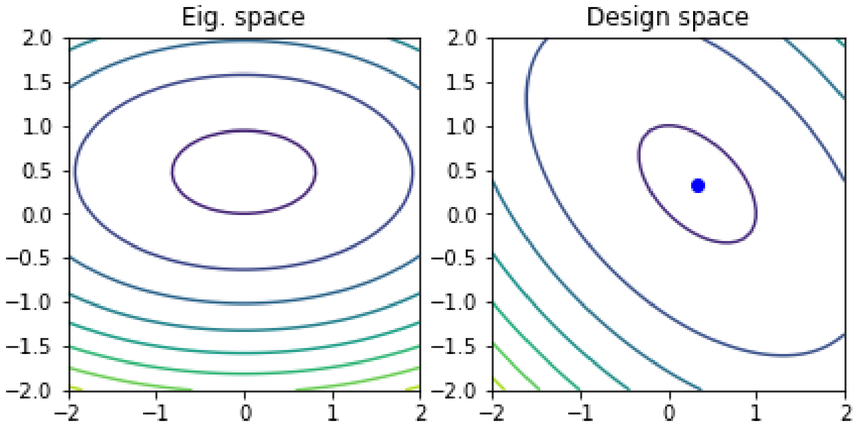
Problem 1

The standard form of quadratic functions is defined as $\mathrm{f(x)} = \frac{1}{2}x^T Ax + x^T b + c$

Part A

The function $\mathrm{f}(x_1, x_2) = {x_1}^2 + x_1 x_2 + {x_2}^2 - x_1 - x_2$ in standard form becomes

$$\mathrm{f}(x_1, x_2) = \frac{1}{2}\begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

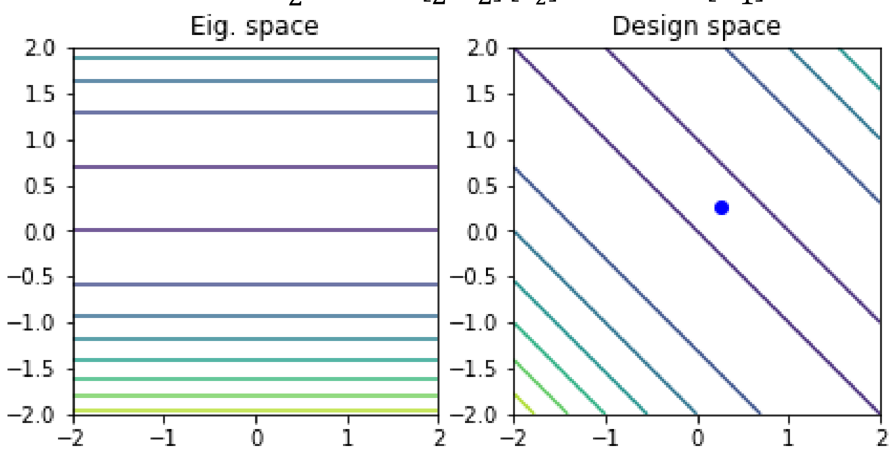


The A matrix is positive definite. Therefore, the local and global minimizers are found through $x^* = -A^{-1}b$. The unique minimizer is located at (x_1^*,x_2^*)= [0.3333, 0.3333]

Part B

The function $\mathrm{f}(x_1, x_2) = {x_1}^2 + 2x_1 x_2 + {x_2}^2 - x_1 - x_2$ in standard form becomes

$$\mathrm{f}(x_1, x_2) = \frac{1}{2}\begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

