Advanced Optimization

Homework 1.

Ali Izadi
810199102

سوال ۱	2
سوال ۲	4
سوال ۳	5
سوال ۴ (پیاده سازی)	8
سوال ۵ (پیاده سازی)	14

سوال ۱

برای این که جهت p یک جهت کاهشی باشد باید ضرب داخلی جهت در بردار گرادیان منفی باشد. یعنی:

$$p^T \nabla f\left(x^T\right) \ < \ 0$$

در نتیجه گرادیان را محاسبه میکنیم:

$$\nabla f\left(x^T\right) \ = \ \begin{bmatrix} 2\left(x_1 + x_2^2\right) \\ 4\left(x_1 + x_2^2\right)x_2 \end{bmatrix}$$

با جایگذاری $x^T = (1, \, 0) \, , \, p^T = (-1, \, 1)$ خواهیم داشت:

$$\begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} \ = \ -2 \ < 0$$

در نتیجه ضرب داخلی جهت در گرادیان منفی به دست آمد و در نتیجه جهت p یک جهت کاهشی است.

در صورت سوال بیان شده برای هر iteration, t در صورتی که برای محاسبه طول پله در هر مرحله تابع متغیری از آلفا است و x^t و p^t مقدار دارند .

بنابراین minimizerتابع شده به ازای مقادیر $x^T = (1, \, 0) \, , \, p^T = (-1, \, 1)$ فعلی را به دست خواهیم آورد.

ابتدا x^{t+1} را به دست می آوریم:

$$x^{t+1} = x^t + \alpha\, p^T$$
$$= \begin{bmatrix} x_1^t \\ x_2^t \end{bmatrix} + \alpha \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$
$$= \begin{bmatrix} 1-\alpha \\ 0+\alpha \end{bmatrix} = \begin{bmatrix} 1-\alpha \\ \alpha \end{bmatrix}$$

در نتیجه $f\left(x^{t+1}\right)$ برابر است با:

$$f\left(x^{t+1}\right) = \left(1-\alpha+\alpha^2\right)^2$$

که هدف پیدا کردن آلفایی است که تابع را minimum کند.

بنابراین مشتق آن به ازای آلفا را برابر صفر قرار میدهیم.

$$\min\ f\left(x^{t+1}\right)$$
$$\alpha > 0$$
$$\implies \frac{df}{d\alpha} = 0$$
$$\frac{df}{d\alpha} = 2\left(1-\alpha+\alpha^2\right)(-1+2\alpha) = 0$$

$$\implies (-1+2\alpha) = 0 \implies \alpha = \frac{1}{2}$$

دو ریشه دیگر معادله بالا عدد مختلط اند. هم چنین شرط لازم برای این که $\alpha = 0.5$ مینیمم باشد باید مشتق دوم تابع به ازای این مقدار نیز مثبت باشد پس داریم:

$$\frac{d^2 f}{d\alpha} = 2\left(6\alpha^2 - 6\alpha + 3\right)$$

$$\frac{d^2 f}{d\alpha}(a=2) = 2\left(\frac{3}{2} - 3 + 3\right) = 3 > 0$$

سوال ۲

ابتدا گرادیان و Hessian تابع F را به دست می آوریم:

$$f(x) = \frac{1}{2}x^TQx - b^Tx$$
$$\nabla f(x) = Qx - b$$
$$\nabla^2 f(x) = Q^T = Q$$

- چون $Q^T > 0$ پس با صفر قرار دادن گرادیان مینیمم تابع به دست می آید.
- هم چنین چون $Q^T > 0$ ماتریس وارون Q وجود خواهد داشت.

در نتیجه خواهیم داشت:

رابطه ۱:

$$\nabla f(x) = Qx - b = 0 \Rightarrow Qx^\star = b \Rightarrow x^\star = Q^{-1}b$$

حال با توجه به گرادیان و Hessian به دست آمده sub-problem نیتون را مینویسیم:

$$\min\ f(x_{k+1})$$
$$\alpha \geqslant 0$$
$$where$$
$$\begin{aligned} x_{k+1} &= x_k - \alpha_k\left(\nabla^2 f(x_k)\right)^{-1}\nabla f(x_k) \\ &= x_k - \alpha_k Q^{-1}(Qx_k - b) \end{aligned}$$

حال اگر $\alpha = 1$ قرار دهیم و با استفاده از به رابطه ۱ خواهیم داشت:

$$\begin{aligned} x_{k+1} &= x_k - Q^{-1}(Qx_k - b) \\ &= x_k - Ix_k + Q^{-1}b \\ &= Q^{-1}b \\ &= x^\star \end{aligned}$$

در نتیجه با $\alpha = 1$ هم مینیمم $f(x_{k+1})$ و هم مینیمم تابع $f(x)$ به دست آمد.

