Introduction to Numerical Optimization
Assignment 4

June 12, 2022

1 Augmented Lagrangian method for Constrained Optimization

1.1 Graph

We need to consider the following quadratic programming problem:

$$
\begin{aligned}
&\min_{x_1,x_2} 2(x_1-5)^2+(x_2-1)^2 \\
&\text{s.t} \quad \begin{array}{l} x_2 \le 1-0.5x_1 \\ x_2 \ge x_1 \\ x_2 \ge -x_1 \end{array}
\end{aligned}
$$

Which can be written as:

$$
\begin{aligned}
&\min_{x_1,x_2} 2x_1^2-20x_1+50+x_2^2-2x_2+1 \\
&\text{s.t} \quad \begin{array}{l} 0.5x_1+x_2-1 \le 0 \\ x_1-x_2 \le 0 \\ -x_1-x_2 \le 0 \end{array}
\end{aligned}
$$

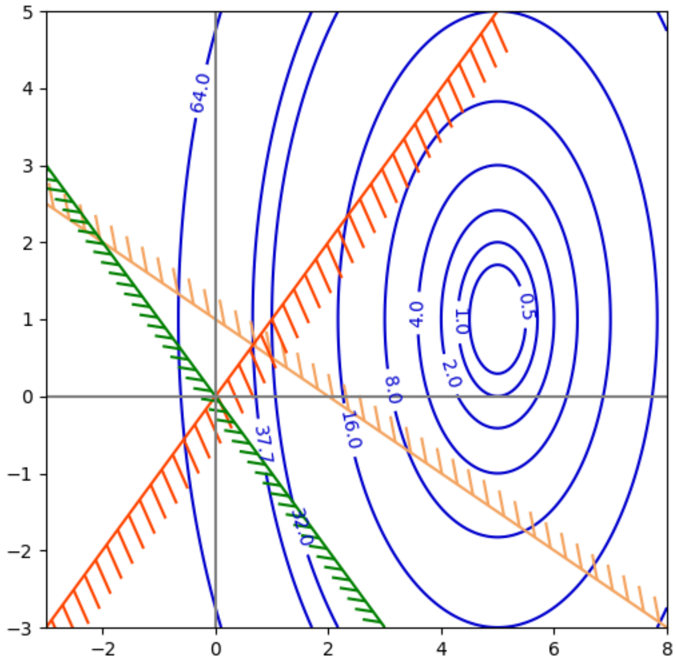
Or with matrices as:

$$
\begin{aligned}
&\min_{x} x^TAx+b^Tx+c \\
&\text{s.t} \quad Bx+d \le 0
\end{aligned}
$$

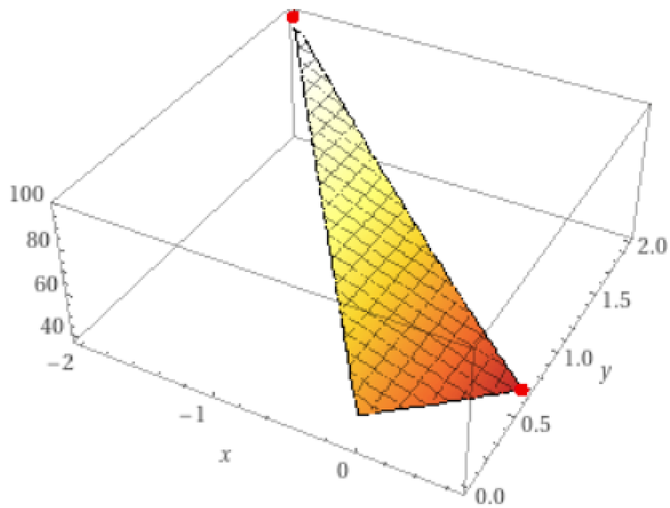
when:

- $A \in \mathbb{R}^{2\times 2}, A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, x, b \in \mathbb{R}^2, b = \begin{bmatrix} -20 \\ -2 \end{bmatrix}, c \in \mathbb{R}, c = 51$
- $B \in \mathbb{R}^{3\times 2}, B = \begin{bmatrix} 0.5 & 1 \\ 1 & -1 \\ -1 & -1 \end{bmatrix}, d \in \mathbb{R}^3, d = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}$

We plot the objective function contours (ellipses as we have different eigen values in our A matrix) and the constraints:



From plot, we can see the active constraints are the red and yellow together (the intersection point) and the blue ($-x_1 - x_2 \leq 0$) is none active.



