AI511 – Machine Learning
Week 1 – Problem Solving

Aniruddh Kishore Budhgavi
aniruddh.kishore@iiitb.ac.in

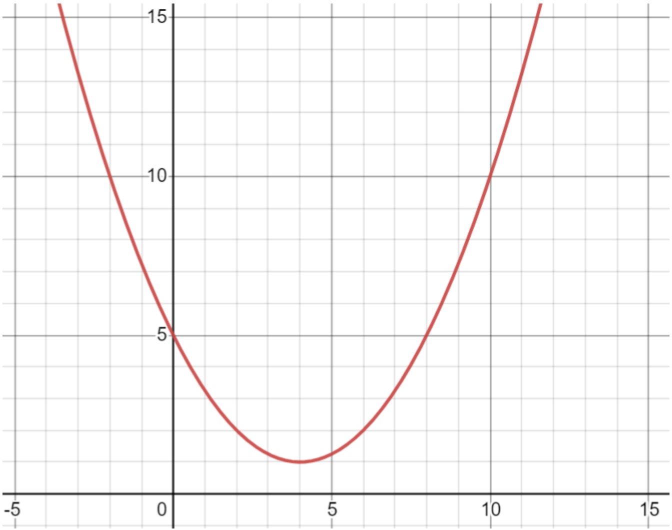
6 Sept 2021

Q1: Statement

Minimize the expression $\frac{x^2}{4} - 2x + 5$ with respect to x using

1. The closed-form solution
2. Gradient descent

Q1: The graph

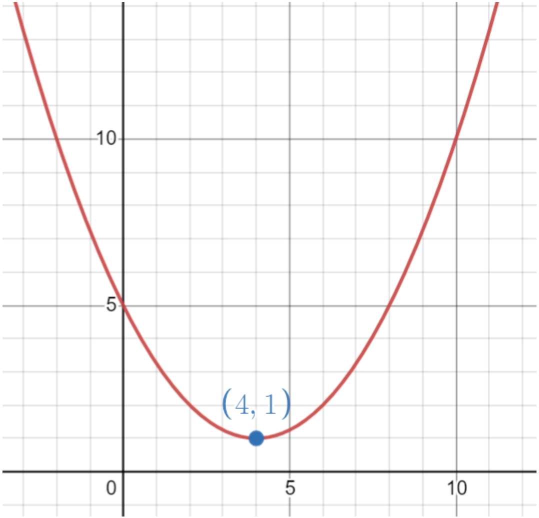


Q1: Closed-form solution

$$f(x) = \frac{x^2}{4} - 2x + 5$$

$$\frac{\mathrm{d}f(x)}{\mathrm{d}x} = \frac{x}{2} - 2 = 0$$

$$x = 4$$



Q1: Recap of Gradient Descent

Algorithm 1 Gradient Descent

$W \leftarrow \text{random}$
Costs $\leftarrow \phi$
for $i = 1$ to n_i **do**
  $\hat{Y} \leftarrow M(W, X)$
  $C \leftarrow J(Y, \hat{Y})$
  $W \leftarrow W - \alpha \nabla_W C$
  Append C to Costs
end for

Q1: Run Through Gradient Descent

Let us set $x = 10$ initially. Let $\alpha = 0.1$

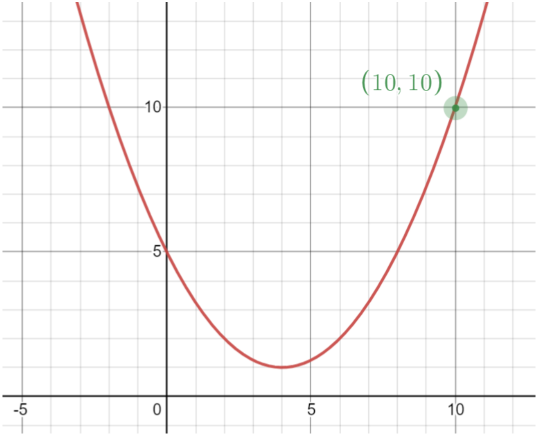
$$f(x) = \frac{x^2}{4} - 2x + 5$$

$$\frac{\mathrm{d}f(x)}{\mathrm{d}x} = \frac{x}{2} - 2$$

$$\begin{aligned} f(10) &= \frac{10^2}{4} - 2 * 10 + 5 \\ &= 25 - 20 + 5 \\ &= 10 \end{aligned}$$

