# Written Assignment 2

Due: Friday 02/16/2024 @ 11:59pm EST

## Disclaimer

I encourage you to work together, I am a firm believer that we are at our best (and learn better) when we communicate with our peers. Perspective is incredibly important when it comes to solving problems, and sometimes it takes talking to other humans (or rubber ducks in the case of programmers) to gain a perspective we normally would not be able to achieve on our own. The only thing I ask is that you report who you work with: this is **not** to punish anyone, but instead will help me figure out what topics I need to spend extra time on/who to help. When you turn in your solution (please use some form of typesetting: do **NOT** turn in handwritten solutions), please note who you worked with.

Remember that if you have a partner, you and your partner should submit only **one** submission on gradescope.

### Question 1: Gradient Descent (25 points)

Given two vectors $\hat{\vec{y}}^{(t)} \in \mathbb{R}^n$ and $\vec{y} \in \mathbb{R}^n$, we can measure the distance between them using the following objective function:

$$L\left(\hat{\vec{y}}^{(t)}, \vec{y}\right) = \frac{1}{2}\sum_{i=1}^{n}\left(\hat{y}_i^{(t)} - y_i\right)^2$$

Assume that $\hat{\vec{y}}^{(t)}$ is the output of an Agent at time $t$ and that we wish to improve using gradient descent by minimizing the distance between $\hat{\vec{y}}^{(t)}$ and $\vec{y}$. Calculate the gradient $\nabla_{\hat{\vec{y}}^{(t)}} L$ that we would need to implement in our code to execute the gradient descent algorithm.

We can separate the summation into each dimension. For example, for $i = 7$, we have

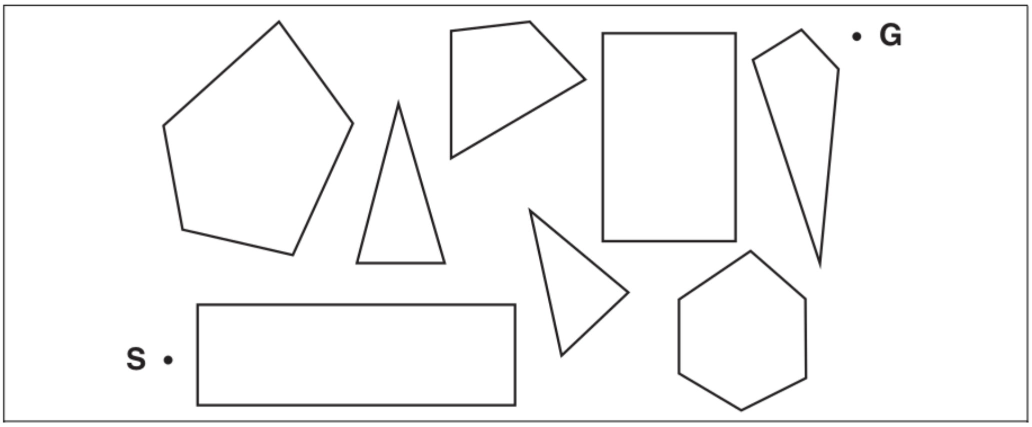
$$\frac{1}{2} \times 2 \times \left(\hat{y}_7^{(t)} - y_7\right)^1 = \hat{y}_7^{(t)} - y_7$$

Thus, our gradient is

$$\left(\hat{y}_1^{(t)} - y_1,\, \hat{y}_2^{(t)} - y_2,\, \hat{y}_3^{(t)} - y_3,\, \ldots,\, \hat{y}_n^{(t)} - y_n\right)$$

**Question 2: Hill Climbing Optimality (25 points)**

Consider the following world (shown from an arial top-down view):




a) Consider discretizing this world (i.e. laying a grid down on top of the world so we see a discrete coordinate system). Assume that the discretization is a fine as required (but not continuous) to preserve the geometry of the shapes present in the world. A *convex* shape is a shape where if we were to draw a line between any two points of the shape, the line would entirely be contained *within* the shape. Given a world that contains convex *obstacles*, is it possible for a vanilla hill climbing agent to get stuck?

b) What if we could prove that the **objective surface** was convex. Would it be possible for a vanilla hill climbing agent to get stuck?

a) **Yes**, because we could be faced with an edge which is perpendicular to the direct line of travel to $G$. In this case, the best we can do is turn 90° to the left or right, neither of which will increase our objective function (they will actually decrease it because the hypotenuese is the longest side of a right triangle; one leg is the direct line from $S$ to $G$ and the other leg is in the direction that we are turning.)

b) **No...** If the objective surface is convex, any local minima are global minima. That is, there is no chance for us to get stuck in a local minima while there is a better state outside of our immediate neighborhood. Even if there were multiple global minima, this would imply that there are several equally desirable states, so the hill climber would find one of them.

**Extra Credit: Gradient Descent and Lipshitz Continuity (50 points)**

Gradient Descent is the continuous version of hill climbing where we use the gradient of the objective function to measure which direction leads uphill (we can then decide whether or not to follow the gradient uphill or downhill). Let us consider going downhill (i.e. we are choosing to *minimize*). The trouble is that gradient descent is a local search algorithm, meaning that it is not guaranteed to:

- Converge at all. There may not be any local minima to find.
- Arrive at the global minima if one exists.

A function $f$ is *Lipschitz Continuous* with constant $k > 0$ if $\forall \vec{x}, \vec{y} \quad ||f(\vec{x}) - f(\vec{y})||_2 \leq k||\vec{x} - \vec{y}||_2$. Think of it this way: for every pair of points, $f$ is bounded on how fast it can change by $k$. Here is a good gif that visualizes lipschitz continuity.

If we know that our objective function $f : \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable, *and* we know that the gradient of $f$ (i.e. $\nabla f$) is Lipschitz continuous with (minimum) constant $c > 0$, prove that **not only** is gradient descent **guaranteed** to converge (for specific values of the step size $\eta$), but that it **will** converge if $\eta \leq \frac{2}{c}$.

Hint: you may find it useful to use a 2nd degree taylor polynomial of $f$ centered around $f(\vec{x})$, which can be expressed as:

$$f(\vec{y}) \leq f(\vec{x}) + \nabla f(\vec{x})^T(\vec{y} - \vec{x}) + \frac{1}{2}(\vec{y} - \vec{x})^T \nabla^2 f(\vec{x})(\vec{y} - \vec{x})$$

Local min is global min for convex surfaces, so we don't have any risk of falling into a small valley (not possible to get stuck)