded by 1 in each component, and as we shall see below, this is the only requirement needed to apply the necessary inequalities to bound the potential increase.

- Since the updates are no longer uniform, the distributions from which we sample, are changed to maintain the (technical) property that the expected changes $A\Delta Y(t)$ and $\Delta X(t)^T A$ are still proportional to $\frac{q}{|q|}$ and $\frac{p}{|p|}$, where q and p are given by

$$p_i(t) = (1-\varepsilon)^{A_i Y(t-1)} \text{ and } q_j(t) = (1+\varepsilon)^{X(t-1)^T A^j}, \tag{6.5}$$

(for purely technical reasons, we use $(1-\varepsilon)^{(\cdot)}$ and $(1+\varepsilon)^{(\cdot)}$ instead of $e^{-\frac{\varepsilon}{2}(\cdot)}$ and $e^{\frac{\varepsilon}{2}(\cdot)}$, respectively). Again, this is needed for an essential cancellation step in the analysis of the potential increase. In particular, the procedure below samples row i and column j, respectively, with probabilities $\tilde{p}_i$ and $\tilde{q}_j$, which are driven from the vectors p and q, and the matrix A in a way that will be specified later (see Exercise 3).

This will be enough to fix the second problem (that is, the dependence of the running time on ρ). To fix the first problem, the algorithm runs until the value of $M(t)$ becomes large enough. To guarantee that the procedure terminates, we only apply the sampling and update steps to the current *active* list $L(t) = \{i \in [m] \ : \ A_i Y(t-1) < T\}$ of dual constraints. As we shall see below, if no constraint is active then we have already reached the termination condition.

Algorithm 1 FPTAS FOR PACKING-COVERING LPS

1. $X(0) := 0$; $Y(0) := 0$; $t := 1$; and $T := \frac{\ln(2nm)}{\varepsilon^2}$
2. **while** $M(t) < T$ **do**
3. $\quad t := t+1$
4. $\quad$ Pick $i \in L(t)$ and $j \in [n]$ with probabilities $\tilde{p}_i(t)$ and $\tilde{q}_j(t)$
5. $\quad X_i(t) := X_i(t-1) + \delta(t)$; $Y_j(t) := Y_j(t-1) + \delta(t)$
6. **return** $(x(t), y(t)) = (\frac{X(t)}{M(t)}, \frac{Y(t)}{m(t)})$

As usual, we denote by $\Delta X(t) = \delta(t)e_i$ and $\Delta Y(t) = \delta(t)e_j$, respectively, the changes in the vectors X and Y in step 5. We can state, in an abstract way, the requirements we need to impose on the distributions $\tilde{p}$ and $\tilde{q}$ and the change $\delta(t)$ as follows

(i) $A_i \Delta Y(t) \leq 1$ for all $i \in L(t)$, and $\Delta X(t)^T A^j \leq 1$ for all $j \in [n]$;

(ii) $\mathbb{E}[A\Delta Y(t)] = \alpha \frac{q(t)}{|q(t)|}$ and $\mathbb{E}[\Delta X(t)^T A] = \alpha \frac{p(t)}{|p(t)|}$, for some constant $\alpha > 0$;

(iii) $\max\{\max_{i \in L(t)} A_i \Delta Y(t), \max_{j \in [n]} \Delta X(t)^T A^j\} \geq \frac{1}{2}$.

In the above and in what follows, we assume that $p_i(t) = 0$ if $i \notin L(t)$.

The facts stated in the following exercise will be useful in the analysis.

Exercise 2. *Show that*

(I) *for all* $\epsilon \in (0,1)$, $x \in [0,1]$, $(1+\epsilon)^x \le 1+\epsilon x$ *and* $(1-\epsilon)^x \le 1-\epsilon x$;

(II) *for all* $\epsilon \in [0,1)$, $\frac{\ln(1+\epsilon)}{\ln\frac{1}{1-\epsilon}} \ge 1-\epsilon$;

(III) *for all* ϵ, $\epsilon \le \ln\frac{1}{1-\epsilon}$;

(IV) *for all* $\epsilon \in [0,1]$, $\frac{\ln(1+\epsilon)}{\epsilon} \ge 1-\epsilon$.