x	$f(x)$
10	10

Q1: Run Through Gradient Descent



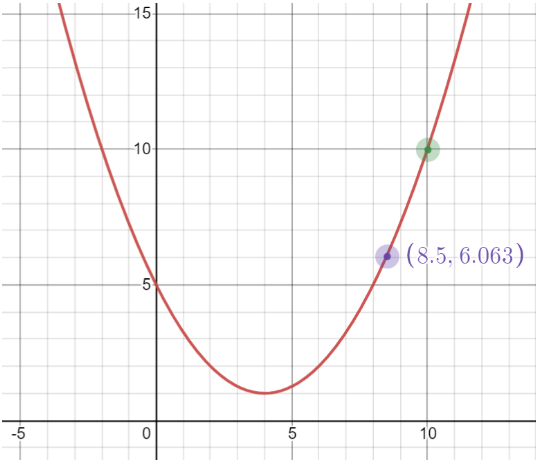
Q1: Run Through Gradient Descent

$$\alpha = 0.5, f(x) = \frac{x^2}{4} - 2x + 5, \frac{\mathrm{d}f(x)}{\mathrm{d}x} = \frac{x}{2} - 2$$

$$\begin{aligned} f'(10) &= \frac{10}{2} - 2 \\ &= 3 \\ x_{next} &= x - \alpha f'(x) \\ &= 10 - 0.5 * 3 \\ &= 8.5 \\ f(8.5) &= \frac{8.5^2}{4} - 2 * 8.5 + 5 \\ &= 6.0625 \end{aligned}$$

x	$f(x)$
10	10
8.5	6.025

Q1: Run Through Gradient Descent



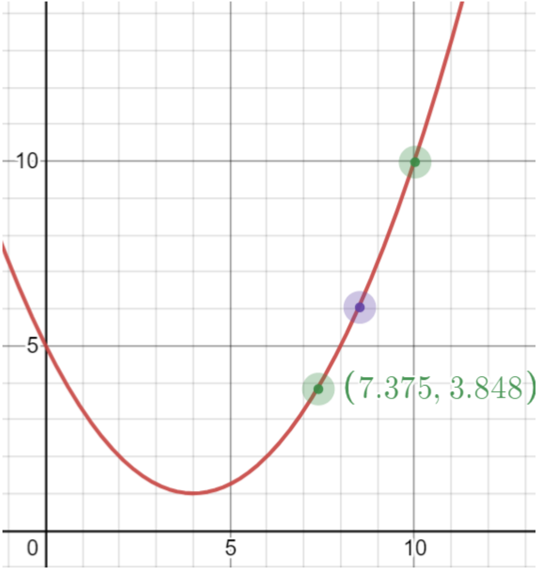
Q1: Run Through Gradient Descent

$$\alpha = 0.5, f(x) = \frac{x^2}{4} - 2x + 5, \frac{\mathrm{d}f(x)}{\mathrm{d}x} = \frac{x}{2} - 2$$

x	$f(x)$
10	10
8.5	6.025
7.375	3.848

$$\begin{aligned}
f'(8.5) &= \frac{8.5}{2} - 2 \\
&= 2.25 \\
x_{next} &= x - \alpha f'(x) \\
&= 8.5 - 0.5 * 2.25 \\
&= 7.375 \\
f(7.375) &= \frac{7.375^2}{4} - 2 * 7.375 + 5 \\
&= 3.848
\end{aligned}$$

Q1: Run Through Gradient Descent



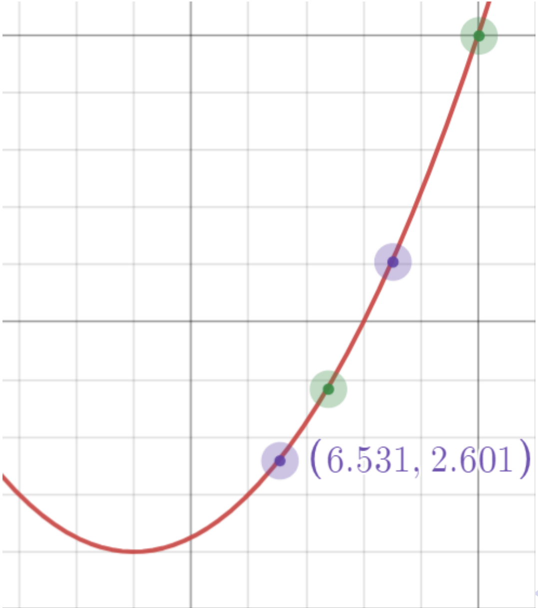
Q1: Run Through Gradient Descent

$$\alpha = 0.5, f(x) = \frac{x^2}{4} - 2x + 5, \frac{\mathrm{d}f(x)}{\mathrm{d}x} = \frac{x}{2} - 2$$

