MIE 1621 Computational Project Part 1

Goal: minimizing a multivariate function f(x) using gradient-based method with backtracking.

(a) Use backtracking as described in class to compute step-lengths (so you need to set the parameters s, γ and β).

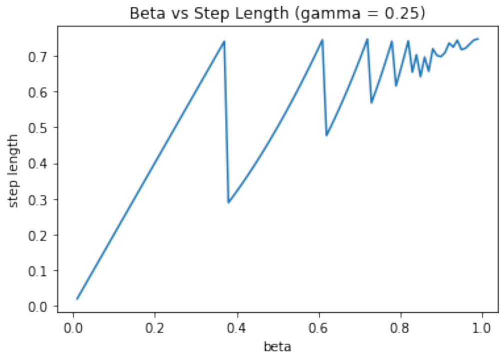
Condition: $s > 0, \gamma \in (0,1), \beta \in (0,1)$

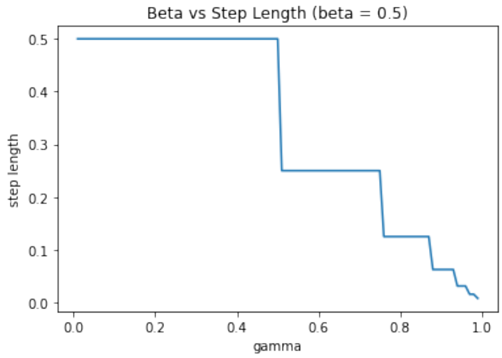
1. Start with initial guess for step length $\alpha^{(k)} = s > 0$
2. If $f(x^k) - f(x^k + \alpha^k d^k) \geq -\gamma\alpha^k \nabla f(x^k)\, d^k$
   $\Rightarrow \alpha^{(k)}$ decrease function sufficiently $\Rightarrow$ current $\alpha^{(k)}$ is chosen
3. Otherwise, repeat reduce $\alpha^{(k)}$ by multiplying $\beta \in (0,1)$ until condition in step 2 is met
   $\Rightarrow \alpha^{(k)} = \beta\alpha^{(k)}$

Observation (take $f_1(x)$ as an example)

$$f_1(x) = x_1^2 + x_2^2 + x_3^2 \;\; with \;\; x^{(0)} = (1,1,1)^T$$

- Let $\gamma = 0.25$ and $s = 2$. The larger the β, the larger the final step length $\alpha^{(k)}$ will be chosen. Larger β reduce smaller amount of $\alpha^{(k)}$, higher chance the larger $\alpha^{(k)}$ will satisfy the condition in step 2.
- Let $\beta = 0.6$ and $s = 2$. The smaller the γ, the larger the final step length $\alpha^{(k)}$ will be chosen. Smaller γ require larger $\alpha^{(k)}$ to decrease the function sufficiently.





- Try different combinations of γ & β.

  Slow convergence when $\alpha^{(k)}$ is small, might miss the value satisfy stop condition when $\alpha^{(k)}$ is large. Set $s = 2, \gamma = 0.25, \beta = 0.5$ in part d for consistence, we could try to adjust the value of β & γ if the function does not meet the termination criterion before hits max number of iterations.

(γ, β)	step length $\alpha^{(1)}$	Iteration	$x^{(k)}$	Gradient at $x^{(k)}$
$(0.1, 0.9)$	0.8609	40	[0.00000218 0.00000218 0.00000218]	[0.00000436 0.00000436 0.00000436]
$(0.25, 0.5)$	0.5	1	[0. 0. 0.]	[0. 0. 0.]
$(0.9, 0.1)$	0.0200	313	[0.00000282 0.00000282 0.00000282]	[0.00000565 0.00000565 0.00000565]

(b) Use as a stopping condition $\frac{||\nabla f(x)||}{1+|f(x)|} \leq \epsilon \;\; with \;\; \epsilon = 10^{-5}$ or stop if the number of iterations hits 1000.

The gradient of the function might not converge to the optimal solution but close to it, set stopping condition to allow some tolerance. The gradient-based method might take a long time to converge, set the number of max iterations to 1000 to avoid computational complexity.

(c) Print the initial point and for each iteration print the search direction, the step length, and the new iterate x(k+1): If the number of iterations is more than 15 then printout the details of the just the first 10 iterations as well as the details of the last 5 iterations before the stopping condition is met. Indicate if the iteration maximum is reached.

