CMO Assignment 1

Biswadeep Debnath BISWADEEPD@IISC.AC.IN
24528

Question 1

REPORT

1. Solution

Table 1: Optimization results for each Q

Question	x^*	$f(x^*)$
Q_a	$\begin{bmatrix}-2.08279486 & -2.07200805\end{bmatrix}$	-2.0774014569849335
Q_b	$\begin{bmatrix}0.00685027 & -0.65260417\end{bmatrix}$	-0.3228769505802093
Q_c	$\begin{bmatrix}-0.48679597 & -0.55464203\end{bmatrix}$	-0.5207190004180156
Q_d	$\begin{bmatrix}1.01844342 & -2.16769472\end{bmatrix}$	-0.5746256495278905
Q_e	$\begin{bmatrix}-2.54759276 & -0.54741552\end{bmatrix}$	-1.5475041386672246

2. Solution

The analytical solutions x^* for the 5 matrices are:

$$\begin{aligned}
\text{Q_a}: \quad & x^* = \begin{bmatrix}-2.08279486 & -2.07200805\end{bmatrix} \\
\text{Q_b}: \quad & x^* = \begin{bmatrix}0.00685027 & -0.65260417\end{bmatrix} \\
\text{Q_c}: \quad & x^* = \begin{bmatrix}-0.48679597 & -0.55464203\end{bmatrix} \\
\text{Q_d}: \quad & x^* = \begin{bmatrix}1.01844342 & -2.16769472\end{bmatrix} \\
\text{Q_e}: \quad & x^* = \begin{bmatrix}-2.54759276 & -0.54741552\end{bmatrix}
\end{aligned}$$

It is observed that the analytical solutions are equal to the solutions we got by solving through exact line search which indicates the convergence. The plots of all five cases are shown in Figure 1.

3. Solution

From the plots of $\|x^{(k)} - x^*\|$ over iterations for the five cases of Q matrices, it is clear that the solution converges for all cases of Q. A tolerance of 10^{-10} was used so that if the value of $\nabla(f(x^{(k)}))$ is less than the tolerance, the algorithm terminates and is used as convergence criteria.

