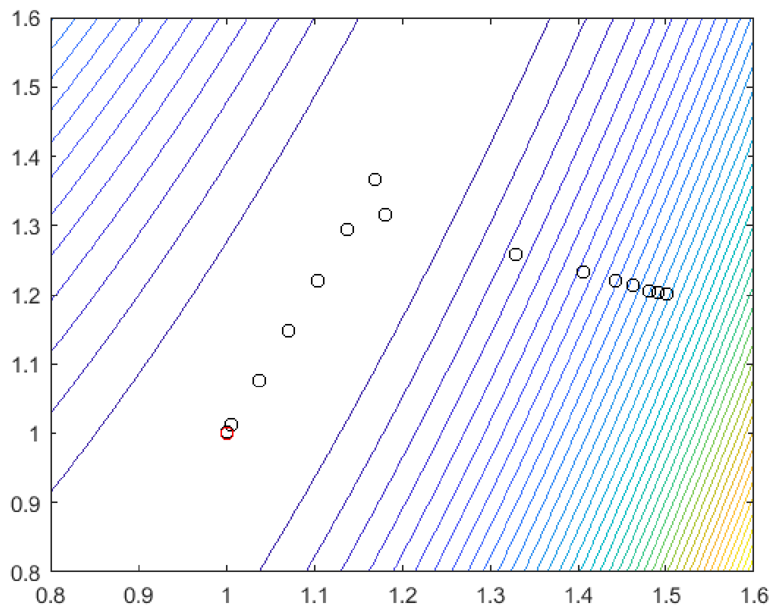
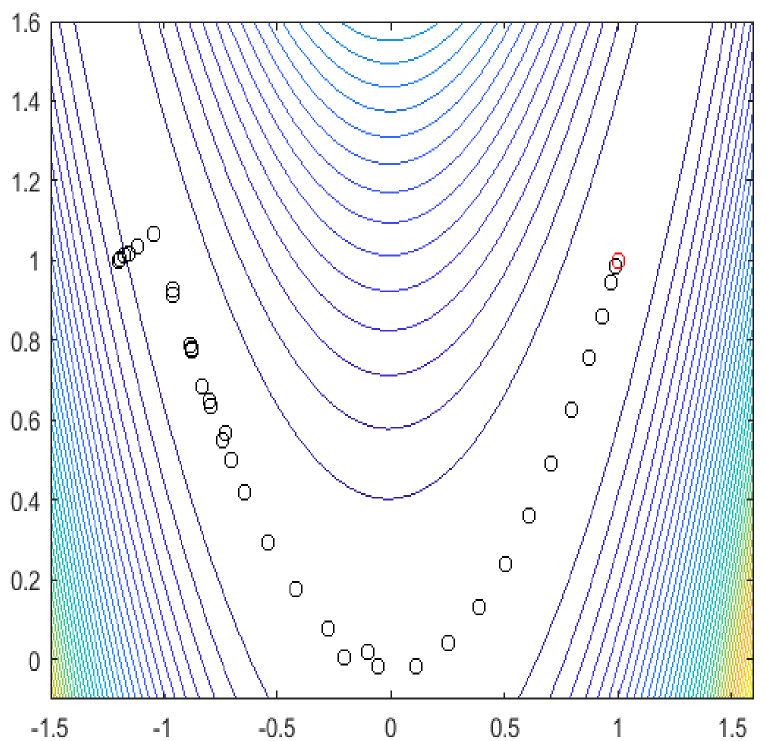
1 Question 1

Do question 4.2 on page 98 of the Nocedal and Wright textbook.
Write a program that implements the dogleg method. Choose B_k to be the exact Hessian. Apply it to solve Rosenbrock's function (2.22). Experiment with the update rule for the trust region by changing the constants in Algorithm 4.1, or by designing your own rules.

I chose $\eta = 1/8$, $\Delta_0 = 0.01$ and $\hat{\Delta} = 100$. I ran the program starting from point $[1.5, 1.2]$ and point $[-1.2, 1]$. The black circles are the x_k generated during the method and the red circle is the minium of the function.





2 Question 2

Do question 4.6 on page 99 of your textbook.
The Cauchy–Schwarz inequality states that for any vectors u and v, we have

$$\left|u^T v\right|^2 \leq (u^T u)(v^T v)$$

with equality only when u and v are parallel. When B is positive definite, use this inequality to show that

$$\gamma = \frac{\|g\|^4}{(g^T B g)(g^T B^{-1} g)}$$

with equality only if g and Bg (and $B^{-1}g$) are parallel.

Can assume that B is symmetric, aside
If B is not symmetric can write

$$B = S + N \qquad S = \frac{B + B^T}{2} \qquad N = \frac{B - B^T}{2} \qquad S^T = S, N^T = -N$$

$$x^T B x = x^T (S + N) x = x^T S x + x^T N x$$
$$(x^T N x)^T = x^T N x \text{ since it is a scalar}$$
$$(x^T N x)^T = x^T N^T (x^T)^T = -x^T N x \qquad = x^T N x$$

So must have that $x^T N x = 0$ So $x^T B x = x^T S x$ where S is symmetric S is also positive definite, $x^T S x = x^T B x > 0$ if $x \neq 0$

Now because B is SPD, can factor it uniquely using Cholesky factorization.
$B = LL^T$ where L is a lower triangular matrix. Then $B^{-1} = (LL^T)^{-1} = (L^T)^{-1} L^{-1} = F^T F$ where $F = L^{-1}$ is also a lower triangular matrix.

$$\begin{aligned}
(g^T B g)(g^T B^{-1} g) &= (g^T L L^T g)(g^T F^T F g) \\
&= [(L^T g)^T (L^T g)][(Fg)^T (Fg)] \\
\text{taking } u = L^T g \text{ and } v = Fg & \\
&= (u^T u)(v^T v) \\
&\geq (u^T v)^2 && \text{by Cauchy inequality} \\
&= [(L^T g)^T (Fg)]^2 \\
&= [g^T L L^{-1} g]^2 \\
&= (g^T g)^2 \\
&= \|g\|^4 && \text{using Euclidean norm}
\end{aligned}$$

So ($g^T B g$ and $g^T B^{-1} g$ are both positive scalars because B and so B^{-1} are SPD)

$$(g^T B g)(g^T B^{-1} g) \geq \|g\|^4 \implies 1 \geq \frac{\|g\|^4}{(g^T B g)(g^T B^{-1} g)} = \gamma$$

In the step using Cauchy inequality we have equality iff for some $\alpha \in \mathbf{R}$

$$u = \alpha v \iff L^T g = \alpha F g \iff L^T = \alpha L^{-1} g \iff L L^T g = \alpha g \iff Bg = \alpha g \iff B^{-1} B g = \alpha B^{-1} g \iff g = \alpha (B^{-1} g)$$

