IOE 611: Homework 5

Yulun Zhuang

November 27, 2024

Problem 1

A convex problem in which strong duality fails. Consider the optimization problem

$$\begin{array}{ll} \text{minimize} & e^{-x} \\ \text{subject to} & x^2/y \leq 0 \end{array}$$

with variables x and y, and domain $\mathcal{D} = \{(x, y) \mid y > 0\}$.

(a) Verify that this is a convex optimization problem. Find the optimal value.

(b) Give the Lagrange dual problem, and find the optimal solution λ^* and optimal value d^* of the dual problem. What is the optimal duality gap?

(c) Does Slater's condition hold for this problem?

(d) What is the optimal value $p^*(u)$ of the perturbed problem

$$\begin{array}{ll} \text{minimize} & e^{-x} \\ \text{subject to} & x^2/y \leq u \end{array}$$

as a function of u? Verify that the global sensitivity inequality

$$p^*(u) \geq p^*(0) - \lambda^* u$$

does not hold.

Solution. (a) The Hessian of the objective function $f_0 = e^{-x}$ is

$$\nabla^2 f_0 = \begin{bmatrix} e^{-x} & 0 \\ 0 & 0 \end{bmatrix} \succeq 0$$

which shows it is convex.

Given $y > 0$ and $x^2 \geq 0$, x^2 must be zero for $x^2/y \leq 0$ to satisfy the constraint. Therefore, the feasible region is the positive y-axis, i.e. $\{(0, y) \mid y > 0\}$, which is convex.

Then we take the Hessian of the constraint function,

$$\nabla^2 f_1 = \begin{bmatrix} 2/y & -2x/y^2 \\ -2x/y^2 & 2x^2/y^3 \end{bmatrix} \succeq 0, \forall y > 0$$

which shows it is also convex.

Since the objective function and constraint are convex, the optimization problem is convex.

Since the only feasible value for x is zero, the optimal value is

$$f_0^* = e^0 = 1$$

(b) The Lagrangian dual function is, for $\lambda \geq 0$,

$$g(\lambda) = \inf_{x,y} L(x, y, \lambda) = \inf_{x,y} e^{-x} + \lambda \frac{x^2}{y} = 0$$

with $x, y \to \infty$ by inspection. The dual problem is

$$d^* = \max_{\lambda \geq 0} g(\lambda) = \max_{\lambda \geq 0} 0 = 0$$

where λ^* is not obtained. The duality gap is

$$p^* - d^* = 1 - 0 = 1$$

(c) Slater's condition does not hold. Since $x^2 \geq 0$ and $y > 0$, there does not exists a feasible (x, y) such that the strict inequality $x^2/y < 0$ hold.

(d) If $u = 0$, then $p^*(u) = 1$.

If $u < 0$, the problem is infeasible since x^2/y is non-negative given that $y > 0$. Therefore, $p^*(u) = \infty$.

If $u > 0$, then the constraint is $x^2 \leq uy$. With $\inf e^{-x} = 0$ as x^* approaches infinity, for any u, there always exists some y^* that satisfies $(x^*)^2 < uy^*$. Therefore, $p^*(u) = 0$.

For the global sensitivity, first find λ^* for the dual problem,

$$\begin{aligned}
\lambda^* &= \operatorname*{argmax}_{\lambda \geq 0} \inf_{x,y} e^{-x} + \lambda \left(\frac{x^2}{y} - u \right) \\
&= \operatorname*{argmax}_{\lambda \geq 0} -u\lambda
\end{aligned}$$

where $\lambda^* = 0$ if $u > 0$, $\lambda^* = \infty$ is $u < 0$. Consider $u > 0$,

$$\begin{aligned}
p^*(u) &= 0 \\
p^*(0) - \lambda^* u &= 1 - 0 = 1
\end{aligned}$$

where $p^*(u) < p^*(0) - \lambda^* u$ and therefore the global sensitivity inequality does not hold.