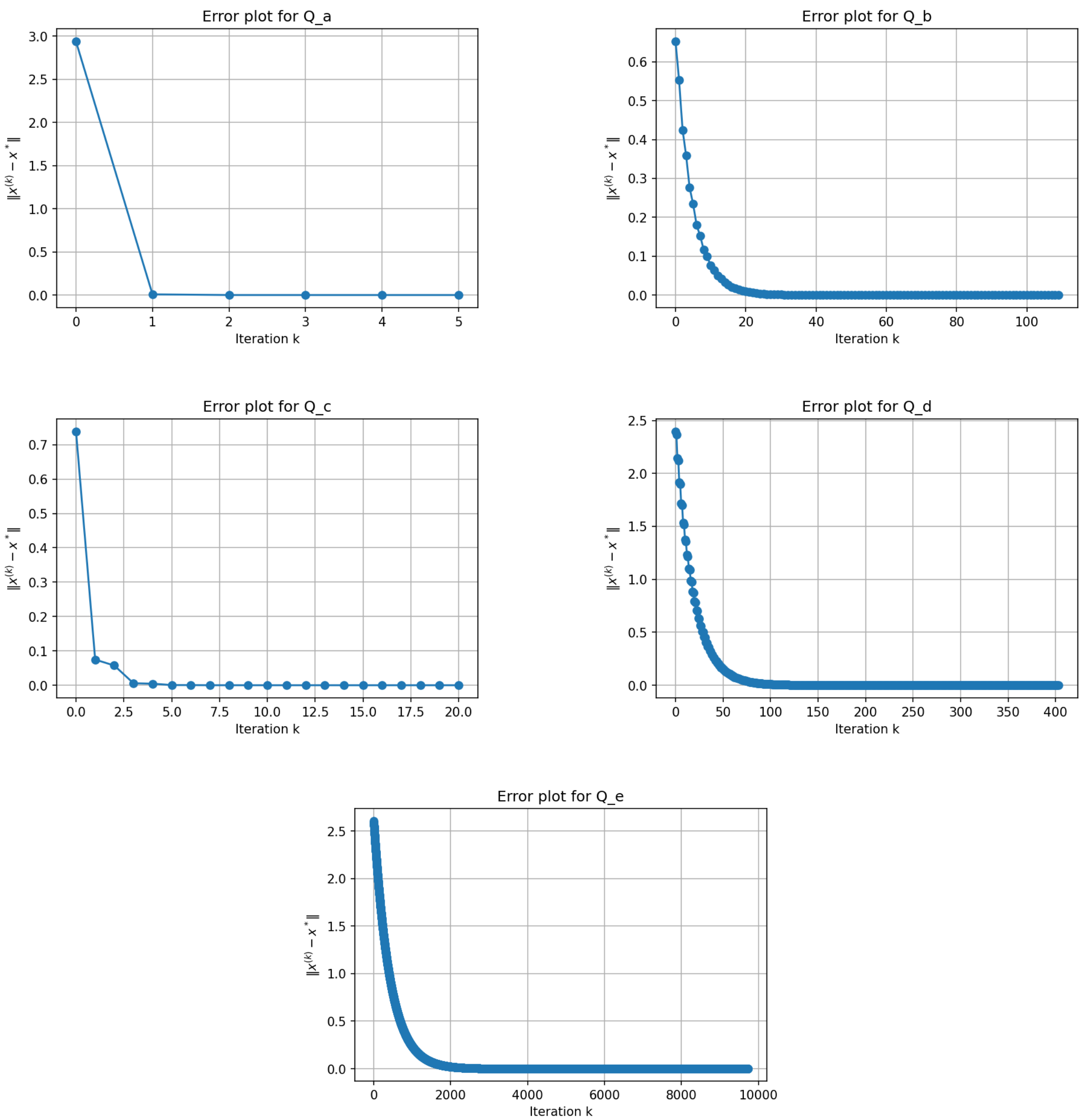

Figure 1: Plots of $\|x^{(k)} - x^*\|$ over iterations for five Q matrices.

The number of iterations taken to converge in all five cases in increasing order is as follows: $Q_a < Q_c < Q_b < Q_d < Q_e$, which is evident from the plots.
The analysis of the condition number of the matrices is done since the optimum gradient method converges slowly when the matrix is ill-conditioned (Akaike (1959)) . It is evident in the observation that in the case of a high condition number the convergence is slower, in this case Q_e is slowest and this is more visible when the difference in condition numbers is larger since Q_c has higher condition number than Q_b but still converges faster than Q_b.

Matrix	Condition Number
Q_a	1.4999
Q_b	10.0
Q_c	19.9999
Q_d	99.9999
Q_e	999.9999

Table 2: Matrix Condition Numbers for Q_a to Q_e

Question 2

REPORT

1. Solution

To terminate the gradient descent algorithm, we can use a very small tolerance value δ, for example 10^{-10} is used in this assignment. When the absolute value of gradient $||\nabla f(x)||$ becomes smaller than δ, we stop the algorithm. This is because near the minimum the gradient gets very close to zero and the closeness or the tolerance we can define it before starting the algorithm.

2. Solution

Table 3 contains the x^* and $f(x^*)$ values for different methods

Table 3: x^* and $f(x^*)$ for different methods

Method	$\mathbf{x^*}$	$\mathbf{f(x^*)}$
Armijo Condition	$\begin{bmatrix}-0.04334144\\0.01173756\\0.03569455\\-0.15490039\\-0.04891529\end{bmatrix}$	-0.07304051470893336
Armijo-Goldstein Condition	$\begin{bmatrix}-0.04334144\\0.01173756\\0.03569455\\-0.15490039\\-0.04891529\end{bmatrix}$	-0.07304051470893341
Wolfe Condition	$\begin{bmatrix}-0.04334144\\0.01173756\\0.03569455\\-0.15490039\\-0.04891529\end{bmatrix}$	-0.07304051470893341
Backtracking Condition	$\begin{bmatrix}-0.04334144\\0.01173756\\0.03569455\\-0.15490039\\-0.04891529\end{bmatrix}$	-0.07304051470893343

3. Solution

Figure 2 shows the step sizes taken by each method at each iteration in a single plot.

4. Solution

Table 4 shows the oracle calls taken by the methods.

