Nonlinear Optimization Fall 2021

HW2 Sample Solutions

Q1 (a) We will make use of the following pair of facts for any ~~positive~~ symmetric M, $s \in \mathbb{R}^d$:

$$s^T M s \leq \lambda_{max}(M) \|s\|_2^2,$$

$$s^T M s \geq \lambda_{min}(M) \|s\|_2^2.$$

(The fact that these inequalities are obtained by the max and min eigenvectors will establish our Lipschitz and strong convexity constants are tight.)

Observing $\nabla^2 f(x) = A^T A$ for any $x \in \mathbb{R}^d$,

$\lambda_{max}(A^T A)$ - Lipschitz gradient $\Leftrightarrow \nabla^2 f(x) \preceq \lambda_{max}(A^T A) I$

$\Leftrightarrow \forall s \quad s^T A^T A s \leq s^T (\lambda_{max}(A^T A)) s$

$\Leftrightarrow \forall s \quad s^T A^T A s \leq \lambda_{max}(A^T A) \|s\|_2^2$

$\Leftrightarrow$ True. (by our first Linear Algebra fact)

$\lambda_{min}(A^T A)$ - strong convexity $\Leftrightarrow \nabla^2 f(x) \succeq \lambda_{min}(A^T A) I$

$\Leftrightarrow \forall s \quad s^T A^T A s \geq s^T (\lambda_{min}(A^T A)) s$

$\Leftrightarrow \forall s \quad s^T A^T A s \geq \lambda_{min}(A^T A) \|s\|_2^2$

$\Leftrightarrow$ True (by our second Linear Algebra fact).

(b) Observe that f is essentially a onedimensional function, only depending on the component of x in c:

$$f(x) = h(c^T x) \quad \text{where } h: \mathbb{R} \to \mathbb{R} \text{ with } h(t) = \frac{e^t}{1+e^t}.$$

Then $\nabla f(x) = h'(c^T x) \cdot c$. Hence it suffices to show h' is L-Lipschitz since ∇f is then $\|c\|_2^2 L$-Lipschitz as

$$\begin{aligned} \|\nabla f(x) - \nabla f(y)\| &= \|h'(c^T x)c - h'(c^T y)c\| \\ &= \|c\|_2 \cdot |h'(c^T x) - h'(c^T y)| \\ &\le \|c\|_2 \cdot L\,|c^T x - c^T y| \\ &\le \|c\|_2^2 L\,\|x - y\|. \end{aligned}$$

A one-dimensional function is L-Lipschitz if its derivative is bounded by L everywhere. So it suffices to bound h''.

Using Wolframalpha, we have

$$\begin{aligned} |h''(t)| &= \left|\frac{e^t}{1+e^t} - \frac{3e^{2t}}{(e^t+1)^2} + \frac{2e^{3t}}{(e^t+1)^3}\right| \\ &\le \left|\frac{e^t}{1+e^t}\right| + 3\left|\frac{e^t}{1+e^t}\right|^2 + 2\left|\frac{e^t}{1+e^t}\right|^3 \\ &\le 6. \end{aligned}$$

$\Rightarrow \nabla f(x)$ is $6\|c\|_2^2$-Lipschitz.

□

Q2 We can prove this directly. For any $x \in \mathbb{R}^d$, we have

$$f(\bar{x}^*) + \frac{\rho}{2}\|\bar{x}^* - \bar{x}\|_2^2 \leq f(x) + \frac{\rho}{2}\|x - \bar{x}\|_2^2 \qquad \text{since } \bar{x}^* \text{ is a minimizer}$$

$$= f(x) + \frac{\rho}{2}\|x - \bar{x}^* + \bar{x}^* - \bar{x}\|_2^2 \qquad \text{add \& subtract } \bar{x}^*$$

$$= f(x) + \rho(\bar{x}^* - \bar{x})^T(x - \bar{x}^*) + \frac{\rho}{2}\|x - \bar{x}^*\|_2^2 + \frac{\rho}{2}\|\bar{x}^* - \bar{x}\|_2^2.$$