The implementation of gradient-based method with backtracking:

1. Set initial point x^0
2. Find the next solution $x^{k+1} = x^k + \alpha^k d^k = x^k - \alpha^k \nabla \mathrm{f}(\mathrm{x}^k)$ with backtracking step size α^k and gradient direction $d^k = -\nabla \mathrm{f}(\mathrm{x}^k)$
3. Repeat step 2 until the termination criteria are met

(d) Test your algorithms on the following test problems (Set $s = 2, \gamma = 0.25, \beta = 0.5$ for consistence)

1. $f_1(x) = x_1^2 + x_2^2 + x_3^2 \ \ with \ \ x^{(0)} = (1,1,1)^T$

```
initial point = [1 1 1], gradient = [2. 2. 2.]

iteration 1, search direction = [-2. -2. -2.], step length = 0.5, x = [0. 0. 0.]

stop at iteration 1, current x = [0. 0. 0.], gradient = [0. 0. 0.], f = 1.4791141972893971e-31
```

2. $f_2(x) = x_1^2 + 2x_2^2 - 2x_1x_2 - 2x_2 \ \ with \ \ x^{(0)} = (0,0)^T$

```
initial point = [0 0], gradient = [ 0. -2.]

iteration 1, search direction = [-0.  2.], step length = 0.25, x = [0.  0.5]
iteration 2, search direction = [ 1. -0.], step length = 0.5, x = [0.5 0.5]
iteration 3, search direction = [-0.  1.], step length = 0.25, x = [0.5  0.75]
iteration 4, search direction = [0.5 0. ], step length = 0.5, x = [0.75 0.75]
iteration 5, search direction = [-0.   0.5], step length = 0.25, x = [0.75  0.875]
iteration 6, search direction = [ 0.25 -0.  ], step length = 0.5, x = [0.875 0.875]
iteration 7, search direction = [-0.    0.25], step length = 0.25, x = [0.875  0.9375]
iteration 8, search direction = [ 0.125 -0.   ], step length = 0.5, x = [0.9375 0.9375]
iteration 9, search direction = [0.    0.125], step length = 0.25, x = [0.9375  0.96875]
iteration 10, search direction = [ 0.0625 -0.    ], step length = 0.5, x = [0.96875 0.96875]
    ......
    ......
iteration 29, search direction = [0.         0.00012207], step length = 0.25, x = [0.99993896 0.99996948]
iteration 30, search direction = [0.00006104 0.        ], step length = 0.5, x = [0.99996948 0.99996948]
iteration 31, search direction = [0.         0.00006104], step length = 0.25, x = [0.99996948 0.99998474]
iteration 32, search direction = [0.00003052 0.        ], step length = 0.5, x = [0.99998474 0.99998474]
iteration 33, search direction = [0.         0.00003052], step length = 0.25, x = [0.99998474 0.99999237]

stop at iteration 33, current x = [0.99998474 0.99999237], gradient = [-0.00001526  0.        ], f = -0.9999999998835
845
```

3. $f_3(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 \ \ with \ \ x^{(0)} = (-1.2, 1)^T$

```
initial point = [-1.2  1. ], gradient = [-215.6  -88. ]

iteration 1, search direction = [215.6  88. ], step length = 0.0009765625, x = [-0.98945313  1.0859375 ]
iteration 2, search direction = [-38.33803031 -21.38400269], step length = 0.0009765625, x = [-1.02689261  1.0650546
8]
iteration 3, search direction = [-0.27816415 -2.10925141], step length = 0.00390625, x = [-1.02797919  1.05681542]
iteration 4, search direction = [ 4.02544229 -0.01484275], step length = 0.5, x = [0.98474196 1.04939405]
iteration 5, search direction = [ 31.41515509 -15.93546343], step length = 0.0009765625, x = [1.01542082 1.03383207]
iteration 6, search direction = [ 1.08718639 -0.55052448], step length = 0.0009765625, x = [1.01648253 1.03329445]
iteration 7, search direction = [-0.00949734 -0.01154359], step length = 0.015625, x = [1.01633413 1.03311408]
iteration 8, search direction = [ 0.0401058  -0.03580224], step length = 0.0009765625, x = [1.0163733  1.03307912]
iteration 9, search direction = [-0.00655043 -0.01288708], step length = 2, x = [1.00327244 1.00730496]
iteration 10, search direction = [ 0.29418475 -0.14987436], step length = 0.0009765625, x = [1.00355973 1.0071586 ]
    ......
    ......
iteration 417, search direction = [ 0.00000298 -0.00001152], step length = 0.001953125, x = [1.00001004 1.0000201 ]
iteration 418, search direction = [-0.00001069 -0.00000469], step length = 0.001953125, x = [1.00001002 1.00002009]
iteration 419, search direction = [ 0.00000239 -0.00001121], step length = 0.001953125, x = [1.00001002 1.00002006]
iteration 420, search direction = [-0.00001011 -0.00000496], step length = 0.00390625, x = [1.00000998 1.00002005]
iteration 421, search direction = [ 0.00001382 -0.00001689], step length = 0.0009765625, x = [1.00000999 1.00002003]

stop at iteration 421, current x = [1.00000999 1.00002003], gradient = [0.0000036  0.00000819], f = 1.000439904048022
8e-10
```