$\square$

Problem 2

Geometric interpretation of duality. For each of the following optimization problems, draw a sketch of the sets

$$\begin{aligned}\mathcal{G} &= \{(u,t) \mid \exists x \in \mathcal{D}, f_0(x) = t, f_1(x) = u\}, \\ \mathcal{A} &= \{(u,t) \mid \exists x \in \mathcal{D}, f_0(x) \leq t, f_1(x) \leq u\},\end{aligned}$$

give the dual problem, and solve the primal and dual problems. Is the problem convex? Is Slater's condition satisfied? Does strong duality hold?

The domain of the problem is $\mathbb{R}$ unless otherwise stated.

(a) Minimize x subject to $x^2 \leq 1$.

(b) Minimize x subject to $x^2 \leq 0$.

(c) Minimize x subject to $|x| \leq 0$.

(d) Minimize x subject to $f_1(x) \leq 0$ where

$$f_1(x) = \begin{cases} -x+2 & x \geq 1 \\ x & -1 \leq x \leq 1 \\ -x-2 & x \leq -1. \end{cases}$$

(e) Minimize x^3 subject to $-x+1 \leq 0$.

(f) Minimize x^3 subject to $-x+1 \leq 0$ with domain $\mathcal{D} = \mathbb{R}_+$.

Solution. (a) $\mathcal{G} = \{(t,u) \mid t \in \mathcal{D}, u = t^2 - 1\}$

The primal problem is

$$\begin{aligned}\min_x \quad & x \\ \text{s.t.} \quad & x^2 - 1 \leq 0\end{aligned}$$

The problem is convex since the objective is affine and the constraint is quadratic over a convex domain.

The Slater's condition is satisfied since there exists feasible $\bar{x} \in (-1,1) \subseteq \textbf{int}\,\mathcal{D}$ which satisfies strict inequality constraint $f_1(\bar{x}) < 0$.

The primal optimal $p^* = f_0(x^*) = -1$ is obtained when $x^* = -1$ is on the boundary of the constraint.

The dual problem is

$$\begin{aligned}&\max_{\lambda \geq 0} \ \min_x \ x + \lambda(x^2 - 1) \\ =&\max_{\lambda \geq 0} \ -\frac{1}{4\lambda} - \lambda\end{aligned}$$

where dual optimal $d^* = g(\lambda^*) = -1$ is obtained when $\lambda^* = 1/2$.

The strong duality holds since $d^* = p^*$. We can also conclude that since the problem is convex and the Slater's condition is satisfied.

(b) $\mathcal{G} = \{(t, u) \mid t \in \mathcal{D}, u = t^2\}$

The primal problem is

$$\begin{aligned} \min_x \quad & x \\ \text{s.t.} \quad & x^2 \leq 0 \end{aligned}$$

The problem is convex since the objective is affine and the constraint is quadratic over a convex domain.

The Slater's condition is not satisfied since there does not exist feasible $\bar{x} \in \textbf{int}\,\mathcal{D}$ which satisfies strict inequality constraint $f_1(\bar{x}) < 0$.

The primal optimal $p^* = f_0(x^*) = 0$ is obtained when $x^* = 0$ is on the boundary of the constraint.

The dual problem is

$$\begin{aligned} & \max_{\lambda \geq 0} \min_x \; x + \lambda x^2 \\ = & \max_{\lambda \geq 0} \; -\frac{1}{4\lambda} \end{aligned}$$

where dual optimal $d^* = g(\lambda^*) = 0$ is obtained when $\lambda^* = \infty$.

The strong duality holds since $d^* = p^*$, but the Slater's condition is not satisfied.

(c) $\mathcal{G} = \{(t, u) \mid t \in \mathcal{D}, u = |t|\}$

The primal problem is

$$\begin{aligned} \min_x \quad & x \\ \text{s.t.} \quad & |x| \leq 0 \end{aligned}$$

The problem is convex since the objective is affine and the constraint is affine over a single feasible point $x = 0$.

The Slater's condition is not satisfied since there does not exist feasible $\bar{x} \in \textbf{int}\,\mathcal{D}$ which satisfies strict inequality constraint $f_1(\bar{x}) < 0$.

The primal optimal $p^* = f_0(x^*) = 0$ is obtained when $x^* = 0$.