$$\begin{aligned}
f'(7.375) &= \frac{7.375}{2} - 2 \\
&= 1.688 \\
x_{next} &= x - \alpha f'(x) \\
&= 7.375 - 0.5 * 1.688 \\
&= 6.531 \\
f(6.531) &= \frac{6.531^2}{4} - 2 * 6.531 + 5 \\
&= 2.601
\end{aligned}$$

x	$f(x)$
10	10
8.5	6.025
7.375	3.848
6.531	2.601

Q1: Run Through Gradient Descent



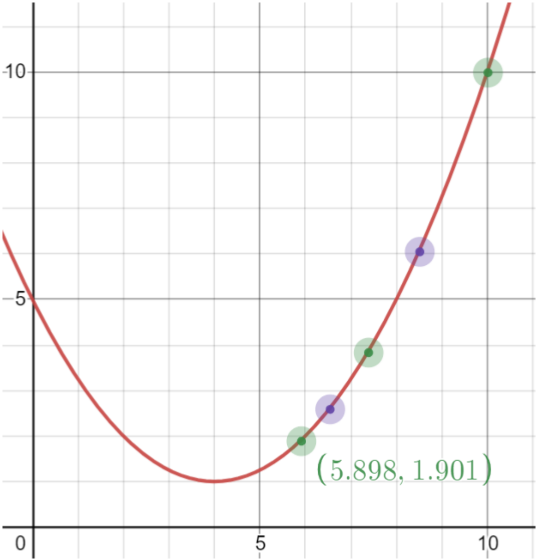
Q1: Run Through Gradient Descent

$$\alpha = 0.5, f(x) = \frac{x^2}{4} - 2x + 5, \frac{\mathrm{d}f(x)}{\mathrm{d}x} = \frac{x}{2} - 2$$

$$\begin{aligned}
f'(6.531) &= \frac{6.531}{2} - 2 \\
&= 1.266 \\
x_{next} &= x - \alpha f'(x) \\
&= 6.531 - 0.5 * 1.266 \\
&= 5.898 \\
f(5.898) &= \frac{5.898^2}{4} - 2 * 5.898 + 5 \\
&= 1.901
\end{aligned}$$

x	$f(x)$
10	10
8.5	6.025
7.375	3.848
6.531	2.601
5.898	1.901

Q1: Run Through Gradient Descent



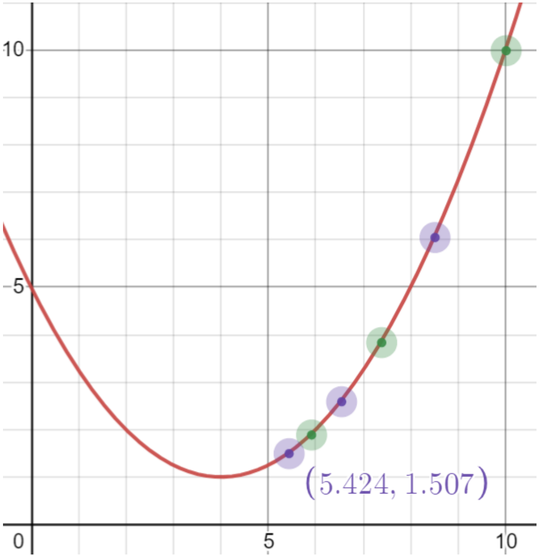
Q1: Run Through Gradient Descent

$$\alpha = 0.5, f(x) = \frac{x^2}{4} - 2x + 5, \frac{\mathrm{d}f(x)}{\mathrm{d}x} = \frac{x}{2} - 2$$

$$\begin{aligned}
f'(5.898) &= \frac{5.898}{2} - 2 \\
&= 0.949 \\
x_{next} &= x - \alpha f'(x) \\
&= 5.898 - 0.5 * 0.949 \\
&= 5.424 \\
f(5.424) &= \frac{5.424^2}{4} - 2 * 5.424 + 5 \\
&= 1.507
\end{aligned}$$

