DDA6010 AS4

November 6, 2024

Problem 1

Proof: If f is convex, then $\text{epi}(f)$ is convex

1. Take any two points $(x_1, \alpha_1), (x_2, \alpha_2) \in \text{epi}(f)$. This implies:

$$f(x_1) \leq \alpha_1, \quad f(x_2) \leq \alpha_2.$$

2. For any $\theta \in [0, 1]$, consider the point:

$$(x, \alpha) = \theta(x_1, \alpha_1) + (1-\theta)(x_2, \alpha_2) = (\theta x_1 + (1-\theta)x_2, \theta\alpha_1 + (1-\theta)\alpha_2).$$

3. Since f is convex, by the definition of convexity:

$$f(\theta x_1 + (1-\theta)x_2) \leq \theta f(x_1) + (1-\theta)f(x_2).$$

4. Using the fact that $f(x_1) \leq \alpha_1$ and $f(x_2) \leq \alpha_2$, we get:

$$\theta f(x_1) + (1-\theta)f(x_2) \leq \theta\alpha_1 + (1-\theta)\alpha_2 = \alpha.$$

5. Thus, $f(x) = f(\theta x_1 + (1-\theta)x_2) \leq \alpha$, which means $(x, \alpha) \in \text{epi}(f)$.
6. Therefore, $\text{epi}(f)$ is a convex set.

Proof: If $\text{epi}(f)$ is convex, then f is convex

1. Take any $x_1, x_2 \in \mathbb{R}^n$ and $\theta \in [0, 1]$. Define $\alpha_1 = f(x_1)$ and $\alpha_2 = f(x_2)$.
2. Since $(x_1, \alpha_1), (x_2, \alpha_2) \in \text{epi}(f)$ and $\text{epi}(f)$ is convex, the point:

$$(x, \alpha) = \theta(x_1, \alpha_1) + (1-\theta)(x_2, \alpha_2)$$

also belongs to $\text{epi}(f)$.

3. Thus, $x = \theta x_1 + (1-\theta)x_2$ and $\alpha = \theta\alpha_1 + (1-\theta)\alpha_2$.

4. Since $(x, \alpha) \in \text{epi}(f)$, we have:

$$f(x) \leq \alpha = \theta f(x_1) + (1 - \theta) f(x_2).$$

5. This is exactly the definition of a convex function, so f is convex.

Conclusion

We have shown that $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a convex function if and only if its epigraph $\text{epi}(f)$ is a convex set.

Problem 2

To show that if x^* is an optimal solution of problem (P) satisfying $g(x^*) < 0$, then x^* is also an optimal solution to the problem

$$\min \quad f(x) \quad \text{s.t.} \quad x \in X,$$

we proceed as follows:

1. Define the Problems:

Consider the following two optimization problems:

- Problem (P):

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g(x) \leq 0 \\ & x \in X, \end{aligned}$$

where f and g are convex functions defined on $\mathbb{R}^n$, and $X \subseteq \mathbb{R}^n$ is a convex set.

- Problem (Q):

$$\min \quad f(x) \quad \text{s.t.} \quad x \in X.$$

2. Assume x^* is Optimal for (P) and $g(x^*) < 0$:

Suppose x^* is an optimal solution to problem (P) and satisfies the strict inequality $g(x^*) < 0$.

3. Analyze the Feasible Region:

Since $g(x^*) < 0$, by the definition of convex sets, there exists an $\epsilon > 0$ such that for all x within the neighborhood $\|x - x^*\| < \epsilon$ and $x \in X$, we have $g(x) < 0$. Therefore, within this neighborhood, any $x \in X$ automatically satisfies $g(x) \leq 0$.

4. Compare Optimal Solutions:

Assume, for the sake of contradiction, that x^* is not an optimal solution to problem (Q). Then there exists some $\hat{x} \in X$ such that $f(\hat{x}) < f(x^*)$. However, since x^* is optimal for problem (P) and the feasible region of (P) includes all $x \in X$ that satisfy $g(x) \leq 0$, and given that $g(x^*) < 0$, $\hat{x}$ must also satisfy $g(\hat{x}) \leq 0$ to be feasible for (P). This leads to a contradiction because $f(\hat{x}) < f(x^*)$ contradicts the optimality of x^* for (P).