Cancelling this last quadratic from both sides gives a weaker result than needed:

$$\Rightarrow \underline{f(x) \geq f(\bar{x}^*) + g^T(x - \bar{x}^*) - \frac{\rho}{2}\|x - \bar{x}^*\|_2^2 \quad \forall_{x \in \mathbb{R}^d} (*)}$$

Using convexity of f, we can strengthen this since for any $\lambda \in [0,1]$, $f(\lambda x + (1-\lambda)\bar{x}^*) \overset{\text{cvx}}{\leq} \lambda f(x) + (1-\lambda) f(\bar{x}^*)$

$$\Rightarrow \underline{f(x) \geq f(\bar{x}^*) + \frac{f(\lambda x + (1-\lambda)\bar{x}^*) - f(\bar{x}^*)}{\lambda}}.$$

Then applying $(*)$ at $\lambda x + (1-\lambda)\bar{x}^*$, we have

$$f(x) \geq f(\bar{x}^*) + \frac{(f(\bar{x}^*) + g^T(\lambda x + (1-\lambda)\bar{x}^* - \bar{x}^*) - \frac{\rho}{2}\|\lambda x + (1-\lambda)\bar{x}^* - \bar{x}^*\|_2^2) - f(\bar{x}^*)}{\lambda}$$

$$= f(x^*) + \frac{g^T(\lambda x - \lambda \bar{x}^*) - \frac{\rho}{2}\|\lambda x - \lambda \bar{x}^*\|_2^2}{\lambda}$$

$$= f(x^*) + g^T(x - \bar{x}^*) - \lambda \frac{\rho}{2}\|x - x^*\|_2^2, \text{ for any } \lambda \in (0,1].$$

Taking the limit as $\lambda \to 0$ gives the claim. □

Importantly, this proof did not require differentiability of f!

Q3 (a) Note we have not assumed f is differentiable.

Let $h(x) := f(x) - \frac{\mu}{2}\|x\|_2^2$, which is convex by assumption.

Then x^* minimizes $f(x) = h(x) + \frac{\mu}{2}\|x - 0\|_2^2$.

This is precisely the shape of the problem considered in Q2. Letting $\rho = \mu$, $\bar{x} = 0$, we then know

$$h(x) \geq h(x^*) + \rho(0 - x^*)^T(x - x^*) \quad \forall x \in \mathbb{R}^d$$

$$= h(x^*) + \frac{\rho}{2}\|x^*\|_2^2 + \frac{\rho}{2}\|x - x^*\|_2^2 - \frac{\rho}{2}\|x\|_2^2.$$

$$\Rightarrow \underbrace{h(x) + \frac{\rho}{2}\|x\|_2^2}_{= f(x)} \geq \underbrace{h(x^*) + \frac{\rho}{2}\|x^*\|_2^2}_{= f(x^*)} + \frac{\rho}{2}\|x - x^*\|_2^2$$

(b) From lecture, we know that the accelerated method has

$$f(y_k) - f(x^*) \leq \frac{2L\|x_0 - x^*\|^2}{k^2} \quad \text{from lecture}$$

$$\leq \frac{2L}{k^2} \cdot \frac{2}{\mu}(f(x_0) - f(x^*)) \quad \text{from part (a)}$$

$$= \frac{4L}{\mu k^2}(f(x_0) - f(x^*))$$

Thus $k \geq \sqrt{16L/\mu}$ has

$$f(y_k) - f(x^*) \leq \frac{4L}{\mu(16L/\mu)}(f(x_0) - f(x^*))$$

$$= \frac{1}{4}(f(x_0) - f(x^*)). \quad \text{(better than needed by a factor of two.)}$$

□

(c) Denote the sequence of points the restarted method restarts at as $\{x_0^{(i)}\}_{i=0}^{\infty}$ where $x_0^{(0)} = x_0$

$x_0^{(2)} = y_{\lceil 4\sqrt{L/\mu} \rceil}$

$\vdots$

From (b), we know that

$$f(x_0^{(i+1)}) - f(x^*) \leq \frac{f(x_0^{(i)}) - f(x^*)}{2}.$$

Inductively applying this, we conclude

$$f(x_0^{(i)}) - f(x^*) \leq \frac{f(x_0) - f(x^*)}{2^i}.$$

Then after $i = \lceil \log_2\left(\frac{f(x_0) - f(x^*)}{\epsilon}\right) \rceil$ restarts, we have $2^i \geq \frac{f(x_0) - f(x^*)}{\epsilon}$ and so $f(x_0^{(i)}) - f(x^*) \leq \epsilon$.