x	$f(x)$
10	10
8.5	6.025
7.375	3.848
6.531	2.601
5.898	1.901
5.424	1.507

Q1: Run Through Gradient Descent



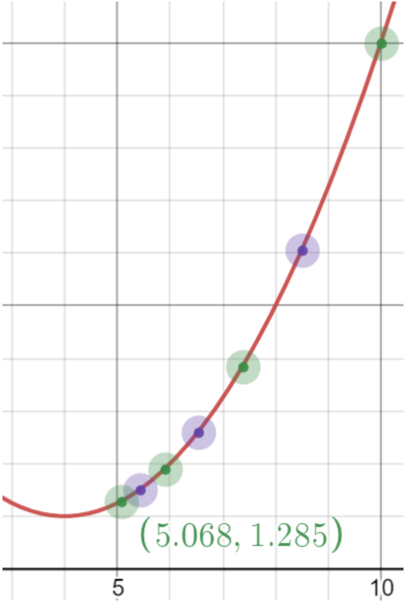
Q1: Run Through Gradient Descent

$$\alpha = 0.5, f(x) = \frac{x^2}{4} - 2x + 5, \frac{\mathrm{d}f(x)}{\mathrm{d}x} = \frac{x}{2} - 2$$

$$\begin{aligned}
f'(5.424) &= \frac{5.424}{2} - 2 \\
&= 0.712 \\
x_{next} &= x - \alpha f'(x) \\
&= 5.424 - 0.5 * 0.712 \\
&= 5.068 \\
f(5.068) &= \frac{5.068^2}{4} - 2 * 5.068 + 5 \\
&= 1.285
\end{aligned}$$

x	$f(x)$
10	10
8.5	6.025
7.375	3.848
6.531	2.601
5.898	1.901
5.424	1.507
5.068	1.285

Q1: Run Through Gradient Descent



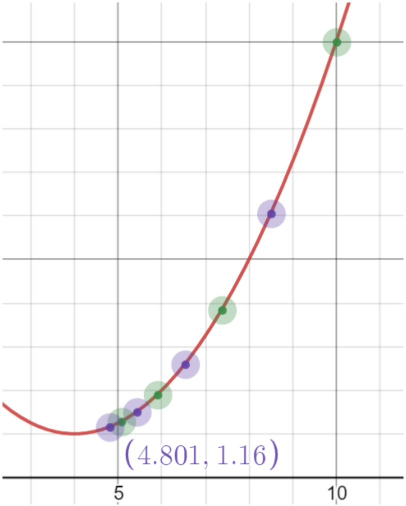
Q1: Run Through Gradient Descent

$$\alpha = 0.5, f(x) = \frac{x^2}{4} - 2x + 5, \frac{\mathrm{d}f(x)}{\mathrm{d}x} = \frac{x}{2} - 2$$

$$\begin{aligned} f'(5.068) &= \frac{5.068}{2} - 2 \\ &= 0.534 \\ x_{next} &= x - \alpha f'(x) \\ &= 5.068 - 0.5 * 0.534 \\ &= 4.801 \\ f(4.801) &= \frac{4.801^2}{4} - 2 * 4.801 + 5 \\ &= 1.16 \end{aligned}$$

x	$f(x)$
10	10
8.5	6.025
7.375	3.848
6.531	2.601
5.898	1.901
5.424	1.507
5.068	1.285
4.801	1.16

Q1: Run Through Gradient Descent



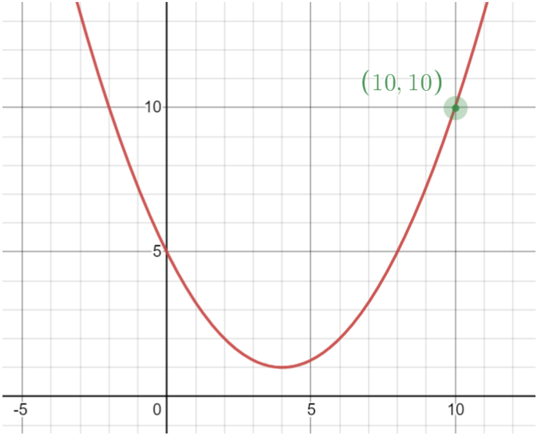
Q1: Gradient Descent With Excess Learning Rate

