Solving General LPs via MWU

Recall the MWU algorithm:

MWU(ε):

$w_i^1 \leftarrow 1$ for every expert i

For $t = 1 \dots T$:

- Follow expert i with prob. $p_i^t = \dfrac{w_i^t}{\sum_j w_j^t}$
- After ℓ^t is revealed, $w_i^{t+1} \leftarrow w_i^t (1 - \varepsilon \ell_i^t)$ for all i

$$L_{MWU} := \sum_{t=1}^{T} \langle p^t, \ell^t \rangle \quad \text{our total expected loss}$$

$$L_i := \sum_{t=1}^{T} \ell_i^t \quad \text{expert } i\text{'s total loss}$$

Thm: For $0 < \varepsilon \le \frac{1}{2}$, MWU($\varepsilon$) guarantees that

$$L_{MWU} \le L_i + \varepsilon \underbrace{T}_{\#\text{days}} + \frac{\lg \overbrace{n}^{\#\text{experts}}}{\varepsilon}. \quad \forall i$$

Let $L = (L_1, \dots, L_n)$ be the loss vector for all n experts.

Cor: For any distribution of experts $z = (z_1, \dots, z_n)$ (i.e., $z_i \ge 0$ & $\sum z_i = 1$) we have

$$L_{MWU} \le \langle L, z \rangle + \varepsilon T + \frac{\lg n}{\varepsilon}.$$

$\hookrightarrow \sum_{i=1}^{n} L_i \cdot z_i \ge \sum_{i=1}^{n} L_j z_i = L_j$

$\hookrightarrow j$ is best expert

Let $\Delta_n = \{ x \in \mathbb{R}^n_{\ge 0} \mid \mathbb{1}^T x = 1 \}$ be the set of all prob. distributions over n objects.

Feasibility Problem

Suppose we are given $A \in [-1,1]^{m \times n}$ and $b \in \mathbb{R}^m$

Find $x \in \Delta_n$

s.t. $Ax \le b$

or return that there is no such x.

Exercise: If we can solve this feasibility problem we can solve any LP. (Hint: use binary search).

We'll solve an approximate version of the feasibility problem. Given $A \in [-1,1]^{m \times n}$ and $b \in \mathbb{R}^m$

Find $x \in \Delta_n$

s.t. $Ax \le b + 2\varepsilon \mathbb{1}$

or return that no $x \in \Delta_n$ satisfies $Ax \le b$

Thm: We can solve this feasibility problem in

$$O\Big(\underbrace{\text{nnz}(A)}_{\#\text{nonzeros in } A} \frac{\lg n}{\varepsilon^2}\Big).$$

A whack-a-Mole Algorithm

- Keep fixing a violated constraint.

Algorithm:

Initialize $w^1 \leftarrow \mathbb{1} \in \mathbb{R}^n$

Maintain $x^t = \dfrac{w^t}{W^t}$ where $W^t = \sum_{j=1}^{n} w_j^t$

For $t = 1 \dots T$ where $T = \dfrac{\lg n}{\varepsilon^2}$:

- If $Ax^t \le b + 2\varepsilon \mathbb{1}$ return $x = x^t$.
- Else there is a constraint $i_t \in [m]$ where $A_{(i_t, \cdot)} x^t > b + 2\varepsilon$:

$$w_j^{t+1} \leftarrow w_j^t (1 - \varepsilon A_{i_t j}) \quad \forall j \in [n]$$

Return no feasible x.

Observe that the algorithm runs in the claimed time.

Lemma: If there exists a feasible $x^* \in \Delta_n$ with $Ax \le b$, then the algorithm returns $x \in \Delta_n$ s.t. $Ax \le b + 2\varepsilon \mathbb{1}$.

pf: Suppose for contradiction that the algorithm returns 'no feasible x'.

Observe that the Whack-a-Mole algorithm implements MWU(ε):

- experts = variables
- We choose the dist'n $x^t = w^t / W^t$
- Take the loss vector to be $\ell^t = A_{(i_t, :)}$

On day t,

The loss of MWU is $\langle x^t, \ell^t \rangle = \langle x^t, A_{i_t} \rangle$
$> b_{i_t} + 2\varepsilon$.

But we know that $\langle x^*, \ell^t \rangle \le b_{i_t}$.

Sum over all T days. We get:

$$L_{MWU} - \langle L, x^* \rangle = \sum_{t=1}^{T} \overbrace{\langle x^t, \ell^t \rangle - \langle x^*, \ell^t \rangle}^{> 2\varepsilon}$$

$$> 2\varepsilon T.$$

On the other hand, we know from the no regret bound for MWU that

$$L_{MWU} - \langle L, x^* \rangle \le \varepsilon T + \frac{\lg n}{\varepsilon} \le 2\varepsilon T$$

$$\frac{\lg n}{\varepsilon} \le \varepsilon T \iff \frac{\lg n}{\varepsilon^2} \le T$$

Suppose $A \in [-\rho, \rho]^{m \times n}$, $x \in \tau \cdot \Delta_n$

The same algorithm works but it runs in time

$$O\left(\mathrm{nnz}(A) \cdot \lg n \left(\frac{\rho\tau}{\varepsilon}\right)^2\right)$$

Remark: There are techniques for removing the dependency on ρ & τ. These are called "width independent MWU". But we don't cover them.

$$A \xrightarrow{1/s} A' \longrightarrow A'x' < b' + 2\varepsilon \mathbb{1}$$

A: $[-\rho, \rho]^{m \times n}$ $\quad$ A': $[-1, 1]^{m \times n}$

$$Ax < b + 2\varepsilon s \mathbb{1}$$