That is iff g, Bg and $B^{-1}g$ are all parallel

3 Question 3

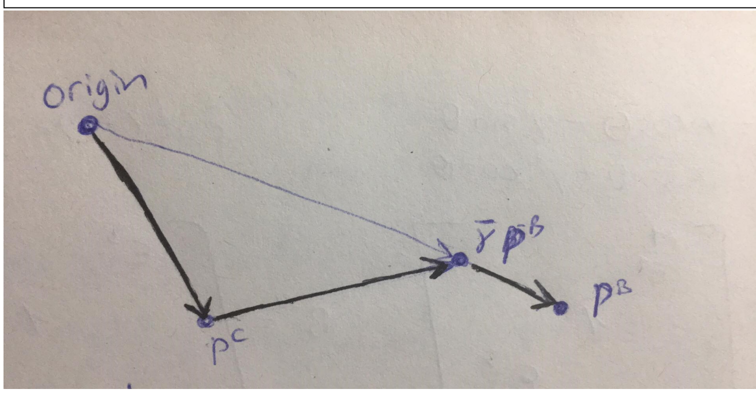
Do question 4.7 on page 99 of the Nocedal and Wright textbook
When B is positive definite, the double-dogleg method constructs a path with three line segments from the origin to the full step. The four points that define the path are

- the origin
- the unconstrained Cauchy step $p^C = -(g^T g)/(g^T B g) g$
- a fraction of the full step $\hat{\gamma} p^B = -\hat{\gamma} B^{-1} g$, for some $\hat{\gamma} \in (\gamma, 1]$, where γ is defined in the previous question; and
- the full step $p^B = B^1 g$.

Show that $\|p\|_2$ increases monotonically along this path.
In addition to showing that $\|p\|_2$ increases along the double-dogleg path, also show that

$$m(p) = f + p^T g + \frac{1}{2} p^T B p$$

decreases along the double-dogleg path.



Can write this trajectory by $p(\tau)$ like in the textbook
The last step is $\hat{\gamma} p^B + \alpha(p^B - \hat{\gamma} p^B)$ where $\alpha : 0 \to 1$

$$p(\tau) = \begin{cases} \tau p^C & 0 \leq \tau \leq 1 \\ p^C + (\tau - 1)(\hat{\gamma} p^B - p^C) & 1 \leq \tau \leq 2 \\ [\hat{\gamma} + (\tau - 2)(1 - \hat{\gamma})] p^B & 2 \leq \tau \leq 3 \end{cases}$$

For $0 \leq \tau \leq 1$
Define $h(\alpha) = \|p(\alpha)\| = \left\|\alpha p^C\right\| = \alpha \left\|p^C\right\|$, don't need absolute value since α is positive
Clearly an increasing function of α for $\alpha : 0 \to 1$

For $1 \leq \tau \leq 2$
Define $h(\alpha) = \frac{1}{2}\|p(1+\alpha)\|^2$ and show $h'(\alpha) \geq 0$ for $\alpha \in (0, 1)$

$$\begin{aligned} h(\alpha) &= \frac{1}{2}\left\|p^C + \alpha(\hat{\gamma} p^B - p^C)\right\|^2 \\ &= \frac{1}{2}\left[p^C + \alpha(\hat{\gamma} p^B - p^C)\right]^T \left[p^C + \alpha(\hat{\gamma} p^B - p^C)\right] \\ &= \frac{1}{2}(p^C)^T p^C + \alpha (p^C)^T(\hat{\gamma} p^B - p^C) + \frac{1}{2}\alpha^2 (\hat{\gamma} p^B - p^C)^T(\hat{\gamma} p^B - p^C) \\ &= \frac{1}{2}(p^C)^T p^C + \alpha (p^C)^T(\hat{\gamma} p^B - p^C) + \frac{1}{2}\alpha^2 \left\|\hat{\gamma} p^B - p^C\right\| \end{aligned}$$

$$
\begin{aligned}
h'(\alpha) &= (p^C)^T(\hat{\gamma}p^B - p^C) + \alpha\left\|\hat{\gamma}p^B - p^C\right\| \\
&\geq -(p^C)^T(p^C - \hat{\gamma}p^B) && \text{since norm and } \alpha \text{ are } \geq 0 \\
&= \frac{g^Tg}{g^TBg}g^T\left(-\frac{g^Tg}{g^TBg}g + \hat{\gamma}B^{-1}g\right) && \text{by def of } p^C \text{ and } p^B \\
&= \frac{g^Tg}{g^TBg}\left(-\frac{g^Tg}{g^TBg}g^Tg + \hat{\gamma}g^TB^{-1}g\right) \\
&= \frac{g^Tg}{g^TBg}g^TB^{-1}g\left(\hat{\gamma} - \frac{(g^Tg)^2}{(g^TB^{-1}g)(g^TBg)}\right) \\
&= \frac{g^Tg}{g^TBg}g^TB^{-1}g\,(\hat{\gamma} - \gamma) && \text{by def of } \gamma \\
&> 0 && \text{since } \hat{\gamma} > \gamma > 0,\ g^TBg > 0,\ g^TB^{-1}g > 0 \text{ and } g^Tg > 0 \text{ because } g \neq 0
\end{aligned}
$$

For $2 \leq \tau \leq 3$
Define $h(\alpha) = \|p(2+\alpha)\| = \left\|[\hat{\gamma} + \alpha(1-\hat{\gamma})]p^B\right\| = [\hat{\gamma} + \alpha(1-\hat{\gamma})]\left\|p^B\right\|$
$h'(\alpha) = (1-\hat{\gamma})\left\|p^B\right\| \geq 0$ since $\hat{\gamma} \in (\gamma, 1]$ and norm is always ≥ 0
So an increasing function of α for $\alpha : 0 \to 1$

Now to show $m(p)$ decreses along the double-dogled path.
For $0 \leq \tau \leq 1$
Define $r(\alpha) = m(p(\alpha))$ and show $r'(\alpha) \leq 0$ for $\alpha \in (0,1)$

$$
\begin{aligned}
r(\alpha) &= m(\alpha p^C) \\
&= f + \alpha g^Tp^C + \frac{1}{2}\alpha^2(p^C)^TBp^C \\
r'(\alpha) &= g^Tp^C + \alpha(p^C)^TBp^C \\
&= -\frac{g^Tg}{g^TBg}g^Tg + \alpha\left(\frac{g^Tg}{g^TBg}\right)^2 g^TBg && \text{by def of } p^C \\
&= -\frac{(g^Tg)^2}{g^TBg} + \alpha\frac{(g^Tg)^2}{g^TBg} \\
&= (\alpha - 1)\frac{(g^Tg)^2}{g^TBg} \\
&< 0 && \text{since } \alpha < 1,\ B \text{ is SPD and } g \neq 0
\end{aligned}
$$