The dual problem is

$$\begin{aligned} & \max_{\lambda \geq 0} \min_x \; x + \lambda |x| \\ = & \max_{\lambda \geq 0} \min_x \begin{cases} (1+\lambda)x & x \geq 0 \\ (1-\lambda)x & x < 0 \end{cases} \end{aligned}$$

where dual optimal $d^* = g(\lambda^*) = 0$ is obtained when $\lambda^* = 1$.

The strong duality holds since $d^* = p^*$, but the Slater's condition is not satisfied.

(d) $\mathcal{G} = \{(t, u) \mid t \in \mathcal{D}, u = f_1(t)\}$

The primal problem is

$$\begin{aligned} \min_x \quad & x \\ \text{s.t.} \quad & f_1(x) \le 0 \end{aligned}$$

The problem is not convex since the objective is affine and the constraint is piece-wise linear (nonlinear) over a convex domain.

The Slater's condition is not useful in this case.

The primal optimal $p^* = f_0(x^*) = -2$ is obtained when $x^* = -2$.

The dual problem is

$$\begin{aligned} & \max_{\lambda \ge 0} \min_x \; x + \lambda f_1(x) \\ = & \max_{\lambda \ge 0} \min_x \; (1-\lambda)x - 2 \end{aligned}$$

where dual optimal $d^* = g(\lambda^*) = -2$ is obtained when $\lambda^* = 1$.

The strong duality holds since $d^* = p^*$, but the problem is not convex.

(e) $\mathcal{G} = \{(t, u) \mid t = (1-u)^3\}$

The primal problem is

$$\begin{aligned} \min_x \quad & x^3 \\ \text{s.t.} \quad & -x + 1 \le 0 \end{aligned}$$

The problem is not convex since the Hessian of the objective function $\nabla^2 f_0 = 6x < 0$ for $x < 0$.

The Slater's condition is not useful in this case.

The primal optimal $p^* = f_0(x^*) = 1$ is obtained when $x^* = 1$.

The dual problem is

$$\max_{\lambda \ge 0} \min_x \; x^3 + \lambda(-x + 1)$$

where dual optimal $d^* = g(\lambda^*) = -\infty$ is not attained.

The strong duality does not hold since $d^* \ne p^*$.

(f) $\mathcal{G} = \{(t, u) \mid t = (1-u)^3, u \le 1\}$

The primal problem is

$$\begin{aligned} \min_{x \ge 0} \quad & x^3 \\ \text{s.t.} \quad & -x + 1 \le 0 \end{aligned}$$

The problem is convex since the Hessian of the objective function $\nabla^2 f_0 = 6x \ge 0$ for $x \ge 0$, and the constraint is affine over a convex domain.

The Slater's condition is satisfied since there exists feasible $\bar{x} \in \mathbb{R}_+ \subseteq \mathbf{int}\, \mathcal{D}$ which satisfies strict inequality constraint $f_1(\bar{x}) < 0$.

The primal optimal $p^* = f_0(x^*) = 1$ is obtained when $x^* = 1$.

The dual problem is

$$\begin{aligned} &\max_{\lambda \geq 0} \min_{x \geq 0} x^3 + \lambda(-x+1) \\ =&\max_{\lambda \geq 0} -\frac{2}{3^{3/2}}\lambda^{3/2} + \lambda \end{aligned}$$

where dual optimal $d^* = g(\lambda^*) = 1$ is obtained when $\lambda^* = 3$.

The strong duality holds since $d^* = p^*$. $\square$

Problem 3

Equality constrained least-squares. Consider the equality constrained least-squares problem

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|_2^2 \\ \text{subject to} & Gx = h \end{array}$$

where $A \in \mathbb{R}^{m\times n}$ with $\mathbf{rank}\, A = n$, and $G \in \mathbb{R}^{p\times n}$ with $\mathbf{rank}\, G = p$.

Give the KKT conditions, and derive expressions for the primal solution x^* and the dual solution ν^*.

Solution. Since the problem is convex and the Slater's condition is satisfied, the strong duality holds such that $p^* = d^*$. Next we will find the dual optimal ν^* and primal optimal x^* respectively.

