Problem 1: Write down the gradient descent update rule for the objective function

$$J(\theta) = \theta^4 - 6\theta^3 + 11\theta^2 - 7\theta + 4$$

from class. Suppose the learning rate is α.

Problem 2: Consider the affine objective function

$$J(\theta) = m\theta + b,$$

for some parameters $m \neq 0$ and $b \in \mathbb{R}$.

(a) Write down the gradient descent update rule. Suppose the learning rate is α.

(b) Find a closed form expression for θ_t.

(c) Using your answer to (b), discuss convergence of the gradient descent algorithm.

Problem 3: Consider the quadratic objective function

$$J(\theta) = \theta^2.$$

(a) Write down the gradient descent update rule. Suppose the learning rate is α.

(b) Find a closed form expression for θ_t.

(c) Using your answer to (b), discuss convergence of the gradient descent algorithm.

Problem 4: Consider again the affine objective function

$$J(\theta) = m\theta + b,$$

for some parameters $m \neq 0$ and $b \in \mathbb{R}$.

(a) Write down the gradient descent update rule with learning rate α and decay rate β.

(b) Find a closed form expression for θ_t.

(c) Using your answer to (b), discuss convergence of the gradient descent algorithm.

Problem 5: Consider the function

$$J : \mathbb{R}^2 \to \mathbb{R}, \quad J(\boldsymbol{\theta}) = 2\theta_1^2 + 4\theta_1\theta_2 + 3\theta_2^2 - 8\theta_1 - 10\theta_2 + 9.$$

(a) Compute the directional first derivative $J'_{\mathbf{v}}(1,1)$ where $\mathbf{v}^\intercal = (1,0)$.

(b) Compute the directional second derivative $J''_{\mathbf{v}}(1,1)$ where $\mathbf{v}^\intercal = (1,0)$.

Problem 6: Re-do the previous problem, but use the relationship between directional first and second derivatives and gradient vectors and Hessian matrices.

Problem 7: Consider again the function J from Problem 5. Here it is, for reference:

$$J : \mathbb{R}^2 \to \mathbb{R}, \quad J(\boldsymbol{\theta}) = 2\theta_1^2 + 4\theta_1\theta_2 + 3\theta_2^2 - 8\theta_1 - 10\theta_2 + 9.$$

(a) Find the direction of maximum rate of change of J at the point $\boldsymbol{\theta}^\intercal = (0, 0)$. What is the rate of change in this direction?

(b) Find the direction of minimum rate of change of J at the point $\boldsymbol{\theta}^\intercal = (0, 0)$. What is the rate of change in this direction?

Problem 8: Yet again, consider the function J from Problem 5:

$$J : \mathbb{R}^2 \to \mathbb{R}, \quad J(\boldsymbol{\theta}) = 2\theta_1^2 + 4\theta_1\theta_2 + 3\theta_2^2 - 8\theta_1 - 10\theta_2 + 9.$$

Find and classify all extremizers of J.

Problem 9: For the fifth time, consider the function J from Problem 5:

$$J : \mathbb{R}^2 \to \mathbb{R}, \quad J(\boldsymbol{\theta}) = 2\theta_1^2 + 4\theta_1\theta_2 + 3\theta_2^2 - 8\theta_1 - 10\theta_2 + 9.$$

Find the directions of extreme curvature at the point $\boldsymbol{\theta} = (0, 0)$. What are the curvatures in these directions?

Problem 10: One last time, consider the function J from Problem 5:

$$J : \mathbb{R}^2 \to \mathbb{R}, \quad J(\boldsymbol{\theta}) = 2\theta_1^2 + 4\theta_1\theta_2 + 3\theta_2^2 - 8\theta_1 - 10\theta_2 + 9.$$

Compute the spectral radius $\rho\left(\nabla^2 J(0,0)\right)$ and the condition number $\kappa\left(\nabla^2 J(0,0)\right)$.

Problem 11: Consider the objective function

$$J : \mathbb{R}^2 \to \mathbb{R}, \quad J(\boldsymbol{\theta}) = 2\theta_1^2 + 4\theta_2^2.$$

(a) Write down the gradient descent update rule. Suppose the learning rate is α, while the decay rate is $\beta = 0$.

(b) Find a closed form expression for $\boldsymbol{\theta}_t$.

(c) Using your answer to (b), discuss convergence of the gradient descent algorithm.

Problem 12: Consider the stochastic objective function

$$J : \mathbb{R}^2 \to \mathbb{R}, \quad J(\boldsymbol{\theta}) = \frac{1}{m}\sum_{i=1}^{m} \frac{1}{2}|\mathbf{x}_i - \boldsymbol{\theta}|^2$$

from class, where $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m \in \mathbb{R}^2$ is an observed dataset.

(a) Compute the update rule in the batch gradient descent algorithm with learning rate α and decay rate β.

(b) Assuming $\beta = 0$, discuss convergence of the batch gradient descent algorithm.