So $m(p(\tau))$ decreases for $\tau : 0 \to 1$

For $1 \leq \tau \leq 2$
Define $r(\alpha) = m(p(1+\alpha))$ and show $r'(\alpha) \leq 0$ for $\alpha \in (0,1)$

$$
\begin{aligned}
r(\alpha) &= m(p^C + \alpha(\hat{\gamma}p^B - p^C)) \\
&= f + g^T[p^C + \alpha(\hat{\gamma}p^B - p^C)] + \frac{1}{2}[p^C + \alpha(\hat{\gamma}p^B - p^C)]^TB[p^C + \alpha(\hat{\gamma}p^B - p^C)] \\
&= f + g^Tp^C + \alpha g^T(\hat{\gamma}p^B - p^C) + \frac{1}{2}(p^C)^TBp^C + \alpha(p^C)^TB(\hat{\gamma}p^B - p^C) + \frac{1}{2}\alpha^2(\hat{\gamma}p^B - p^C)^TB(\hat{\gamma}p^B - p^C)
\end{aligned}
$$

$$
\begin{aligned}
r'(\alpha) &= g^T(\hat{\gamma}p^B - p^C) + (p^C)^T B(\hat{\gamma}p^B - p^C) + \alpha(\hat{\gamma}p^B - p^C)^T B(\hat{\gamma}p^B - p^C) \\
&= (\hat{\gamma}p^B - p^C)^T(g + Bp^C) + \alpha(\hat{\gamma}p^B - p^C)^T B(\hat{\gamma}p^B - p^C) \\
&< (\hat{\gamma}p^B - p^C)^T(g + Bp^C) + (\hat{\gamma}p^B - p^C)^T B(\hat{\gamma}p^B - p^C) && \text{since } 0 < \alpha < 1 \\
&= (\hat{\gamma}p^B - p^C)^T(g + Bp^C + B\hat{\gamma}p^B - Bp^C) \\
&= (\hat{\gamma}p^B - p^C)^T(g + \hat{\gamma}p^B) \\
&= (\hat{\gamma}p^B - p^C)^T(g - \hat{\gamma}BB^{-1}g) && \text{by def } p^B \\
&= (\hat{\gamma}p^B - p^C)^T(g - \hat{\gamma}g) \\
&= (1-\hat{\gamma})(\hat{\gamma}p^B - p^C)^T g \\
&= (1-\hat{\gamma})(\hat{\gamma}g^T p^B - g^T p^C) \\
&= (1-\hat{\gamma})\left(-\hat{\gamma}g^T B^{-1} g - \frac{g^T g}{g^T B g} g^T g\right) \\
&= (1-\hat{\gamma})(g^T B^{-1} g)\left(-\hat{\gamma} - \frac{(g^T g)^2}{(g^T B g)(g^T B^{-1} g)}\right) \\
&= -(1-\hat{\gamma})(g^T B^{-1} g)(\hat{\gamma} + \gamma) \\
&\le 0 && \text{because } B \text{ is SPD, } 0 < \gamma < \hat{\gamma} \le 1
\end{aligned}
$$

So $m(p(\tau))$ decreases for $\tau : 1 \to 2$

For $2 \le \tau \le 3$
Define $r(\alpha) = m(p(2+\alpha))$ and show $r'(\alpha) \le 0$ for $\alpha \in (0,1)$

$$
\begin{aligned}
r(\alpha) &= m([\hat{\gamma} + \alpha(1-\hat{\gamma})]p^B) \\
&= f + [\hat{\gamma} + \alpha(1-\hat{\gamma})]g^T p^B + \frac{1}{2}[\hat{\gamma} + \alpha(1-\hat{\gamma})]^2 (p^B)^T B p^B \\
&= f + [\hat{\gamma} + \alpha(1-\hat{\gamma})]g^T p^B + \frac{1}{2}[\hat{\gamma}^2 + 2\hat{\gamma}\alpha(1-\hat{\gamma}) + \alpha^2(1-\hat{\gamma})^2](p^B)^T B p^B \\
r'(\alpha) &= (1-\hat{\gamma})g^T p^B + [\hat{\gamma}(1-\hat{\gamma}) + \alpha(1-\hat{\gamma})^2](p^B)^T B p^B \\
&= (1-\hat{\gamma})[g^T p^B + \hat{\gamma}(p^B)^T B p^B + \alpha(1-\hat{\gamma})(p^B)^T B p^B] \\
&< (1-\hat{\gamma})[g^T p^B + \hat{\gamma}(p^B)^T B p^B + (1-\hat{\gamma})(p^B)^T B p^B] \\
&= (1-\hat{\gamma})(g^T p^B + (p^B)^T B p^B) \\
&= (1-\hat{\gamma})(p^B)^T(g + Bp^B) \\
&= (1-\hat{\gamma})(p^B)^T(g - BB^{-1}g) \\
&= (1-\hat{\gamma})(p^B)^T(g - g) \\
&= 0
\end{aligned}
$$

So $m(p(\tau))$ decreases for $\tau : 2 \to 3$

4 Question 4

Do question 4.8 on page 99 of the Nocedal and Wright textbook.
Show that (4.43) and (4.44) are equivalent

$$\lambda^{(l+1)} = \lambda^{(l)} + \frac{\phi_2(\lambda^{(l)})}{\phi_2'(\lambda^{(l)})} \tag{4.43}$$

$$\lambda^{(l+1)} = \lambda^{(l)} + \left(\frac{\|p_l\|}{\|q_l\|}\right)^2 \left(\frac{\|p_l\| - \Delta}{\Delta}\right) \tag{4.44}$$

where

$$\phi_2(\lambda^{(l)}) = \frac{1}{\Delta} - \frac{1}{\|p_l\|}$$

$$p_l = -(B + \lambda^{(l)} I)^{-1} g = -(R^T R)^{-1} g$$

$$q_l = R^{-T} p_l$$

Given hint

$$\|q_l\|^2 = \left\|R^{-T} p_l\right\|^2 = (R^{-T}p_l)^T(R^{-T}p_l) = p_l^T R^{-1} R^{-T} p_l = p_l^T (R^T R)^{-1} p_l = p_l^T (B + \lambda^{(l)} I)^{-1} p_l$$

$$p = -(B+\lambda I)^{-1} g = -\sum_{i=1}^{n} \frac{q_i^T g}{\lambda_i + \lambda} q_i \text{ so } (B+\lambda I)^{-1} p = \sum_{i=1}^{n} \frac{q_i^T p}{\lambda_i + \lambda} q_i = \sum_{i=1}^{n} \frac{q_i^T \left(-\sum_{j=1}^{n} \frac{q_j^T g}{\lambda_j + \lambda} q_j\right)}{\lambda_i + \lambda} q_i = -\sum_{i=1}^{n} \frac{q_i^T g}{(\lambda_i + \lambda)^2} q_i$$