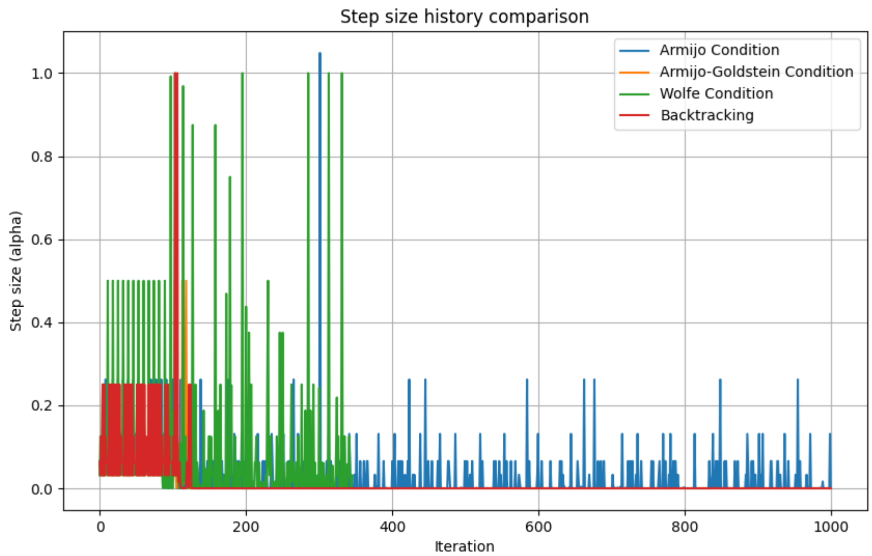

Figure 2: Step size history plot

Table 4: Number of oracle calls for each optimization method

Method	**Number of oq2f calls**	**Number of oq2g calls**
Armijo Condition	8403	1001
Armijo-Goldstein Condition	68554	1001
Wolfe Condition	9618	716
Backtracking Condition	28819	1001

5. Solution

All four methods seemed to converge and they all have same x^* values. Wolfe's condition got terminated in less than 1000 iterations (the maximum iterations were set to 1000) and so it look less $oq2g$ oracle calls. Armijo-Goldstein took the most number of $oq2f$ oracle calls and backtrackking also took comparatively more oracle calls.

Question 3

REPORT

1. SOLUTION

We can write the problem of solving $Ax = b$ as a convex optimization problem.

$$\min_x \; f(x) = \tfrac{1}{2}\|Ax - b\|^2$$

Expanding,

$$f(x) = \tfrac{1}{2}(Ax - b)^T(Ax - b) = \tfrac{1}{2}\left(x^T A^T Ax - 2b^T Ax + b^T b\right)$$

For Gradient,

$$\nabla f(x) = A^T Ax - A^T b$$

And Hessian,

$$\nabla^2 f(x) = A^T A$$

Now, for any $y \in \mathbb{R}^n$,

$$y^T(A^T A)y = (Ay)^T(Ay) = \|Ay\|^2 \geq 0$$

Hence, the Hessian is positive semidefinite. Therefore, this is a valid convex optimization problem.

3. SOLUTION

(a) $m < n$: The system is underdetermined and $\text{rank}(A) \leq m < n$. If $\text{rank}(A) = m$, the system is consistent when $\mathbf{b}$ lies in the column space of A. In this case, there exist infinitely many solutions.

(b) $m = n$: If $\text{rank}(A) = m$ and A is invertible, then the system has a unique solution given by

$$\mathbf{x}^* = A^{-1}\mathbf{b}.$$

(c) $m > n$: The system is overdetermined and if $\text{rank}(A) = n$, then a unique least squares solution exists and is given by

$$\mathbf{x}^* = (A^T A)^{-1}A^T\mathbf{b},$$

since $\text{rank}(A^T A) = n$.

4. SOLUTION

(a) For $x = A^{-1}b$, matrix inversion have complexity of $O(n^3)$ using methods like Gaussian elimination (Golub and Van Loan (2013)) .

Table 5: Timing and residuals for different values of m

m	Direct Time (s)	Optimized Time (s)	Residuals (Inversion)	Residuals (Optimization)
2	0.00001842	0.00474738	0.00000000	0.00000000
4	0.00001971	0.03694146	0.00000000	0.00000000
8	0.00006275	0.04136983	0.00000000	0.05414338
16	0.00003829	0.03928112	0.00000000	0.11920511
32	0.00003992	0.03683267	0.00000000	1.48519236
64	0.00007008	0.03880537	0.00000000	0.83926671
128	0.00009871	0.06576971	0.00000000	1.80731671
256	0.00029588	0.10904283	0.00000000	1.30562962
512	0.00158846	0.22842421	0.00000000	3.34189888
1024	0.01057808	0.63716129	0.00000000	3.23242270
2048	0.07721712	6.57854812	0.00000000	3.29743787
4096	0.48554225	24.49881342	0.00000000	6.47320368

5. Solution

Table 5 shows the time taken using matrix inversion and optimization for different $m \times m$ random matrices. Residuals $||Ax - b||$ are also shown to access the solution obtained.

References

Hirotugu Akaike. On a successive transformation of probability distribution and its application to the analysis of the optimum gradient method. *Annals of the Institute of Statistical Mathematics*, 11(1):1–16, 1959. doi: https://doi.org/10.1007/BF01831719.

Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, 4th edition, 2013.