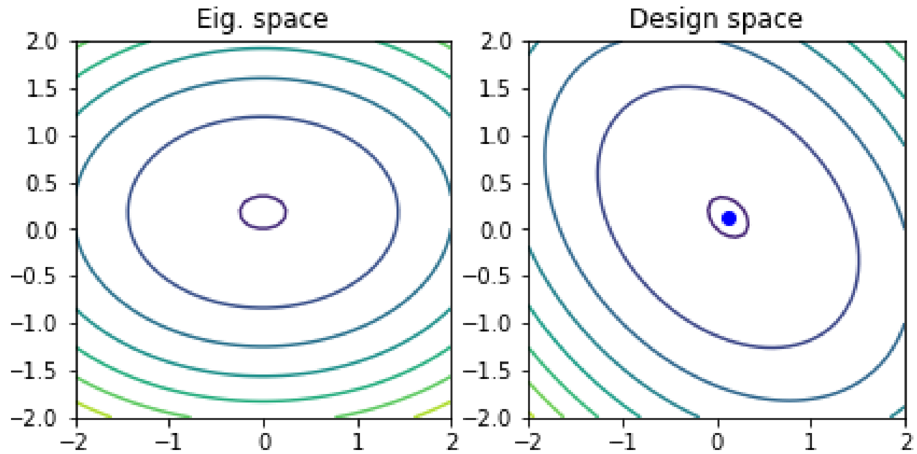


The A matrix is positive semi-definite so there are an infinite number of minimizers along a line. A possible critical point is (x_1^*,x_2^*)= [0.25, 0.25]

Part C

The function $f(x_1, x_2) = 3{x_1}^2 + 2x_1x_2 + 3{x_2}^2 - x_1 - x_2$ in standard form becomes

$$f(x_1, x_2) = \frac{1}{2}\begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 6 & 2 \\ 2 & 6 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

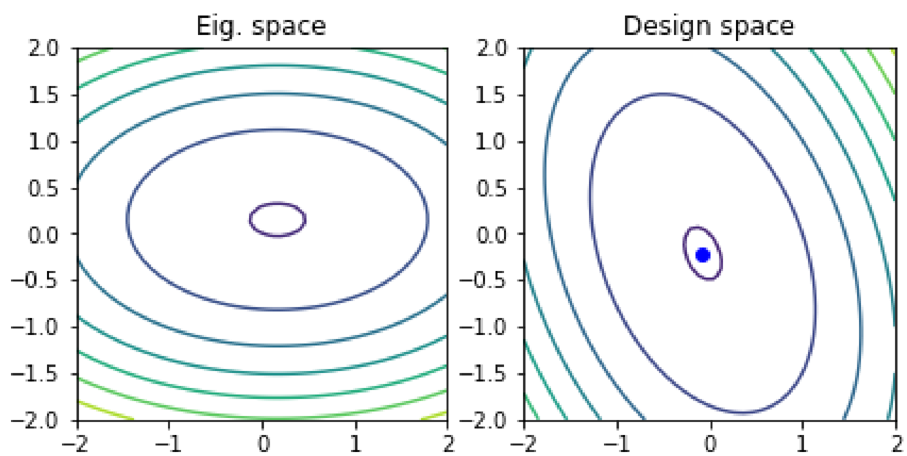


The A matrix is positive definite. The unique global minimizer is located at (x_1^*, x_2^*)= [0.125, 0.125]

Part D

The function $f(x_1, x_2) = 4{x_1}^2 + 6x_1x_2 + 2{x_2}^2 + x_1 + x_2$ in standard form becomes

$$f(x_1, x_2) = \frac{1}{2}\begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 8 & 2 \\ 2 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$



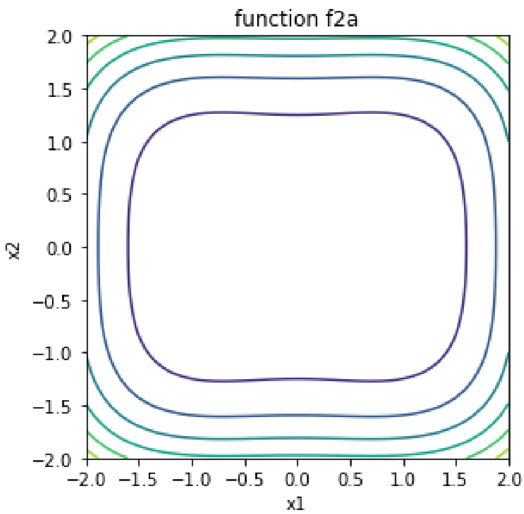
The A matrix is positive definite. The unique global minimizer is located at (x_1^*, x_2^*)= [-0.07142857, -0.21428571]