$$p^T (B+\lambda I)^{-1} p = \left(-\sum_{j=1}^{n} \frac{q_j^T g}{\lambda_j + \lambda} q_j\right)^T \left(-\sum_{i=1}^{n} \frac{q_i^T g}{(\lambda_i + \lambda)^2} q_i\right) = \sum_{j=1}^{n} \frac{(q_j^T g)^2}{(\lambda_j - \lambda)^3}$$

$$\frac{d}{d\lambda}\|p\|^2 = \frac{d}{d\lambda} p^T p = \frac{d}{d\lambda}\left(-\sum_{i=1}^{n} \frac{q_i^T g}{\lambda_i + \lambda} q_i^T\right)\left(-\sum_{j=1}^{n} \frac{q_j^T g}{\lambda_j + \lambda} q_j^T\right) = \frac{d}{d\lambda}\sum_{i=1}^{n} \frac{(q_i^T g)^2}{(\lambda_i + \lambda)^2} = -2\sum_{i=1}^{n} \frac{(q_i^T g)^2}{(\lambda_i + \lambda)^3}$$

$$\begin{aligned}
\phi_2'(\lambda^{(l)}) &= -\frac{d}{d\lambda^{(l)}} \frac{1}{\|p_l\|} \\
&= -\frac{d}{d\lambda^{(l)}} \left(\|p_l\|^2\right)^{-1/2} \\
&= -\frac{1}{2}\left(\|p_l\|^2\right)^{-3/2} \frac{d}{d\lambda^{(l)}} \|p_l\|^2 \\
&= -\frac{1}{2}\left(\|p_l\|^2\right)^{-3/2} \left(-2\sum_{j=1}^{n} \frac{(q_j^T g)^2}{(\lambda_j - \lambda^{(l)})^3}\right) && \text{by given hint} \\
&= -\frac{1}{2}\left(\|p_l\|^2\right)^{-3/2} \left(-2\|q_l\|^2\right) = \frac{\|q_l\|^2}{\|p_l\|^3} && \text{by given hint above}
\end{aligned}$$

$$\begin{aligned}
\frac{1}{\phi_2'(\lambda^{(l)})} &= \frac{\|p_l\|^3}{\|q_l\|^2} \\
\frac{\phi_2(\lambda^{(l)})}{\phi_2'(\lambda^{(l)})} &= \left(\frac{1}{\Delta} - \frac{1}{\|p_l\|}\right)\left(\frac{\|p_l\|^3}{\|q_l\|^2}\right) \\
&= \left(\frac{\|p_l\|}{\Delta} - 1\right)\left(\frac{\|p_l\|^2}{\|q_l\|^2}\right) \\
&= \left(\frac{\|p_l\| - \Delta}{\Delta}\right)\left(\frac{\|p_l\|^2}{\|q_l\|^2}\right)
\end{aligned}$$

5 Question 5

Do question 4.12 on page 100 of the Nocedal and Wright textbook.
The following example shows that the reduction in the model function m achieved by the two-dimensional minimization strategy can be much smaller than that achieved by the exact solution of (4.5). In (4.5), set

$$g = \left(-\frac{1}{\epsilon}, -1, -\epsilon^2\right)^T$$

where is a small positive number. Set

$$B = \text{diag}\left(\frac{1}{\epsilon^3}, 1, \epsilon^3\right), \qquad \Delta = 0.5$$

Show that the solution of (4.5) has components $\left(O(\epsilon), \frac{1}{2} + O(\epsilon), O(\epsilon)\right)^T$ and that the reduction in the model m is $\frac{3}{8} + O(\epsilon)$. For the two-dimensional minimization strategy, show that the solution is a multiple of $B^{-1}g$ and that the reduction in m is $O(\epsilon)$.

$$B = \begin{bmatrix} \frac{1}{\epsilon^3} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \epsilon^3 \end{bmatrix}, B^{-1} = \begin{bmatrix} \epsilon^3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{\epsilon^3} \end{bmatrix}, g = \begin{bmatrix} -\frac{1}{\epsilon} \\ -1 \\ -\epsilon^2 \end{bmatrix}, B^{-1}g = \begin{bmatrix} -\epsilon^2 \\ -1 \\ -\frac{1}{\epsilon} \end{bmatrix}, Bg = \begin{bmatrix} -\frac{1}{\epsilon^4} \\ -1 \\ -\epsilon^5 \end{bmatrix}$$

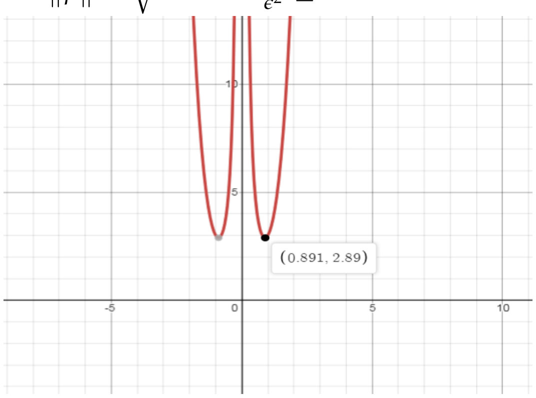
$$B + \lambda I = \begin{bmatrix} \frac{1}{\epsilon^3} + \lambda & 0 & 0 \\ 0 & 1 + \lambda & 0 \\ 0 & 0 & \epsilon^3 + \lambda \end{bmatrix}, (B + \lambda I)^{-1} = \begin{bmatrix} \frac{\epsilon^3}{1+\epsilon^3\lambda} & 0 & 0 \\ 0 & \frac{1}{1+\lambda} & 0 \\ 0 & 0 & \frac{1}{\epsilon^3+\lambda} \end{bmatrix}, -(B + \lambda I)^{-1}g = \begin{bmatrix} \frac{\epsilon^2}{1+\epsilon^3\lambda} \\ \frac{1}{1+\lambda} \\ \frac{\epsilon^2}{\epsilon^3+\lambda} \end{bmatrix}$$

$$g^T B^{-1} g = \epsilon + 1 + \epsilon = 1 + 2\epsilon \qquad g^T B g = \frac{1}{\epsilon^5} + 1 + \epsilon^7 \qquad g^T g = \frac{1}{\epsilon^2} + 1 + \epsilon^4$$

First will find the exact solution using Theorem 4.1 in the textbook.
B is a diagonal matrix, so the eigenvalues are the diagonal entries, since they are all positive, B is positive definite. We know the unique minimizer of

$$m(p) = f + g^T p + \frac{1}{2} p^T B p$$

is $p^* = -B^{-1}g = \left(\epsilon^2, 1\frac{1}{\epsilon}\right)^T$
But $\|p\| = \sqrt{\epsilon^4 + 1 + \frac{1}{\epsilon^2}} \geq 0.891 \qquad$ for all $\epsilon \in \mathbf{R}$



