Linear and Convex Optimization Homework 07

Qiu Yihang, 2021/11/15-11/18

0. Preparation

Complete gd.py. The completed code (with `gd_const_ss` and `gd_armijo` function) is enclosed in the zip file.

1.(a) ***Solution:***

$$\nabla f = (e^{x_1+3x_2-0.1} + e^{x_1-3x_2-0.1} - e^{-x_1-0.1}, 3e^{x_1+3x_2-0.1} - 3e^{x_1-3x_2-0.1}) = \boldsymbol{0}$$

$$\Leftrightarrow \begin{cases} e^{x_1+3x_2-0.1} + e^{x_1-3x_2-0.1} - e^{-x_1-0.1} = 0 \\ x_1 + 3x_2 - 0.1 = x_1 - 3x_2 - 0.1 \end{cases} \Leftrightarrow \begin{cases} 2e^{x_1} = e^{-x_1} \\ x_2 = 0 \end{cases} \Leftrightarrow \begin{cases} x_1 = -\frac{1}{2}\ln 2 \\ x_2 = 0 \end{cases}$$

Thus, the optimal solution $\boldsymbol{x}^\star = \left(-\frac{1}{2}\ln 2, 0\right)^T$ and the optimal value is $f(\boldsymbol{x}^\star) = 2\sqrt{2}e^{-0.1}$. ■

(b) ***Solution:***

I made some adjustments to the given p1.py and utils.py. The code is enclosed in the zip file.

Use gradient descent with backtracking line search to solve the problem numerically.

The solution, the number of iterations in the outer loop and the total number of iterations in the inner loop is given below.

```
In [1]: runfile('D:/Textbooks/2021-2022-1/Linear and Convex
Optimization/hw7/p1.py', wdir='D:/Textbooks/2021-2022-1/Linear and
Convex Optimization/hw7')

gradient descent with Armijo
  number of iterations in outer loop: 28
  total number of iterations in inner loop: 151
  solution: [-3.46574284e-01  3.04072749e-07]
  value: 2.5592666966593645
```

Fig.01. Results of Program 1(b)

The visualization of the trajectory of x_k and the change of error $f(x_k) - f(x^\star)$ and the step size t_k is as follows.