در نتیجه در یک iteration توانستیم با روش نیوتن به جواب بهینه $x^\star$ دست پیدا کنیم.

سوال ۳

برای استفاده از روش نیوتن نیاز به محاسبه گرادیان و وارون Hessian تابع f داریم.

$$f(x) = ||x||^B = \left(\sum_{n=1}^{N} x_n^2\right)^{\frac{B}{2}}$$

ابتدا به محاسبه گرادیان میپردازیم:

$$\frac{\partial f(x)}{x_i} = \frac{B}{2}(2x_i)\left(\left(\sum_{n=1}^{N} x_k^2\right)^{\frac{1}{2}}\right)^{B-2} = \frac{B}{2}(2x_i)||x||^{B-2}$$

$$\Rightarrow$$

$$\nabla f(x) = B||x||^{B-2}x$$

حال به محاسبه Hessian میپردازیم:

$$\nabla^2 f(x) = B(B-2)||x||^{B-4}\, xx^T + B||x||^{B-2}$$

$$= B||x||^{B-2}\left(\frac{B-2}{||x||^2}xx^T + I\right)$$

حال به محاسبه وارون Hessian میپردازیم:

$$\nabla^2 f(x)^{-1} = \frac{1}{B||x||^{B-2}}\left(\frac{B-2}{||x||^2}xx^T + I\right)^{-1}$$

حال رابطه به روزرسانی روش نیوتن را نوشته و با گرادیان و وارون Hessian به دست آمده جایگذاری میکنیم:

$$\begin{aligned}
x_{k+1} &= x_k - \alpha\left(\nabla^2 f(x_k)\right)^{-1}\left(\nabla f(x_k)\right) \\
&= x_k - \alpha \frac{1}{\cancel{B||x_k||^{B-2}}}\left(I + \frac{B-2}{||x_k||^2}x_k x_k^T\right)^{-1} \cancel{B||x_k||^{B-2}}\, x_k \\
&= x_k - \alpha\left(I + \frac{B-2}{||x_k||^2}x_k x_k^T\right)^{-1} x_k \\
Hint: \quad &= x_k - \alpha x_k\left(I + \frac{B-2}{||x_k||^2}x_k^T x_k\right)^{-1} \\
&= x_k - \alpha x_k\left(I + \frac{B-2}{\cancel{||x_k||^2}}\,\cancel{||x_k||^2}\right)^{-1} \\
&= x_k - \alpha x_k(1 + B - 2)^{-1} \\
&= \left(1 - \frac{\alpha}{B-1}\right)x^k
\end{aligned}$$

چون روش pure newton است بنابراین $\alpha = 1$ است و در نهایت داریم.

$$x_{k+1} = \left(1 - \frac{1}{B-1}\right)x_k$$

- به ازای B>2 رابطه:

$$\left(1-\frac{1}{B-1}\right) < 1$$

و در نتیجه طبق رابطه به روزرسانی x به سمت صفر که minimum تابع است همگرا خواهد شد.

- به ازای B=2 در یک مرحله همگرا میشود و x=0 بهینه به دست می آید

- به ازای B=1 نیز چون مخرج کسر صفر میشود الگوریتم پایان نمی‌یابد و دچار خطا میشود.

- به ازای B<2 و B !=1 نیز x به بی نهایت واگرا میشود و minimum به دست نمی‌آید.

سوال ۴ (پیاده سازی)

هدف از این سوال بهینه سازی تابع داده شده با استفاده از روش newton و steepest descent و استفاده از backtracking line search برای محاسبه طول پله است.

