1 Best Response Dynamics

While the current outcome is not a Pure Nash equilibirum (PNE), we can pick an arbitrary player i and an arbitrary beneficial deviation s'_i for player i and move to outcome $(s'_i, \mathbf{s_{-i}})$.
Recall that the definition of a potential game is one where there exists a function $\Phi : \mathcal{S} \to \mathbb{R}$ where $\mathcal{S}$ is the finite set of strategies with

$$\Phi(s'_i, s_{-i}) - \Phi(s_i, s_{-i}) = c_i(s'_i, s_{-i}) - c_i(s_i, s_{-i})$$

Proposition 1.1. In a finite potential game from any arbitrary outcome, best-response dynamics converge to a PNE.

Proof. In a best-response dynamics approach, every iteration has $\Phi(\mathbf{s^{t+1}}) < \Phi(\mathbf{s^t})$, i.e. the potential decreases. Unless the $\mathbf{s^t}$ is a PNE, our Φ is lower bounded by $\min_{s\in\mathcal{S}} \Phi(s)$ and hence the process must terminate. □

Definition 1.2 ($\epsilon-$Pure Nash Equilibrium)**.** For $\epsilon \in [0, 1]$, and outcome $\mathbf{s}$ is an $\epsilon-$pure NE if for every agent i and deviations $s'_i \in S_i$

$$c_i(s'_i, s_{-i}) \geq (1-\epsilon) c_i(s_i, s_{-i})$$

An $\epsilon-$best response dynamics is one which permits moves when there is significant improvements (substential lowering of cost or increasing of utility) which is an important factor to for a state to converge to near optimal equilibrium. While a current outcome $\mathbf{s}$ is not an $\epsilon-$PNE, we pick an arbitary player i that has an $\epsilon-$move, i.e. a deviation to s'_i:

$$c_i(s'_i, s_{-i}) < (1-\epsilon) c_i(\mathbf{s})$$

Lemma 1.3. For $x \in (0, 1)$

$$(1-x)^{1/x} \leq (e^{-x})^{1/x} = e^{-1}$$

Theorem 1.4 (Fast convergence of $\epsilon-$Best Response Dynamics)**.** Consider an atomic selfish routing game where:

1. All players have the same source s and destination t vertex.
2. Cost function satisfy the "$\alpha-$bound jump condition"

   $$c_e(x) \leq c_e(x+1) \leq \alpha \cdot c_e(x)$$

   for all edges e.
3. The MaxGain variant of $\epsilon-$BR dynamics is used: in every iteration, amongst all players with an $\epsilon-$move available, the player who can obtain the biggest absolute cost decrease gets to move.

Then an $\epsilon-$PNE is reached in at most

$$\frac{k \cdot \alpha}{\epsilon} \log \frac{\Phi(\mathbf{s^0})}{\Phi_{min}}$$

iterations, where k is the number of agents, $\mathbf{s^0}$ is the initial state of the system.

Proof. Using lemma 1.5 we pick the agent i with the highest cost to get

$$\begin{aligned} \Phi(\mathbf{s}) - \Phi(s'_i, s_{-i}) = c_i(\mathbf{s}) - c_i(s'_i, s_{-i}) &\geq \frac{\epsilon}{\alpha k} \cdot c_i(\mathbf{s}), \quad \text{by Lemma 1.6} \\ &\geq \frac{\epsilon}{\alpha k} \cdot \Phi(\mathbf{s}) \end{aligned}$$

thus we have

$$\left(1 - \frac{\epsilon}{\alpha k}\right) \Phi(\mathbf{s^t}) \geq \Phi(\mathbf{s^{t+1}})$$

thus using Lemma 1.3 we obtain that an $\epsilon-$PNE is reached in $\frac{\alpha k}{\epsilon} \log \frac{\Phi(\mathbf{s^0})}{\Phi_{min}}$ iterations. □

The two lemmas below are the ones used in the proof.

Lemma 1.5. For all $\mathbf{s} \in \mathcal{S}$ there exists an agent such that

$$c_i(s) \geq \frac{\Phi(\mathbf{s})}{k}$$

Proof. Recall that $\Phi(\mathbf{s}) \leq cost(\mathbf{s})$, then pick the agent that realizes the highest cost, $i = \text{argmax}_i\, c_i(\mathbf{s})$, then

$$c_i(\mathbf{s}) \geq \frac{cost(\mathbf{s})}{k} \geq \frac{\Phi(\mathbf{s})}{k}$$

□

Lemma 1.6. Suppose player i is chosen at outcome s by MaxGain $\epsilon-$best response dynamics and he takes the $\epsilon-$ move s'_i,then

$$c_i(\mathbf{s}) - c_i(s'_i, s_{-i}) \geq \frac{\epsilon}{\alpha} c_j(\mathbf{s}) \tag{1}$$

for any other agent j.