Theorem 6.2. *Assume that $\tilde{p}(t)$, $\tilde{q}(t)$ and $\delta(t)$ satisfy (i), (ii), and (iii) for all t. Then the above procedure terminates in at most $2(n+m)T$ iterations with a primal-dual feasible pair. At termination, it holds with probability at least $\frac{1}{2}$ that*

$$e^T x(t) \ge (1-2\varepsilon)e^T y(t). \tag{6.6}$$

Proof: Note that, at any iteration t, we maintain the invariant, $e^T X(t) = e^T Y(t)$. Thus, if at a certain time t we have $m(t) \ge T$ then $M(t) \ge T$ also holds, for otherwise, $m(t) \ge T > M(t)$, and hence, $\frac{e^T x(t)}{e^T y(t)} = \frac{m(t)}{M(t)} > 1$, in contradiction with weak duality.

Now, the termination time follows from (iii) since, for each column $j \in [n]$, $X(t)^T A^j$ can be updated, as the maximizer in (iii), at most $2T$ times before $M(t)$ becomes at least T; similarly, for each row $i \in L(t)$, $A_i Y(t)$ can be updated (as the maximizer in (iii)) at most $2T$ times before i to goes out of the active list.

To show (6.6), we analyze, as usual, the increase in the potential function $\Phi(t) = |p(t+1)||q(t+1)|$, conditioned on $X(t-1)$ and $Y(t-1)$:

$$\begin{aligned} |p(t+1)| &= \sum_{i=1}^{m} p_i(t+1) = \sum_{i \in L(t+1)} (1-\varepsilon)^{A_i Y(t)} = \sum_{i \in L(t+1)} (1-\varepsilon)^{A_i(Y(t-1)+\Delta Y)} \\ &= \sum_{i \in L(t+1)} p_i(t)(1-\varepsilon)^{A_i \Delta Y}. \end{aligned} \tag{6.7}$$

By (i), $0 \le A_i \Delta Y \le 1$, and hence Fact (I) of Exercise 2 implies that $(1-\varepsilon)^{A_i\Delta Y} \le 1 - \varepsilon A_i \Delta Y$. Plugging this into (6.7), we get

$$|p(t+1)| \le \sum_{i \in L(t+1)} p_i(t)(1-\varepsilon A_i \Delta Y) \le \sum_{i \in L(t)} p_i(t)(1-\varepsilon A_i \Delta Y) = |p(t)| \left(1 - \frac{\varepsilon p(t)^T A \Delta Y}{|p(t)|}\right).$$

Similarly, we can derive that

$$|q(t+1)| \le |q(t)| \left(1 + \frac{\varepsilon \Delta X^T A q(t)}{|q(t)|}\right).$$

Thus,

$$\begin{aligned} \Phi(t) &= |p(t+1)||q(t+1)| \le |p(t)||q(t)| \left[1 + \varepsilon\left(\frac{\Delta X^T A q(t)}{|q(t)|} - \frac{p(t)^T A \Delta Y}{|p(t)|}\right) \right. \\ &\qquad \left. - \varepsilon^2 \frac{\Delta X^T A q(t)}{|q(t)|} \cdot \frac{p(t)^T A \Delta Y}{|p(t)|}\right] \\ &\le \Phi(t-1) \left[1 + \varepsilon\left(\frac{\Delta X^T A q(t)}{|q(t)|} - \frac{p(t)^T A \Delta Y}{|p(t)|}\right)\right], \end{aligned}$$

where the last inequality follows from the non-negativity of A, ΔX and ΔY. Now taking the expectation with respect to ΔX and ΔY, and using (ii), we get

$$\mathbb{E}[\Phi(t)] \leq \Phi(t-1)\left[1+\varepsilon\left(\frac{\alpha p(t)^T A q(t)}{|p(t)||q(t)|} - \frac{\alpha p(t)^T A q(t)}{|p(t)||q(t)|}\right)\right] = \Phi(t-1),$$

and taking the expectations with respect to $X(t-1)$ and $Y(t-1)$ we get that $\mathbb{E}[\Phi(t)] \leq \mathbb{E}[\Phi(t-1)]$. Iterating, we get $\mathbb{E}[\Phi(t)] \leq \Phi(0) = nm$, and thus with probability at least $\frac{1}{2}$, $\Phi(t) \leq 2nm$. This gives

$$(1-\varepsilon)^{A_i Y(t)} \cdot (1+\varepsilon)^{X(t)^t A^j} \leq 2nm \quad \text{for all } i \in L(t+1) \text{ and } j \in [n],$$

and after some algebraic manipulation we get

$$X(t)^t A^j \frac{\ln(1+\varepsilon)}{\ln\frac{1}{1-\varepsilon}} \leq A_i Y(t) + \frac{\ln(2nm)}{\ln\frac{1}{1-\varepsilon}} \quad \text{for all } i \in L(t+1) \text{ and } j \in [n]. \tag{6.8}$$