ابتدا به محاسبه گرادیان و Hessian تابع داده شده میپردازیم:

$$\nabla f(x) = \begin{bmatrix} 400x_1^3 - 400x_1x_2 + 2x_1 - 2 \\ 200x_2 - 200x_1^2 \end{bmatrix}$$

$$\nabla^2 f(x) = \begin{bmatrix} 1200x_1^2 - 400x_2 + 2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix}$$

- دو روش تنها در جهت P با یکدیگر تفاوت دارند که روش نیوتن از ماتریس Hessian نیز برای محاسبه جهت استفاده میکند. و سوال پرسیده شده در صورت تمرین را متوجه نشدم که چگونه میتوان از ترکیب این دو روش استفاده کرد. در کل میتوان از اطلاعات گرادیان حساب شده در روش steepest descent برای مثلا مقداردهی اولیه steplength یا جاهای دیگر استفاده کرد.

- پیاده سازی line search برای محاسبه طول پله مطابق با الگوریتم backtracking زیر انجام شده است.

Algorithm 3.1 (Backtracking Line Search).
Choose $\bar{\alpha} > 0$, $\rho \in (0, 1)$, $c \in (0, 1)$; Set $\alpha \leftarrow \bar{\alpha}$;
repeat until $f(x_k + \alpha p_k) \le f(x_k) + c\alpha \nabla f_k^T p_k$
$\alpha \leftarrow \rho\alpha$;
end (repeat)
Terminate with $\alpha_k = \alpha$.

مقادیر $c = 10^{-4},\ \rho = 0.5,\ \bar{\alpha} = 1$ در نظر گرفته شده اند

برای انتخاب نقطه اولیه طول پله علاوه بر حالت بالا و انتخاب عدد 1 سه حالت زیر نیز پیاده سازی و تست شده اند.

- طول پله‌ای که تقریب دو بعدی تابع را در مرحله قبل کمینه می‌کند

$$\alpha_0 = \frac{2(f_k - f_{k-1})}{\nabla f_{k-1}^T p_{k-1}}$$

- طول پله با مقدار درجه اول برابر

$$\alpha_0 = \frac{\alpha_{k-1} \nabla f_{k-1}^T p_{k-1}}{\nabla f_k^T p_k}$$

- یا ترکیبی از دو مورد که جواب بهتری می‌دهد

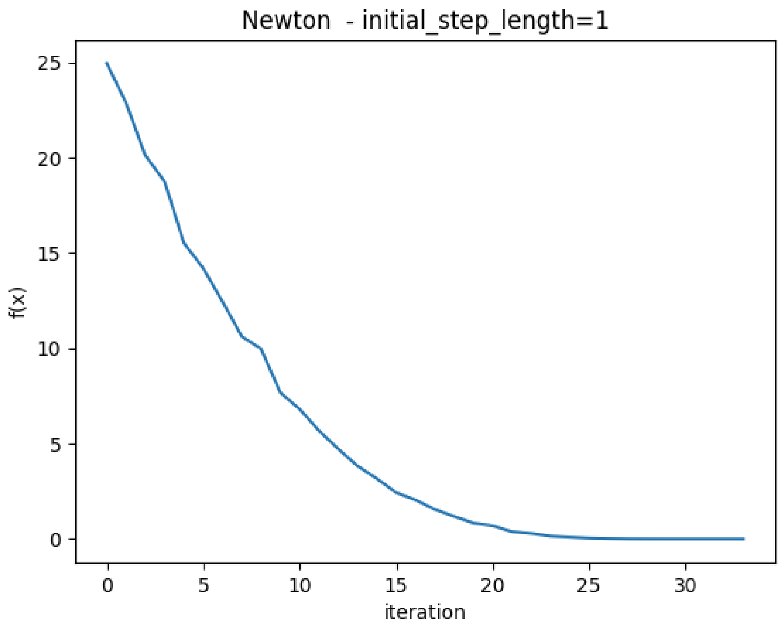
$$\alpha_0 = \frac{2(f_k - f_{k-1})}{\nabla f_k^T p_k}$$