So the unique minimizer is definitely not in the trust region defined by $\Delta = 0.5$.
Using Theorem 4.1 in the textbook we know that for the global solution to

$$\min_{\|p\| \leq \Delta} m(p)$$

We must have $\lambda(\Delta - \|p^*\|) = 0$ and since $\lambda = 0$ doesn't work, λ must be positive and nonzero.
Because $B + \lambda I$ is positive semidefinite for any positive λ, we don't need to worry about that condition.
Need λ such that $\Delta = \|p^*\|$ (or equivalently $\Delta^2 = \|p^*\|^2$ since both are positive) where $p^* = -(B + \lambda I)^{-1}g$
This gives the equation

$$\|p^*\|^2 = \left\|-(B + \lambda I)^{-1}g\right\|^2 = \frac{\epsilon^4}{(1 + \epsilon^3\lambda)^2} + \frac{1}{(1 + \lambda)^2} + \frac{\epsilon^4}{(\epsilon^3 + \lambda)^2} = \Delta^2 = 0.5^2 = 0.25 \qquad \textbf{(i)}$$

Note that both $\frac{\epsilon^4}{(1+\epsilon^3\lambda)^2} = O(\epsilon)$ and $\frac{\epsilon^4}{(\epsilon^3+\lambda)^2} = O(\epsilon)$. The below arguments show why, presented in reverse order for clarity

$$\begin{aligned}
\frac{\epsilon^4}{(1+\epsilon^3\lambda)^2} = \frac{\epsilon^4}{1+2\epsilon^3\lambda+\epsilon^6\lambda^2} &< \epsilon \\
\epsilon^4 &< \epsilon + 2\epsilon^4\lambda + \epsilon^7\lambda^2 \\
\epsilon^3 &< 1 + 2\epsilon^3\lambda + \epsilon^6\lambda^2 \\
0 &< 1 + 2\epsilon^3\lambda - \epsilon^3 + \epsilon^6\lambda^2
\end{aligned}$$

The above is true for all $\lambda > 0$ because $1 - \epsilon^3 > 0$ since ϵ is small, all the other terms are positive. So the this term is $O(\epsilon)$

$$\begin{aligned}
\frac{\epsilon^4}{(\epsilon^3+\lambda)^2} = \frac{\epsilon^4}{\epsilon^6+2\lambda\epsilon^3+\lambda^2} &< \epsilon \\
\epsilon^4 &< \epsilon^7 + 2\lambda\epsilon^4 + \lambda^2\epsilon \\
\epsilon^3 &< \epsilon^6 + 2\lambda\epsilon^3 + \lambda^2 \\
0 &< \epsilon^6 + 2\lambda\epsilon^3 - \epsilon^3 + \lambda^2 \\
0 &< \epsilon^6 + 2\epsilon\epsilon^3 - \epsilon^3 + \epsilon^2
\end{aligned}$$

if $\epsilon < \lambda$ which is a valid assumption since ϵ is small and λ close to 0 doesn't satisfy conditions of theorem 4.1

$$0 < \epsilon^6 + 2\epsilon^4 - \epsilon^3 + \epsilon^2$$

The above is true because $\epsilon^2 - \epsilon^3 > 0$ since ϵ is small. So this term is also $O(\epsilon)$
So equation **(i)** becomes

$$\begin{aligned}
O(\epsilon) + \frac{1}{(1+\lambda)^2} + O(\epsilon) &= 0.25 \\
\frac{1}{1+2\lambda+\lambda^2} + O(\epsilon) &= 0.25 \\
1 + O(\epsilon) &= 0.25 + 0.5\lambda + 0.25\lambda^2 \\
0 &= 0.25\lambda^2 + 0.5\lambda - [0.75 + O(\epsilon)]
\end{aligned}$$

$$\lambda = \frac{-0.5 \pm \sqrt{(0.5)^2 - 4(0.25)(-1)(0.75+O(\epsilon))}}{0.5} = \frac{-0.5 \pm \sqrt{0.25+0.75+O(\epsilon)}}{0.5} = \frac{-0.5 \pm \sqrt{1+O(\epsilon)}}{0.5}$$

Since $\lambda > 0$ we take the positive root and since ϵ is small, by binomial approximation $(1+O(\epsilon))^{0.5} \approx 1 + 0.5 \times O(\epsilon)$
Therefore $\lambda = \frac{-0.5+1+O(\epsilon)}{0.5}$

$$\boxed{\lambda = 1 + O(\epsilon)}$$

Plugging that into p^* we get

$$p^* = -(B+\lambda I)^{-1}g = \begin{bmatrix} \frac{\epsilon^2}{1+\epsilon^3\lambda} \\ \frac{1}{1+\lambda} \\ \frac{\epsilon^2}{\epsilon^3+\lambda} \end{bmatrix} = \begin{bmatrix} \frac{\epsilon^2}{1+\epsilon^3+\epsilon^4} \\ \frac{1}{1+1+\epsilon} \\ \frac{\epsilon^2}{\epsilon^3+1+\epsilon} \end{bmatrix} = \begin{bmatrix} O(\epsilon) \\ \frac{1}{2} + O(\epsilon) \\ O(\epsilon) \end{bmatrix}$$

Set $f(x) = \frac{1}{2+x}$, then $f'(x) = -\frac{1}{(2+x)^2}$, $f(0) = \frac{1}{2}$ and $f'(0) = -\frac{1}{4}$
By Taylor's theorem, $f(x)$ about the point a is $f(x) \approx f(a) + f'(a)(x-a)$, taking $a = 0$ and evaluating at $x = \epsilon$ is a good approximation since ϵ is small (close to 0).
So $f(\epsilon) = \frac{1}{2+\epsilon} \approx f(0) + f'(0)\epsilon = \frac{1}{2} - \frac{1}{4}\epsilon = \frac{1}{2} + O(\epsilon)$
Also $\frac{\epsilon^2}{1+\epsilon^3+\epsilon^4} < \epsilon$ and $\frac{\epsilon^2}{\epsilon^3+1+\epsilon} < \epsilon$, again in reverse order.

$$\begin{aligned}
\frac{\epsilon^2}{1+\epsilon^3+\epsilon^4} &< \epsilon \\
\epsilon^2 &< \epsilon + \epsilon^4 + \epsilon^5 \\
\epsilon &< 1 + \epsilon^3 + \epsilon^4 \\
0 &< 1 - \epsilon + \epsilon^3 + \epsilon^4
\end{aligned}$$

The last statement is true since $1-\epsilon > 0$ because ϵ is small. So this term is $O(\epsilon)$.

$$\begin{aligned}
\frac{\epsilon^2}{\epsilon^3+1+\epsilon} &< \epsilon \\
\epsilon^2 &< \epsilon^4+\epsilon+\epsilon^2 \\
0 &< \epsilon^4+\epsilon
\end{aligned}$$

Clearly holds since ϵ is positive. So this term is also $O(\epsilon)$
Now the reduction in the model.