Conclusion

Therefore, our assumption is false, and x^* must be an optimal solution to problem (Q).

Problem 3

(a)

The function $f : \mathbb{R}^d \to \mathbb{R}$ is μ-strongly convex if, for all $x, y \in \mathbb{R}^d$, it holds that:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|y - x\|^2.$$

By substituting $y = x^*$ into the strong convexity inequality, we obtain:

$$f(x^*) \geq f(x) + \langle \nabla f(x), x^* - x \rangle + \frac{\mu}{2}\|x^* - x\|^2.$$

$$f(x) - f^* \leq \langle \nabla f(x), x - x^* \rangle - \frac{\mu}{2}\|x^* - x\|^2.$$

Using the Cauchy-Schwarz inequality:

$$\langle \nabla f(x), x - x^* \rangle \leq \|\nabla f(x)\| \cdot \|x - x^*\|.$$

and a basic inequality $a \cdot b \leq \frac{a^2}{2\mu} + \frac{\mu b^2}{2}$ (which is a generalization of the inequality $ab \leq \frac{a^2}{2} + \frac{b^2}{2}$), we obtain:

$$\langle \nabla f(x), x - x^* \rangle \leq \frac{1}{2\mu}\|\nabla f(x)\|^2 + \frac{\mu}{2}\|x - x^*\|^2.$$

Substituting back to the inequality before yields

$$f(x) - f^* \leq \frac{1}{2\mu}\|\nabla f(x)\|^2.$$

(b)

Step 1: Using the smoothness of the function

According to the definition of L-smoothness, for all $x, y \in \mathbb{R}^d$:

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2}\|y - x\|^2.$$

Taking $x = x^k$ and $y = x^{k+1}$, we have:

$$f(x^{k+1}) \leq f(x^k) + \langle \nabla f(x^k), x^{k+1} - x^k \rangle + \frac{L}{2}\|x^{k+1} - x^k\|^2.$$

Step 2: Substituting the update rule

Based on the update rule:

$$x^{k+1} = x^k - \alpha_k \left(\nabla f(x^k) + e^k\right).$$

Thus:

$$x^{k+1} - x^k = -\alpha_k \left(\nabla f(x^k) + e^k \right).$$

Calculating terms:

1. Inner product term:

$$\langle \nabla f(x^k), x^{k+1} - x^k \rangle = -\alpha_k \left(\|\nabla f(x^k)\|^2 + \langle \nabla f(x^k), e^k \rangle \right).$$

Since $\langle \nabla f(x^k), e^k \rangle = 0$, we have:

$$\langle \nabla f(x^k), x^{k+1} - x^k \rangle = -\alpha_k \|\nabla f(x^k)\|^2.$$

2. Norm square difference:

$$\|x^{k+1} - x^k\|^2 = \alpha_k^2 \|\nabla f(x^k) + e^k\|^2.$$

Similarly, since $\langle \nabla f(x^k), e^k \rangle = 0$ and $\|e^k\|^2 \le \sigma^2$, we get:

$$\|x^{k+1} - x^k\|^2 \le \alpha_k^2 \left(\|\nabla f(x^k)\|^2 + \sigma^2 \right).$$

Step 3: Substituting the results

Substitute the above results into the smoothness inequality:

$$f(x^{k+1}) \le f(x^k) - \alpha_k \|\nabla f(x^k)\|^2 + \frac{L}{2} \alpha_k^2 \left(\|\nabla f(x^k)\|^2 + \sigma^2 \right).$$

This simplifies to:

$$f(x^{k+1}) \le f(x^k) - \left(\alpha_k - \frac{L\alpha_k^2}{2} \right) \|\nabla f(x^k)\|^2 + \frac{L\alpha_k^2 \sigma^2}{2}.$$

Step 4: Using the step size condition

Since $\alpha_k \le \frac{1}{L}$, we have:

$$\alpha_k - \frac{L\alpha_k^2}{2} \ge \alpha_k \left(1 - \frac{L\alpha_k}{2} \right) \ge \alpha_k \left(1 - \frac{1}{2} \right) = \frac{\alpha_k}{2}.$$

Thus:

$$f(x^{k+1}) \le f(x^k) - \frac{\alpha_k}{2} \|\nabla f(x^k)\|^2 + \frac{L\alpha_k^2 \sigma^2}{2}.$$