4. $f_4(x) = (x_1 + x_2)^4 + x_2^2 \ with \ x^{(0)} = (2, -3)^T$

```
initial point = [ 2 -2], gradient = [ 0. -4.]

iteration 1, search direction = [-0.  4.], step length = 0.25, x = [ 2. -1.]
iteration 2, search direction = [-4. -2.], step length = 0.0625, x = [ 1.75  -1.125]
iteration 3, search direction = [-0.9765625  1.2734375], step length = 1.0, x = [0.7734375 0.1484375]
iteration 4, search direction = [-3.13383484 -3.43070984], step length = 0.0625, x = [ 0.57757282 -0.06598186]
iteration 5, search direction = [-0.5355852  -0.40362147], step length = 0.25, x = [ 0.44367652 -0.16688723]
iteration 6, search direction = [-0.08482187  0.2489526 ], step length = 0.5, x = [ 0.40126559 -0.04241093]
iteration 7, search direction = [-0.18484842 -0.10002655], step length = 0.5, x = [ 0.30884138 -0.09242421]
iteration 8, search direction = [-0.0405448   0.14430362], step length = 0.5, x = [ 0.28856898 -0.0202724 ]
iteration 9, search direction = [-0.07725123 -0.03670644], step length = 0.5, x = [ 0.24994336 -0.03862562]
iteration 10, search direction = [-0.03774574  0.03950549], step length = 1.0, x = [0.21219763 0.00087988]
    ......
    ......
iteration 613, search direction = [-0.00001014 -0.00001734], step length = 1.0, x = [ 0.01362032 -0.00001374]
iteration 614, search direction = [-0.00001008  0.0000174 ], step length = 1.0, x = [0.01361024 0.00000366]
iteration 615, search direction = [-0.00001009 -0.00001742], step length = 0.5, x = [ 0.01360519 -0.00000505]
iteration 616, search direction = [-0.00001006  0.00000003], step length = 2, x = [ 0.01358507 -0.00000499]
iteration 617, search direction = [-0.00001002 -0.00000005], step length = 2, x = [ 0.01356503 -0.00000508]

stop at iteration 617, current x = [ 0.01356503 -0.00000508], gradient = [ 0.00000997 -0.00000019], f = 3.38348546392
72314e-08
```

5. $f_5(x) = (x_1 - 1)^2 + (x_2 - 1)^2 + c(x_1^2 + x_2^2 - 0.25)^2 \ with \ x^{(0)} = (1, -1)^T$

c =1

```
initial point = [ 1 -1], gradient = [  7. -11.]

iteration 1, search direction = [-7. 11.], step length = 0.125, x = [0.125 0.375]
iteration 2, search direction = [1.796875 1.390625], step length = 0.25, x = [0.57421875 0.72265625]
iteration 3, search direction = [-0.5310626  -1.18535089], step length = 0.125, x = [0.50783592 0.57448739]
iteration 4, search direction = [0.29786991 0.07447204], step length = 0.125, x = [0.54506966 0.58379639]
iteration 5, search direction = [ 0.0640888  -0.07345603], step length = 0.25, x = [0.56109186 0.56543239]
iteration 6, search direction = [ 0.01477202 -0.00058542], step length = 0.25, x = [0.56478487 0.56528603]
iteration 7, search direction = [-0.00731374 -0.00909495], step length = 0.125, x = [0.56387065 0.56414916]
iteration 8, search direction = [0.00115887 0.00017157], step length = 0.125, x = [0.56401551 0.56417061]
iteration 9, search direction = [ 0.00022217 -0.00032777], step length = 0.25, x = [0.56407105 0.56408867]
iteration 10, search direction = [0.00009246 0.00003   ], step length = 0.125, x = [0.56408261 0.56409242]
iteration 11, search direction = [ 0.00001252 -0.00002226], step length = 0.25, x = [0.56408574 0.56408685]

stop at iteration 11, current x = [0.56408574 0.56408685], gradient = [-0.00000762 -0.00000367], f = 0.52933619558020
93
```