Let us set $x = 10$ initially. Let $\alpha = 5$

$$
\begin{aligned}
f(x) &= \frac{x^2}{4} - 2x + 5 \\
\frac{\mathrm{d}f(x)}{\mathrm{d}x} &= \frac{x}{2} - 2 \\
f(10) &= \frac{10^2}{4} - 2 * 10 + 5 \\
&= 25 - 20 + 5 \\
&= 10
\end{aligned}
$$

x	$f(x)$
10	10

Q1: Gradient Descent With Excess Learning Rate



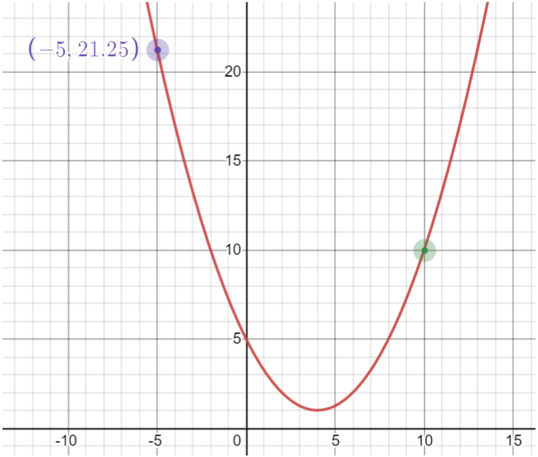
Q1: Run Through Gradient Descent

$$\alpha = 5, f(x) = \frac{x^2}{4} - 2x + 5, \frac{\mathrm{d}f(x)}{\mathrm{d}x} = \frac{x}{2} - 2$$

x	$f(x)$
10	10
-5	21.25

$$\begin{aligned}
f'(10) &= \frac{10}{2} - 2 \\
&= 3 \\
x_{next} &= x - \alpha f'(x) \\
&= 10 - 5 * 3 \\
&= -5 \\
f(-5) &= \frac{(-5)^2}{4} - 2 * (-5) + 5 \\
&= 21.25
\end{aligned}$$

Q1: Gradient Descent With Excess Learning Rate



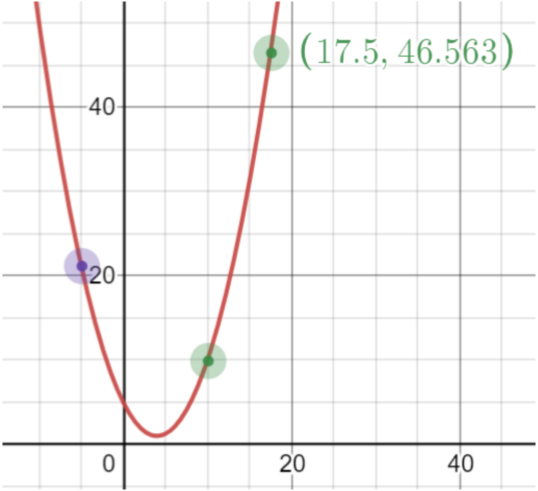
Q1: Run Through Gradient Descent

$$\alpha = 5, f(x) = \frac{x^2}{4} - 2x + 5, \frac{\mathrm{d}f(x)}{\mathrm{d}x} = \frac{x}{2} - 2$$

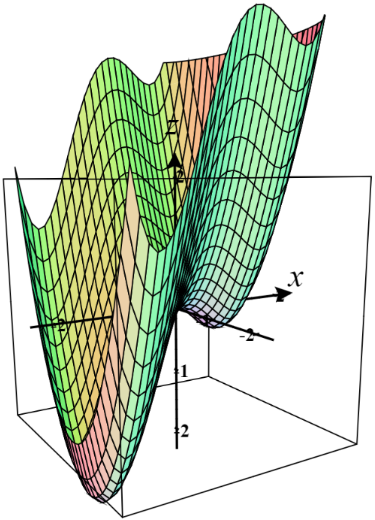
x	$f(x)$
10	10
-5	21.25
17.5	46.563

$$\begin{aligned} f'(-5) &= \frac{-5}{2} - 2 \\ &= -4.5 \\ x_{next} &= x - \alpha f'(x) \\ &= -5 - 5 * (-4.5) \\ &= 17.5 \\ f(17.5) &= \frac{17.5^2}{4} - 2 * 17.5 + 5 \\ &= 46.563 \end{aligned}$$

Q1: Gradient Descent With Excess Learning Rate

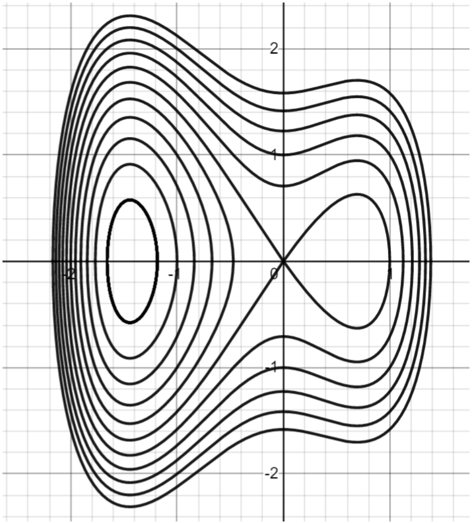


Q2: Statement



Find the minimum of $f(x, y) = x^4 + x^3 - 2x^2 + y^2$, using the closed-form solution and using gradient descent. Interact with the 3D plot here.