Step 5: Applying the result from part (a)

From part (a), for all $x \in \mathbb{R}^d$:

$$\|\nabla f(x)\|^2 \ge 2\mu \left(f(x) - f^* \right).$$

Substitute this into the inequality:

$$f(x^{k+1}) \le f(x^k) - \alpha_k \mu \left(f(x^k) - f^* \right) + \frac{L\alpha_k^2 \sigma^2}{2}.$$

Step 6: Rearranging the inequality

Rearrange the inequality:

$$f(x^{k+1}) - f^* \leq \left(f(x^k) - f^*\right) - \alpha_k \mu \left(f(x^k) - f^*\right) + \frac{L\alpha_k^2 \sigma^2}{2}.$$

This simplifies to:

$$f(x^{k+1}) - f^* \leq (1 - \alpha_k \mu) \left(f(x^k) - f^*\right) + \frac{L\alpha_k^2 \sigma^2}{2}.$$

Conclusion

Therefore, for all $k \geq 0$ and $\alpha_k \leq \frac{1}{L}$, we have:

$$f(x^{k+1}) - f^* \leq (1 - \alpha_k \mu) \left(f(x^k) - f^*\right) + \frac{L\alpha_k^2 \sigma^2}{2}.$$

(c)

Strong convexity: For all $x, y \in \mathbb{R}^d$, there exists

$$\langle \nabla f(y) - \nabla f(x), x - y \rangle \geq \mu \|x - y\|^2.$$

Lipschitz continuous gradient (smoothness): For all $x, y \in \mathbb{R}^d$, we have

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|.$$

This can be rewritten as

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \leq L \|x - y\|^2.$$

From the smoothness property, we have

$$\|\nabla f(x) - \nabla f(y)\|^2 \leq L \langle \nabla f(x) - \nabla f(y), x - y \rangle.$$

Similarly, from the strong convexity property,

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \frac{1}{L} \|\nabla f(x) - \nabla f(y)\|^2.$$

Set a weight parameter $\theta \in [0, 1]$, and consider the convex combination of the inequalities of strong convexity and smoothness:

$$\theta \left[\langle \nabla f(x) - \nabla f(y), x - y \rangle - \mu \|x - y\|^2\right] + (1 - \theta) \left[\langle \nabla f(x) - \nabla f(y), x - y \rangle - \frac{1}{L} \|\nabla f(x) - \nabla f(y)\|^2\right] \geq 0.$$

Expanding the above combination yields:

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle - \theta \mu \|x - y\|^2 - (1 - \theta) \frac{1}{L} \|\nabla f(x) - \nabla f(y)\|^2 \geq 0.$$

Simplifying, we get:

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \theta\mu\|x - y\|^2 + \frac{1-\theta}{L}\|\nabla f(x) - \nabla f(y)\|^2.$$

To maximize the lower bound on the right side, we choose $\theta = \frac{L}{L+\mu}$, which yields:

$$\theta = \frac{L}{L+\mu}, \quad 1 - \theta = \frac{\mu}{L+\mu}.$$

Substituting the value of θ into the simplified inequality, we obtain:

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \frac{L\mu}{L+\mu}\|x - y\|^2 + \frac{\mu}{(L+\mu)L}\|\nabla f(x) - \nabla f(y)\|^2.$$

Simplifying the coefficient of the second term:

$$\frac{\mu}{(L+\mu)L} = \frac{1}{L+\mu}.$$

Thus, the inequality becomes:

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \frac{L\mu}{L+\mu}\|x - y\|^2 + \frac{1}{L+\mu}\|\nabla f(x) - \nabla f(y)\|^2.$$

Conclusion: Therefore, for all $x, y \in \mathbb{R}^d$, we have

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \frac{L\mu}{L+\mu}\|x - y\|^2 + \frac{1}{L+\mu}\|\nabla f(x) - \nabla f(y)\|^2.$$

(d)

According to the iterative update rule:

$$x^{k+1} = x^k - \alpha_k \left(\nabla f(x^k) + e^k\right).$$

Therefore, the difference is:

$$x^{k+1} - x^* = x^k - x^* - \alpha_k \left(\nabla f(x^k) + e^k\right).$$

Taking the squared norm on both sides:

$$\|x^{k+1} - x^*\|^2 = \|x^k - x^* - \alpha_k \left(\nabla f(x^k) + e^k\right)\|^2.$$