Problem 2

Part A

$f(x_1, x_2) = {x_1}^4 + {x_2}^4 + 1 - {x_1}^2 + {x_2}^2$

$$\nabla f(x_1, x_2) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 4{x_1}^3 - 2x_1 \\ 4{x_2}^3 + 2x_2 \end{bmatrix}; \quad H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial {x_1}^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial {x_2}^2} \end{bmatrix} = \begin{bmatrix} 12{x_1}^2 - 2 & 0 \\ 0 & 12{x_2}^2 + 2 \end{bmatrix}$$



The gradient is zero at $x_1 = \pm\frac{1}{\sqrt{2}}, 0$ and $x_2 = 0, \pm\frac{j}{\sqrt{2}}$

X1_cp	X2_cp	λ of H(x1, x2)	Notes
0.7071	0.7071j	[-4, 4]	Indefinite, saddle point
0.7071	-0.7071j	[-4, 4]	Indefinite, saddle point
0.7071	0	[2, 4]	Positive definite, minimum
-0.7071	0.7071j	[-4, 4]	Indefinite, saddle point
-0.7071	-0.7071j	[-4, 4]	Indefinite, saddle point
-0.7071	0	[2, 4]	Positive definite, minimum
0	0.7071j	[-4, -2]	Negative definite, maximum
0	-0.7071j	[-4, -2]	Negative definite, maximum
0	0	[-2, 2]	Indefinite, saddle point

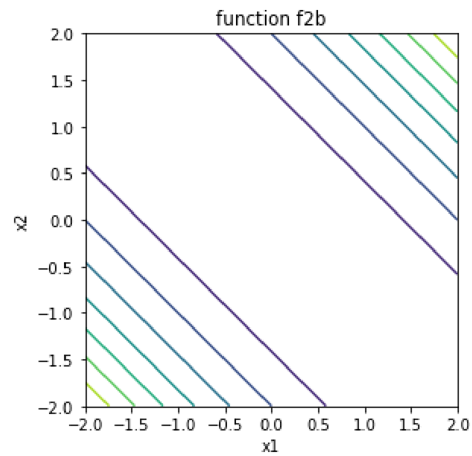
Part B

$f(x_1, x_2) = {x_1}^2 + {x_2}^2 + 2x_1x_2$

$$f(x_1, x_2) = \frac{1}{2}\begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 + 2x_2 \\ 2x_2 + 2x_1 \end{bmatrix}; \; H(f) = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$$

A is semi-definite; the gradient is zero for x1=-x2. This indicates that the critical points lie on a line of infinite minimizers.

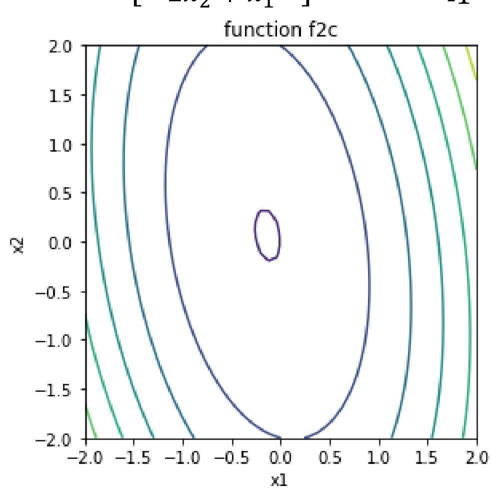


Part C

$f(x_1, x_2) = 4{x_1}^2 + {x_2}^2 + x_1x_2 + x_1$

$$f(x_1, x_2) = \frac{1}{2}\begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 8 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\nabla f(x_1, x_2) = \begin{bmatrix} 8x_1 + x_2 + 1 \\ 2x_2 + x_1 \end{bmatrix}; \; H(f) = \begin{bmatrix} 8 & 1 \\ 1 & 2 \end{bmatrix}$$