Proof. For the case when $j = i$, when $\alpha = 1$, it is exactly the definition of the $\epsilon-$move. Now consider when $i \neq j$ with j having an $\epsilon-$move, by MaxGain dynamics and $\alpha = 1$,

$$c_i(\mathbf{s}) - c_i(s'_i, s_{-i}) \geq c_j(\mathbf{s}) - c_j(s'_j, s_{-j}) > \epsilon \cdot c_j(\mathbf{s})$$

the proof is completed by with the case where j does not have an $\epsilon-$move, which we consider $-$ since s'_i is such a great deviation for player i, why isn't it good for player j? That is

$$c_i(s'_i, s_{-i}) < (1-\epsilon)c_i(\mathbf{s})$$

while

$$c_j(s'_i, s_{-j}) \geq (1-\epsilon)c_j(\mathbf{s})$$

and here we used the condition that the agents have the same source and sink vertex, i.e. they have the same set of strategies. An observation made here is that (s'_i, s_{-i}) and (s'_i, s_{-j}) have at least $k-1$ strategies in common (note that s'_i is played by agent i in the former and agent j in the latter and the $k-2$ players other than i and j are playing fixed strategies.) Thus by the "$\alpha-$bound jump condition", we have

$$c_j(s'_i, s_{-j}) \leq \alpha c_i(s'_i, s_{-i})$$

using the inequality above

$$c_i(\mathbf{s}) - c_i(s'_i, s_{-i}) > \epsilon c_i(\mathbf{s}) \geq \frac{\epsilon}{\alpha} c_j(\mathbf{s})$$

which completes the proof. □

Theorem 1.7. Consider a $(\lambda, \mu)-$cost minimization game with a positive potential function Φ such that $\Phi(\mathbf{s}) \leq \mathit{cost}(\mathbf{s})$ for every outcome $\mathbf{s}$. Let $\mathbf{s^0}, \mathbf{s^1}, \dots, \mathbf{s^T}$ be a sequence generated by MaxGain best response dynamics, $\mathbf{s}^*$ a minimum cost outcome and $1 > \gamma > 0$ is a parameter, Then for all but

$$\frac{k}{\gamma(1-\mu)} \log \frac{\Phi(\mathbf{s^0})}{\Phi_{min}} \tag{2}$$

outcomes $\mathbf{s}^t$ satisfy

$$\mathit{cost}(\mathbf{s}^t) \leq \left(\frac{\lambda}{(1-\mu)(1-\gamma)} \right) \cdot \mathit{cost}(\mathbf{s}^*) \tag{3}$$

Proof.

$$\begin{aligned}
\mathit{cost}(\mathbf{s^t}) &\leq \sum_i c_i(\mathbf{s^t}) \\
&= \sum_i \left[c_i(s_i^*, s_{-i}^t) + \delta_i(\mathbf{s^t}) \right], \quad \delta_i(\mathbf{s^t}) = c_i(\mathbf{s^t}) - c_i(s_i^*, s_{-i}^t) \\
&\leq \lambda \cdot \mathit{cost}(\mathbf{s}^*) + \mu \cdot \mathit{cost}(\mathbf{s^t}) + \sum_i \delta_i(\mathbf{s^t}) \\
\mathit{cost}(\mathbf{s^t}) &\leq \frac{\lambda}{1-\mu} \cdot \mathit{cost}(\mathbf{s}^*) + \frac{1}{1-\mu} \cdot \sum_i \delta_i(\mathbf{s^t}) \qquad (4)
\end{aligned}$$

we shall let $\Delta(\mathbf{s^t}) = \sum_i \delta_i(\mathbf{s^t})$ in the remaining parts of the proof. We shall now define a state $\mathbf{s^t}$ to be bad if it does not satisfy (3) and by (4), when $\mathbf{s^t}$ is bad we get

$$\Delta(\mathbf{s^t}) \geq \gamma(1-\mu) \cdot \mathit{cost}(\mathbf{s^t})$$

By the MaxGain definition and the inequality relating the potential function and cost,

$$\max_i \delta_i(\mathbf{s^t}) \geq \frac{\Delta(\mathbf{s^t})}{k} \geq \frac{\gamma(1-\mu)}{k} \cdot \mathit{cost}(\mathbf{s^t}) \geq \frac{\gamma(1-\mu)}{k} \cdot \Phi(\mathbf{s^t})$$

and we get what we desire as

$$\Phi(\mathbf{s^t}) - \Phi(s_i^*, s_{-i}^t) = c_i(\mathbf{s^t}) - c_i(s_i^*, s_{-i}^t) = \delta_i(\mathbf{s^t})$$