Expanding the right side:

$$\|x^{k+1} - x^*\|^2 = \|x^k - x^*\|^2 - 2\alpha_k\langle x^k - x^*, \nabla f(x^k) + e^k\rangle + \alpha_k^2\|\nabla f(x^k) + e^k\|^2.$$

Given that $\mathbb{E}\langle e^k, x^k - x^*\rangle = 0$ from the problem's conditions, we have:

$$\mathbb{E}\langle x^k - x^*, \nabla f(x^k) + e^k\rangle = \langle x^k - x^*, \nabla f(x^k)\rangle.$$

Expanding the squared norm term:

$$\|\nabla f(x^k) + e^k\|^2 = \|\nabla f(x^k)\|^2 + 2\langle \nabla f(x^k), e^k\rangle + \|e^k\|^2.$$

Since $\mathbb{E}\langle \nabla f(x^k), e^k\rangle = 0$ and $\mathbb{E}\|e^k\|^2 \leq \sigma^2$, we have:

$$\mathbb{E}\|\nabla f(x^k) + e^k\|^2 \leq \|\nabla f(x^k)\|^2 + \sigma^2.$$

Substituting the result back into the expansion, We obtain:

$$\mathbb{E}\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - 2\alpha_k\langle x^k - x^*, \nabla f(x^k)\rangle + \alpha_k^2\|\nabla f(x^k)\|^2 + \alpha_k^2\sigma^2.$$

According to part (c), for all $x, y \in \mathbb{R}^d$:

$$\langle \nabla f(x) - \nabla f(y), x - y\rangle \geq \frac{L\mu}{L+\mu}\|x - y\|^2 + \frac{1}{L+\mu}\|\nabla f(x) - \nabla f(y)\|^2.$$

Take $y = x^*$, and since $\nabla f(x^*) = 0$, we get:

$$\langle \nabla f(x^k), x^k - x^*\rangle \geq \frac{L\mu}{L+\mu}\|x^k - x^*\|^2 + \frac{1}{L+\mu}\|\nabla f(x^k)\|^2.$$

Substitute this inequality into Step 5:

$$\mathbb{E}\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - 2\alpha_k\left(\frac{L\mu}{L+\mu}\|x^k - x^*\|^2 + \frac{1}{L+\mu}\|\nabla f(x^k)\|^2\right) + \alpha_k^2\|\nabla f(x^k)\|^2 + \alpha_k^2\sigma^2.$$

Simplifying terms:
- Term involving $\|x^k - x^*\|^2$:

$$\|x^k - x^*\|^2 - 2\alpha_k\frac{L\mu}{L+\mu}\|x^k - x^*\|^2 = \left(1 - 2\alpha_k\frac{L\mu}{L+\mu}\right)\|x^k - x^*\|^2.$$

- Term involving $\|\nabla f(x^k)\|^2$:

$$-2\alpha_k\frac{1}{L+\mu}\|\nabla f(x^k)\|^2 + \alpha_k^2\|\nabla f(x^k)\|^2 = \left(-\frac{2\alpha_k}{L+\mu} + \alpha_k^2\right)\|\nabla f(x^k)\|^2.$$

Since $\alpha_k \leq \frac{1}{L+\mu}$, we have:

$$-\frac{2\alpha_k}{L+\mu} + \alpha_k^2 \leq -\frac{\alpha_k}{L+\mu}.$$

Thus,

$$\left(-\frac{2\alpha_k}{L+\mu} + \alpha_k^2\right)\|\nabla f(x^k)\|^2 \leq -\frac{\alpha_k}{L+\mu}\|\nabla f(x^k)\|^2.$$

Combining the results, we get:

$$\mathbb{E}\|x^{k+1} - x^*\|^2 \le \left(1 - \frac{2\alpha_k L\mu}{L+\mu}\right)\|x^k - x^*\|^2 - \frac{\alpha_k}{L+\mu}\|\nabla f(x^k)\|^2 + \alpha_k^2\sigma^2.$$

Conclusion:

Therefore, for all $k \ge 0$ and $\alpha_k \le \frac{1}{L+\mu}$, the above inequality holds.