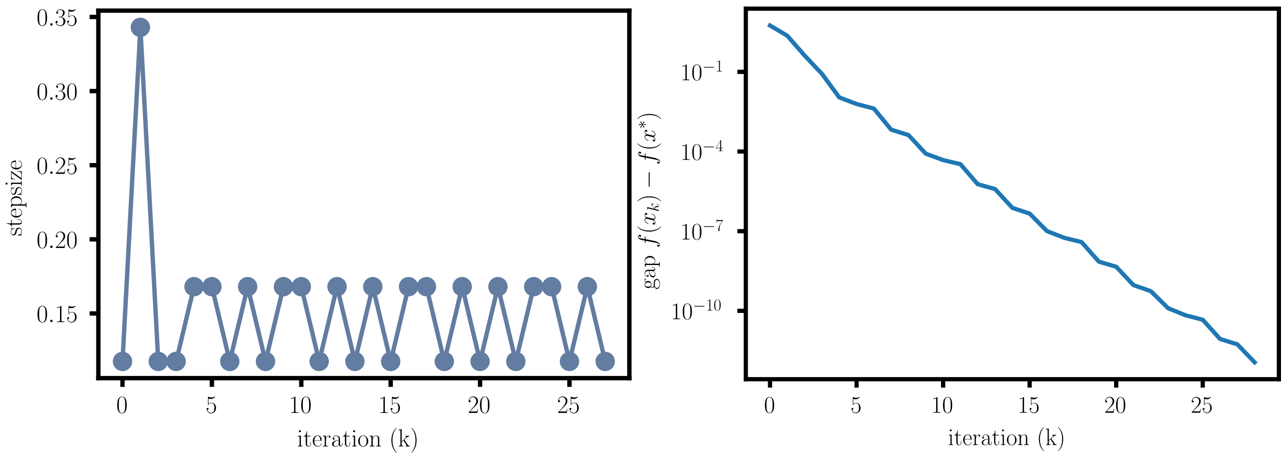

Fig.02. Change of Step Size and The Divergence of Gap

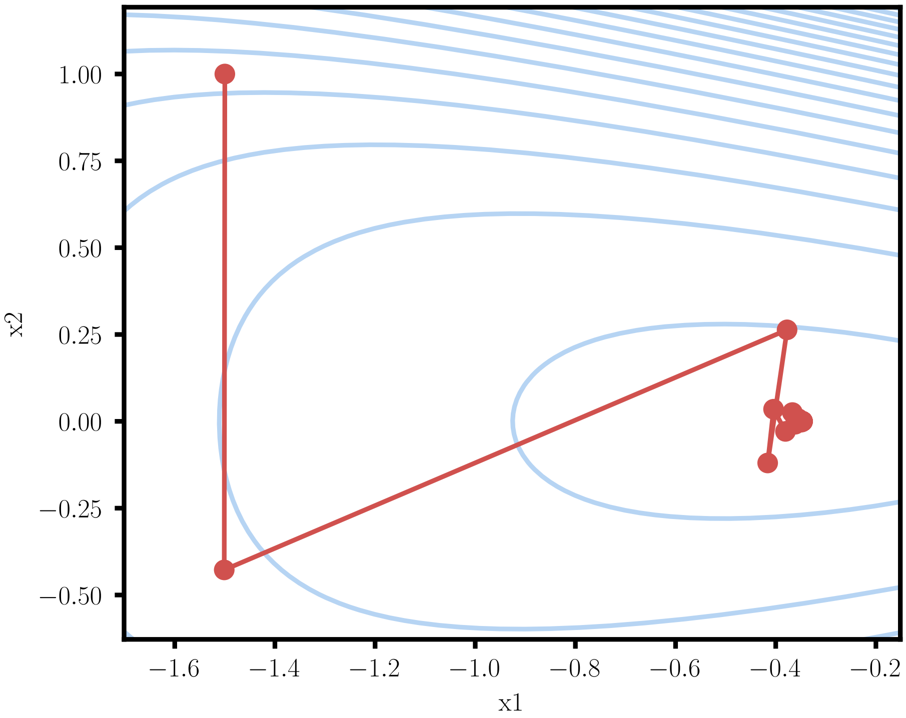

Fig.03. Trajectory of x_k Produced by Gradient Descent with Backtracking Line Search

(c) ***Solution:***

Use gradient descent with constant step size to solve the problem numerically.

The solution and the number of iterations in the case where step size is 0.1 or 0.01 respectively is as follows.

```
gradient descent with constant stepsize  0.1
  number of iterations: 44
  solution: [-3.46576607e-01  3.21465960e-18]
  value: 2.559266696669859

gradient descent with constant stepsize  0.01
  number of iterations: 489
  solution: [-3.46577419e-01  8.65140907e-18]
  value: 2.559266696676969
```

Fig.04. Results of Program 1(c)

Compare the result in part (c) and part (b). We find that Gradient Descent with backtracking line search converges faster (the number of iterations in the outer loop is smaller than Gradient Descent with constant step size). Meanwhile, the numerical errors between solution calculated by Gradient Descent with constant step size and the solution calculated by hand in part (a) is smaller than Gradient Descent with backtracking line search (the former one being 10^{-7}, the latter one being 10^{-18}). ■

(d) ***Solution:***

Change the initial point to $x_0 = (1.5,1)^T$. The solution, the number of iterations in the outer loop and the total number of iterations in the inner loop is given below.

```
In [2]: runfile('D:/Textbooks/2021-2022-1/Linear and Convex
Optimization/hw7/p1.py', wdir='D:/Textbooks/2021-2022-1/Linear and
Convex Optimization/hw7')
Reloaded modules: gd, utils

gradient descent with Armijo
  number of iterations in outer loop: 32
  total number of iterations in inner loop: 197
  solution: [-3.4657238e-01  6.5447655e-07]
  value: 2.5592666966625575
```

Fig.05. Results of Program 1(d)

The visualization of the trajectory of x_k and the change of error $f(x_k) - f(x^{\star})$ and the step size t_k is as follows.

