AMSC660 Homework #13

Kecai Xuan

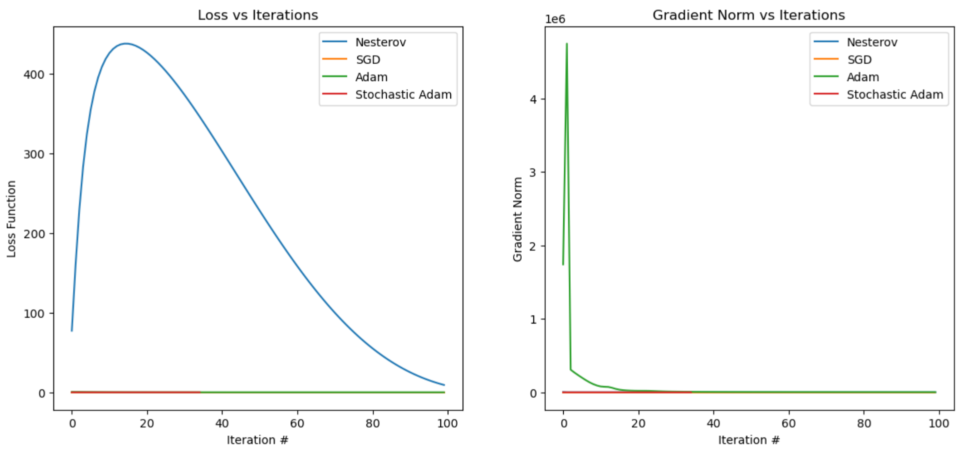
December 8, 2024

Problem 1

the code can be found at https://github.com/Bessgendre/homework-13-AMSC660-2024

(a)

Using the Deterministic and Stochastic Nesterov method, the loss function will first increase and then decrease:



(b)

The Stochastic Adam method is the most efficient among the four methods, with the least number of iterations and the smallest loss value: For this problem, the best batch size is around 16.

Problem 2

$$\begin{bmatrix} G & A^\top \\ A & 0 \end{bmatrix} = \begin{bmatrix} I & 0 \\ X & I \end{bmatrix} \begin{bmatrix} G & 0 \\ 0 & S \end{bmatrix} \begin{bmatrix} I & X^\top \\ 0 & I \end{bmatrix} = \begin{bmatrix} G & GX^\top \\ XG & XGX^\top + S \end{bmatrix}$$

Since G is symmetric, as a result:

$$X = AG^{-1}, \quad S = -AG^{-1}A^\top$$

where S is the negative congruence matrix of G^{-1}. When the eigenvalues of G^{-1} are all positive, then S is symmetric negative definite with the number of negative eigenvalues equal to m.

Since $K = \begin{bmatrix} G & A^\top \\ A & 0 \end{bmatrix}$ and $\begin{bmatrix} G & 0 \\ 0 & S \end{bmatrix}$ are related by congruence, they have the same number of positive, zero and negative eigenvalues, which is the result of Sylvester's law of inertia. Therefore, matrix K has d positive eigenvalues and m negative eigenvalues.

Problem 3

(a)

Introduce the Lagrange function for this problem:

$$\mathscr{L}(x,\lambda) = \frac{1}{2}x^\top Gx + c^\top x + \lambda^\top(Ax - b)$$

where λ is the Lagrange multiplier. The necessary conditions for optimality are:

$$\begin{aligned}\nabla_x \mathscr{L} &= Gx + c + A^\top\lambda = 0\\ \nabla_\lambda \mathscr{L} &= Ax - b = 0\end{aligned}$$

To write it in the matrix form:

$$\begin{bmatrix} G & A^\top \\ A & 0 \end{bmatrix}\begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}$$

where the matrix on the left-hand side $K = \begin{bmatrix} G & A^\top \\ A & 0 \end{bmatrix}$ is the KKT matrix.