1.2 Optimal Solution

The intersection of the 2 active constraints when both are satisfied:

$$x_1 = x_2, 0.5x_1 + x_2 = 1 \Rightarrow x = (x_1, x_2) = (\frac{2}{3}, \frac{2}{3}), f^*(x) = 37\frac{2}{3}$$

1.3 KKT conditions

The Lagrangian is defined as follows:

$$L(x, \lambda) = f(x) + \lambda^T g(x) = f(x) + \Sigma_{i=1}^3 \lambda_i g_i(x)$$

According to KKT, if x^* is an optimal solution and $\nabla g_i(x)$ are independent (and we can see they are for any x according to plot), we have $\lambda^* \in \mathbb{R}^3$ for which:

1. $\nabla_x L(x^*, \lambda^*) = 0$
2. $g(x^*) \leq 0$
3. $\lambda_i^* \geq 0, i \in I_a$
4. $\lambda_i^* = 0, i \notin I_a$

$\Rightarrow$ And the complementary slackness:

5. $\Sigma_{i=1}^3 \lambda_i^* g_i(x^*) = 0$

Note: since the constraints ($g_i(x)$) are affine functions then we stand with LCQ condition and no need for the independence condition for the above KKT first order conditions to hold.

Let's find the optimal point and multipliers:

1. $\nabla_x L(x^*, \lambda^*) = 0 \Rightarrow$

$$\nabla_{x_1} L(x^*, \lambda^*) = \nabla_{x_1} f(x^*) + \Sigma_{i=1}^3 \lambda_i^* \nabla_{x_1} g_i(x^*) = 0 \Rightarrow$$
$$\nabla_{x1} L(x^*, \lambda^*) = 4(x_1 - 5) + 0.5\lambda_1^* + \lambda_2^* - \lambda_3^* = 0$$
$$\nabla_{x2} L(x^*, \lambda^*) = \nabla_{x_2} f(x^*) + \Sigma_{i=1}^3 \lambda_i^* \nabla_{x_2} g_i(x^*) = 0 \Rightarrow$$
$$\nabla_{x_2} L(x^*, \lambda^*) = 2(x_2 - 1) + \lambda_1^* - \lambda_2^* - \lambda_3^* = 0$$

2. $g(x^*) \leq 0$

$$g_1(x^*) = 0.5x_1 + x_2 - 1 \leq 0$$
$$g_2(x^*) = x_1 - x_2 \leq 0$$
$$g_3(x^*) = -x_1 - x_2 \leq 0$$

3. $\lambda_i^* \geq 0, i \in I_a$

4. $\lambda_i^* = 0, i \notin I_a$

5. $\Sigma_{i=1}^3 \lambda_i^* g_i(x^*) = 0$

$$\lambda_1^* g_1(x^*) = \lambda_1^*(0.5x_1 + x_2 - 1) = 0$$
$$\lambda_2^* g_2(x^*) = \lambda_2^*(x_1 - x_2) = 0$$
$$\lambda_3^* g_3(x^*) = \lambda_3^*(-x_1 - x_2) = 0$$

From the last 2 equations of comp. slackness:
We have:
1. $x_1 = x_2$ or $\lambda_2^* = 0$
2. $x_1 = -x_2$ or $\lambda_3^* = 0$

If we assume $x_1 = x_2 = -x_2$ we get the optimal point is $x_1 = x_2 = 0 \Rightarrow \lambda_1^* = 0, \lambda_3^* = \frac{-22}{2} = -11 < 0$ (from $\lambda_1^*(0.5x_1 + x_2 - 1) = 0$ and

$\nabla_x L(x^*, \lambda^*) = 0)$, and we get a contradiction.