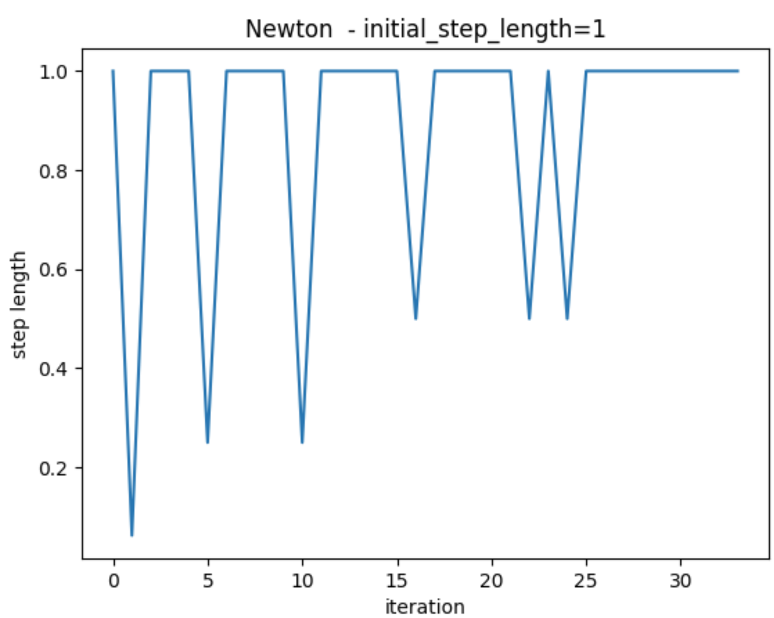
با توجه به تابع داده شده مقدار بهینه $x^\star = (1,\, 1)$ و $f(x^\star) = 0$ است.

۱- ابتدا به بررسی نتایج با استفاده از روش newton میپردازیم:

- ابتدا حالت step-length اولیه ثابت ۱ را در نظر میگیریم.

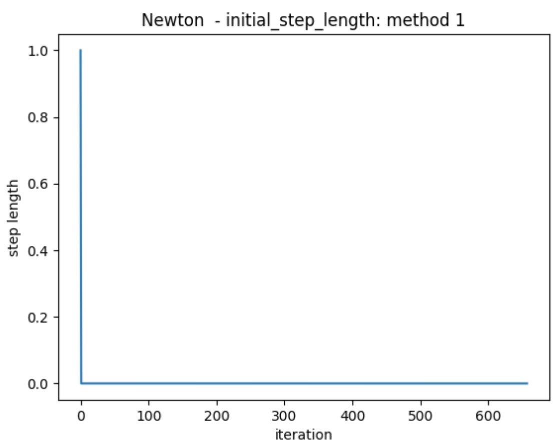
در زیر نمودارهای مقدار تابع (فاصله از نقطه بهینه چون صفر است نیز برابر همین نمودار میشود) و step-length انتخاب شده توسط line search در هر مرحله نمایش داده شده است.