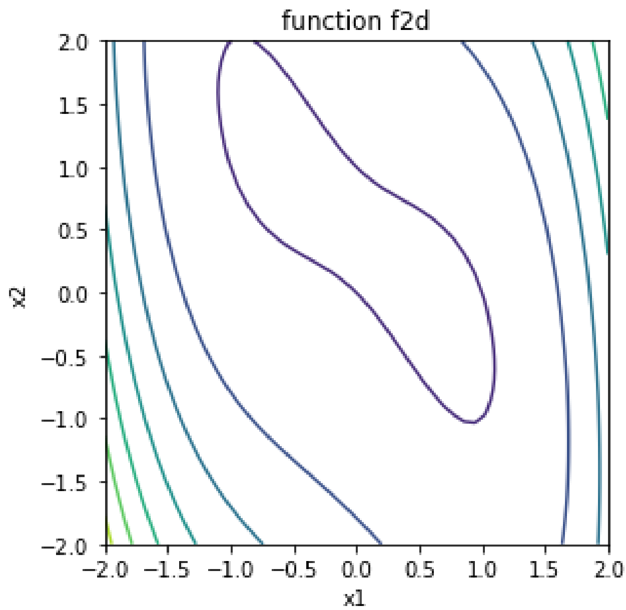


A is positive definite with a unique minimizer at the critical points are (x1, x2)= [0,0]

Part D

$f(x_1, x_2) = {x_1}^4 + {x_2}^2 + 2x_1x_2 - x_1 - x_2$

$$\nabla f(x_1, x_2) = \begin{bmatrix} 4{x_1}^3 + 2x_2 - 1 \\ 2x_2 + x_1 - 1 \end{bmatrix};\ H(f) = \begin{bmatrix} 12{x_1}^2 & 2 \\ 1 & 2 \end{bmatrix}$$



The gradient is zero at (x1,x2)= [-0.5, 0.75], [0, 0.5], [0.5, 0.25]

X1_cp	X2_cp	λ of H(x1, x2)	Notes
-0.5	0.75	[1.38196601 3.61803399]	Positive definite, minimum
0	0.5	[-0.41421356 2.41421356]	Nonsingular, Indefinite, saddle point
0.5	0.25	[1.38196601 3.61803399]	Positive definite, minimum

Problem 3

Consider the function f(x)= x(1-x)^2(x-3) on the interval $x \in [-1,4]$. To find a descent direction from the point x=0, we need to consider the following information:

- f(0)= 0
- f'(x)= 4x^3-15x^2+14x-3
- f'(0)= -3
- $\phi'(0) = \left.\frac{d\phi}{d\alpha}\right|_{\alpha=0} = \nabla f^T(x_k)p < 0$

Since the first derivative evaluated at x=0 is negative, we need to choose p>0.

Pick p= -f'(x)/|f'(x)= 1

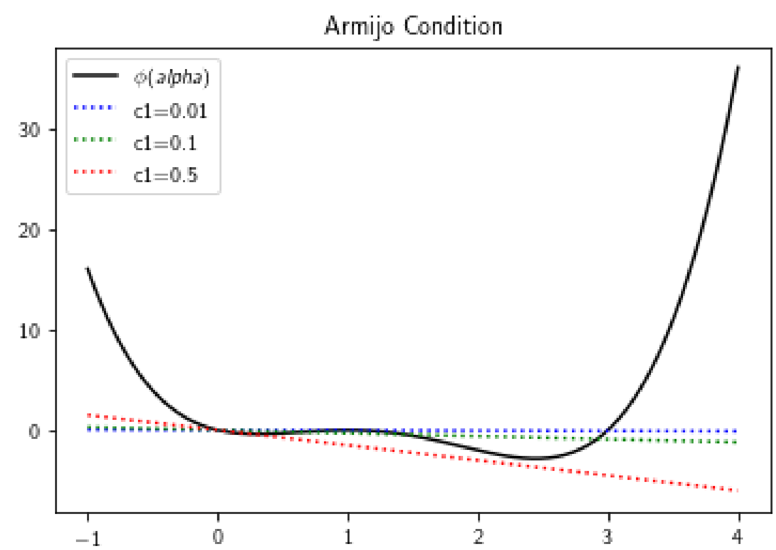
$$\phi(\alpha) = f(0 + 1\alpha) = (\alpha)(1-\alpha)^2(\alpha-3)$$

Part A

The sufficient decrease condition states that

$$\phi(\alpha) \le \phi(0) + c_1\alpha\phi'(0)$$

Graphically, this is represented below for several values of c1.