Problem 4

Defining the constraints of problem (P):

Δ_n is the standard simplex, i.e.,

$$\Delta_n = \left\{ x \in \mathbb{R}^n \mid x_i \ge 0, \ \sum_{i=1}^n x_i = 1 \right\}.$$

Therefore, the constraints of problem (P) include:

- Inequality constraints: $x_i \ge 0$, for all i.
- Equality constraint: $\sum_{i=1}^n x_i = 1$.

Writing the Lagrangian function:

Introduce Lagrange multipliers $\lambda \in \mathbb{R}^n$ (for the inequality constraints) and $\mu \in \mathbb{R}$ (for the equality constraint). The Lagrangian function is:

$$L(x,\lambda,\mu) = f(x) - \sum_{i=1}^n \lambda_i x_i + \mu\left(\sum_{i=1}^n x_i - 1\right).$$

A stationary point x^* satisfies the following KKT conditions:

- Gradient condition:

$$\nabla f(x^*) - \lambda + \mu\mathbf{1} = 0,$$

where $\mathbf{1} = (1,1,\ldots,1)^\top$.

- Complementary slackness:

$$\lambda_i x_i^* = 0, \quad \forall i = 1,2,\ldots,n.$$

- Primal feasibility:

$$x_i^* \ge 0, \quad \lambda_i \ge 0, \quad \forall i = 1,2,\ldots,n, \quad \sum_{i=1}^n x_i^* = 1.$$

From the gradient condition, we obtain:

$$\frac{\partial f}{\partial x_i}(x^*) - \lambda_i + \mu = 0, \quad \forall i.$$

Rearranging, we have:

$$\lambda_i = \frac{\partial f}{\partial x_i}(x^*) + \mu.$$

Based on the complementary slackness condition: - If $x_i^* > 0$, then $\lambda_i = 0$. Therefore:

$$0 = \frac{\partial f}{\partial x_i}(x^*) + \mu \implies \frac{\partial f}{\partial x_i}(x^*) = -\mu.$$

- If $x_i^* = 0$, then $\lambda_i \geq 0$. Therefore:

$$\lambda_i = \frac{\partial f}{\partial x_i}(x^*) + \mu \geq 0 \implies \frac{\partial f}{\partial x_i}(x^*) + \mu \geq 0.$$

From the above results, we obtain for all i:

$$\frac{\partial f}{\partial x_i}(x^*) \begin{cases} = -\mu, & x_i^* > 0 \\ \geq -\mu, & x_i^* = 0. \end{cases}$$

Note that μ is a scalar, so we can redefine $\mu' = -\mu$, giving:

$$\frac{\partial f}{\partial x_i}(x^*) \begin{cases} = \mu', & x_i^* > 0 \\ \geq \mu', & x_i^* = 0. \end{cases}$$

To maintain consistency in notation, we simply use μ to represent this, reassigning $\mu = -\mu'$, so the results remain unchanged.

Conclusion: Therefore, $x^* \in \Delta_n$ is a stationary point of problem (P) if and only if there exists $\mu \in \mathbb{R}$ such that for all i:

$$\frac{\partial f}{\partial x_i}(x^*) \begin{cases} = \mu, & x_i^* > 0 \\ \geq \mu, & x_i^* = 0. \end{cases}$$

Problem 5

(i)

To prove that the vector $x^* = \left(\frac{17}{7}, 0, \frac{6}{7}\right)^\top$ is an optimal solution of problem (Q), we can verify that it satisfies the KKT conditions.

Step 1: Write down the objective function and constraints

Objective function:

$$f(x) = 2x_1^2 + 3x_2^2 + 4x_3^2 + 2x_1x_2 - 2x_1x_3 - 8x_1 - 4x_2 - 2x_3$$

Constraints:

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0$$

Step 2: Calculate the gradient of $f(x)$:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 4x_1 + 2x_2 - 2x_3 - 8 \\ 6x_2 + 2x_1 - 4 \\ 8x_3 - 2x_1 - 2 \end{bmatrix}$$

Step 3: Verify the KKT conditions

The KKT conditions include: 1. Stationarity: For each i:

- - If $x_i^* > 0$, then $\frac{\partial f}{\partial x_i}(x^*) = 0$
- - If $x_i^* = 0$, then $\frac{\partial f}{\partial x_i}(x^*) \geq 0$