The Lagrangian is

$$L(x, \nu) = \|Ax - b\|_2^2 + \nu^\top (Gx - h)$$

The KKT conditions are

$$\begin{aligned} \nabla L(x^*, \nu^*) = 2A^\top (Ax^* - b) + G^\top \nu^* = 0 \\ Gx^* = h \end{aligned}$$

The dual function is

$$\begin{aligned} g(\nu) &= \min_x \ L(x, \nu) \\ &= \min_x \ \|Ax - b\|_2^2 + \nu^\top (Gx - h) \\ &= \min_x \ x^\top A^\top Ax + (v^\top G - 2b^\top A)x + b^\top b - v^\top h \\ &= -\frac{1}{4}(G^\top \nu - 2A^\top b)^\top (A^\top A)^{-1}(G^\top \nu - 2A^\top b) - \nu^\top h \end{aligned}$$

where the minimizer is found from its derivative

$$\begin{aligned} 2A^\top Ax + (v^\top G - 2b^\top A)^\top = 0 \\ x = -\frac{1}{2}(A^\top A)^{-1}(G^\top v - 2A^\top b) \end{aligned}$$

Take the derivative of $g(\nu)$ w.r.t. ν, we have

$$\begin{aligned} 0 &= -\frac{1}{2}G(A^\top A)^{-1}G^\top v + G(A^\top A)^{-1}A^\top b - h^\top \\ v^* &= -2(G(A^\top A)^{-1}G^\top)^{-1}(h - GA^\dagger b) \end{aligned}$$

where $A^\dagger = (A^\top A)^{-1}A^\top$.

From the first KKT condition, plug in ν^*

$$\begin{aligned} 0 &= 2(A^\top A)x - 2A^\top b + G^\top v^* \\ 0 &= 2(A^\top A)x^* - 2A^\top b - 2G^\top (G(A^\top A)^{-1}G^\top)^{-1}(h - GA^\dagger b) \\ x^* &= A^\dagger b + (A^\top A)^{-1}G^\top (G(A^\top A)^{-1}G^\top)^{-1}(h - GA^\dagger b) \end{aligned}$$

Verify the second KKT condition,

$$\begin{aligned}
h &= Gx \\
&= GA^{\dagger}b + G(A^{\top}A)^{-1}G^{\top}(G(A^{\top}A)^{-1}G^{\top})^{-1}(h - GA^{\dagger}b) \\
&= GA^{\dagger}b - GA^{\dagger}b + h \\
&= h
\end{aligned}$$

$\square$

Problem 4

Derive the KKT conditions for the problem

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}\, X - \log\det X \\ \text{subject to} & Xs = y, \end{array}$$

with variable $X \in \mathbb{S}^n$ and domain $\mathbb{S}^n_{++}$. $y \in \mathbb{R}^n$ and $s \in \mathbb{R}^n$ are given, with $s^\top y = 1$.

Verify that the optimal solution is given by

$$X^* = I + yy^\top - \frac{1}{s^\top s} ss^\top.$$

Solution. The Lagrangian is

$$L(X,\nu) = \mathbf{tr}\, X - \log\det X + \nu^\top (Xs - y)$$

The KKT conditions are

$$\begin{aligned} 0 &= \nabla L(X^*, \nu^*) = \nabla(\mathbf{tr}\, X^* - \log\det X^*) + (\nu^*)^\top \nabla(X^* s - y) \\ 0 &= X^* s - y \end{aligned}$$

To find and verify the optimal solution X^*, find (X^*, ν^*) that satisfy the KKT conditions. Since the problem is convex, satisfying the KKT conditions guarantees X^* is optimal.

From the first KKT condition,

$$\begin{aligned} 0 &= \nabla(\mathbf{tr}\, X - \log\det X) + \nu^\top \nabla(Xs - y) \\ 0 &= \nabla(\mathbf{tr}\, X) - \nabla(\log\det X) + \nabla(\nu^\top Xs) \\ 0 &= I - X^{-1} + \frac{1}{2}\nabla(\nu^\top Xs + s^\top X\nu) \\ 0 &= I - X^{-1} + \frac{1}{2}(\nu s^\top + s\nu^\top) \\ (X^*)^{-1} &= I + \frac{1}{2}(\nu s^\top + s\nu^\top) \end{aligned}$$

where $\nabla(\log\det X) = X^{-1}$ is given from the textbook.