Using Facts (II) and (III) of Exercise 2, (6.8) reduces to

$$(1-\varepsilon)X(t)^t A^j \leq A_i Y(t) + \frac{\ln(2nm)}{\varepsilon} \quad \text{for all } i \in L(t+1) \text{ and } j \in [n],$$

or, in other words,

$$(1-\varepsilon)M(t) \leq A_i Y(t) + \varepsilon T \quad \text{for all } i \in L(t+1). \tag{6.9}$$

By the stopping criterion, we have $M(t) \geq T$, where t is the stopping time. Thus,

$$(1-2\varepsilon)M(t) \leq A_i Y(t) \text{ for } i \in L(t+1) \text{ (by (6.9))} \tag{6.10}$$

$$T \leq A_i Y(t) \text{ for } i \notin L(t+1) \text{ (by definition of } L(t+1)). \tag{6.11}$$

Since $M(t) \leq T+1$ by assumption (i) (since $M(t-1) < T$), we get by (6.11), for $i \notin L(t+1)$,

$$A_i Y(t) \geq M(t) - 1 \geq (1-2\varepsilon)M(t) \text{ (since } M(t) \geq T = \tfrac{\ln(2nm)}{\varepsilon^2} \geq \tfrac{1}{2\varepsilon}). \tag{6.12}$$

(6.10) and (6.12) imply that $m(t) \geq (1-2\varepsilon)M(t)$ and this implies (6.6) since $e^T X(t) = e^T Y(t)$. ☐

The following exercise establishes the existence of distributions $\tilde{p}(t)$ and $\tilde{q}(t)$ and change $\delta(t)$, satisfying the condition (i), (ii) and (iii) above.

Exercise 3. *Let $A \in \mathbb{R}_+^{m\times n}$ be a matrix with rows $A_1, \ldots, A_m$ and columns $A^1, \ldots, A^n$. Assume none of the rows or columns is identically 0. For $i \in [m]$ and $j \in [n]$, define $v_i = \max_j\{a_{ij}\}$, $u_j = \max_i\{a_{ij}\}$, and $\delta_{ij} = \frac{1}{v_i+u_j}$. Let $p = (p_1, \ldots, p_m)$ and $q = (q_1, \ldots, q_n)$ be two strictly positive vectors. Denote respectively by pv and qu the vectors $(p_1v_1, \ldots, p_mv_m)$ and $(q_1u_1, \ldots, q_nu_n)$. For a vector u denote by $|u|$ the sum of the components of u. Let $s = |pv||q| + |p||qu|$. Suppose that we pick an i and j, respectively, with probabilities $\tilde{p}_i$ and $\tilde{q}_j$, defined as follows: With probability $\frac{|pv||q|}{s}$ let $\tilde{p}_i = \frac{p_iv_i}{|pv|}$, $\tilde{q}_j = \frac{q_j}{|q|}$, and with probability $\frac{|p||qu|}{s}$ let $\tilde{p}_i = \frac{p_i}{|p|}$, $\tilde{q}_j = \frac{q_ju_j}{|qu|}$. Let $\Delta x \in \mathbb{R}^m$ be the vector with all components equal to zero except at position i in which $(\Delta x)_i = \delta_{ij}$, and $\Delta y \in \mathbb{R}^n$ be the vector with all components equal to zero except at position j in which $(\Delta y)_j = \delta_{ij}$. Show the following*

(i) *for all i, $A_i\Delta y \leq 1$, and for all j, $(\Delta x)^T A^j \leq 1$;*

(ii) $\mathbb{E}[A\Delta y] = \alpha\frac{q}{|q|}$ *and* $\mathbb{E}[A^T\Delta x] = \alpha\frac{p}{|p|}$, *where* $\alpha = \frac{|p||q|}{s}$;

(iii) $\max\{\max_i\{A_i\Delta y\}, \max_j\{(\Delta x)^T A^j\}\} \geq \frac{1}{2}$.

Bibliography

[GK95] Michael D. Grigoriadis and Leonid G. Khachiyan. A sublinear-time randomized approximation algorithm for matrix games. *Operations Research Letters*, 18(2):53 – 58, 1995.

[GK98] Naveen Garg and Jochen Könemann. Faster and simpler algorithms for multicommodity flow and other fractional packing problems. In *39th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 300–309, 1998.

[KY07] Christos Koufogiannakis and Neal E. Young. Beating simplex for fractional packing and covering linear programs. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 494–504, 2007.