Lecture 7 (Constantinos Daskalakis)

Problem 1

Anna applies gradient descent-ascent with a small step-size to the following minimax problem:

$$\min_{x\in\mathbb{R}} \max_{y\in\mathbb{R}} (y \cdot (x - 0.5))$$

and start at $x_0 = 0, y_0 = 0$. What does she observe?

- **(a)** The sequence (x_k, y_k), $k = 1, 2, \ldots$ converge to $(0, 0)$. **[False]**
- **(b)** The sequence (x_k, y_k), $k = 1, 2, \ldots$ converge to $(0.5, 0)$. **[False]**
- **(c)** The sequence $(\sum_{j=1}^{k} x_j/k, \sum_{j=1}^{k} y_j/k)$, $k = 1, 2, \ldots$ converges to $(0, 0)$. **[False]**
- **(d)** The sequence $(\sum_{j=1}^{k} x_j/k, \sum_{j=1}^{k} y_j/k)$, $k = 1, 2, \ldots$ converges to $(0.5, 0)$. **[True]**
- **(e)** None of the above. **[False]**

Berta applies the optimistic gradient descent-ascent algorithm (with a small step-size) to the same problem. She starts again from $x_0 = 0$, $y_0 = 0$. What does she observe?

- **(a)** The sequence (x_k, y_k), $k = 1, 2, \ldots$ converge to $(0, 0)$. **[False]**
- **(b)** The sequence (x_k, y_k), $k = 1, 2, \ldots$ converge to $(0.5, 0)$. **[True]**
- **(c)** The sequence $(\sum_{j=1}^{k} x_j/k, \sum_{j=1}^{k} y_j/k)$, $k = 1, 2, \ldots$ converges to $(0, 0)$. **[False]**
- **(d)** The sequence $(\sum_{j=1}^{k} x_j/k, \sum_{j=1}^{k} y_j/k)$, $k = 1, 2, \ldots$ converges to $(0.5, 0)$. **[True]**
- **(e)** None of the above. **[False]**

Note: Gradient descent-ascent is defined through the following update rules:

$$x_{k+1} = x_k - \eta \cdot \nabla_x f(x_k, y_k) \ ,$$
$$y_{k+1} = y_k + \eta \cdot \nabla_y f(x_k, y_k) \ ,$$

where $f(x, y)$ is the objective function and $\eta > 0$ the step-size.

The optimistic gradient descent-ascent algorithm is defined via:

$$x_{k+1} = x_k - 2\eta \cdot \nabla_x f(x_k, y_k) + \eta \cdot \nabla_x f(x_{k-1}, y_{k-1}) \ ,$$
$$y_{k+1} = y_t + 2\eta \cdot \nabla_{\mathrm{y}} f(x_k, y_k) - \eta \cdot \nabla_y f(x_{k-1}, y_{k-1}) \ ,$$

where f is again the objective function and $\eta > 0$ the step-size.