Next, substitute $(X^*)^{-1}$ into the second KKT condition to find ν^*,

$$\begin{aligned} Xs &= y \\ X^{-1}Xs &= (I + \frac{1}{2}(\nu s^\top + s\nu^\top))y \\ y^\top s &= y^\top y + \frac{1}{2}(y^\top \nu + (\nu^\top y) y^\top s) \qquad \Leftarrow \text{multiply } y^\top \\ \nu^* &= -2y + (1 + y^\top y)s \end{aligned}$$

Plug in ν^* to find $(X^*)^{-1}$,

$$\begin{aligned} X^{-1} &= I + \frac{1}{2}(\nu s^\top + s\nu^\top) \\ X^{-1} &= I + \frac{1}{2}\left(-2ys^\top + (1 + y^\top y)ss^\top - 2sy^\top + (1 + y^\top y)ss^\top\right) \\ (X^*)^{-1} &= I - ys^\top - sy^\top + (1 + y^\top y)ss^\top \end{aligned}$$

To show that

$$X^* = I + yy^\top - \frac{1}{s^\top s} ss^\top$$

is the optimal is equivalent to showing $(X^*)^{-1}X^* = I$,

$$\begin{aligned}
&\left(I + (1 + y^\top y)ss^\top - ys^\top - sy^\top\right)\left(I - \frac{1}{s^\top s}ss^\top + yy^\top\right) \\
=&I + yy^\top + (1 + y^\top y)(sy^\top) - yy^\top - (sy^\top + sy^\top yy^\top) \\
=&I + (1 + y^\top y)(sy^\top) - (1 + y^\top y)(sy^\top) \\
=&I
\end{aligned}$$

$\square$

Problem 5

Dual of SOCP. Show that the dual of the SOCP

$$\begin{array}{ll} \text{minimize} & f^\top x \\ \text{subject to} & \|A_i x + b_i\|_2 \le c_i^\top x + d_i, \quad i = 1, \dots, m, \end{array}$$

with variables $x \in \mathbb{R}^n$, can be expressed as

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{i=1}^m (b_i^\top u_i - d_i v_i) \\ \text{subject to} & \displaystyle\sum_{i=1}^m (A_i^\top u_i - c_i v_i) + f = 0 \\ & \|u_i\|_2 \le v_i, \quad i = 1, \dots, m, \end{array}$$

with variables $u_i \in \mathbb{R}^{n_i}, v_i \in \mathbb{R}, i = 1, \dots, m$. The problem data are $f \in \mathbb{R}^n$, $A_i \in \mathbb{R}^{n_i \times n}$.

Derive the dual in the following two ways.

(a) Introduce new variables $y_i \in \mathbb{R}^{n_i}$ and $t_i \in \mathbb{R}$ and equalities $y_i = A_i x + b_i, t_i = c_i^\top x + d_i$, and derive the Lagrange dual.

(b) Start from the conic formulation of the SOCP and use the conic dual. Use the fact that the second-order cone is self-dual.

Proof. Based on y_i, t_i, the constraints can be rewritten as

$$\begin{aligned} \|y_i\|_2 &\le t_i \\ A_i x + b_i &= y_i \\ c_i^\top x + d_i &= t_i \end{aligned}$$

The Lagrangian $L(x, y, t, \lambda, \nu, \mu)$ is

$$L = f^\top x + \sum_{i=1}^m \lambda_i(\|y_i\|_2 - t_i) + \sum_{i=1}^m \nu_i^\top (y_i - A_i x - b_i) + \sum_{i=1}^m \mu_i (t_i - c_i^\top x - d_i)$$