Q2: Contour Plots



It's difficult to visualize the gradient and the path on the 3D plot.

Every curve on the contour plot is a locus of constant height.

To obtain the contour of $f(x, y)$ at height h, on the x-y plane, plot the curve of $h = f(x, y)$. Interact with the contour plot here.

Q2: Closed-form solution

$$f(x,y) = x^4 + x^3 - 2x^2 + y^2$$

$$\frac{\partial f(x,y)}{\partial x} = 4x^3 + 3x^2 - 4x = 0$$

$$\therefore x = -1.443, 0, 0.693$$

$$\frac{\partial f(x,y)}{\partial y} = 2y = 0$$

$$\therefore y = 0$$

$$f(-1.443, 0) = -2.833$$

$$f(0,0) = 0$$

$$f(0.693, 0) = -0.39$$

$$\therefore \text{the minimum of } f(x,y) \text{ is at} (-1.443, 0)$$

Q2: Run Through Gradient Descent

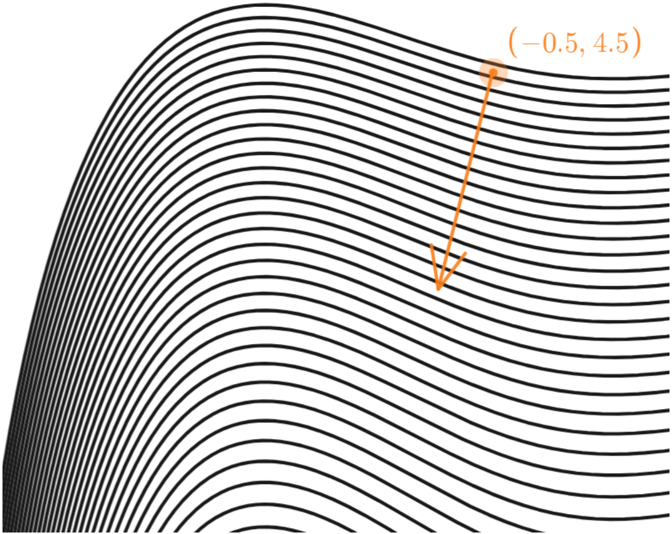
Let's start from $(-0.5, 4.5)$ with $\alpha = 0.1$.

x	y	$f(x,y)$
-0.5	4.5	19.688
-0.725	3.6	11.80

$$\begin{aligned}
\alpha &= 0.1 \\
f(x,y) &= x^4 + x^3 - 2x^2 + y^2 \\
\frac{\partial f(x,y)}{\partial x} &= 4x^3 + 3x^2 - 4x \\
\frac{\partial f(x,y)}{\partial y} &= 2y
\end{aligned}$$

$$\begin{aligned}
f(-0.5, 4.5) &= 19.688 \\
(\nabla_{x,y} f)(-0.5, 4.5) &= (2.25, 9) \\
(x_{next}, y_{next}) &= (-0.5, 4.5) - \alpha(\nabla_{x,y} f)(-0.5, 4.5) \\
&= (-0.725, 3.6) \\
f(-0.725, 3.6) &= 11.80
\end{aligned}$$