Since each restart requires $\lceil 4\sqrt{L/\mu} \rceil$ steps, the total iteration bound to reach ϵ-accuracy is

$$\lceil 4\sqrt{L/\mu} \rceil \log\left(\frac{f(x_0) - f(x^*)}{\epsilon}\right).$$

↑ faster than our gradient descent guarantee under strong convexity by a squareroot.

□

In [1]:
```
import numpy as np
import matplotlib.pyplot as plt
import math

n = 1000
m = 2000
A = np.random.normal(0, 1, size=(m, n)) #Draw normal random entries for A
b = np.random.normal(0, 1, size=m)      #Draw normal random entries for b
```

In [2]:
```
#Computing smoothness and strong convexity (eigenvalues of A^TA) for part (a)
eigenvalues = np.linalg.eigvals(np.matmul(A.transpose(),A))
print("Lipschitz constant (Maximum Eigenvalue of A^TA): ", max(eigenvalues))
 #Should be approximately 5800
print("Strong Convexity constant (Minimum Eigenvalue of A^TA): ", min(eigenvalues))
 #Should be approximately 170
```

```
Lipschitz constant (Maximum Eigenvalue of A^TA):  5842.2015293678205
Strong Convexity constant (Minimum Eigenvalue of A^TA):  175.78811021349685
```

In [3]:
```
#Helper functions for evaluating the objective and gradient of the objective
def f(x):
 y=np.dot(A,x)-b
 return (np.linalg.norm(y)**2)/2
def grad_f(x):
 y=np.dot(A,x)-b
 return np.dot(A.transpose(),y)
```