The dual function is

$$g(\lambda, \nu, \mu) = \inf_{x,y,t} L(x, y, t, \lambda, \nu, \mu)$$

The minimum over x is

$$\inf_x \left(f^\top x - \sum_{i=1}^m \nu_i^\top A_i x - \sum_{i=1}^m \mu_i c_i^\top x \right) = \inf_x \left(f^\top - \sum_{i=1}^m \nu_i^\top A_i + \mu_i c_i^\top \right) x$$

which is only bounded and equals 0 if

$$f^\top - \sum_{i=1}^m \nu_i^\top A_i + \mu_i c_i^\top = 0$$

$$-\sum_{i=1}^m \left(A_i^\top \nu_i + \mu_i c_i \right) + f = 0$$

The minimum over each y_i is

$$\inf_{y_i} \left(\lambda_i \|y_i\|_2 + \nu_i^\top y_i \right)$$

which is only bounded and equals 0 if $\|\nu_i\|_2 \leq \lambda_i$.

The minimum over each t_i is

$$\inf_{t_i} -\lambda_i t_i + \mu_i t_i = \inf_{t_i} (\mu_i - \lambda_i) t_i$$

which is only bounded and equals 0 if $\mu_i = \lambda_i$.

Plug in above results into the dual function, where all terms that depend on (x, y, t) have been minimized to zero,

$$\begin{aligned} g(\lambda, \nu, \mu) &= \inf_{x,y,t} L(x, y, t, \lambda, \nu, \mu) \\ &= -\sum_{i=1}^m \nu_i^\top b_i - \sum_{i=1}^m \lambda_i d_i \end{aligned}$$

with the conditions

$$\begin{aligned} -\sum_{i=1}^m \left(A_i^\top \nu_i + \lambda_i c_i \right) + f = 0 \\ \|\nu_i\|_2 \leq \lambda_i \end{aligned}$$

Let $u_i = -\nu_i, v_i = \lambda_i$, the dual problem $\max g(u_i, v_i)$ is

$$\begin{aligned} \max \quad & \sum_{i=1}^m b_i^\top u_i - d_i v_i \\ \text{s.t.} \quad & \sum_{i=1}^m A_i^\top u_i - c_i v_i + f = 0 \\ & \|u_i\|_2 \leq v_i \quad i = 1, \ldots, m \end{aligned}$$

(b) The conic formulation is

$$\begin{aligned} \min \quad & f^\top x \\ \text{s.t.} \quad & -(A_i x + b, c_i^\top x + d_i) \leq_{K_i} 0 \end{aligned}$$

The conic dual is

$$\begin{aligned} \max \quad & \sum_{i=1}^m b_i^\top u_i - d_i v_i \\ \text{s.t.} \quad & \sum_{i=1}^m A_i^\top u_i - c_i v_i + f = 0 \\ & (u_i, v_i) \geq_{K_i} 0 \quad i = 1, \ldots, m \end{aligned}$$

□

Problem 6

Gradient and Newton methods. Consider the unconstrained problem

$$\text{minimize} \quad f(x) = -\sum_{i=1}^{m} \log(1 - a_i^\top x) - \sum_{i=1}^{n} \log(1 - x_i^2),$$

with variable $x \in \mathbb{R}^n$, and $\mathbf{dom}\, f = \{x \mid a_i^\top x < 1, i = 1, \ldots, m, |x_i| < 1, i = 1, \ldots, n\}$. This is the problem of computing the analytic center of the set of linear inequalities

$$a_i^\top x \leq 1, \quad i = 1, \ldots, m, \quad |x_i| \leq 1, \quad i = 1, \ldots, n.$$

Note that we can choose $x^{(0)} = 0$ as our initial point. You can generate instances of this problem by choosing a_i from some distribution on $\mathbb{R}^n$.

(a) Use the gradient method to solve the problem, using reasonable choices for the backtracking parameters, and a stopping criterion of the form $\|\nabla f(x)\|_2 \leq \eta$. Plot the objective function and step length versus iteration number. (Once you have determined p^* to high accuracy, you can also plot $f - p^*$ versus iteration.) Experiment with the backtracking parameters α and β to see their effect on the total number of iterations required. Carry these experiments out for several instances of the problem, of different sizes.