c = 10

```
initial point = [ 1 -1], gradient = [ 70. -74.]

iteration 1, search direction = [-70.  74.], step length = 0.0078125, x = [ 0.453125 -0.421875]
iteration 2, search direction = [-1.32232666  5.09320068], step length = 0.125, x = [0.28783417 0.21477509]
iteration 3, search direction = [2.81771562 2.61016016], step length = 0.0625, x = [0.46394139 0.3779101 ]
iteration 4, search direction = [-0.9331796  -0.38926338], step length = 0.03125, x = [0.43477953 0.36574562]
iteration 5, search direction = [-0.1356909   0.20341224], step length = 0.25, x = [0.40085681 0.41659868]
iteration 6, search direction = [-0.1524509  -0.23697885], step length = 0.03125, x = [0.39609272 0.40919309]
iteration 7, search direction = [ 0.03017671 -0.03497323], step length = 0.25, x = [0.40363689 0.40044978]
iteration 8, search direction = [0.00954105 0.0252577 ], step length = 0.03125, x = [0.40393505 0.40123908]
iteration 9, search direction = [-0.00604371  0.00734515], step length = 0.25, x = [0.40242412 0.40307537]
iteration 10, search direction = [-0.00270279 -0.00594378], step length = 0.03125, x = [0.40233966 0.40288963]
    ......
    ......
iteration 12, search direction = [0.00064954 0.00131034], step length = 0.03125, x = [0.40266514 0.40255271]
iteration 13, search direction = [-0.0002455   0.00031298], step length = 0.25, x = [0.40260376 0.40263096]
iteration 14, search direction = [-0.00015928 -0.00029438], step length = 0.03125, x = [0.40259879 0.40262176]
iteration 15, search direction = [ 0.00004929 -0.00006483], step length = 0.25, x = [0.40261111 0.40260555]
iteration 16, search direction = [0.00003847 0.00006607], step length = 0.03125, x = [0.40261231 0.40260761]

stop at iteration 16, current x = [0.40261231 0.40260761], gradient = [ 0.00000987 -0.00001345], f = 0.76879062761683
07
```

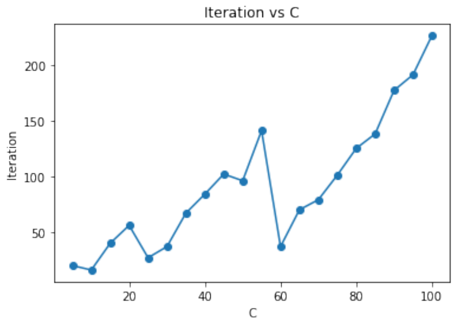
c = 100

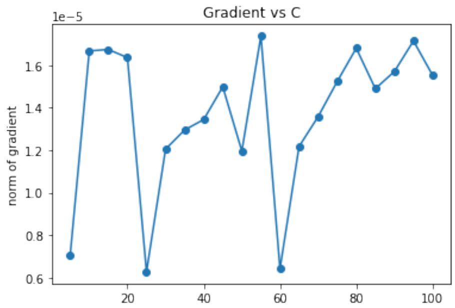
```
initial point = [ 1 -1], gradient = [ 700. -704.]

iteration 1, search direction = [-700.  704.], step length = 0.0009765625, x = [ 0.31640625 -0.3125    ]
iteration 2, search direction = [ 7.97765255 -3.90385437], step length = 0.0078125, x = [ 0.37873166 -0.34299886]
iteration 3, search direction = [-0.43689437  4.20697682], step length = 0.0078125, x = [ 0.37531842 -0.31013186]
iteration 4, search direction = [3.1941601  1.01324565], step length = 0.015625, x = [ 0.42522717 -0.29429989]
iteration 5, search direction = [-1.81523636  4.64052656], step length = 0.0078125, x = [ 0.41104564 -0.25804578]
iteration 6, search direction = [3.55438672 1.02418882], step length = 0.015625, x = [ 0.46658293 -0.24204283]
iteration 7, search direction = [-3.83870021  5.02886242], step length = 0.0078125, x = [ 0.43659309 -0.20275484]
iteration 8, search direction = [4.31864987 0.92321363], step length = 0.0078125, x = [ 0.47033254 -0.19554223]
iteration 9, search direction = [-0.71842115  3.13019215], step length = 0.015625, x = [ 0.45910721 -0.14663298]
iteration 10, search direction = [4.33581607 1.25397017], step length = 0.0078125, x = [ 0.49298077 -0.13683634]
    ......
    ......
iteration 222, search direction = [0.00000264 0.00002604], step length = 0.0078125, x = [0.35979159 0.35978756]
iteration 223, search direction = [-0.00002068  0.0000017 ], step length = 0.015625, x = [0.35979127 0.35978759]
iteration 224, search direction = [0.00001182 0.00003226], step length = 0.0078125, x = [0.35979136 0.35978784]
iteration 225, search direction = [-0.00002436 -0.00000481], step length = 0.0078125, x = [0.35979117 0.3597878 ]
iteration 226, search direction = [0.0000003 0.000019 ], step length = 0.0078125, x = [0.35979117 0.35978795]

stop at iteration 226, current x = [0.35979117 0.35978795], gradient = [ 0.00001533 -0.00000256], f = 0.8276545736982
954
```