(b)

Consider $Kz = 0$:

$$\begin{bmatrix} G & A^\top \\ A & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

which can be written as:

$$\begin{aligned}Gx + A^\top y &= 0\\ Ax &= 0\end{aligned}$$

Since Z is the basis of the null space of A, then x can be written as the linear combination of the columns of Z: $x = Zv$. Substitute it into the first equation:

$$GZv + A^\top y = 0$$

Multiply $Z^\top$ on both sides, because $AZ = 0$, we have:

$$Z^\top G Z v = 0$$

Since $Z^\top G Z$ is positive definite with full rank, then $v = 0$. Therefore, $x = 0$.

Back to $Gx + A^\top y = 0$, we have $A^\top y = 0$. Since A has full row rank, then $y = 0$. Therefore, $z = 0$ is the only solution to $Kz = 0$, which means K is invertible.

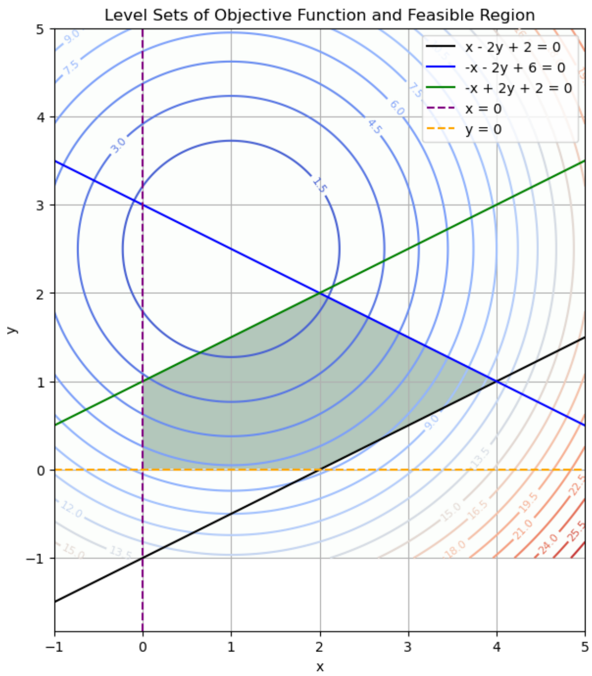
(c)

When K is invertible, there exist unique x^* and λ^* that satisfy $\begin{bmatrix} G & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}$. Therefore, the solution to the optimization problem is unique.

Problem 4

(a)

The level sets and the feasible set is shown in the following figure shaded in green:



(b)

The feasible region is a polygon, and the level sets are concentric circles around $(1, 2.5)$. So the optimal solution should have the minimum distance to the center, which is $(1.4, 1.7)$.

(c)

Iteration 1:

Current iterate: $(x_1, y_1) = (2, 0)$

Active set: $\mathcal{W} = \{3, 5\}$

Step 1: Form KKT conditions with current active set

We treat constraints 3 and 5 as equalities. Define the gradients:

$$\begin{aligned}
&\nabla f(x,y) = (2(x-1), 2(y-2.5)), \quad \text{At } (2,0): \ \nabla f = (2(2-1), 2(0-2.5)) = (2, -5) \\
&\nabla g_3 = \nabla(-x + 2y + 2) = (-1, 2) \\
&\nabla g_5 = \nabla(y) = (0, 1)
\end{aligned}$$

The KKT system for an active-set QP at a stationary point requires:

$$\nabla f(x,y) + \lambda_3 \nabla g_3 + \lambda_5 \nabla g_5 = 0$$

and the active constraints $g_3(x,y) = 0, g_5(x,y) = 0$.

Substitute:

$$(2, -5) + \lambda_3(-1, 2) + \lambda_5(0, 1) = (0, 0) \quad \Rightarrow \quad \lambda_3 = 2, \ \lambda_5 = 1$$

No search direction was needed here because we directly solved the KKT conditions at the given point. Since $\lambda_3, \lambda_5 > 0$, $(2, 0)$ is a KKT point for the subproblem defined by these two active constraints, but it is not the global minimum.