Part B

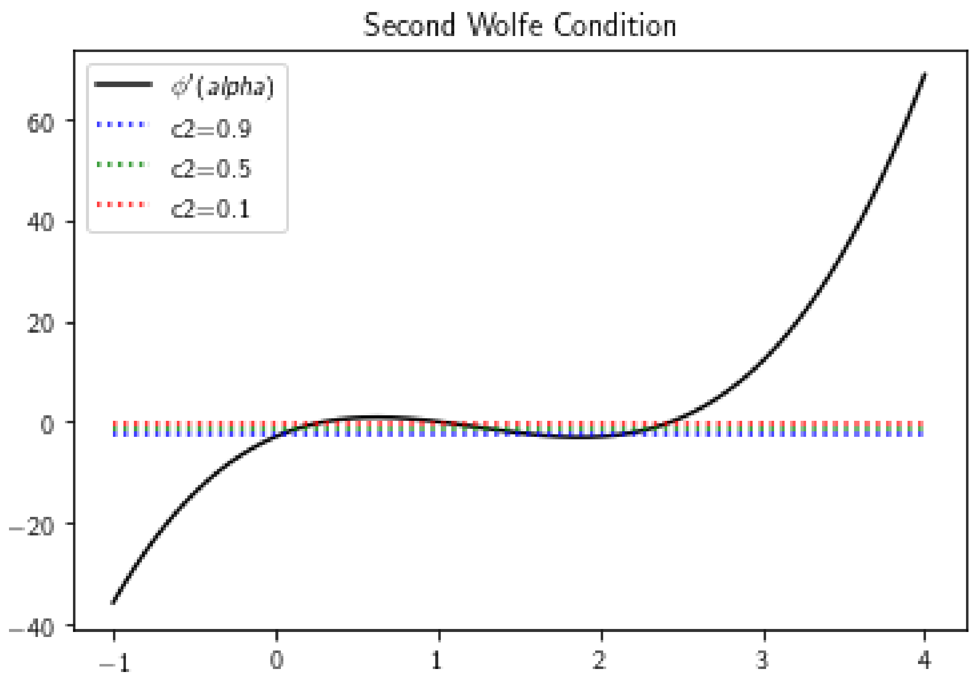
The second Wolfe condition states that

$$\phi'(\alpha) \geq c_2\phi'(0)$$

Recall

$$\phi(\alpha) = (\alpha)(1-\alpha)^2(\alpha-3) \implies \phi'(\alpha) = (\alpha-1)(4\alpha^2-11\alpha+3)$$

Graphically, this is represented below for several values of c2.



Part C:

The strong Wolfe conditions state that

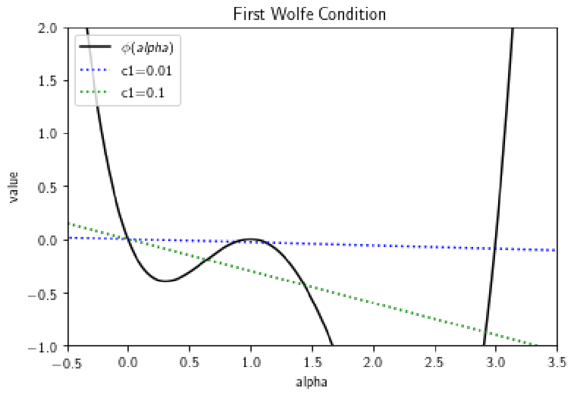
$$\phi(\alpha) \leq \phi(0) + c_1\alpha\phi'(0)$$