In [4]:
```
#Run gradient descent for part (b)
x = np.zeros(n) #initialize at the origin
for i in range(100):
 x = x - grad_f(x)/eigenvalues[0]
 print("After ", i, " steps, the grad size is ", np.linalg.norm(grad_f(x)))
```

```
After  0  steps, the grad size is  768.2560147219825
After  1  steps, the grad size is  538.1960776794093
After  2  steps, the grad size is  414.1109433036309
After  3  steps, the grad size is  336.29030217485206
After  4  steps, the grad size is  282.76361396295414
After  5  steps, the grad size is  243.44842186598027
After  6  steps, the grad size is  213.14533524755112
After  7  steps, the grad size is  188.94355569142425
After  8  steps, the grad size is  169.0978268159752
After  9  steps, the grad size is  152.49646363778177
After  10  steps, the grad size is  138.3928408040189
After  11  steps, the grad size is  126.26241355839205
After  12  steps, the grad size is  115.72268717121553
After  13  steps, the grad size is  106.4862113389045
After  14  steps, the grad size is  98.33166750917631
After  15  steps, the grad size is  91.08530035602274
After  16  steps, the grad size is  84.60851595788874
After  17  steps, the grad size is  78.78930744034271
After  18  steps, the grad size is  73.5361478037148
After  19  steps, the grad size is  68.77352908885578
After  20  steps, the grad size is  64.43863458184006
After  21  steps, the grad size is  60.47881212640191
After  22  steps, the grad size is  56.84962716216343
After  23  steps, the grad size is  53.51334366546755
After  24  steps, the grad size is  50.437726270838546
After  25  steps, the grad size is  47.5950869204029
After  26  steps, the grad size is  44.961519951550486
After  27  steps, the grad size is  42.51628391886762
After  28  steps, the grad size is  40.241298715108854
After  29  steps, the grad size is  38.12073401595004
After  30  steps, the grad size is  36.14067057605034
After  31  steps, the grad size is  34.28882001689253
After  32  steps, the grad size is  32.55429185626027
After  33  steps, the grad size is  30.927398903298048
After  34  steps, the grad size is  29.399493971554637
After  35  steps, the grad size is  27.96283228147071
After  36  steps, the grad size is  26.610455032687486
After  37  steps, the grad size is  25.336090498524978
After  38  steps, the grad size is  24.13406968456632
After  39  steps, the grad size is  22.99925414154405
After  40  steps, the grad size is  21.926973960805174
After  41  steps, the grad size is  20.912974332332634
After  42  steps, the grad size is  19.953369328931554
After  43  steps, the grad size is  19.04460180991226
After  44  steps, the grad size is  18.183408524439596
After  45  steps, the grad size is  17.366789647282403
After  46  steps, the grad size is  16.591982104758337
After  47  steps, the grad size is  15.856436151563626
After  48  steps, the grad size is  15.157794744141453
After  49  steps, the grad size is  14.493875326642264
After  50  steps, the grad size is  13.862653704052612
After  51  steps, the grad size is  13.262249725883665
After  52  steps, the grad size is  12.690914544643265
After  53  steps, the grad size is  12.147019247582309
After  54  steps, the grad size is  11.629044689037425
```

```
After  55  steps, the grad size is  11.135572375026221
After  56  steps, the grad size is  10.66527627233528
After  57  steps, the grad size is  10.216915431804644
After  58  steps, the grad size is  9.78932733036333
After  59  steps, the grad size is  9.381421849029163
After  60  steps, the grad size is  8.992175814902446
After  61  steps, the grad size is  8.620628044448305
After  62  steps, the grad size is  8.265874833311013
After  63  steps, the grad size is  7.92706584474408
After  64  steps, the grad size is  7.603400354635671
After  65  steps, the grad size is  7.294123816199627
After  66  steps, the grad size is  6.998524711811016
After  67  steps, the grad size is  6.715931663287556
After  68  steps, the grad size is  6.445710775240239
After  69  steps, the grad size is  6.187263189007672
After  70  steps, the grad size is  5.940022827213635
After  71  steps, the grad size is  5.703454311191358
After  72  steps, the grad size is  5.477051035449343
After  73  steps, the grad size is  5.260333385047243
After  74  steps, the grad size is  5.052847083241996
After  75  steps, the grad size is  4.854161658071106
After  76  steps, the grad size is  4.663869017701194
After  77  steps, the grad size is  4.481582125389325
After  78  steps, the grad size is  4.306933765810651
After  79  steps, the grad size is  4.139575395308383
After  80  steps, the grad size is  3.979176069335135
After  81  steps, the grad size is  3.8254214409904876
After  82  steps, the grad size is  3.6780128251226816
After  83  steps, the grad size is  3.5366663229715476
After  84  steps, the grad size is  3.401112002778901
After  85  steps, the grad size is  3.2710931321985277
After  86  steps, the grad size is  3.1463654587029666
After  87  steps, the grad size is  3.026696534508644
After  88  steps, the grad size is  2.9118650828383195
After  89  steps, the grad size is  2.801660402601745
After  90  steps, the grad size is  2.695881808819416
After  91  steps, the grad size is  2.594338106327448
After  92  steps, the grad size is  2.4968470945006094
After  93  steps, the grad size is  2.403235100908463
After  94  steps, the grad size is  2.313336541980671
After  95  steps, the grad size is  2.226993508906814
After  96  steps, the grad size is  2.14405537712845
After  97  steps, the grad size is  2.0643784379056025
After  98  steps, the grad size is  1.9878255505507045
After  99  steps, the grad size is  1.9142658140263373
```

