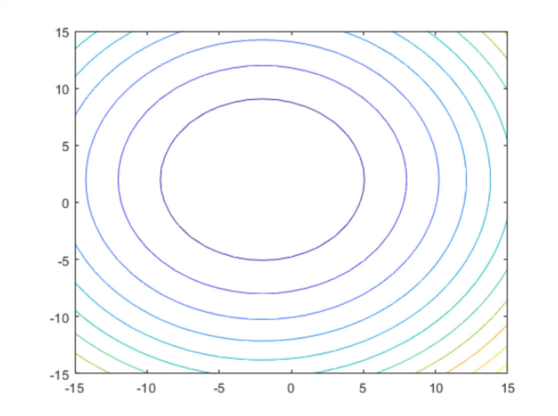
Q.1 $J(w) = (w-w_0)^T \cdot A \cdot (w - w_0)$; $w \in R^2$

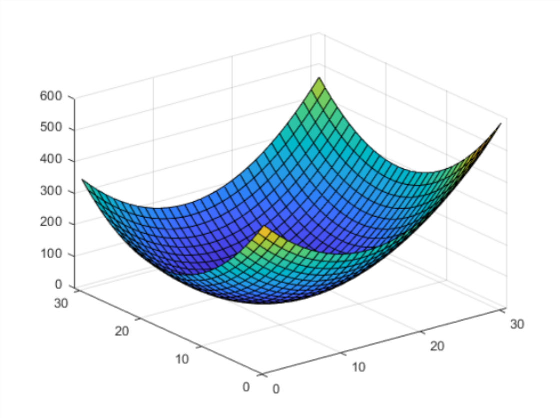
$w_0 = (-2,2)^T$

$A = I$.

(a) To the right is the contour plot showing iso-contours of $J(w)$.

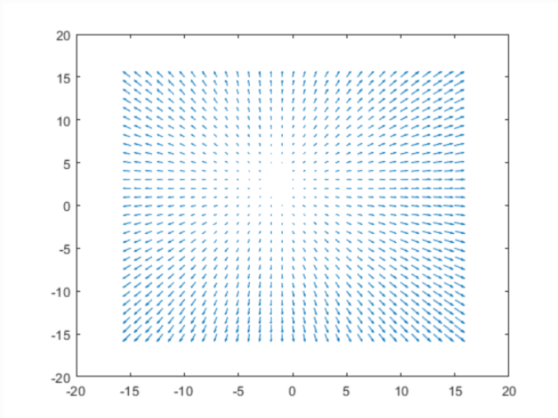


& below is the surface plot of $J(w)$.