2. Primal feasibility:

$$x_i^* \geq 0$$

3. Complementary slackness:

For $x_i^* = 0$, $\frac{\partial f}{\partial x_i}(x^*) \geq 0$ Verifying the gradient condition:

Calculate the gradient components at x^*:

- For x_1:

$$x_1^* = \frac{17}{7} > 0, \quad \text{we need to verify } \frac{\partial f}{\partial x_1}(x^*) = 0$$

$$\frac{\partial f}{\partial x_1}(x^*) = 4\left(\frac{17}{7}\right) + 2 \cdot 0 - 2\left(\frac{6}{7}\right) - 8 = \frac{68}{7} - \frac{12}{7} - 8 = \frac{56}{7} - 8 = 8 - 8 = 0$$

- For x_2:

$$x_2^* = 0, \quad \text{we need to verify } \frac{\partial f}{\partial x_2}(x^*) \geq 0$$

$$\frac{\partial f}{\partial x_2}(x^*) = 6 \cdot 0 + 2\left(\frac{17}{7}\right) - 4 = \frac{34}{7} - 4 = \frac{6}{7} > 0$$

- For x_3:

$$x_3^* = \frac{6}{7} > 0, \quad \text{we need to verify } \frac{\partial f}{\partial x_3}(x^*) = 0$$

$$\frac{\partial f}{\partial x_3}(x^*) = 8\left(\frac{6}{7}\right) - 2\left(\frac{17}{7}\right) - 2 = \frac{48}{7} - \frac{34}{7} - 2 = \frac{14}{7} - 2 = 2 - 2 = 0$$

Conclusion:

All KKT conditions are satisfied at x^*, so x^* is an optimal solution for problem (Q).

(ii)

```python
# Define objective function
def f(x):
    x1, x2, x3 = x
    return 2 * x1 ** 2 + 3 * x2 ** 2 + 4 * x3 ** 2 + 2 * x1 * x2 - 2 * x1 * x3 - 8 * x1
    - 4 * x2 - 2 * x3

# Define gradient
def grad_f(x):
```

```
    x1, x2, x3 = x
    df_dx1 = 4 * x1 + 2 * x2 - 2 * x3 - 8
    df_dx2 = 6 * x2 + 2 * x1 - 4
    df_dx3 = 8 * x3 - 2 * x1 - 2
    return np.array([df_dx1, df_dx2, df_dx3])

# Parameters
L = 10  # Lipschitz constant
alpha = 1 / L  # Step size
x = np.array([1.0, 1.0, 1.0])  # Initial point
num_iters = 100  # Number of iterations
f_values = []  # Store function values

# Gradient projection method iteration
for k in range(num_iters):
    grad = grad_f(x)
    x = x - alpha * grad
    x = np.maximum(x, 0)  # Projection onto non-negative orthant
    f_values.append(f(x))

# Output final solution and function value
final_x = x
final_f_value = f(x)

# Plot function value convergence
plt.plot(range(1, num_iters + 1), f_values)
plt.xlabel('Iteration')
plt.ylabel('Function value f(x^k)')
plt.title('Function Value Convergence Curve for Gradient Projection Method')
plt.grid(True)
plt.show()

final_x, final_f_value
```

Listing 1: Gradient Projection Method

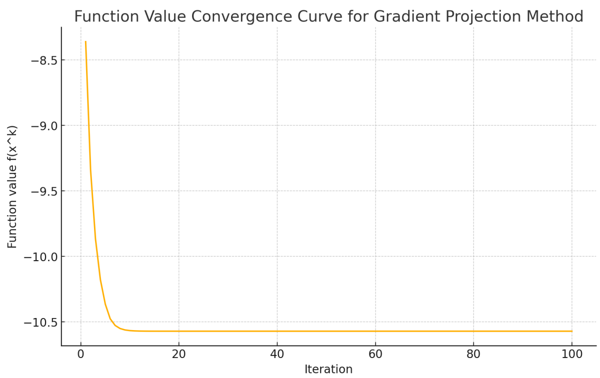

Figure 1: Enter Caption

Final solution x^*: [2.4286, 0, 0.8571]

Final function value $f(x^*)$: -10.5714