In [5]:
```
#Run accelerated gradient descent for part (c)
x = np.zeros(n) #initialize at the origin
y = np.zeros(n)       #auxillary sequence for the accelerated method
y_prev = np.zeros(n)  #tracking previous iterations value of y_prev
lam = 0
lam_prev=lam
for i in range(100):
 lam_prev = lam
 lam = (1+math.sqrt(1+4*lam**2))/2
 y = x - grad_f(x)/eigenvalues[0]
 x = y + (lam_prev-1.0)/lam*(y-y_prev) #You can also replace the lambda sequence by (i/(i+3
 y_prev = y
 print("After ", i, " steps, the grad size is ", np.linalg.norm(grad_f(x)))
```

```
After  0  steps, the grad size is  1403.1075003197154
After  1  steps, the grad size is  768.2560147219825
After  2  steps, the grad size is  483.20265074900544
After  3  steps, the grad size is  328.15342575650715
After  4  steps, the grad size is  236.73381915533665
After  5  steps, the grad size is  177.32191549208628
After  6  steps, the grad size is  134.8514284762135
After  7  steps, the grad size is  103.22048447859387
After  8  steps, the grad size is  79.87221182386696
After  9  steps, the grad size is  62.96370412534777
After  10  steps, the grad size is  50.64724718217208
After  11  steps, the grad size is  41.3420244273921
After  12  steps, the grad size is  34.02194455621862
After  13  steps, the grad size is  28.178866719423983
After  14  steps, the grad size is  23.60937071726631
After  15  steps, the grad size is  20.207989442569048
After  16  steps, the grad size is  17.831374323596197
After  17  steps, the grad size is  16.251287557379644
After  18  steps, the grad size is  15.194078779664489
After  19  steps, the grad size is  14.413726544845723
After  20  steps, the grad size is  13.73364105298297
After  21  steps, the grad size is  13.044668782772895
After  22  steps, the grad size is  12.285563201094204
After  23  steps, the grad size is  11.427898628871283
After  24  steps, the grad size is  10.469952143271861
After  25  steps, the grad size is  9.434703258193135
After  26  steps, the grad size is  8.36674941791428
After  27  steps, the grad size is  7.325932370002859
After  28  steps, the grad size is  6.377624371160065
After  29  steps, the grad size is  5.580002754109604
After  30  steps, the grad size is  4.9691199347025465
After  31  steps, the grad size is  4.546267827455944
After  32  steps, the grad size is  4.277069953350048
After  33  steps, the grad size is  4.106868308669589
After  34  steps, the grad size is  3.9822498518826737
After  35  steps, the grad size is  3.8646468381919106
After  36  steps, the grad size is  3.732734198930985
After  37  steps, the grad size is  3.5786090000929813
After  38  steps, the grad size is  3.4027561772586257
After  39  steps, the grad size is  3.2101006074737097
After  40  steps, the grad size is  3.0075552631762443
After  41  steps, the grad size is  2.8026808136321506
After  42  steps, the grad size is  2.6028979130205503
After  43  steps, the grad size is  2.4148165797466175
After  44  steps, the grad size is  2.243503181430563
After  45  steps, the grad size is  2.091785029383136
After  46  steps, the grad size is  1.9598811394193607
```

```
After  47  steps, the grad size is  1.845630109364283
After  48  steps, the grad size is  1.7453391682595059
After  49  steps, the grad size is  1.6549601711402526
After  50  steps, the grad size is  1.571148351900048
After  51  steps, the grad size is  1.4918678844655988
After  52  steps, the grad size is  1.4164522604632126
After  53  steps, the grad size is  1.3452377365630221
After  54  steps, the grad size is  1.2789908009476507
After  55  steps, the grad size is  1.2183574121758032
After  56  steps, the grad size is  1.163503207471251
After  57  steps, the grad size is  1.1140142547993477
After  58  steps, the grad size is  1.0690191426485065
After  59  steps, the grad size is  1.0274192858906739
After  60  steps, the grad size is  0.9881059835845605
After  61  steps, the grad size is  0.9500917939662665
After  62  steps, the grad size is  0.9125510592835315
After  63  steps, the grad size is  0.8748113416042079
After  64  steps, the grad size is  0.8363481194872966
After  65  steps, the grad size is  0.7968160981573145
After  66  steps, the grad size is  0.7561197366359774
After  67  steps, the grad size is  0.7144998869773015
After  68  steps, the grad size is  0.6726011211043353
After  69  steps, the grad size is  0.6314842903569615
After  70  steps, the grad size is  0.5925552770373892
After  71  steps, the grad size is  0.5573906142270802
After  72  steps, the grad size is  0.5274591849739109
After  73  steps, the grad size is  0.5037782963081013
After  74  steps, the grad size is  0.486600426183482
After  75  steps, the grad size is  0.47526288633910996
After  76  steps, the grad size is  0.468286362264461
After  77  steps, the grad size is  0.4636855483716834
After  78  steps, the grad size is  0.4593556445127243
After  79  steps, the grad size is  0.45340156670758813
After  80  steps, the grad size is  0.44435098602843537
After  81  steps, the grad size is  0.43126036682606483
After  82  steps, the grad size is  0.4137505087015838
After  83  steps, the grad size is  0.3920055046628369
After  84  steps, the grad size is  0.3667561255225672
After  85  steps, the grad size is  0.3392548941459453
After  86  steps, the grad size is  0.31123425961165113
After  87  steps, the grad size is  0.2848137283408678
After  88  steps, the grad size is  0.262283025659029
After  89  steps, the grad size is  0.2456694180320073
After  90  steps, the grad size is  0.23610515719192604
After  91  steps, the grad size is  0.23330071335759855
After  92  steps, the grad size is  0.23557404108793764
After  93  steps, the grad size is  0.2404777396354275
After  94  steps, the grad size is  0.24555601705146865
After  95  steps, the grad size is  0.24882688574261116
After  96  steps, the grad size is  0.2489491658508212
After  97  steps, the grad size is  0.24520903976206063
After  98  steps, the grad size is  0.2374412698274858
After  99  steps, the grad size is  0.225940992952322
```