(b) $\nabla J(w) = 2 \times A \times (w - w_0)$

Alongside is the quiver plot of $\nabla J(w)$

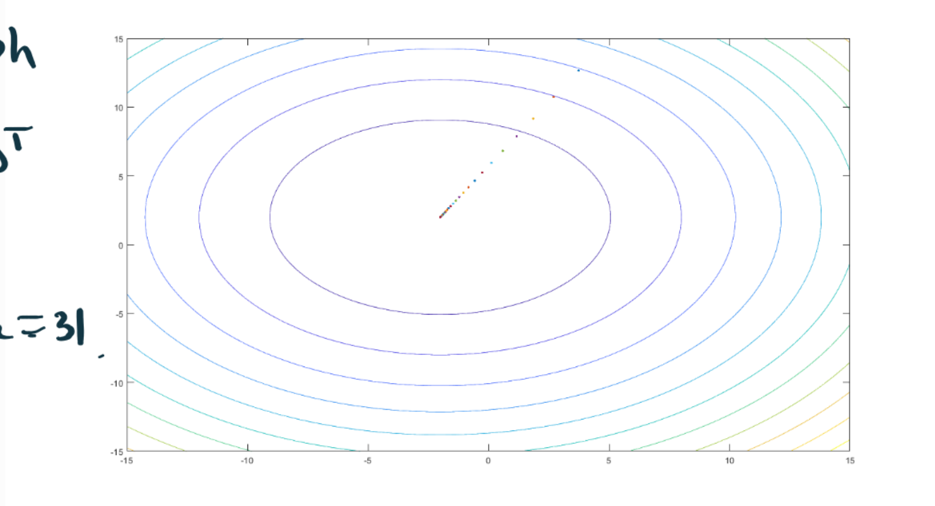


(c) $\boxed{w^{(n+1)} = w^{(n)} - \mu \cdot (\nabla(J(w^{(n)})))}$

Given, $w^{(0)} = (5,15)^T$, $\mu = 0.09$

Condition of convergence $= \|w^{(n)} - w_0\|^2 < 0.001$

Alongside is the graph showing all points converging to $w_0 = (-2,2)^T$



Number of iterations required for convergance = 31.

(d) For $\mu = 5$, the algorithm doesn't converge. It keeps running forever since $5 > \frac{2}{\lambda_{max}} = 2$.

For line search, $\mu = 0.5$ & it converges in 1 steps.

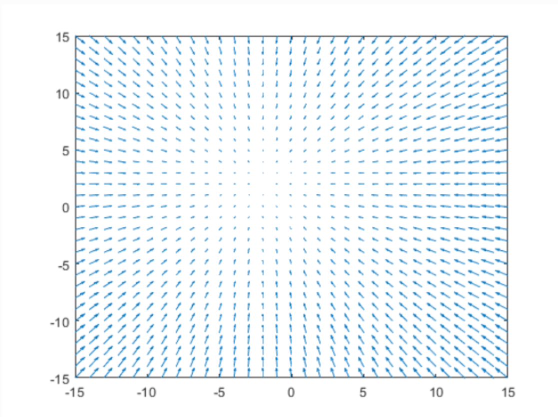
(e) For Newton's method, $w^{(n+1)} = w^{(n)} + F(w^{(n)})$

here, $F(w^{(n)}) = -\left[\nabla^2(J(w))\right]^{-1} \cdot \nabla_w J(w)$

$= w_0 - w$

The quiverplot is as shown in the image here.

The difference between this plot & the one in (b) is that the vectors in this case are pointing towards w_0 & in the previous one were pointing away from w_0.



(f) Newton's method converges in 1 step alone,

$$w^{(n+1)} = w^{(n)} + F(w^{(n)})$$

in this case, $w' = w^0 + (w_0 - w^0)$

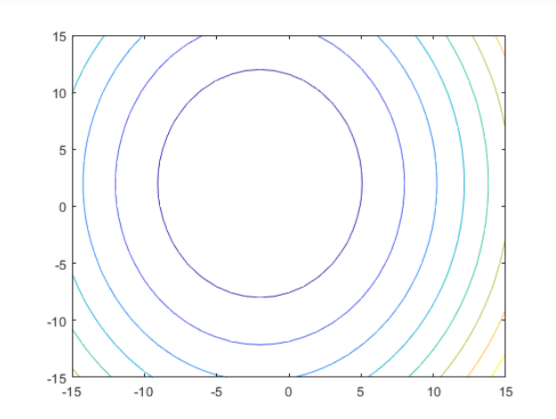
$$\boxed{w' = w_0}$$

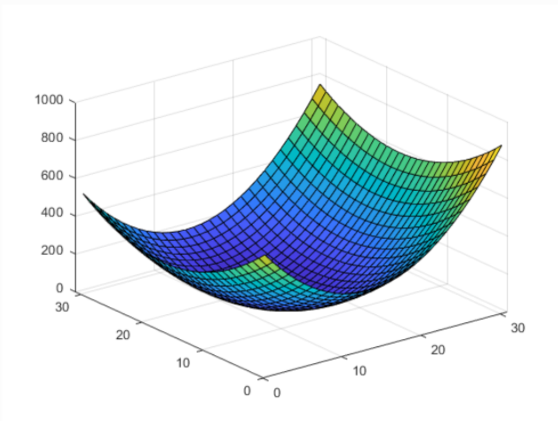
No matter the starting w^0, we always converge in 1 step when using Newton's method.