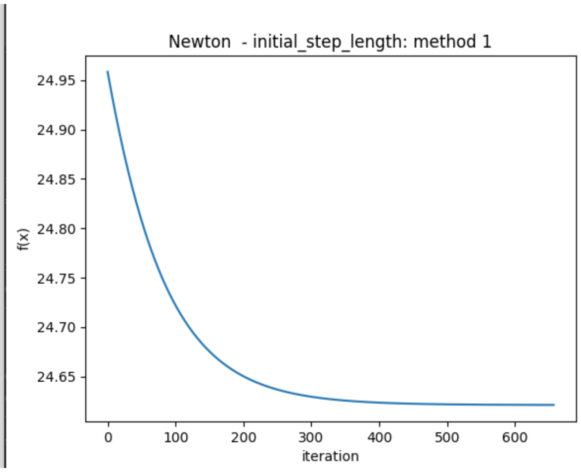


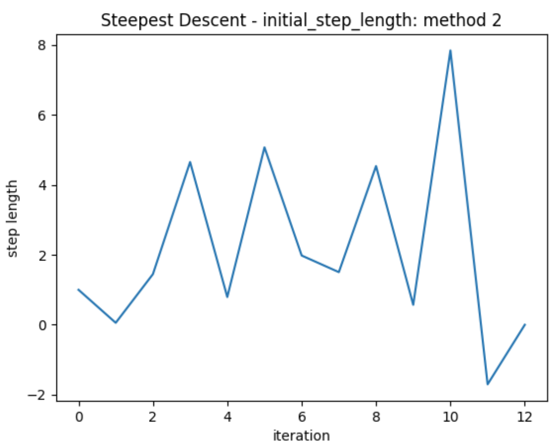


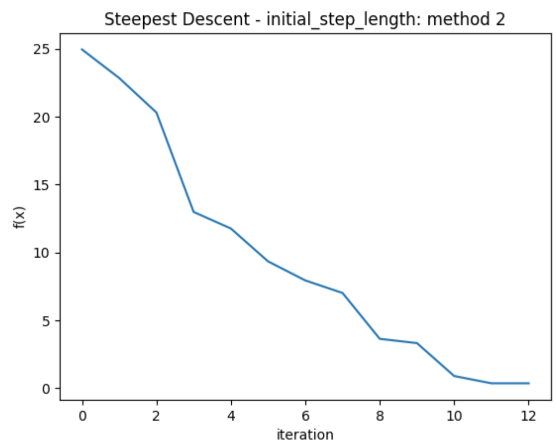
همان طور که مشاهده میشود این روش بعد از ۳۳ iteraation توانسته است همگرا شود.

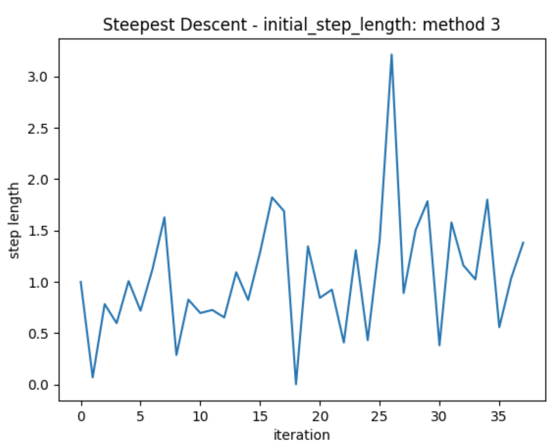
نمودارهای سه استراتژی انتخاب step-length اولیه مطابق ترتیب گفته شده در زیر آورده شده است.

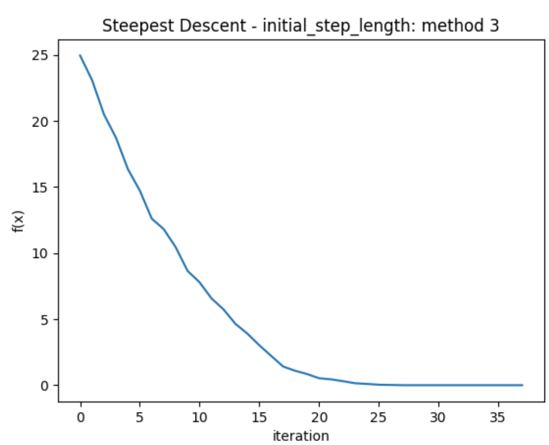










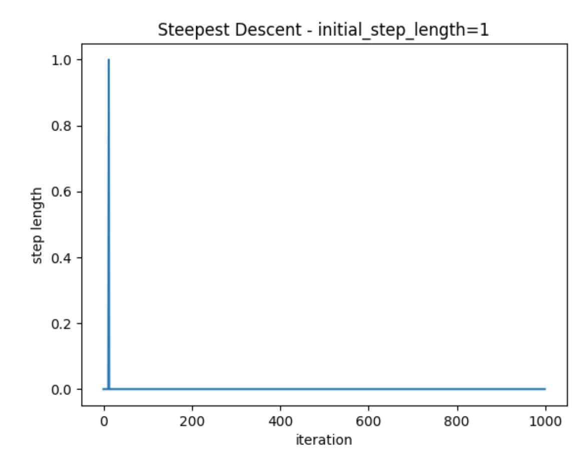


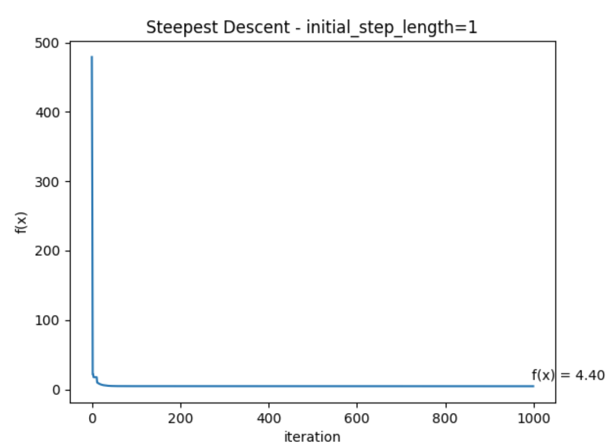
همان طور که مشاهده میشود روش اول نتوانسته است نقطه بهینه را پیدا کند. که احتمالا به خاطر این است که step-lengthهای بسیار نزدیک به صفر توسط line search به دست می آید.