$$B = \begin{bmatrix} \frac{1}{\epsilon^3} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \epsilon^3 \end{bmatrix}, g = \begin{bmatrix} -\frac{1}{\epsilon} \\ -1 \\ -\epsilon^2 \end{bmatrix}, p^* = \begin{bmatrix} \frac{\epsilon^2}{1+\epsilon^3\lambda} \\ \frac{1}{1+\lambda} \\ \frac{\epsilon^2}{\epsilon^3+\lambda} \end{bmatrix}, Bp^* = \begin{bmatrix} \frac{1/\epsilon}{1+\epsilon^3\lambda} \\ \frac{1}{1+\lambda} \\ \frac{\epsilon^5}{\epsilon^3+\lambda} \end{bmatrix}$$

$$\begin{aligned}
m(0)-m(p^*) &= f - f - g^Tp^* - \frac{1}{2}p^{*T}Bp^* \\
&= -g^Tp^* - \frac{1}{2}p^{*T}Bp^* \\
&= \frac{\epsilon}{1+\epsilon^3\lambda} + \frac{1}{1+\lambda} + \frac{\epsilon^4}{\epsilon^3+\lambda} - \frac{1}{2}\left(\frac{\epsilon}{(1+\epsilon^3\lambda)^2} + \left(\frac{1}{1+\lambda}\right)^2 + \frac{\epsilon^7}{(\epsilon^3+\lambda)^2}\right) \\
&= O(\epsilon) + \left(\frac{1}{2}+O(\epsilon)\right) + O(\epsilon) - \frac{1}{2}\left(O(\epsilon) + \left(\frac{1}{2}+O(\epsilon)\right)^2 + O(\epsilon)\right) \\
&= \frac{1}{2} - \frac{1}{2}\left(\frac{1}{2}\right)^2 + O(\epsilon) \\
&= \frac{3}{8} + O(\epsilon)
\end{aligned}$$

$$\boxed{\textbf{REDUCTION IS } \frac{3}{8} + O(\epsilon)}$$

$\frac{1}{1+\lambda} = \frac{1}{2} + O(\epsilon)$ was shown before and all these other terms that appear above are $O(\epsilon)$

$$\frac{\epsilon}{1+\epsilon^3\lambda} < \epsilon, \frac{\epsilon^4}{\epsilon^3+\lambda} = \frac{\epsilon}{1+\lambda/\epsilon^3} < \epsilon, \frac{\epsilon}{(1+\epsilon^3\lambda)^2} < \epsilon, \frac{\epsilon^7}{(\epsilon^3+\lambda)^2} = \frac{\epsilon^7}{\epsilon^6+2\lambda\epsilon^3+\lambda^2} = \frac{\epsilon}{1+2\lambda/\epsilon^3+\lambda^2/\epsilon^6} < \epsilon$$

Because the denominator is > 1, that is $x > 1 \implies \frac{1}{x} < 1 \implies \frac{\epsilon}{x} < \epsilon$

Now the two-dimensional minimization strategy, the solution p will be of the form

$$p = xg + yB^{-1}g \qquad \text{where } x, y \in \mathbf{R}$$

$$\min_{\|p\| \le \Delta} g^Tp + \frac{1}{2}p^TBp$$

$$\begin{aligned}
g^Tp + \frac{1}{2}p^TBp &= g^T(xg+yB^{-1}g) + \frac{1}{2}(xg+yB^{-1}g)^TB(xg+yB^{-1}g) \\
&= xg^Tg + yg^TB^{-1}g + \frac{1}{2}x^2g^TBg + xyg^TB(B^{-1}g) + \frac{1}{2}y^2(B^{-1}g)^TB(B^{-1}g) \\
&= xg^Tg + yg^TB^{-1}g + \frac{1}{2}x^2g^TBg + xyg^Tg + \frac{1}{2}y^2g^TB^{-1}g && \text{B is symmetric, } (B^{-1})^T = B^{-1} \\
&= (x+xy)g^Tg + (\frac{1}{2}y^2+y)g^TB^{-1}g + \frac{1}{2}x^2g^TBg \\
&= x(1+y)\left(\frac{1}{\epsilon^2}+1+\epsilon^4\right) + (\frac{1}{2}y^2+y)(1+2\epsilon) + \frac{1}{2}x^2\left(\frac{1}{\epsilon^5}+1+\epsilon^7\right) && \text{by equations from the start}
\end{aligned}$$

Now note that $\frac{1}{x}x^2$ is always positive and $x(1+y)$ could be negative if x is negative. So no matter what $x \le 0$
The y also has to be negative, if y is positive the second term will be positive (worse than $y = 0$), so $y \le 0$.
The second term can be made negative with any negative y.
Now the term $\left(\frac{1}{\epsilon^5} + 1 + \epsilon^7\right)$ attached to x^2 is much bigger, in fact blows up for small ϵ, than $\left(\frac{1}{\epsilon^2} + 1 + \epsilon^4\right)$. So to minimize, need $x = 0$ (or very close to it). Is it possible for these terms to be negative.

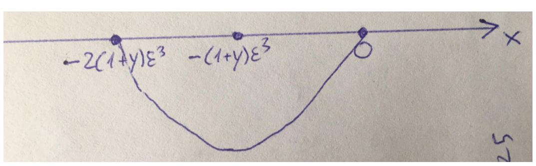
$$\begin{aligned} f(x,y) = x(1+y)\left(\frac{1}{\epsilon^2} + 1 + \epsilon^4\right) + \frac{1}{2}x^2\left(\frac{1}{\epsilon^5} + 1 + \epsilon^7\right) &\le 0 \\ \frac{1}{2}x^2\left(\frac{1}{\epsilon^5} + 1 + \epsilon^7\right) &\le -x(1+y)\left(\frac{1}{\epsilon^2} + 1 + \epsilon^4\right) \\ \frac{1}{2}x\left(\frac{1}{\epsilon^5} + 1 + \epsilon^7\right) &\ge -(1+y)\left(\frac{1}{\epsilon^2} + 1 + \epsilon^4\right) && \text{since } x \le 0 \\ x &\ge \frac{-2(1+y)\left(\frac{1}{\epsilon^2} + 1 + \epsilon^4\right)}{\frac{1}{\epsilon^5} + 1 + \epsilon^7} \\ x &\ge -2(1+y)\frac{\frac{1+\epsilon^2+\epsilon^6}{\epsilon^2}}{\frac{1+\epsilon^5+\epsilon^{12}}{\epsilon^5}} \\ x \ge -2(1+y)\epsilon^3\frac{1+\epsilon^2+\epsilon^6}{1+\epsilon^5+\epsilon^{12}} &\approx -2(1+y)\epsilon^3 \end{aligned}$$