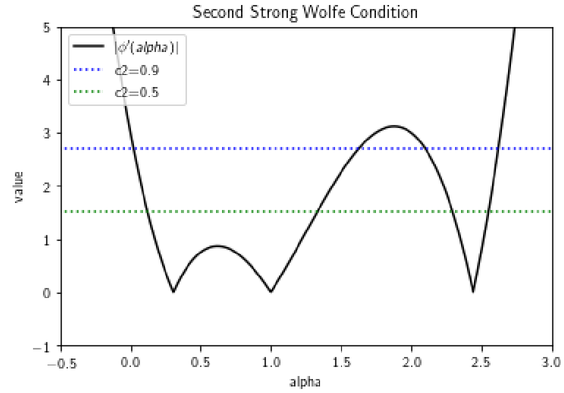
$$|\phi'(\alpha)| \leq |c_2\phi'(0)|$$

In other words, we need to find the values for alpha which satisfy BOTH conditions. This can be found through iterating through the given alphas and then comparing the associated value with the strong Wolfe conditions:

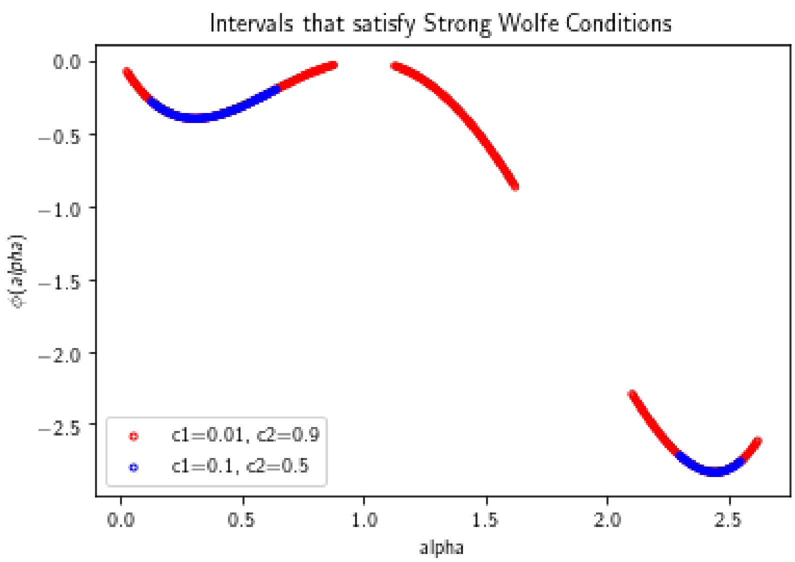
```
    #c1=0.01; c2=0.9
    alpha_good_1=[]
    phi_good_1=[]
    i=0
    while i<len(alpha):
        if (merit_func[i]<= suff_dec_1[i]) and (abs(merit_func_der[i])<=abs(wolfe2_1)
            #print(alpha[i],"satisfies both conditions")
            alpha_good_1.append(alpha[i])
            phi_good_1.append(merit_func[i])
        else:
            pass
        i+=1
```

Note that the variables on line 144 have already been calculated at previous steps in the assignment. Therefore, the actual value of that point is compared instead of substituting in alpha[i]. Simply put, we want an intersection of the alphas that satisfy each of the conditions (see figure below).





This leads to the following graph:



Furthermore, we can solve for the intervals of alpha which satisfy the strong Wolfe conditions, i.e.

$$(\alpha)(1-\alpha)^2(\alpha-3) \leq \; 0 + c_1\alpha(\text{-}3)$$

$$|(\alpha-1)(4\alpha^2-11\alpha+3)| \leq \; |c_2(\text{-}3)|$$

Problem 4

Consider the following functions:

$$f(x_1, x_2) = x_1^4 + x_2^4 + 1 - x_1^2 + x_2^2$$
$$f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

At Prof. Kennedy's suggestion, the Python implementation uses the code he developed for the line search methods in the notebook "Line Search Algorithms", with some modifications.