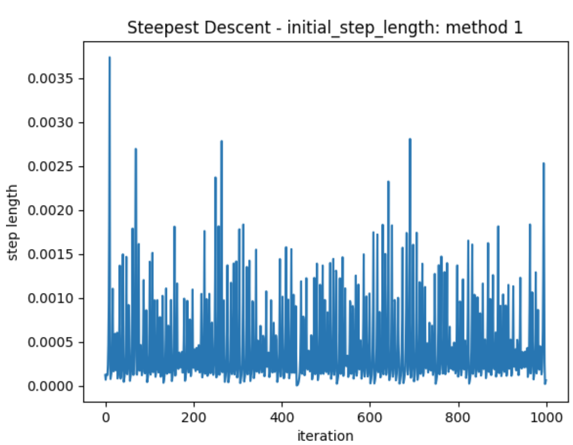
روش دوم سریع ترین روش در مقایسه با روش های دیگر و روش default بوده است و تنها بعد از ۱۲ مرحله همگرا شده است.

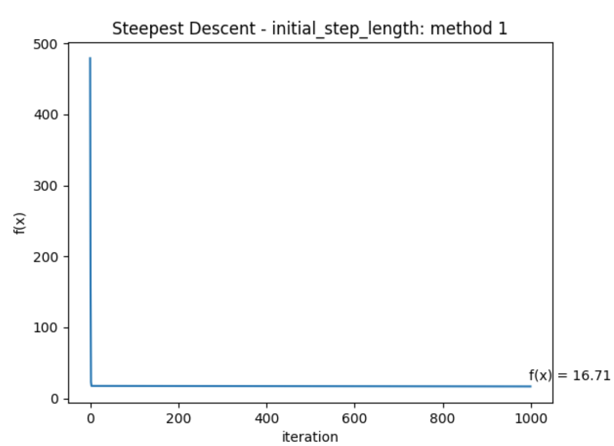
۲- حال به بررسی روش steepest descent میپردازیم.

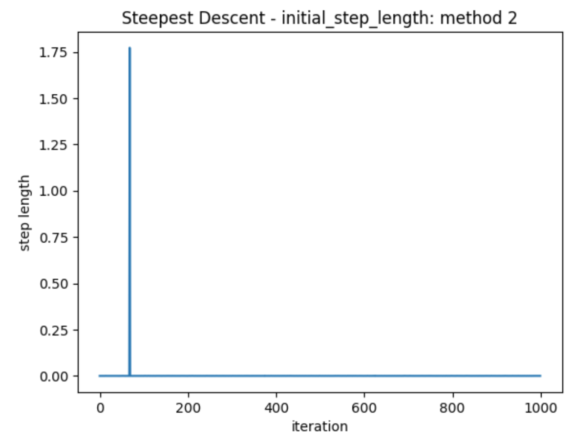
به ترتیب نمودارهای f و step length برای حالت اولیه یک و سه استراژی بیان شده در زیر آورده شده است.