So, $x_1 = x_2, \lambda_3^* = 0$ or $x_1 = -x_2, \lambda_2^* = 0$

if $x_1 = -x_2, \lambda_2^* = 0$
From

$$\lambda_1^*(0.5x_1 + x_2 - 1) = 0 \Rightarrow$$
$$\lambda_1^*(-0.5x_1 - 1) = 0$$

We get $\lambda_1^* = 0$ or $x_1 = -2, x_2 = 2$

if $x_1 = -2, x_2 = 2$
from first equations we get:

$$4(x_1 - 5) + 0.5\lambda_1^* - \lambda_3^* = 0$$
$$2(-x_1 - 1) + \lambda_1^* - \lambda_3^* = 0 \Rightarrow$$
$$6x_1 - 18 - 0.5\lambda_1^* = 0 \Rightarrow$$

$\lambda_1^* < 0$ and we get a contradiction again as $\lambda_1^* \geq 0$

if $\lambda_1^* = 0$
from first equations we get:

$$4(x_1 - 5) - \lambda_3^* = 0$$
$$2(-x_1 - 1) - \lambda_3^* = 0 \Rightarrow$$
$$x_1 = 3, \lambda_3^* < 0$$

and again we get a contradiction as $\lambda_3^* \geq 0$

Thus, we are left with:

$$x_1 = x_2, \lambda_3^* = 0$$

From

$$\lambda_1^*(0.5x_1 + x_2 - 1) = 0 \Rightarrow$$
$$\lambda_1^*(1.5x_1 - 1) = 0$$

We get $\lambda_1^* = 0$ or $x_1 = \frac{2}{3}, x_2 = \frac{2}{3}$

if $\lambda_1^* = 0$
from first equations we get:

$$4(x_1 - 5) + \lambda_2^* = 0$$
$$2(x_1 - 1) - \lambda_2^* = 0 \Rightarrow$$
$$x_1 = \frac{22}{6}, \lambda_2^* = \frac{16}{3}$$

But in this case we get a contradiction from:

$$g_1(x^*) = 0.5x_1 + x_2 - 1 \leq 0$$

Bottom line, we are left with:

$$x_1 = x_2 = \frac{2}{3}, \lambda_3^* = 0$$

From first equations we get:

$$4(x_1 - 5) + 0.5\lambda_1^* + \lambda_2^* = 0$$
$$2(x_1 - 1) + \lambda_1^* - \lambda_2^* = 0 \Rightarrow$$
$$\frac{-52}{3} + 0.5\lambda_1^* + \lambda_2^* = 0$$
$$-\frac{2}{3} + \lambda_1^* - \lambda_2^* = 0 \Rightarrow$$
$$\frac{3}{2}\lambda_1^* = \frac{54}{3} = 18 \Rightarrow$$

$$\boxed{\lambda_1^* = 12, \lambda_2^* = 11\tfrac{1}{3}, \lambda_3^* = 0}$$

1.4 Dual Problem

Since we have KKT conditions with a convex problem (convex objective function, convex constraints (even linear in our case)), then strong duality holds. Meaning:

$$f^*(x^*, \lambda^*) = min_x\{max_{\lambda \geq 0} L(x, \lambda)\} = max_{\lambda \geq 0}\{min_x L(x, \lambda)\}$$

Let's define the dual function $\eta(\lambda) = min_x L(x, \lambda)$:

First we find x^* by taking the derivative of the Lagrangian and comparing it to 0. This is enough for finding the optimal point as the Lagrangian is a convex function.