So

$$-2(1+y)\epsilon^3 \le x \le 0$$

Want to find the x that makes f(x, y) smallest. Find critical point in the function

$$\begin{aligned} \frac{d}{dx}f(x,y) = (1+y)\left(\frac{1}{\epsilon^2} + 1 + \epsilon^4\right) + x\left(\frac{1}{\epsilon^5} + 1 + \epsilon^7\right) &= 0 \\ x\left(\frac{1}{\epsilon^5} + 1 + \epsilon^7\right) &= -(1+y)\left(\frac{1}{\epsilon^2} + 1 + \epsilon^4\right) \\ x &= -(1+y)\frac{\frac{1}{\epsilon^2} + 1 + \epsilon^4}{\frac{1}{\epsilon^5} + 1 + \epsilon^7} \\ &= -(1+y)\frac{\frac{1+\epsilon^2+\epsilon^6}{\epsilon^2}}{\frac{1+\epsilon^5+\epsilon^{12}}{\epsilon^5}} \\ &= -(1+y)\epsilon^3\frac{1+\epsilon^2+\epsilon^6}{1+\epsilon^5+\epsilon^{12}} \approx -(1+y)\epsilon^3 \\ \frac{d}{dx^2}f(x,y) &= \frac{1}{\epsilon^5} + 1 + \epsilon^7 \ge 0 \end{aligned}$$



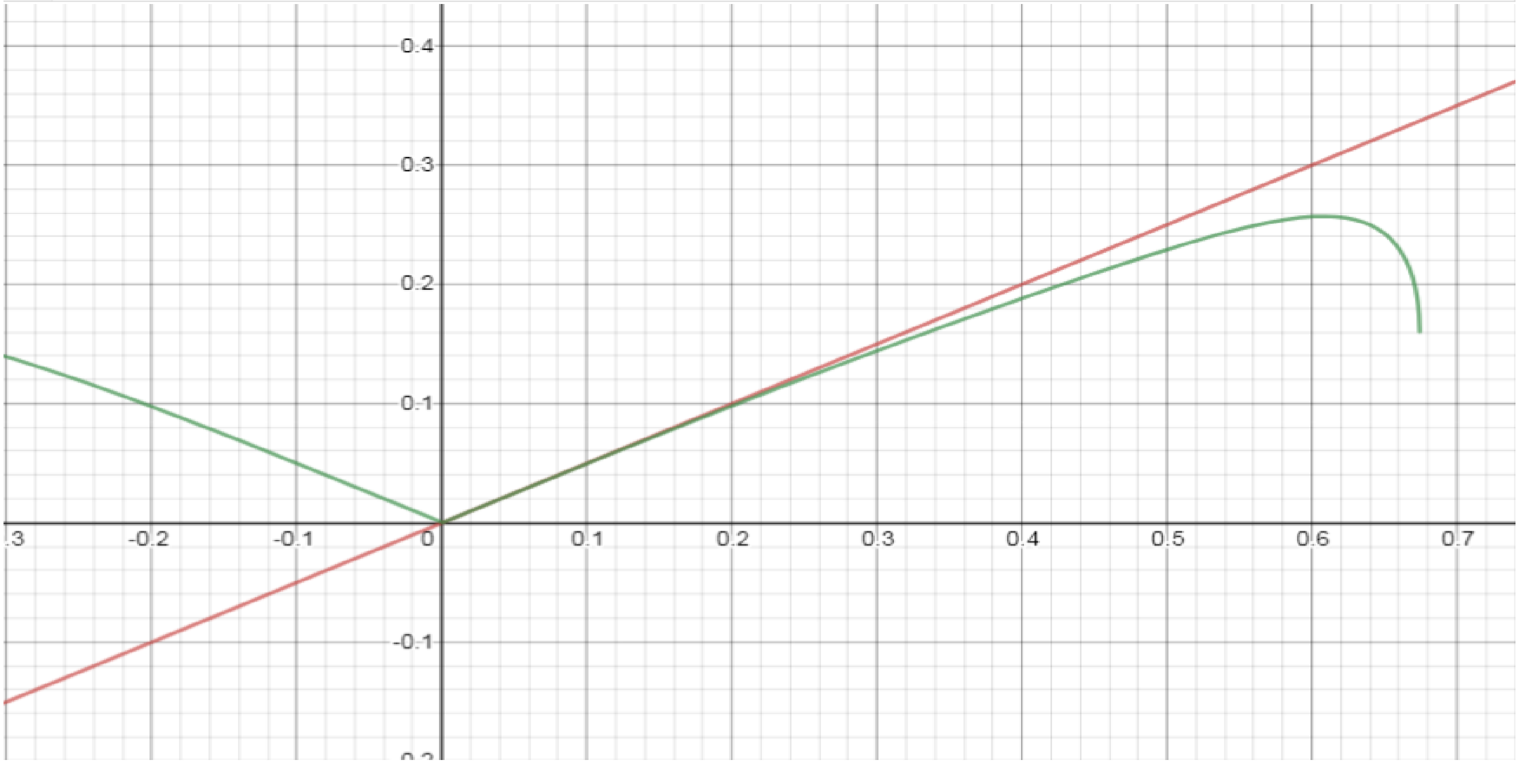
It is most negative at the critical point x, so $x = -(1+y)\epsilon^3$
So $p^* = -(1+y)\epsilon^3 g + yB^{-1}g$ for some $y \in \mathbf{R}$, need to determine it so $\|p^*\| \le \Delta = 0.5$ or $\|p^*\|^2 \le \Delta = 0.5^2 = 0.25$. Take the largest step possible.

$$\begin{aligned} \|p\| = p^T p &= (xg + yB^{-1}g)^T(xg + yB^{-1}g) \\ &= x^2 g^T g + 2xyg^T B^{-1} g + y^2 (B^{-1}g)^T(B^{-1}g) \\ &= x^2 g^T g + 2xyg^T B^{-1} g + y^2 g^T B^{-2} g && B^{-1} \text{ is symmetric} \\ &= x^2 g^T g + 2xyg^T B^{-1} g + y^2 g^T g && g^T B^{-2} g = g^T g \\ &= (1+y)^2\epsilon^6(1 + \frac{1}{\epsilon^2} + \epsilon^4) - 2(1+y)\epsilon^3 y(1+2\epsilon) + y^2(1 + \frac{1}{\epsilon^2} + \epsilon^4) \\ &= (1+2y+y^2)(\epsilon^6 + \epsilon^4 + \epsilon^{10}) - 2(y+y^2)(\epsilon^3 + 2\epsilon^4) + y^2(1 + \frac{1}{\epsilon^2} + \epsilon^4) \\ &= y^2\left(\epsilon^6 + \epsilon^4 + \epsilon^{10} - 2\epsilon^3 - 4\epsilon^4 + 1 + \frac{1}{\epsilon^2} + \epsilon^4\right) + y\left(2\epsilon^6 + 2\epsilon^4 + 2\epsilon^{10} - 2\epsilon^3 - 4\epsilon^4\right) + \left(\epsilon^6 + \epsilon^4 + \epsilon^{10}\right) \\ 0.25 &= y^2\left(\epsilon^6 + \epsilon^{10} - 2\epsilon^3 - 2\epsilon^4 + 1 + \frac{1}{\epsilon^2}\right) + y\left(2\epsilon^6 + 2\epsilon^{10} - 2\epsilon^3 - 2\epsilon^4\right) + \left(\epsilon^6 + \epsilon^4 + \epsilon^{10}\right) \end{aligned}$$