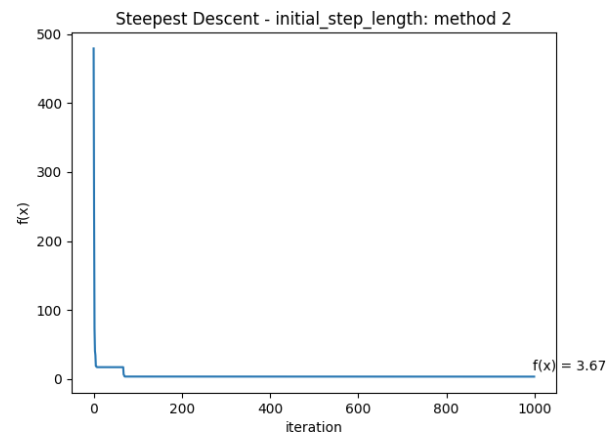


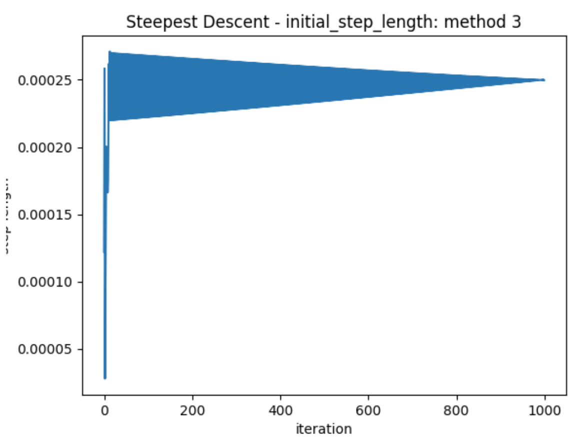


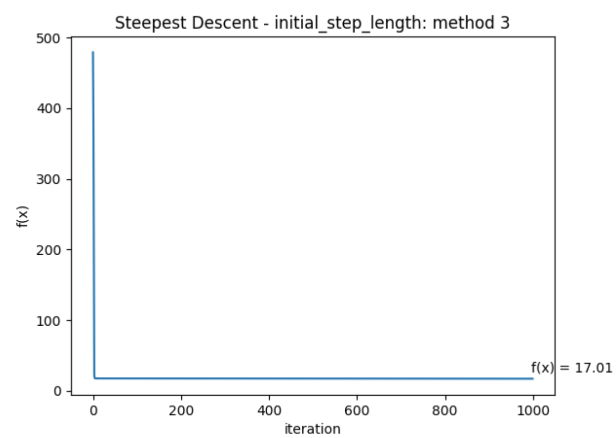












همان طور که مشاهده میشود (عدد نهایی به دست آمده روی نمودار مشخص شده است) هیچ کدام از استراتژی ها نتوانسته است تابع را بهینه کند گرچه روش اول و سوم نسبت به دو روش دیگر بهتر عمل کرده است ولی هیچ کدام نتوانسته اند به مقدار بهینه صفر برسند.

دلیل ناکارآمدی الگوریتم به انتخاب step-length نامناسب و بسیار نزدیک صفر توسط الگوریتم backtracking line search برمیگردد که با این که ابتدا توانسته است تابع را به درستی کاهش دهد ولی نزدیک نقطه بهینه دیگر عملکرد الگوریتم صحیح نیست و باید از الگوریتم zooming و interpolation برای linesearch صحیح تر استفاده کرد تا steepest descent بتواند تابع داده شده را با استفاده از line search بهینه کند.

سوال ۵ (پیاده سازی)

هدف از این سوال بهینه سازی تابع داده شده با استفاده از روش simulated annealing است.

پیاده سازی انجام شده مطابق با روش گفته شده انجام شده است به این صورت که:

- عدد x به صورت یک عدد ۵ بیتی نمایش داده شده است.
- در هر لحظه یک بیت به صورت رندوم flip میشود.
- اگر f عدد جدید به دست آمده موجب کاهش تابع شد این عدد به عنوان عدد بعدی جایگزین میشود.
- در غیر این صورت بر اساس احتمال زیر accept میشود:

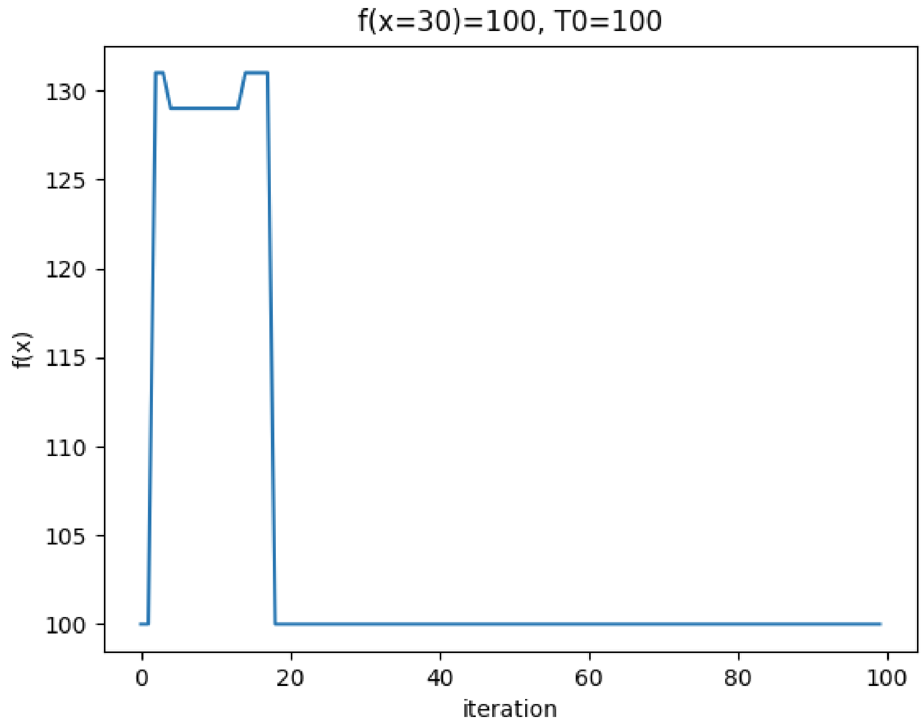
$$\exp\left(-\frac{f_{new} - f_{old}}{T}\right)$$