The Lagrangian:

$$L(x, \lambda) = x^T Ax + b^T x + c + \lambda^T (Bx + d)$$

When:

- $A \in \mathbb{R}^{2\times 2}, A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, x, b \in \mathbb{R}^2, b = \begin{bmatrix} -20 \\ -2 \end{bmatrix}, c \in \mathbb{R}, c = 51$
- $B \in \mathbb{R}^{3\times 2}, B = \begin{bmatrix} 0.5 & 1 \\ 1 & -1 \\ -1 & -1 \end{bmatrix}, d \in \mathbb{R}^3, d = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}$

Therefore:

$$\nabla_x L(x, \lambda) = 2Ax + b + B^T \lambda = 0 \Rightarrow$$
$$x^* = -\frac{1}{2} A^{-1}(B^T \lambda + b)$$

We next put it in the Lagrangian to get the dual function:
Note: A is diagonal in our case so it's also symmetric (and its inverse as well).

$$\eta(\lambda) = L(x^*, \lambda) = \frac{1}{4}(B^T\lambda+b)^T A^{-1}(B^T\lambda+b) - \frac{1}{2}b^T A^{-1}(B^T\lambda+b) + \lambda^T(-\frac{1}{2}BA^{-1}(B^T\lambda+b)+d) =$$

$$= \frac{1}{4}\lambda^T BA^{-1}B^T\lambda + \frac{1}{2}b^T A^{-1}B^T\lambda + \frac{1}{4}b^T A^{-1}b - \frac{1}{2}b^T A^{-1}B^T\lambda - \frac{1}{2}b^T A^{-1}b - \frac{1}{2}\lambda^T BA^{-1}B^T\lambda$$

$$-\frac{1}{2}\lambda^T BA^{-1}b + \lambda^T d =$$

$$= -\frac{1}{4}\lambda^T BA^{-1}B^T\lambda - \frac{1}{2}\lambda^T BA^{-1}b - \frac{1}{4}b^T A^{-1}b + \lambda^T d$$

As expected, this is a concave function in λ
So, the dual problem is:

$$\boxed{\max_{\lambda\geq 0}\eta(\lambda) = \max_{\lambda\geq 0} -\tfrac{1}{4}\lambda^T BA^{-1}B^T\lambda - \tfrac{1}{2}\lambda^T BA^{-1}b - \tfrac{1}{4}b^T A^{-1}b + \lambda^T d}$$

When:

- $A \in \mathbb{R}^{2\times 2}, A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, x, b \in \mathbb{R}^2, b = \begin{bmatrix} -20 \\ -2 \end{bmatrix}, c \in \mathbb{R}, c = 101$
- $B \in \mathbb{R}^{3\times 2}, B = \begin{bmatrix} 0.5 & 1 \\ 1 & -1 \\ -1 & -1 \end{bmatrix}, d \in \mathbb{R}^3, d = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}$

1.5 Dual Optimum

Let's check that optimal x^* is achieved when setting the optimal multipliers:

$$x^* = -\frac{1}{2}A^{-1}(B^T\lambda + b)$$

When:
$\lambda^* = \begin{bmatrix} 12 \\ 11\frac{1}{3} \\ 0 \end{bmatrix}$ $A^{-1} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix}, B^T = \begin{bmatrix} 0.5 & 1 & -1 \\ 1 & -1 & -1 \end{bmatrix}, b = \begin{bmatrix} -20 \\ -2 \end{bmatrix}$

Therefore:

$$x^* = -\frac{1}{2}A^{-1}(B^T\lambda + b) = -\frac{1}{2}\begin{bmatrix}\frac{1}{2} & 0\\ 0 & 1\end{bmatrix}(\begin{bmatrix}0.5 & 1 & -1\\ 1 & -1 & -1\end{bmatrix}\begin{bmatrix}12\\ \frac{34}{3}\\ 0\end{bmatrix} + \begin{bmatrix}-20\\ -2\end{bmatrix}) =$$

$$= -\frac{1}{2}\begin{bmatrix}\frac{1}{2} & 0\\ 0 & 1\end{bmatrix}(\begin{bmatrix}\frac{52}{3}\\ \frac{2}{3}\end{bmatrix} + \begin{bmatrix}-20\\ -2\end{bmatrix}) =$$