and hence

$$\left(1 - \frac{\gamma(1-\mu)}{k}\right) \Phi(\mathbf{s^t}) \geq \Phi(\mathbf{s^{t+1}}) \tag{5}$$

whenever $\mathbf{s^t}$ is a bad state. The equation in (5) says that for every MaxGain best response dynamics, if the state is bad, the new state $\mathbf{s^{t+1}}$ is smaller than the previous state $\mathbf{s^t}$ by a factor of $1 - \frac{\gamma(1-\mu)}{k}$. By Lemma 1.3, the potential decreases by a factor of e for every $\frac{k}{\gamma(1-\mu)}$ bad states encountered. Thus solving

$$e^{-n}\Phi(\mathbf{s^0}) \geq \Phi_{min}$$

shows (2). □

2 No Regret Learning

Consider a set A of actions with $|A| = n$, then at time $t = 1, 2, \dots, T$

- A decision maker picks a mixed strategy p^t (i.i. a probability distribution function over its actions A)
- An adversary (nature) picks a cost vector $c^t : A \to [0, 1]$
- An action a^t is chosen accordingly to the distribution p^t and the decision maker incurs cost $c^t(a^t)$. The decision maker learns the entire cost vector s^t and not just the realised cost.

Definition 2.1 (Time Average of Regret)**.** The time average of regret of the action sequence $a^1, a^2, ldots, a^T$ with respect to $a \in A$ is

$$\frac{1}{T}\sum_{t=1}^{T} c^t(a^t) - \sum_{t=1}^{T} c^t(a)$$

Definition 2.2 (No Regret Algorithm)**.** Let $\mathcal{A}$ be an online decision making algorithm.

(a) An adversary for $\mathcal{A}$ is a function that takes an input the day t, the history of mixed strategies $p^1, p^2, \ldots, p^t$ produced by $\mathcal{A}$ on the first t days and the realised actions $a^1, a^2, \ldots, a^{t-1}$ of the first $t-1$ days and outputs a cost vector $c^t : A \to [0, 1]$.

(b) An online decision making algorithm has no (external) regret if for every adversary, the expected regret with respect to every action $a \in A$ converges to 0 as $T \to \infty$.

$$R = \frac{1}{t}\left(\sum_{t} c^t(a^t) - \min_{a \in A}\sum_{t} c^t(a)\right) \leq \mathcal{O}(1)$$

Theorem 2.3. There exists simple no-regret algorithms such that the expected regret of every action is

$$\mathcal{O}\left(\sqrt{\frac{\ln n}{T}}\right)$$

An easy consequence of the theorem above is presented in the following collary.

Corollary 2.4. There exists an online learning algorithm such that for every $\epsilon > 0$, has expected regret of at most ϵ after $\mathcal{O}\left(\ln n/\epsilon^2\right)$

2.1 Algorithm

1. Initialize $w^1(a) = 1$ for every $a \in A$.
2. For $t = 1, 2, \ldots, T$
   (a) Play an action according to the distribution $p^t := w^t/\Gamma^t$, where $\Gamma^t = \sum_{a\in A} w^t(a)$ is the sum of the weights.
   (b) Given the cost vector c^t, decrease the weights using the formula

   $$w^{t+1}(a) = w^t(a) \cdot (1-\epsilon)^{c^t(a)}$$

   for every action $a \in A$.

The construction of the algorithm is such that when the cost is 0, then the weight remains the same and when the cost is 1, the weight gets reduced by a factor of $1 - \epsilon$. Considering what happens to the distribution of the weights when we vary ϵ, for small values of ϵ, the weights are slowly eroded and the distribution p^t tends to the uniform distribution (needs a more rigor argument as to why).When ϵ is close to 1, we see that the weight is concentrated on the action that has accumulated the least cost so far. This can be observed by the analysis of the algorithm below. For all $a \in A$, we start with $w^1(a) = 1$

$$\begin{aligned} w^{t+1}(a) &= f(w^t(a), c^t(a)) \\ &= w^t(a) \cdot (1-\epsilon)^{c^t(a)} \\ &= \quad \vdots \\ &= w^1(a) \cdot (1-\epsilon)^{\sum_{k=1}^{t-1} c^k(a)} \end{aligned}$$

Now we would like to connect the expected performance at day t denoted by V^t, the optimal performance (OPT) and $w^t(a)$ together. Let $a^* = \max_{a\in A} w^T(a)$, then

$$\begin{aligned} \Gamma^T = \sum_{a\in A} w^T(a) \geq w^T(a^*) \\ &= w^1(a^*) \cdot (1-\epsilon)^{\sum_{k=1}^{t-1} c^k(a^*)} \\ &= (1-\epsilon)^{\text{OPT}} \end{aligned} \tag{6}$$

the expected performance at time t to be

$$V^t = \sum_{a\in A} p^t(a) \cdot c^t(a) = \sum_{a\in A} \frac{w^t(a)}{\Gamma^t} \cdot c^t(a)$$

with the prelimaries done, we are ready to connect them together. We now try to find a recursive equation relating