Q2: Run Through Gradient Descent



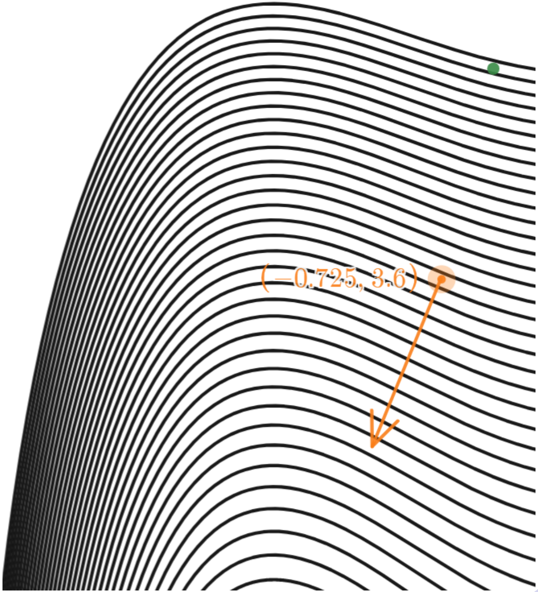
Q2: Run Through Gradient Descent

$$\begin{aligned}
\alpha &= 0.1 \\
f(x, y) &= x^4 + x^3 - 2x^2 + y^2 \\
\frac{\partial f(x, y)}{\partial x} &= 4x^3 + 3x^2 - 4x \\
\frac{\partial f(x, y)}{\partial y} &= 2y
\end{aligned}$$

x	y	$f(x, y)$
-0.5	4.5	19.688
-0.725	3.6	11.80
-1.02	2.88	6.23

$$\begin{aligned}
f(-0.725, 3.6) &= 11.80 \\
(\nabla_{x,y} f)(-0.725, 3.6) &= (2.952, 7.2) \\
(x_{next}, y_{next}) &= (-0.725, 3.6) - \alpha(\nabla_{x,y} f)(-0.725, 3.6) \\
&= (-1.02, 2.88) \\
f(-1.02, 2.88) &= 6.23
\end{aligned}$$

Q2: Run Through Gradient Descent



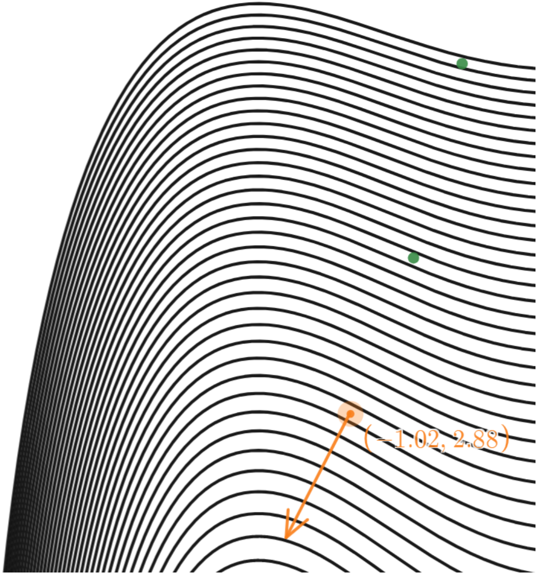
Q2: Run Through Gradient Descent

$$
\begin{aligned}
\alpha &= 0.1 \\
f(x, y) &= x^4 + x^3 - 2x^2 + y^2 \\
\frac{\partial f(x, y)}{\partial x} &= 4x^3 + 3x^2 - 4x \\
\frac{\partial f(x, y)}{\partial y} &= 2y
\end{aligned}
$$

x	y	$f(x, y)$
-0.5	4.5	19.688
-0.725	3.6	11.80
-1.02	2.88	6.23
-1.31	2.30	2.56

$$
\begin{aligned}
f(-1.02, 2.88) &= 6.23 \\
(\nabla_{x,y} f)(-1.02, 2.88) &= (2.96, 5.76) \\
(x_{next}, y_{next}) &= (-1.02, 2.88) - \alpha(\nabla_{x,y} f)(-1.02, 2.88) \\
&= (-1.31, 2.30) \\
f(-1.31, 2.30) &= 2.56
\end{aligned}
$$

Q2: Run Through Gradient Descent



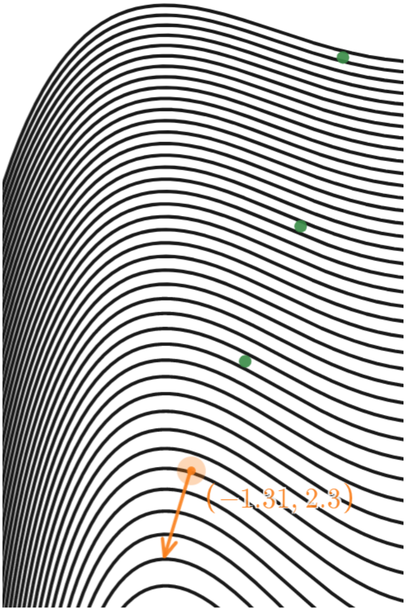
Q2: Run Through Gradient Descent

$$\begin{aligned}
\alpha &= 0.1 \\
f(x,y) &= x^4 + x^3 - 2x^2 + y^2 \\
\frac{\partial f(x,y)}{\partial x} &= 4x^3 + 3x^2 - 4x \\
\frac{\partial f(x,y)}{\partial y} &= 2y
\end{aligned}$$