In [6]:
```
#Run restarted accelerated gradient descent for part (d)
x = np.zeros(n) #initialize at the origin
for j in range(4):
 y = x         #auxillary sequence for the accelerated method
 y_prev = x    #tracking previous iterations value of y_prev
 lam = 0
 lam_prev=lam
 for i in range(25):
  lam_prev = lam
  lam = (1+math.sqrt(1+4*lam**2))/2
  y = x - grad_f(x)/eigenvalues[0]
  x = y + (lam_prev-1.0)/lam*(y-y_prev) #You can also replace the lambda sequence by (i/(i+
  y_prev = y
  print("After ", 25*j+i, " steps, the grad size is ", np.linalg.norm(grad_f(x)))
```

```
After  0  steps, the grad size is  1403.1075003197154
After  1  steps, the grad size is  768.2560147219825
After  2  steps, the grad size is  483.20265074900544
After  3  steps, the grad size is  328.15342575650715
After  4  steps, the grad size is  236.73381915533665
After  5  steps, the grad size is  177.32191549208628
After  6  steps, the grad size is  134.8514284762135
After  7  steps, the grad size is  103.22048447859387
After  8  steps, the grad size is  79.87221182386696
After  9  steps, the grad size is  62.96370412534777
After  10  steps, the grad size is  50.64724718217208
After  11  steps, the grad size is  41.3420244273921
After  12  steps, the grad size is  34.02194455621862
After  13  steps, the grad size is  28.178866719423983
After  14  steps, the grad size is  23.60937071726631
After  15  steps, the grad size is  20.207989442569048
After  16  steps, the grad size is  17.831374323596197
After  17  steps, the grad size is  16.251287557379644
After  18  steps, the grad size is  15.194078779664489
After  19  steps, the grad size is  14.413726544845723
After  20  steps, the grad size is  13.73364105298297
After  21  steps, the grad size is  13.044668782772895
After  22  steps, the grad size is  12.285563201094204
After  23  steps, the grad size is  11.427898628871283
After  24  steps, the grad size is  10.469952143271861
After  25  steps, the grad size is  10.469952143271861
After  26  steps, the grad size is  9.869247438336366
After  27  steps, the grad size is  9.165763564907936
After  28  steps, the grad size is  8.394945865340645
After  29  steps, the grad size is  7.587256403090467
After  30  steps, the grad size is  6.76813989418699
After  31  steps, the grad size is  5.958061249106315
After  32  steps, the grad size is  5.172835727992778
After  33  steps, the grad size is  4.424222908758966
After  34  steps, the grad size is  3.720696130837436
After  35  steps, the grad size is  3.0682976999376974
After  36  steps, the grad size is  2.4715229748807745
After  37  steps, the grad size is  1.9342542169168746
After  38  steps, the grad size is  1.4609428166162273
After  39  steps, the grad size is  1.0586832603092278
After  40  steps, the grad size is  0.7418097475837306
After  41  steps, the grad size is  0.5395917761352829
After  42  steps, the grad size is  0.4821152213791013
After  43  steps, the grad size is  0.5336739834952844
After  44  steps, the grad size is  0.6171887704507006
After  45  steps, the grad size is  0.6888323771555729