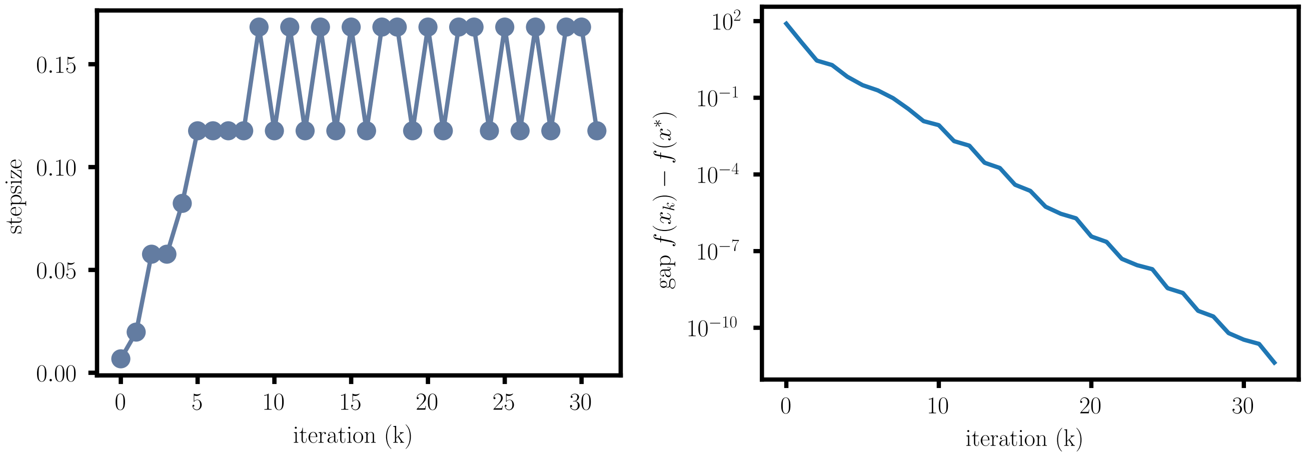

Fig.06. Change of Step Size and The Divergence of Gap

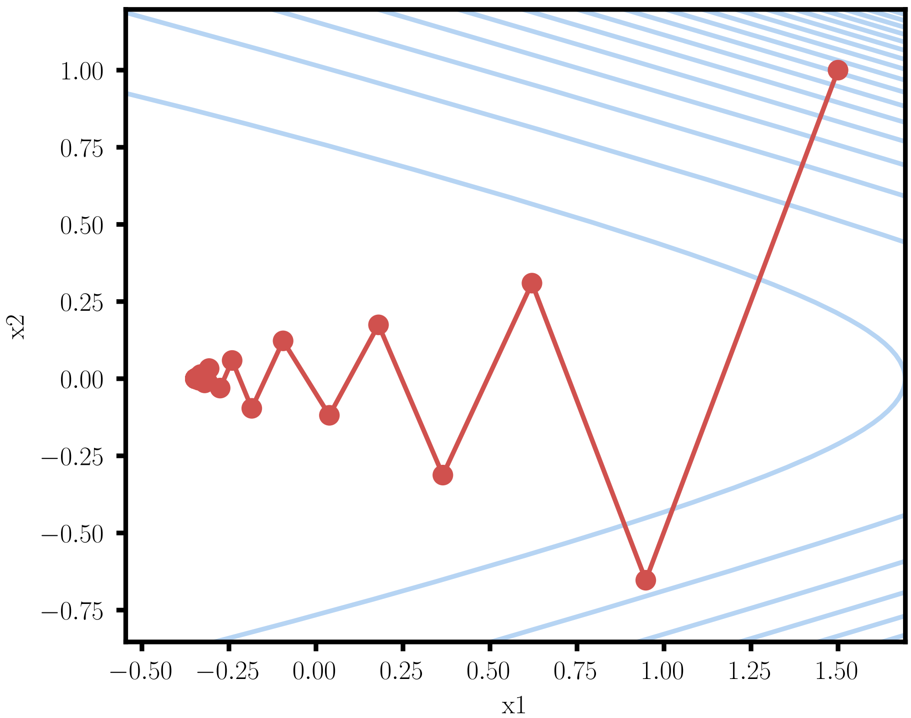

Fig.07. Trajectory of x_k Produced by Gradient Descent with Backtracking Line Search

(e) ***Solution:***

Change the initial point to $x_0 = (1.5,1)^T$.

The solution and the number of iterations when step size is 0.005 respectively is as follows.

```
gradient descent with constant stepsize  0.005
  number of iterations: 984
  solution: [-3.46569713e-01 -7.62280416e-18]
  value: 2.559266696677449
```

Fig.08. Results of Program 1(e)

Compare the result of 1(d) and 1(e). We find that in this case, Gradient Descent with backtracking line search **converges** much faster than Gradient Descent with constant step size.

Set step size to 0.1 or 0.01. The results are as follows.

```
gradient descent with constant stepsize  0.1
  number of iterations: 2
  solution: [-3.35077129e+26  1.00523139e+27]
  value: inf

gradient descent with constant stepsize  0.01
  number of iterations: 4
  solution: [-1.82571291e+32  5.47713874e+32]
  value: inf
```

Fig.09. Results of Program 1(e)

We can see that in these cases, Gradient Descent wit constant step size **diverges**. ■

2.(a) ***Proof:***