x	y	$f(x,y)$
-0.5	4.5	19.688
-0.725	3.6	11.80
-1.02	2.88	6.23
-1.31	2.30	2.56
-1.45	1.84	0.55

$$\begin{aligned}
f(-1.31, 2.30) &= 2.56 \\
(\nabla_{x,y} f)(-1.31, 2.30) &= (1.39, 4.6) \\
(x_{next}, y_{next}) &= (-1.31, 2.30) - \alpha(\nabla_{x,y} f)(-1.31, 2.30) \\
&= (-1.45, 1.84) \\
f(-1.45, 1.84) &= 0.55
\end{aligned}$$

Q2: Run Through Gradient Descent



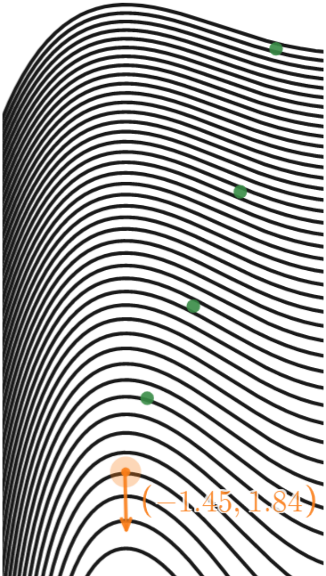
Q2: Run Through Gradient Descent

$$\alpha = 0.1$$
$$f(x, y) = x^4 + x^3 - 2x^2 + y^2$$
$$\frac{\partial f(x, y)}{\partial x} = 4x^3 + 3x^2 - 4x$$
$$\frac{\partial f(x, y)}{\partial y} = 2y$$

x	y	$f(x, y)$
-0.5	4.5	19.688
-0.725	3.6	11.80
-1.02	2.88	6.23
-1.31	2.30	2.56
-1.45	1.84	0.55
-1.44	1.47	-0.67

$$f(-1.45, 1.84) = 0.55$$
$$(\nabla_{x,y} f)(-1.45, 1.84) = (-0.087, 3.68)$$
$$(x_{next}, y_{next}) = (-1.45, 1.84) - \alpha(\nabla_{x,y} f)(-1.45, 1.84)$$
$$= (-1.44, 1.47)$$
$$f(-1.44, 1.47) = -0.67$$

Q2: Run Through Gradient Descent



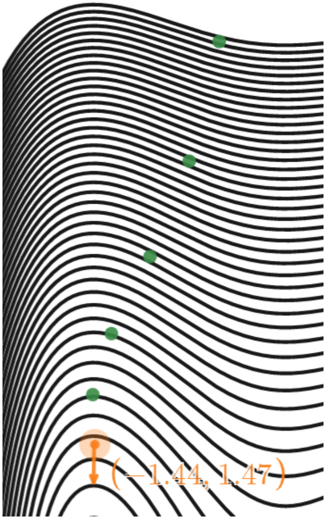
Q2: Run Through Gradient Descent

$$
\begin{aligned}
\alpha &= 0.1 \\
f(x,y) &= x^4 + x^3 - 2x^2 + y^2 \\
\frac{\partial f(x,y)}{\partial x} &= 4x^3 + 3x^2 - 4x \\
\frac{\partial f(x,y)}{\partial y} &= 2y
\end{aligned}
$$

x	y	$f(x,y)$
-0.5	4.5	19.688
-0.725	3.6	11.80
-1.02	2.88	6.23
-1.31	2.30	2.56
-1.45	1.84	0.55
-1.44	1.47	-0.67
-1.44	1.17	-1.45

$$
\begin{aligned}
f(-1.44, 1.47) &= -0.67 \\
(\nabla_{x,y} f)(-1.44, 1.47) &= (0.037, 2.94) \\
(x_{next}, y_{next}) &= (-1.44, 1.47) - \alpha(\nabla_{x,y} f)(-1.44, 1.47) \\
&= (-1.44, 1.17) \\
f(-1.44, 1.17) &= -1.45
\end{aligned}
$$

Q2: Run Through Gradient Descent



Q2: Run Through Gradient Descent

For happens next, open the interactive graph and play with it. To see the effect of a large learning rate, change α and see what happens!