$$= -\frac{1}{2}\begin{bmatrix}\frac{1}{2} & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}-\frac{8}{3}\\ -\frac{4}{3}\end{bmatrix} =$$

$$= \begin{bmatrix}\frac{2}{3} & \frac{2}{3}\end{bmatrix}^T \Rightarrow$$

$$\boxed{x^* = \begin{bmatrix}\frac{2}{3} & \frac{2}{3}\end{bmatrix}^T}$$

Exactly like we got at the primal problem solution. So we conclude that the dual optimum is achieved at that point as expected due to strong duality in this case.

1.6 Augmented Lagrangian Solver

In code.
We've defined 3 main classes:
1. OptimizationProblem - objective functions pointers' to value,gradient and hessian + the same for constraints functions (w/ or w/o constraints)
2. Penalty - manages the penalty function value and derivatives calculations + considers penalty parameter and multiplier
3. AugmentedLagarangianSolver - the main class

Note: we've limited the multipliers increase/decrease relative to previous multiplier by factor of 3 (as suggested in lectures).

1.7 Testing solver

Some snippets from code outputs:
Newton Method has converged after 6 iterations

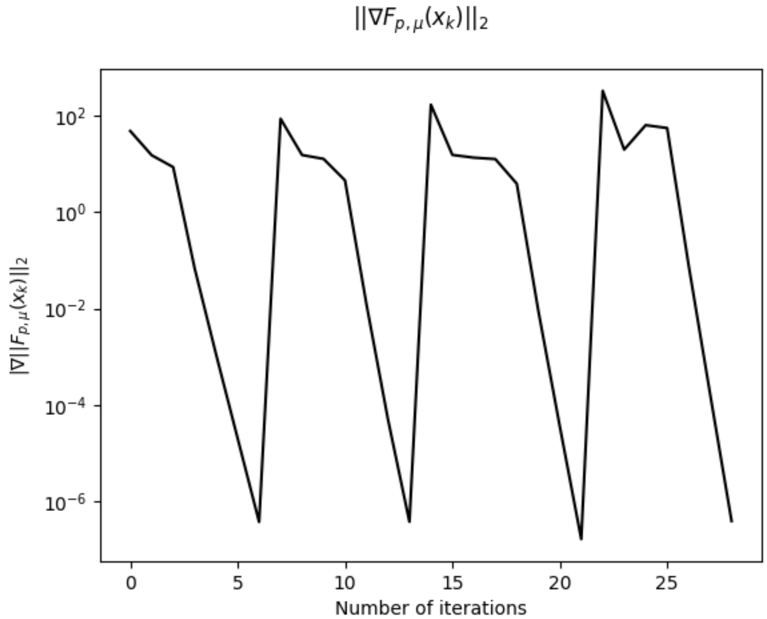
Newton Method has converged after 6 iterations
Newton Method has converged after 7 iterations
Newton Method has converged after 6 iterations

We can see the x values have converged to the optimal x under constraints as calculates:
x_traj=array([[0.], [-0.2]]),
array([[0.09900358], [0.02088282]]), array([[0.82696752], [0.75244632]]), array([[0.80335486], [0.70356693]]), array([[0.80387324], [0.70393907]]), array([[0.80386339], [0.70393477]]), array([[0.80386357], [0.70393483]]), array([[0.], [-0.2]]), array([[0.04311619], [0.01379573]]), array([[0.64303387], [0.61516818]]), array([[0.71960123], [0.69190039]]), array([[0.72342876], [0.6824236]]), array([[0.72340545], [0.68244142]]), array([[0.72340564], [0.68244143]]), array([[0.], [-0.2]]), array([[0.01385387], [0.01412047]]), array([[0.45665664], [0.45744669]]), array([[0.65601656], [0.65706519]]), array([[0.67112535], [0.67215757]]), array([[0.67549307], [0.66969653]]), array([[0.6754751], [0.66969512]]), array([[0.67547518], [0.66969514]]), array([[0.], [-0.2]]), array([[0.00544623], [0.00846268]]), array([[0.61151288], [0.5594119]]), array([[0.70594954], [0.65477842]]), array([[0.66662328], [0.66666689]]), array([[0.66670718], [0.66668043]]),
array([[0.66670699],[0.66668037]])