$$\begin{aligned}\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star\| &= \|\boldsymbol{x}_k - t(\nabla f(\boldsymbol{x}_k) + \varepsilon_k) - \boldsymbol{x}^\star\| = \|\boldsymbol{x}_k - t\nabla f(\boldsymbol{x}_k) - t\varepsilon_k - \boldsymbol{x}^\star\| \\ &= \|\boldsymbol{x}_k - t\nabla f(\boldsymbol{x}_k) - \boldsymbol{x}^\star\| + t\|\varepsilon_k\| \le \|\tilde{\boldsymbol{x}}_{k+1} - \boldsymbol{x}^\star\| + tE\end{aligned}$$

Qed. ■

(b) ***Proof:***

$$\begin{aligned}\|\tilde{\boldsymbol{x}}_{k+1} - \boldsymbol{x}^\star\|^2 &= \|\boldsymbol{x}_k - t\nabla f(\boldsymbol{x}_k) - \boldsymbol{x}^\star\|^2 \\ &= \|\boldsymbol{x}_k - \boldsymbol{x}^\star\|^2 + t^2\|\nabla f(\boldsymbol{x}_k)\|^2 + 2t\nabla f(\boldsymbol{x}_k)^T(\boldsymbol{x}^\star - \boldsymbol{x}_k)\end{aligned}$$

By L-smoothness of f, $\|\nabla f(\boldsymbol{x}_k)\|^2 \le \frac{2}{t}\big(f(\boldsymbol{x}_k) - f(\tilde{\boldsymbol{x}}_{k+1})\big)$.

By m-strong convexity of f, $\nabla f(\boldsymbol{x}_k)^T(\boldsymbol{x}^\star - \boldsymbol{x}_k) \le f(\boldsymbol{x}^\star) - f(\boldsymbol{x}_k) - \frac{m}{2}\|\boldsymbol{x}_k - \boldsymbol{x}^\star\|$.

Since $f(\boldsymbol{x}^\star) < f(\tilde{\boldsymbol{x}}_{k+1})$,

$$\|\tilde{\boldsymbol{x}}_{k+1} - \boldsymbol{x}^\star\|^2 \le (1 - mt)\|\boldsymbol{x}_k - \boldsymbol{x}^\star\|^2 + 2t\big(f(\boldsymbol{x}^\star) - f(\tilde{\boldsymbol{x}}_{k+1})\big) \le (1 - mt)\|\boldsymbol{x}_k - \boldsymbol{x}^\star\|^2$$

Since f is L-smooth and m-strongly convex, $L \ge m \ge 0$. Thus, $mt \le \frac{m}{L} \le 1$, **i.e.** $1 - mt \ge 0$.