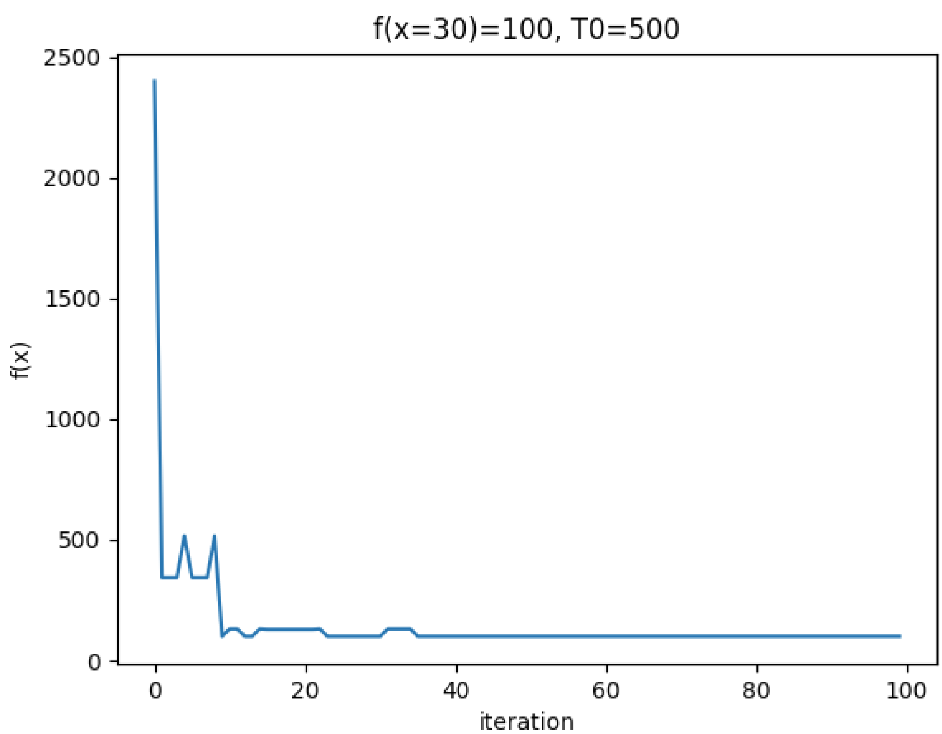
و چون T در هر مرحله ۱۰ درصد کاهش می یابد(در ۰.۹ ضرب میشود) پس در مراحل ابتدایی احتمال accept کردن زیاد و در مراحل پایانی کم است.

نمودار f عددهای accept شده در زیر برای دو حالت دمای T ابتدایی 500 و 100 آورده شده است.

x و f بهینه به دست آمده در عنوان نمودار مشخص شده است. که در هر دو حالت عدد ۳۰ به درستی به عنوان عدد minimum به دست آمده است.

زمانی که T=500 بوده است سریع تر به جواب رسیده ایم به خاطر این که احتمال acceptance بیشتری داشته ایم و exploration بیشتر ابتدای کار منجر به پیداکردن سریع تر عدد بهینه شده است گرچه این exploration منجر به این شده که بعد از پیدا کردن عدد بهینه همگرایی کامل نداشته باشیم و بعد از این که دما کاهش یافت الگوریتم همگرا میشود.


f(x=30)=100, T0=100
130
125
120
115
110
105
100
f(x)
0
20
40
60
80
100
iteration


f(x=30)=100, T0=500
2500
2000
1500
1000
500
0
f(x)
0
20
40
60
80
100
iteration