We can also see the multipliers have converged close to thier optimal calculated values:
multipliers=
array([[1.],[1.],[1.]]),
array([[3.], [3.], [0.33333333]]),
array([[9.], [9.], [0.11111111]]),
array([[11.97309228], [11.31201791], [0.03703704]])

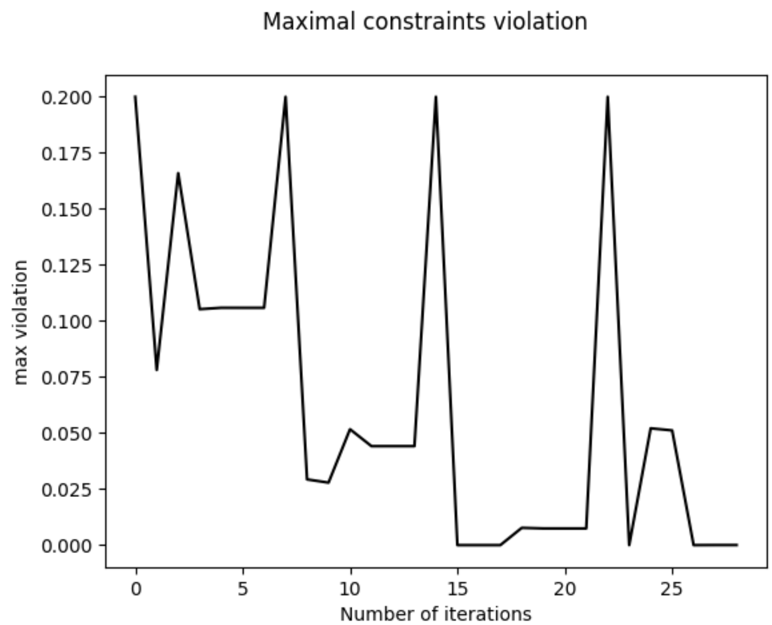
1.8 Plots

1.8.1



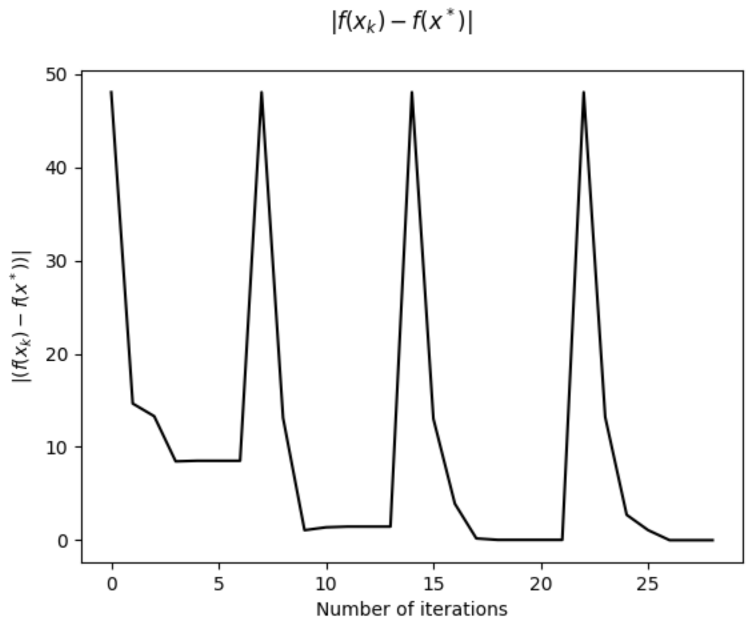
We can see in above plot that the gradients on each newton iteration converged in a polynomial convergence rate. In between, we ruin our current point a bit as we change the penalty parameter (increase by $\alpha = 2$) and multipliers (derivative at active points).