Writing the solution for this quadratic and graphing it in terms of ϵ we get something which is almost identical to $\frac{1}{2}\epsilon$ for small ϵ (which is the case here)

$$0.5x$$

$$\frac{-\left(2x^6+2x^{10}-2x^3-2x^4\right)+\sqrt{\left(2x^6+2x^{10}-2x^3-2x^4\right)^2-4\left(x^6+x^{10}-2x^3-2x^4+1+\frac{1}{x^2}\right)\left(x^6+x^4+x^{10}-0.25\right)}}{2\left(x^6+x^{10}-2x^3-2x^4+1+\frac{1}{x^2}+x^4\right)}$$



Back to minimization problem

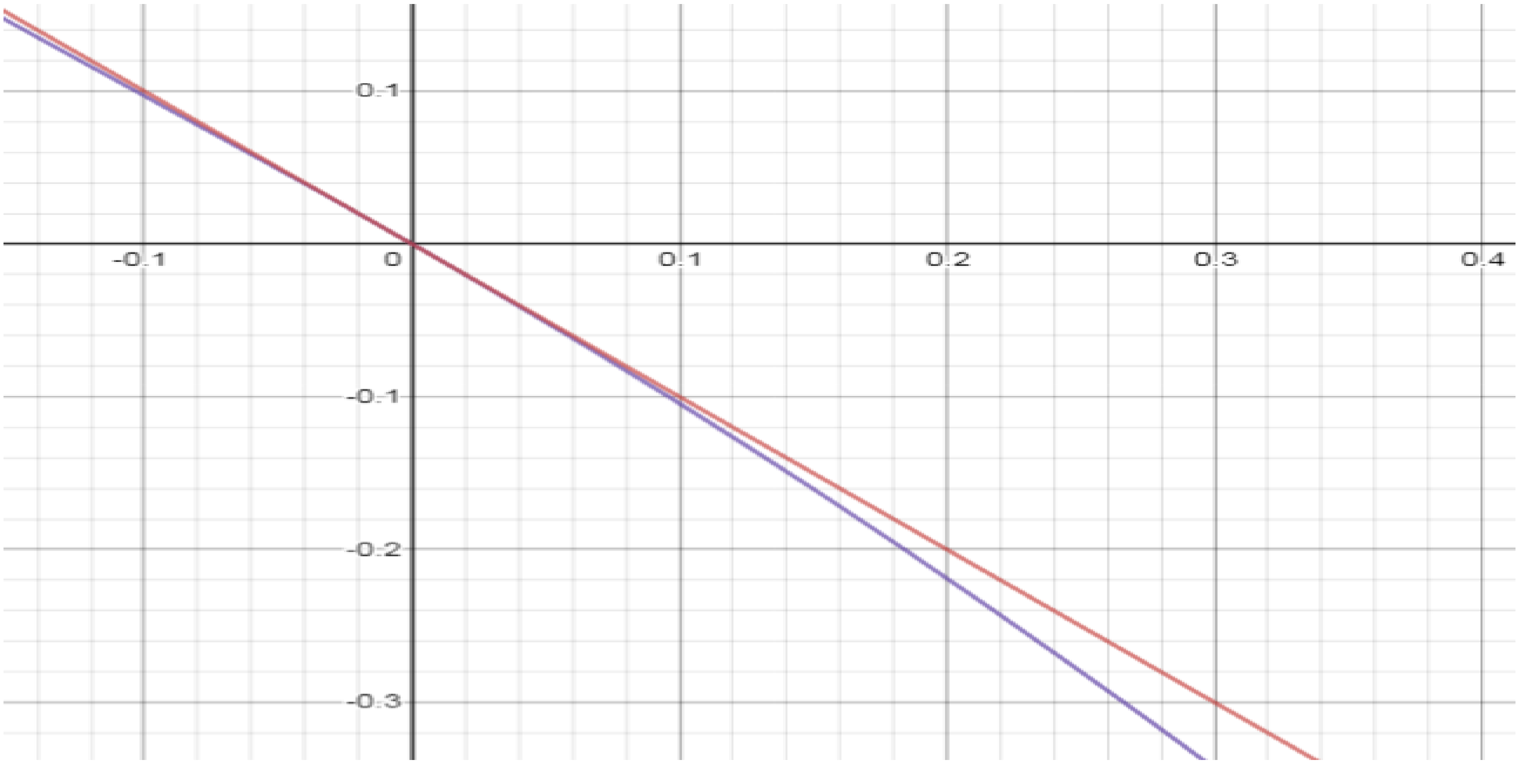
$$g^T p + \frac{1}{2} p^T B p$$

So to minimize, $y = -\frac{1}{2}\epsilon$ and $x = -(1+y)\epsilon^4 = -(1-0.5\epsilon)\epsilon^3 = -\epsilon^3 + \epsilon^4$

$$\begin{aligned} g^T p + \frac{1}{2} p^T B p &= x(1+y)\left(\frac{1}{\epsilon^2}+1+\epsilon^4\right) + (\frac{1}{2}y^2 + y)(1+2\epsilon) + \frac{1}{2}x^2\left(\frac{1}{\epsilon^5}+1+\epsilon^7\right) \\ &= (-\epsilon^3+\epsilon^4)(1-0.5\epsilon)\left(\frac{1}{\epsilon^2}+1+\epsilon^4\right) + (\frac{1}{2}(0.5\epsilon)^2 - 0.5\epsilon)(1+2\epsilon) + \frac{1}{2}(-\epsilon^3+\epsilon^4)^2\left(\frac{1}{\epsilon^5}+1+\epsilon^7\right) \\ &\approx -\epsilon \end{aligned}$$

$$\left(-x^3+x^4\right)(1-0.5x)\left(\frac{1}{x^2}+1+x^4\right)+\left(\frac{1}{2}(0.5x)^2-0.5x\right)(1+2x)+\frac{1}{2}\left(-x^3+x^4\right)^2\left(\frac{1}{x^5}+1+x^7\right)$$

$$-x$$



$$m(0) - m(p) = -(g^T p + \frac{1}{2} p^T B p) = O(\epsilon)$$

Where $p = (-\epsilon^3 + \epsilon^4)g - \left(\frac{1}{2}\epsilon\right) B^{-1}g$, almost a multiple of $B^{-1}g$