(g) The convergence of Newton's method is fastest i.e. in 1 step, whereas others take longer. Gradient descent took 49 iterations to converge & when using $\mu = 5$, the algorithm did not converge at all.

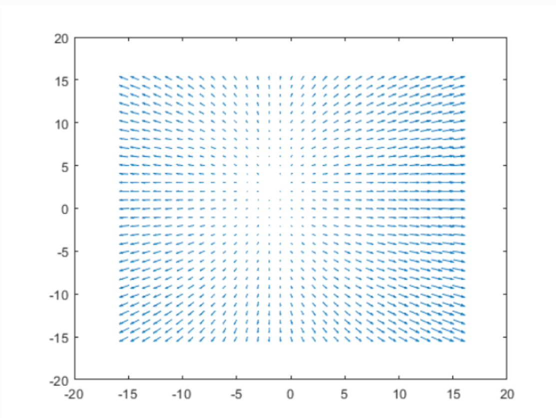
Q.2 When $A = \text{diag}(2,1)$, we get the following:

On convergence using different algorithms: the graph steps = 31 for 0.09 & 58 for 0.05

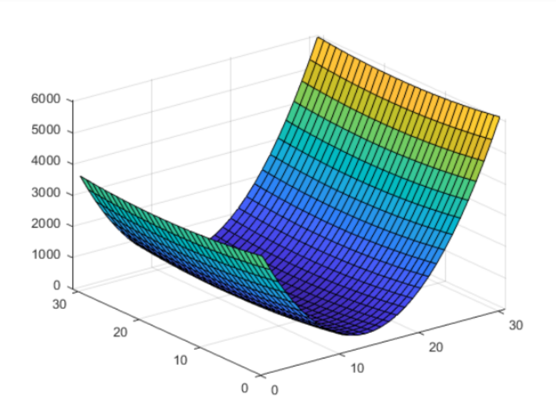


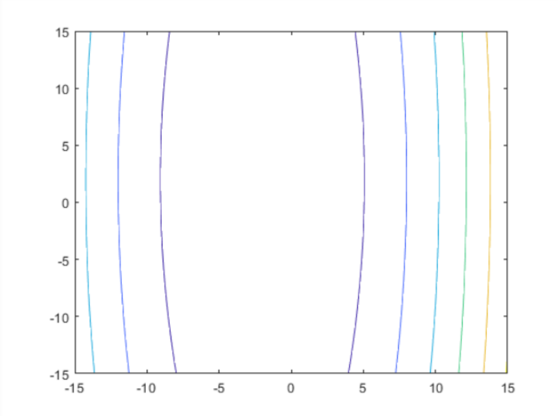


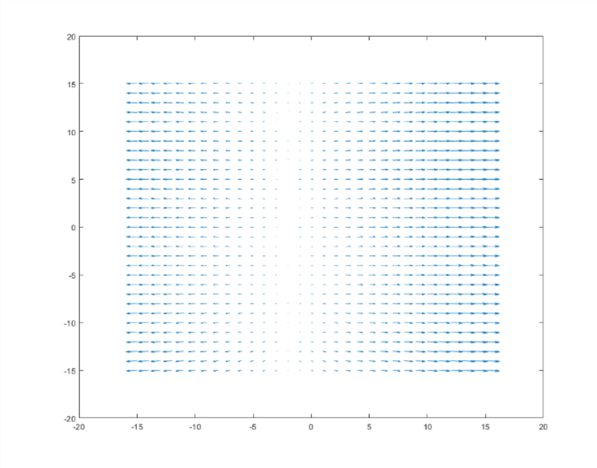
For this value of A, we see that the surface is shallower compared to the identity matrix & the convergence is slower.