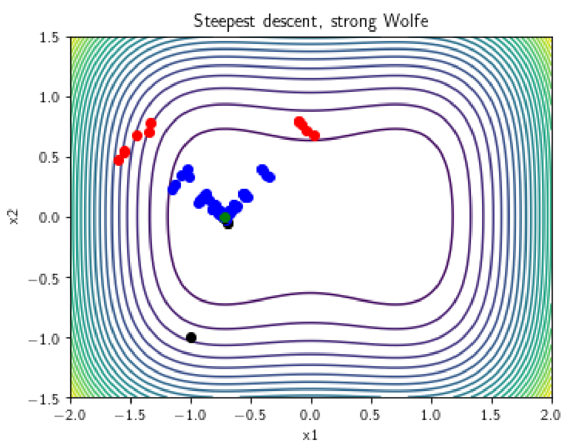
Part A:

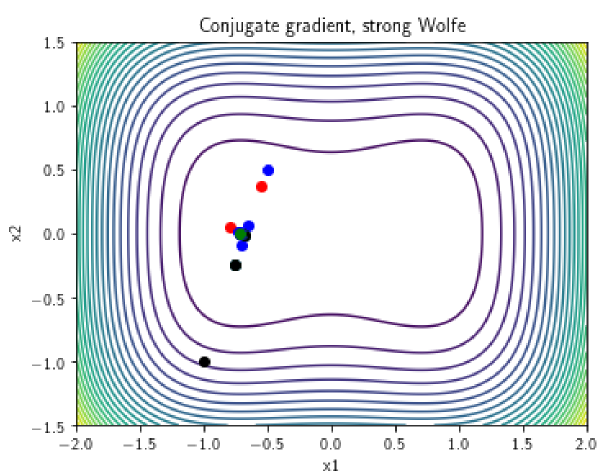
Function 1

The gradient of this function is given by

$$\nabla \mathrm{f}(x_1, x_2) = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 4{x_1}^3 - 2x_1 \\ 4{x_2}^3 + 2x_2 \end{bmatrix}$$

With a starting point of [-1, -1] both searches were able to find a first order point.