(b) Repeat using Newton's method, with stopping criterion based on the Newton decrement λ^2. Look for quadratic convergence. You do not have to use an efficient method to compute the Newton step; you can use a general purpose dense solver, although it is better to use one that is based on a Cholesky factorization.

Solution. (i) To generate instances of the problem, we randomly sampled the rows of the matrix $A = \begin{bmatrix} a_1^\top \\ a_2^\top \\ \vdots \\ a_m^\top \end{bmatrix}$ from a uniform distribution over $[-1, 1]^n$, where each row vector $a_i \in \mathbb{R}^n$. For the constraint $a_i^\top x < 1$, we ensured that a_i is scaled such that the constraint bounds are consistent within the domain $|x_i| < 1$ for $i = 1, \ldots, n$. Problem sizes (m, n) were varied to analyze performance for different dimensions.

(ii) Gradient and Hessian formulations

The objective function is given as:

$$f(x) = -\sum_{i=1}^{m} \log(1 - a_i^\top x) - \sum_{i=1}^{n} \log(1 - x_i^2).$$

The gradient of $f(x)$ is:

$$\nabla f(x) = \sum_{i=1}^{m} \frac{a_i}{1 - a_i^\top x} + \sum_{i=1}^{n} \frac{2x_i}{1 - x_i^2}.$$

The Hessian of $f(x)$ is:

$$\nabla^2 f(x) = \sum_{i=1}^{m} \frac{a_i a_i^\top}{(1 - a_i^\top x)^2} + 2\text{diag}\left(\frac{1 + x_i^2}{(1 - x_i^2)^2}\right).$$

(iii) Parameter choices in implementation

- Backtracking parameters: $\alpha = 0.25$, $\beta = 0.5$.
- Gradient method stopping criterion: $\|\nabla f(x)\|_2 \leq \eta$, with $\eta = 10^{-6}$.
- Newton's method stopping criterion: $\lambda^2/2 \leq \eta$, with $\eta = 10^{-9}$.

(iv) Summary of observations and results

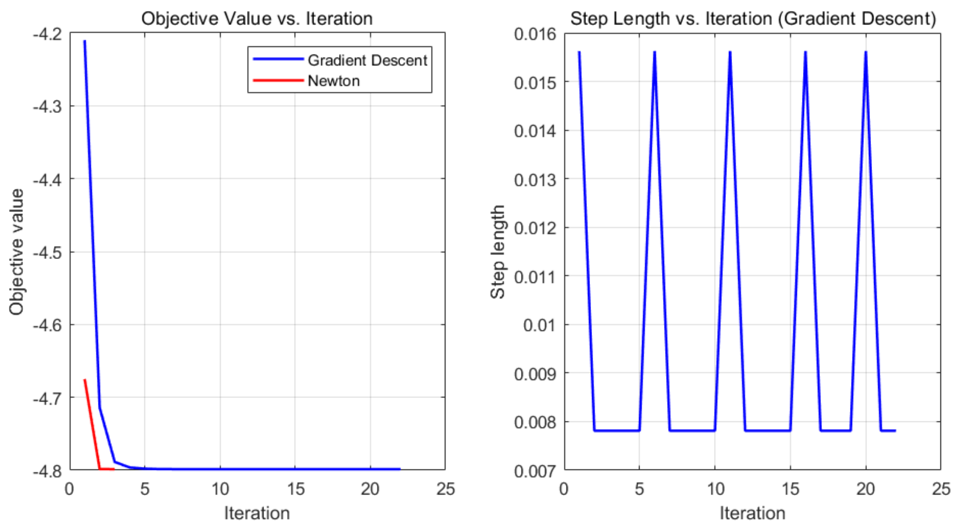

Figure 1: Comparisons of Gradient method and Newton's method

As shown in Fig. 1,

- The gradient method converged linearly. Increasing the backtracking parameters α or β reduced the total number of iterations but required longer computation per iteration due to larger step size.
- Newton's method exhibited quadratic convergence near the solution. It was significantly faster in terms of iteration count, although each iteration was more computationally intensive due to Hessian inversion.
- The step size t_k for the gradient method decreased rapidly in early iterations but stabilized near the optimal solution.

□