```

```
After  46  steps, the grad size is  0.7336165917283759
After  47  steps, the grad size is  0.7486580060359111
After  48  steps, the grad size is  0.7359923505814131
After  49  steps, the grad size is  0.6998441757292064
After  50  steps, the grad size is  0.6998441757292064
After  51  steps, the grad size is  0.6702015215131653
After  52  steps, the grad size is  0.63398113804602
After  53  steps, the grad size is  0.5924075917671329
After  54  steps, the grad size is  0.5466684105618221
After  55  steps, the grad size is  0.4979249366560311
After  56  steps, the grad size is  0.447301493895110 76
After  57  steps, the grad size is  0.3958696074258252
After  58  steps, the grad size is  0.34463266040076185
After  59  steps, the grad size is  0.2945132213053958
After  60  steps, the grad size is  0.24634445703594057
After  61  steps, the grad size is  0.20086750839052817
After  62  steps, the grad size is  0.15873897597168438
After  63  steps, the grad size is  0.1205596817851686
After  64  steps, the grad size is  0.086958843692532
After  65  steps, the grad size is  0.05885093871986429
After  66  steps, the grad size is  0.03825069254399303
After  67  steps, the grad size is  0.029494542419263605
After  68  steps, the grad size is  0.03328042980086804
After  69  steps, the grad size is  0.041573294734837195
After  70  steps, the grad size is  0.04912390571179536
After  71  steps, the grad size is  0.05441538634643795
After  72  steps, the grad size is  0.05718927524852605
After  73  steps, the grad size is  0.0575984147501797
After  74  steps, the grad size is  0.055954434897099184
After  75  steps, the grad size is  0.055954434897099184
After  76  steps, the grad size is  0.053708116647108214
After  77  steps, the grad size is  0.050949203791640696
After  78  steps, the grad size is  0.04776446718788497
After  79  steps, the grad size is  0.04423926405676116
After  80  steps, the grad size is  0.04045868235901789
After  81  steps, the grad size is  0.03650683352895011
After  82  steps, the grad size is  0.0324655755086287
After  83  steps, the grad size is  0.02841311163932139
After  84  steps, the grad size is  0.024422680214020047
After  85  steps, the grad size is  0.020561481609160735
After  86  steps, the grad size is  0.01689000109652605
After  87  steps, the grad size is  0.01346199907383044
After  88  steps, the grad size is  0.01032581864515885
After  89  steps, the grad size is  0.007528933846398174
After  90  steps, the grad size is  0.00513249662830888
After  91  steps, the grad size is  0.003262315376450469
After  92  steps, the grad size is  0.002251377217442368
After  93  steps, the grad size is  0.00239998917630202
After  94  steps, the grad size is  0.0031037093459684033
After  95  steps, the grad size is  0.003796421353864154
After  96  steps, the grad size is  0.004311580485487332
After  97  steps, the grad size is  0.004617573967375408
After  98  steps, the grad size is  0.004722740877417635
After  99  steps, the grad size is  0.004649964346689991
```