CSE 543t

HW1

Wenzhen Zhu (445518)

1.1.4 Use optimality conditions to show that $\forall x>0$, we have

$$\frac{1}{x} + x \geqslant 2$$

let $f(x) = \frac{1}{x} + x$

$$f'(x) = -x^{-2} + 1 = -\frac{1}{x^2} + 1$$

Since $f''(x) = 2x^{-3} = \frac{2}{x^3} > 0$ when $x>0$, $f(x)$ is strictly convex on $(0, +\infty)$

By Prop 1.1.2 convex cost function, $\exists$ at most 1 global minimum of f

Since domain of $f(x)$ is open, then $f'(x^*) = 0$ is a necessary and sufficient condition for x^* be a global minimum.

Thus, solve for $f'(x) = -\frac{1}{x^2} + 1 = 0$

$$x^2 = 1 \Rightarrow x_{min} = 1.$$

$$f(x) = \frac{1}{x} + x \geqslant 1 + 1 = 2 \quad \forall x \in \mathbb{R}^+ \quad \blacksquare$$

Question 2

In[4]:=
```
f[{x_, y_}] := 3 x^2 + y^4
f[x_, y_] := 3 x^2 + y^4
```

In[6]:=
```
df[x_, y_] = Grad[f[x, y], {x, y}]
```

Out[6]= $\{6x, 4y^3\}$

(a)

In[7]:=
```
s = 1; σ = 0.1; β = 0.5;
```

In[8]:=
```
x0 = 1; y0 = -2;
```

In[9]:=
```
df[x0, y0]
```

Out[9]= {6, -32}

In[10]:=
```
d0 = -1 * df[x0, y0]
```

Out[10]= {-6, 32}

In[11]:=
```
α0 = s
f[x0, y0] - f[{x0, y0} + α0 * d0]
-σ α0 {6, -32}.{-6, 32}
```

Out[11]= 1

Out[12]= -810 056

Out[13]= 106.

In[14]:=
```
α1 = β s
f[x0, y0] - f[{x0, y0} + α1 * d0]
-σ α1 {6, -32}.{-6, 32}
```

Out[14]= 0.5

Out[15]= -38 409.

Out[16]= 53.

In[17]:=
```
α2 = β^2 s
f[x0, y0] - f[{x0, y0} + α2 * d0]
-σ α2 {6, -32}.{-6, 32}
```

Out[17]= 0.25

Out[18]= -1277.75

Out[19]= 26.5

In[20]:=
```
α3 = β^3 s
f[x0, y0] - f[{x0, y0} + α3 * d0]
-σ α3 {6, -32}.{-6, 32}
```
Out[20]= 0.125

Out[21]= 2.8125

Out[22]= 13.25

In[23]:=
```
α4 = β^4 s
f[x0, y0] - f[{x0, y0} + α4 * d0]
-σ α4 {6, -32}.{-6, 32}
```
Out[23]= 0.0625

Out[24]= 17.8281

Out[25]= 6.625

Thus, step-size = $\alpha 4$ = 0.0625

In[26]:=
```
{x1, y1} = {x0, y0} - α4 df[x0, y0]
```
Out[26]= {0.625, 0.}

In[27]:=
```
f[0.625, 0]
```
Out[27]= 1.17188

(b) Repeat (a) using different parameters

In[99]:=
```
s = 1; σ = 0.1; β = 0.1;
```
In[113]:=
```
d0 = -1 * df[x0, y0]
```
Out[113]= {-6, 32}

In[114]:=
```
α1 = β s
f[x0, y0] - f[{x0, y0} + α1 * d0]
-σ α1 {6, -32}.{-6, 32}
```
Out[114]= 0.1

Out[115]= 16.4464

Out[116]= 10.6

Thus, α = 0.1

In[119]:=
```
{x1, y1} = {x0, y0} - α1 df[x0, y0]
```
Out[119]= {0.4, 1.2}

In[120]:=
```
f[x1, y1]
```
Out[120]= 2.5536

(c) Newton's method

Hessian Matrix:

In[28]:=
```
(hessian[x_, y_] = D[f[x, y], {{x, y}, 2}]) // MatrixForm
```

Out[28]//MatrixForm=

$$\begin{pmatrix} 6 & 0 \\ 0 & 12\,y^2 \end{pmatrix}$$

In[29]:=
$$\left(\texttt{hessianInverse[x\_, y\_] = Inverse}\left[\begin{pmatrix} 6 & 0 \\ 0 & 12\,y^2 \end{pmatrix}\right]\right) \texttt{ // MatrixForm}$$