Γ^{t+1} and Γ^t:

$$\begin{aligned}
\Gamma^{t+1} &= \sum_{a\ inA} w^{t+1}(a) \\
&= \sum_{a\ inA} w^t(a)(1-\epsilon)^{c^t(a)} \\
&\leq \sum_{a\in A} w^t(a)(1-\epsilon\cdot c^t(a)), \quad \text{by lemma 2.5} \\
&= \Gamma^t \sum_{a\in A} \frac{w^t(a)}{\Gamma^t}(1-\epsilon\cdot c^t(a)) \\
&= \Gamma^t \sum_{a\in A} p^t(a)(1-\epsilon\cdot c^t(a)) \\
&= \Gamma^t \sum_{a\in A} p^t(a) - \epsilon \sum_{a\in A} p^t(a)c^t(a) \\
&= \Gamma^t \left(1-\epsilon V^t\right) \\
\implies \Gamma^{t+1} &\leq \Gamma^1 \prod_{i=1}^{t}(1-\epsilon\cdot V^i) \\
\text{By (6) , } \quad (1-\epsilon)^{\text{OPT}} \leq \Gamma^{t+1} &\leq \Gamma^1 \prod_{i=1}^{t}(1-\epsilon\cdot V^i)
\end{aligned}$$

Taking ln,

$$\begin{aligned}
\text{OPT}\cdot\ln(1-\epsilon) &\leq \ln n + \sum_{i=1}^{t}\ln(1-\epsilon\cdot V^i) \\
\text{OPT}\cdot(-\epsilon-\epsilon^2) &\leq \ln n - \epsilon\sum_{i=1}^{t} V^i, \quad \text{by Lemma 2.6} \\
\sum_{i=1}^{t} V^i &\leq \frac{\ln n}{\epsilon} + (1+\epsilon)\text{OPT} \\
&\leq \text{OPT} + \frac{\ln n}{\epsilon} + \epsilon\text{OPT}
\end{aligned}$$

Here we realize that in the last eqution decreasing ϵ increases $\ln n/\epsilon$ and increasing ϵ increases ϵOPT. Thus to keep the upper bound low, we keep them equal, which then we obtain $\epsilon = \sqrt{\ln n/T}$. Hence the cumulative expected cost of the no regret algorithm is at most $2\sqrt{T\ln n}$ more than the cumulative cost of the optimal.

Lemma 2.5. For $\epsilon \in [0,1]$ and $x \in [0,1]$,

$$(1-\epsilon)^x \leq (1-\epsilon x)$$

Lemma 2.6. If $x \in [0,1/2]$

$$-x - x^2 \leq \ln(1-x) \leq -x$$

Next we show that in a game, i.e. routing game where everyone is using a no regret algorithm, the state of the game converges to a coarse correlated equilibrium. In each time step $t = 1,2,\dots,T$ of no regret dynamics:

1. Each player i chooses simultaneously and independently a mixed strategy p_i^t using a no regret algorithm of their choice.

2. Each player receives a cost vector c_i^t where $c_i^t(s_i)$ is the expected cost of strategy s_i when the other players play their chosen mixed strategies, i.e. $c_i^t(s_i) = \mathbf{E}_{s_{-i}\sim\sigma_{-i}}[c_i(s_i,s_{-i})]$ where $\sigma_{-i} = \prod_{j\neq i}\sigma_j$.

Theorem 2.7. Suppose after T iterations of no-regret dynamics, every player of a cost minimization game has a regret of at most ϵ for each of its strategies. Let $\sigma = \prod_{i=1}^{k} p_i^t$ denote the outcome distribution at time t and $\sigma = \frac{1}{T}\sum_{t=1}^{T}\sigma^t$

the time average history of these distributions. Then σ is an $\epsilon-$approximate coarse correlated equilibrium, in the sense that

$$\mathbf{E}[c_i(\mathbf{s})] \leq \mathbf{E}[c_i(s'_i, s_{-i})] + \epsilon$$

for every player i and unilateral decision s'_i.

Corollary 2.8. Suppose after T iterations of no regret dynamics, player i has expected regret at most R_i for each of its actions. Then the time average expected objective function value $\frac{1}{T}\mathbf{E}_{\mathbf{s}\sim\sigma_i}$ is at most

$$\frac{\lambda}{1-\mu} cost(s^*) + \frac{\sum_{i=1}^{k} R_i}{1-\mu}$$

In particular, as $T \rightarrow \infty, \sum_{i=1}^{k} R_i \rightarrow 0$ and the guarantee to converge to the standard PoA bound $\frac{\lambda}{1-\mu}$.