Q.3 In this value of A, we see that the surface is very shallow. When we start with the value of $w^{(0)} = (5,15)$ the gradient vanishes, and is unable to converge at all.



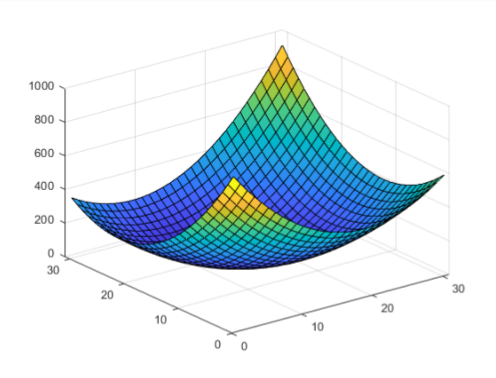


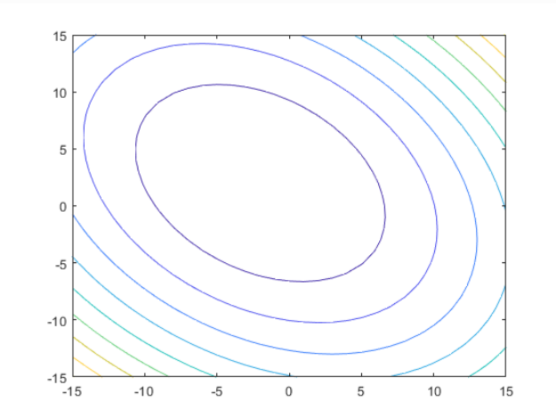


we can also see that the quiver plot & the contour plot have distinct lines that are almost parallel to each other.

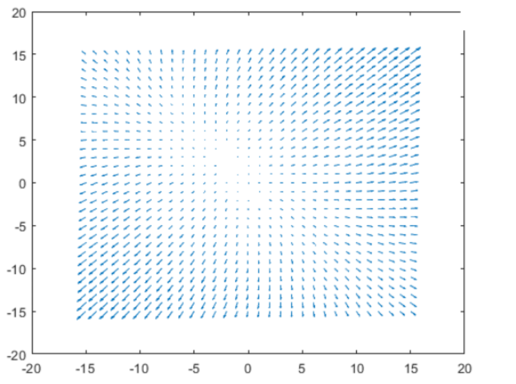
Q.4 For the given eigenvalues, we have eigenvectors $\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}$ & $\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}$ which gives us

$$A = \begin{bmatrix}1.5 & 0.5\\0.5 & 1.5\end{bmatrix}$$

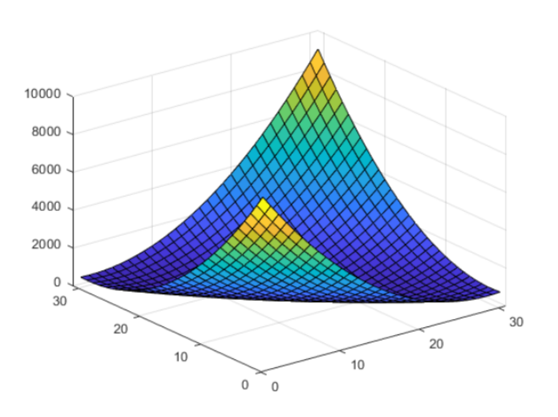




for $\mu = 0.09$, the gradient descent converges in 25 steps & for $\mu = 0.05$, it converges in 47 steps



Q.5 For the given eigenvalues & eigenvectors, we get $\longrightarrow$ $A = \begin{bmatrix} 10.5 & 9.5 \\ 9.5 & 10.5 \end{bmatrix}$



As can be seen in the graphs here, the gradient almost vanishes after a point & our algorithm is not able to converge for any values of μ that I tried.