Thus, $\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star\| \le \|\tilde{\boldsymbol{x}}_{k+1} - \boldsymbol{x}^\star\| + tE \le \sqrt{1 - mt}\|\boldsymbol{x}_k - \boldsymbol{x}^\star\| + tE = q\|\boldsymbol{x}_k - \boldsymbol{x}^\star\| + tE$.

Qed. ■

(c) ***Proof:***

BASE STEP. $k = 0$. $\|\boldsymbol{x}_k - \boldsymbol{x}^\star\| = \|\boldsymbol{x}_0 - \boldsymbol{x}^\star\| + 0$. Obvious.

INDUCTIVE STEP.

Suppose when $k = n$, $\|\boldsymbol{x}_k - \boldsymbol{x}^\star\| \le q^k \|\boldsymbol{x}_0 - \boldsymbol{x}^\star\| + \frac{1-q^k}{1-q} tE$ holds.

When $k = n + 1$, from (b), we know

$$\|\boldsymbol{x}_{n+1} - \boldsymbol{x}^\star\| \le q\|\boldsymbol{x}_n - \boldsymbol{x}^\star\| + tE \le q \cdot q^n \|\boldsymbol{x}_0 - \boldsymbol{x}^\star\| + \left(q\frac{1-q^n}{1-q} + 1\right) tE$$
$$\le q^{n+1}\|\boldsymbol{x}_0 - \boldsymbol{x}^\star\| + \frac{1-q^{n+1}}{1-q} tE$$

i.e. $\|\boldsymbol{x}_k - \boldsymbol{x}^\star\| \le q^k \|\boldsymbol{x}_0 - \boldsymbol{x}^\star\| + \frac{1-q^k}{1-q} tE$ still holds.

In conclusion,

$$\|\boldsymbol{x}_k - \boldsymbol{x}^\star\| \le q^k \|\boldsymbol{x}_0 - \boldsymbol{x}^\star\| + \frac{1-q^k}{1-q} tE.$$

Qed. ■

Note: In fact, the proposition can be proved without induction, considering

$$\|\boldsymbol{x}_k - \boldsymbol{x}^\star\| \le q\|\boldsymbol{x}_{k-1} - \boldsymbol{x}^\star\| + tE \Leftrightarrow \|\boldsymbol{x}_k - \boldsymbol{x}^\star\| - \frac{t}{1-q}E \le q\left(\|\boldsymbol{x}_{k-1} - \boldsymbol{x}^\star\| - \frac{t}{1-q}E\right)$$

(d) ***Proof:***

$$\limsup_{k\to\infty} \|\boldsymbol{x}_k - \boldsymbol{x}^\star\| = \limsup_{k\to\infty}\left(q^k\|\boldsymbol{x}_0 - \boldsymbol{x}^\star\| + \frac{1-q^k}{1-q} tE\right)$$
$$\le \lim_{k\to\infty} q^k \|\boldsymbol{x}_0 - \boldsymbol{x}^\star\| + \lim_{k\to\infty}\frac{1-q^k}{1-q} tE = 0 + \frac{tE}{1-q} = \frac{tE}{1-q}$$

(Since $0 \le q \le 1$)

Meanwhile,

$$\frac{t}{1-q} \le \frac{1}{L\left(1 - \sqrt{1 - \frac{m}{L}}\right)} \le \frac{1}{L\left(1 - \left(1 - \frac{1}{2}\frac{m}{L}\right)\right)} = \frac{2}{m}$$

Thus,

$$\limsup_{k\to\infty} \|\boldsymbol{x}_k - \boldsymbol{x}^\star\| \le \frac{tE}{1-q} \le \frac{2E}{m}$$

Qed. ■