Comment on how larger c affects the performance of the algorithm.

c	Iteration	$x^{(k)}$	Gradient at $x^{(k)}$
1	11	[0.56408574 0.56408685]	[-0.00000762, -0.00000367]
10	16	[0.40261231 0.40260761]	[0.00000987, -0.00001345]
100	226	[0.35979117 0.35978795]	[0.00001533, -0.00000256]

The larger the c, more iterations are required to converge to optimal solution most case. Use the norm of gradients as metric of performance. The larger the norm of gradient, the final gradient is more far away from 0. The plot in the right side illustrates the value of c won't affect the performance too much.





(e) Are your computational results consistent with the theory of the gradient-based methods?

Under the assumption that $f(x)$ is bounded below and the gradient of $f(x)$ is Lipchitz continuous over R^n. The computational results are consistent with the theory of the gradient-based methods with backtracking. For functions in part d, show

$$\nabla \mathrm{f}\left(x^{(k)}\right) \rightarrow 0 \ as \ k \rightarrow \infty$$

1. $f_1(x) = x_1^2 + x_2^2 + x_3^2 \ with \ x^{(0)} = (1,1,1)^T$

   $k = 1$:

   $$\nabla \mathrm{f}_1\left(x^{(k)}\right) = 0 \ , x = x^* = [0,0,0]$$

   $$f_1(x^*) = 0 \ \Rightarrow x^* \text{ is global minimum since } f_1(x) \geq 0$$

2. $f_2(x) = x_1^2 + 2x_2^2 - 2x_1x_2 - 2x_2 \ with \ x^{(0)} = (0,0)^T$

   $k = 33$:

   $$\nabla \mathrm{f}_2\left(x^{(k)}\right) = [-0.00001526, 0] \ , x = [0.99998474, 0.99999237]$$

   $k \rightarrow \infty$ (if there's no stopping criteria)

   $$x \rightarrow x^* = [1,1], \nabla \mathrm{f}_2\left(x^{(k)}\right) \rightarrow 0$$
   $$H(x) = \begin{bmatrix} 2 & -2 \\ -2 & 4 \end{bmatrix}, \text{Hessian PD} \Rightarrow \mathrm{f}_2(x) \text{ is convex}$$
   $$\Rightarrow x^* \text{ is global minimum}$$

3. $f_3(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 \ with \ x^{(0)} = (-1.2, 1)^T$

   $k = 421$:

   $$\nabla \mathrm{f}_3\left(x^{(k)}\right) = [0.0000036, 0.00000819], x = [1.00000999, 1.00002003]$$

   $k \rightarrow \infty$ (if there's no stopping criteria)

$x \to x^* = [1,1], \nabla \mathrm{f}_2(x^{(k)}) \to 0$
$f_3(x^*) = 0 \;\Rightarrow x^*$ is global minimum since $f_3(x) \geq 0$

4. $f_4(x) = (x_1 + x_2)^4 + x_2^2 \;\; with \;\; x^{(0)} = (2, -3)^T$

$k = 617$:

$\nabla \mathrm{f}_4(x^{(k)}) = [\, 0.00000997, -0.00000019], x = [\, 0.01356503, -0.00000508]$

$k \to \infty$ (if there's no stopping criteria)

$x \to x^* = [0,0], \nabla \mathrm{f}_4(x^{(k)}) \to 0$
$f_4(x^*) = 0 \;\Rightarrow x^*$ is global minimum since $f_4(x) \geq 0$

5. $f_5(x) = (x_1 - 1)^2 + (x_2 - 1)^2 + c(x_1^2 + x_2^2 - 0.25)^2 \;\; with \;\; x^{(0)} = (1, -1)^T$

c	Iteration	$\boldsymbol{x}^{(\mathbf{k})}$	Gradient at $\boldsymbol{x}^{(\mathbf{k})}$
1	11	[0.56408574 0.56408685]	[-0.00000762, -0.00000367]
10	16	[0.40261231 0.40260761]	[0.00000987, -0.00001345]
100	226	[0.35979117 0.35978795]	[0.00001533, -0.00000256]

$k \to \infty$ (if there's no stopping criteria)
$\nabla \mathrm{f}_5(x^{(k)}) \to 0$