Out[29]//MatrixForm=

$$\begin{pmatrix} \frac{1}{6} & 0 \\ 0 & \frac{1}{12\,y^2} \end{pmatrix}$$

In[30]:=
```
dk[x_, y_] = -hessianInverse[x, y].df[x, y]
```

Out[30]= $\left\{-x, -\frac{y}{3}\right\}$

In[31]:=
```
d0 = dk[x0, y0]
```

Out[31]= $\left\{-1, \frac{2}{3}\right\}$

Armijo rule

In[32]:=
```
s = 1; σ = 0.1; β = 0.5;
```

In[33]:=
```
α = s
```

Out[33]= 1

In[34]:=
```
f[x0, y0] - f[{x0, y0} + α * d0] // N
```

Out[34]= 15.8395

In[35]:=
```
-σ α df[x0, y0].d0
```

Out[35]= 2.73333

So $\alpha = 1$

In[36]:=
```
{x1, y1} = {x0, y0} + α d0
```

Out[36]= $\left\{0, -\frac{4}{3}\right\}$

In[37]:=
```
f[x1, y1] // N
```

Out[37]= 3.16049

We can observe that
1. Newton's method converge much faster than steepest descent.
Since with same α and β, newton method converge for one single step which took steepest descent 5 steps to do so.

2. Newton's method is more expensive since it requires to compute Hessian matrix and it's inverse,

while steepest decent only needs 1st derivative.

Question 3

L-BFGS stands for Limited-memory BFGS.

BFGS is a algorithm in the family of quasi-Newton methods for optimization of smooth functions. It is a line search method. The search direction is of type d:

$$d = -H_k \nabla f(x_k)$$

where H_k is the k-th approximation to the inverse Hessian of f.

This k - th step approximation for the inverse Hessian is calculated via the BFGS formula is as followings:

$$H_{k+1} = \left(I - \delta x_k\, y_k^T \,/\, y_k^T\, \delta x_k\right) H_k \left(I - y_k\, \delta x_k^T \,/\, y_k^T\, \delta x_k\right) + \frac{\delta x_k\, \delta x_k^T}{y_k^T\, \delta x_k}$$

Tasks	Solution Time	Solution Quality
dqrtic.mod	0.113981	L-BFGS-B: CONVERGENCE: NORM OF PROJECTED GRADIENT <= PGTO 0 iterations, 140720308486145 function and gradient evaluations. Objective = 1.998169937e-14. Projected gradient maxnorm = 9.06316e-13.
eigenbls.mod	0.12198	L-BFGS-B: CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMC 328 iterations, 140728898421077 function and gradient evaluations. Objective = 5.19644528e-13. Projected gradient maxnorm = 4.27133e-07.
freuroth.mod	0.225965	L-BFGS-B: CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMC 1 iterations, 140733193388035 function and gradient evaluations. Objective = 608159.189. Projected gradient maxnorm = 3.87825e-05.

Question 4

Rastrigin

Out[38]= http://tracer.lcc.uma.es/problems/rastrigin/rastrigin.html

```
param n  := 10;
param pi := 4*atan(1);

var x{1..n};

minimize f:
    10*n+sum {i in 1..n} (x[i]^2-10*cos(2*pi*x[i]));

subject to bounds {i in 1..n}:
    -5.12 <= x[i] <= 5.12;

solve; display f; display x;
display _solve_time;
display _ampl_time;
display _total_solve_time;
```

Tasks	n	Solution Time	Solution Quality
Rastrigin	10	0	L-BFGS-B: CONVERGENCE: NORM OF PROJECTED GRADIENT <= PGTO 0 iterations, 140728898420737 function and gradient evaluations. Objective = 0. Projected gradient maxnorm = 0.
	20	0.000999	L-BFGS-B: CONVERGENCE: NORM OF PROJECTED GRADIENT <= PGTO 0 iterations, 140728898420737 function and gradient evaluations. Objective = 0. Projected gradient maxnorm = 0.
	50	0.000999	L-BFGS-B: CONVERGENCE: NORM OF PROJECTED GRADIENT <= PGTO 0 iterations, 140728898420737 function and gradient evaluations. Objective = 0. Projected gradient maxnorm = 0.
	100	0.000999	L-BFGS-B: CONVERGENCE: NORM OF PROJECTED GRADIENT <= PGTO 0 iterations, 140733193388033 function and gradient evaluations. Objective = 0. Projected gradient maxnorm = 0.
	1000	0.002998	L-BFGS-B: CONVERGENCE: NORM OF PROJECTED GRADIENT <= PGTO 0 iterations, 140720308486145 function and gradient evaluations. Objective = 0. Projected gradient maxnorm = 0.
	10000	0.013997	L-BFGS-B: CONVERGENCE: NORM OF PROJECTED GRADIENT <= PGTO 0 iterations, 140733193388033 function and gradient evaluations. Objective = 0. Projected gradient maxnorm = 0.