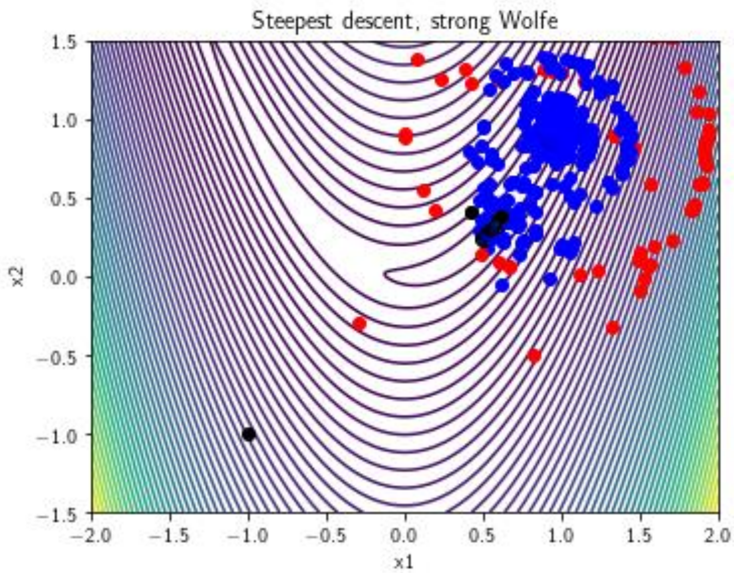


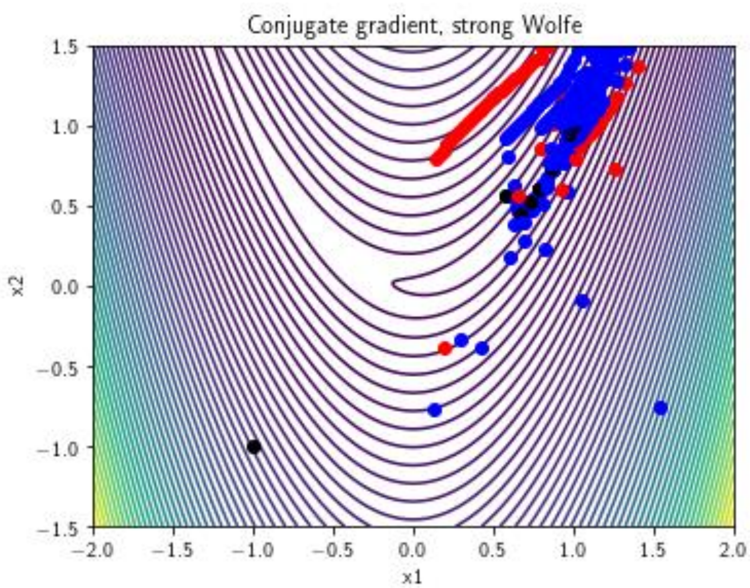
Function 2

This function is the Rosenbrock function. The gradient is given by

$$\begin{bmatrix} -2(1-x_1) - 200(x_2 - {x_1}^2)x_1 \\ 200(x_2 - {x_1}^2) \end{bmatrix}$$

With the starting point of [-1, -1] over 100 iterations produces the following figures. Note that neither method found a first order point, requiring all of the iterations and still failing to converge.



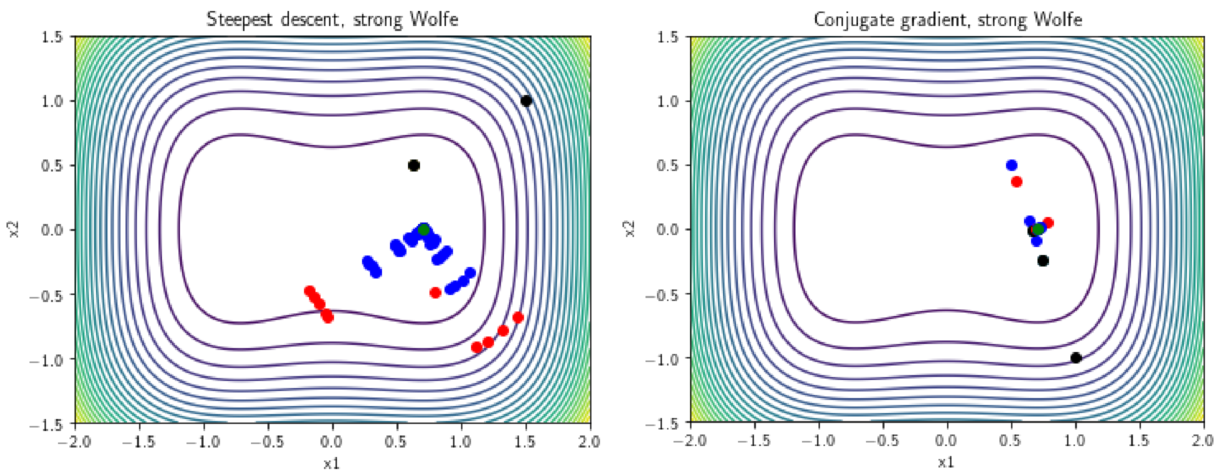


Part B:

Consider the first function where the:

- Steepest descent start at [1.5, 1.0] -> x0= [7.07106668e-01, 3.64133402e-07]
- Conjugate gradient start at [1.0, -1.0] -> x0= [7.07106720e-01, -9.29942394e-09]

This produces the following figures- both methods were able to find a first order point. However, the conjugate gradient was able to find a minimum in less steps, similar to Part A.



Consider the second function where the:

- Steepest descent start at [1.5, 1.0] -> x0= [1.03426082, 1.06979839]
- Conjugate gradient start at [1.0, -1.0] -> x0= [0.9995105, 0.99897048]

This produces the following figures- neither method was able to find a first order point. However, the steepest descent gradient tends to search in a “circle” around the in minimum, while the conjugate gradient searches along a “valley”.