Now we should drop constraint #5 from the working set.

Iteration 2:

Step 1: Compute gradient at $(2, 0)$

$$\nabla f(x,y) = (2(x-1), 2(y-2.5)), \ \text{At } (2,0): \nabla f = (2(2-1), 2(0-2.5)) = (2, -5)$$

Step 2: Set up KKT with active set $\mathcal{W} = \{3\}$

We must find a search direction $d = (d_x, d_y)$ that satisfies:

$$2\,(d_x, d_y) + (2, -5) + \lambda_3(-1, 2) = 0 \quad \text{and} \quad (-1, 2)d = 0$$

where we find $d = (0.2, 0.1)$.

Step 3: Perform line search

We increase α from 0 until another constraint becomes active or we hit the minimum along that line. After checking feasibility, no immediate blocking constraints appear for small α. The best step (from a simple analytic line search) is found by minimizing f along the direction and it comes out $\alpha = 1$.

Step 4: Update iterate

$$(x_2, y_2) = (2, 0) + 1(0.2, 0.1) = (2.2, 0.1)$$

The following iterations are similar to iteration 1 and 2.

Iteration 3:

Still active set $\mathcal{W} = \{3\}$ at $(2.2, 0.1)$.

No descent direction is found, $\lambda_3 = 2.4 > 0$.

Drop constraint #3.

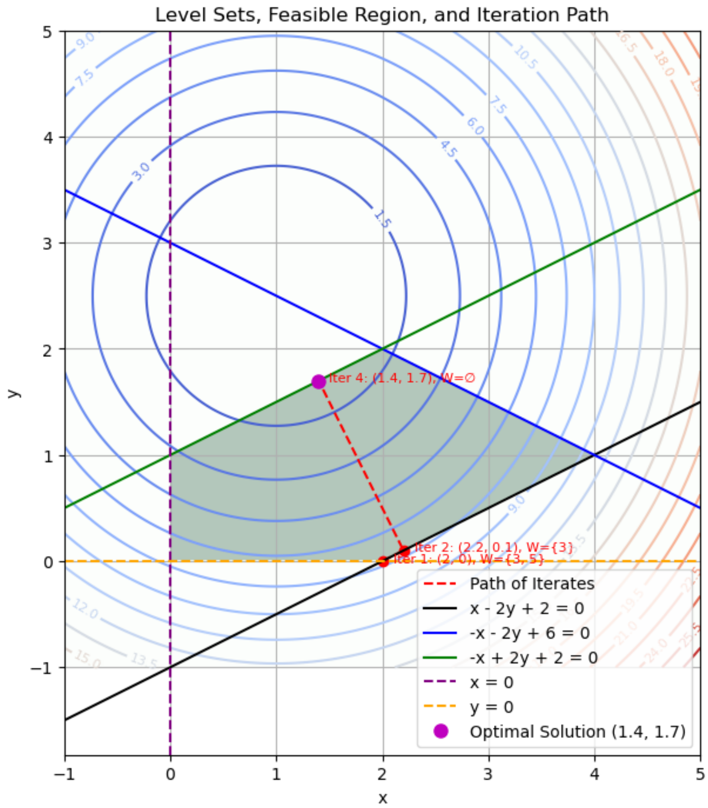
Iteration 4:

Unconstrained direction $d = (-1.2, 2.4)$.

Move to $(1.4, 1.7)$.

Iteration 5:

This is the known optimal solution. No further improvement.



Level Sets, Feasible Region, and Iteration Path
Iter 4: (1.4, 1.7), W=∅
Iter 2: (2.2, 0.1), W={3}
Iter 1: (2, 0), W={3, 5}
Path of Iterates
x - 2y + 2 = 0
-x - 2y + 6 = 0
-x + 2y + 2 = 0
x = 0
y = 0
Optimal Solution (1.4, 1.7)
x
y