1.8.2



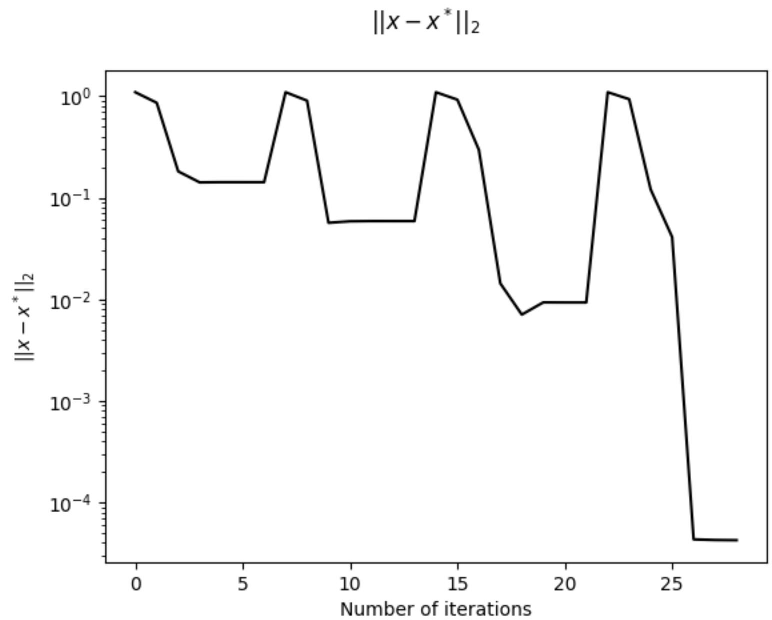
Above is the plot of the maximum constraint violation $\max_i(g_i(x)) \geq 0$ at each x in the trajectory we've followed.

1.8.3

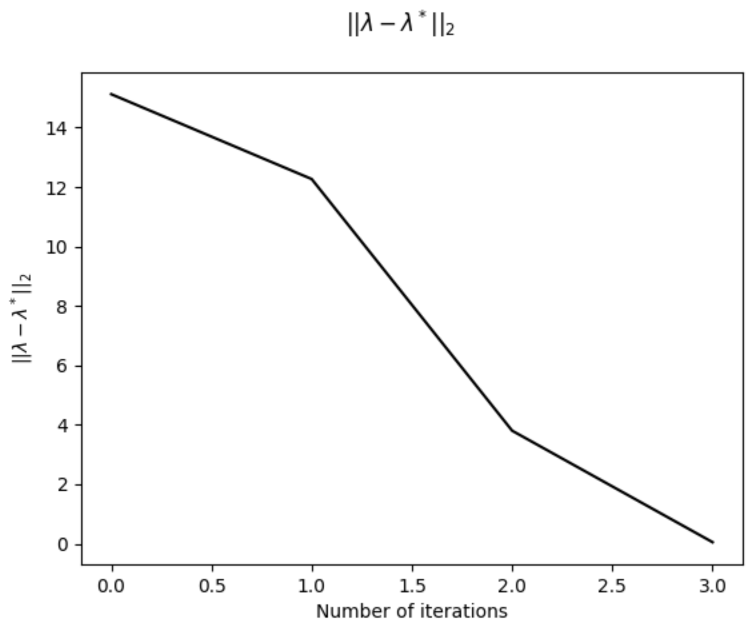


The absolute value of the diff between the x trajectory function's values and the optimal x function value (for the constrained problem). we can see we eventually converge to the optimal value (diff is zero)

1.8.4



We can see the distance between trajectory x values and optimal x converges to 0. The same bumps as we saw due to restating a new unconstrained optimization problem with different penalty and multipliers.



On multipliers, we can see we converge to the optimal monotonically as we update them between new unconstrained optimization problems (defined on new penalty function aggregate)