HOMEWORK-6

NUMERICAL ALGORITHMS

1

To prove $f: \mathbb{R}^n \to \mathbb{R}$ is convex iff $\tilde{f}(t) = f(x + tv)$ is convex $\forall x, v \in \mathbb{R}^n$

\# lets assume f is convex, then to prove that $\tilde{f}(t)$ is convex

$\Rightarrow$ $\tilde{f}(\alpha t_1 + (1-\alpha)t_2) = f(x + (\alpha t_1 + (1-\alpha)t_2)v)$ for any $t_1, t_2 \in$ domain$(\tilde{f})$

$= f(x + \alpha t_1 v + (1-\alpha)t_2 v)$

$= f(\alpha x + \alpha t_1 v + (1-\alpha)x + (1-\alpha)t_2 v)$

$= f(\alpha(x + t_1 v) + (1-\alpha)(x + t_2 v))$

$\leq \alpha f(x + t_1 v) + (1-\alpha) f(x + t_2 v)$

[$\because f$ is assumed to be convex]

$\Rightarrow$ $\tilde{f}(\alpha t_1 + (1-\alpha)t_2) \leq \alpha \tilde{f}(t_1) + (1-\alpha)\tilde{f}(t_2)$

$\Rightarrow$ $\tilde{f}(t)$ is convex if $f(x)$ is convex.

Now, lets assume $\tilde{f}(t)$ is convex

And $f(\alpha x_1 + (1-\alpha)x_2)$ can be written exactly in terms of $\tilde{f}(t)$ as:

$$\tilde{f}(t) = f(x+tv)$$

$$\tilde{f}(t) = f(x_1 + t(x_2 - x_1)) \quad \text{if } x = x_1 \text{ and } v = x_2 - x_1$$

$$\Rightarrow \tilde{f}(0) = f(x_1)$$

$$\tilde{f}(1) = f(x_2)$$

and
$$\tilde{f}(1-\alpha) = f(x_1 + (1-\alpha)(x_2 - x_1))$$

$$= f(x_1 + (1-\alpha)x_2 - (1-\alpha)x_1)$$

$$\Rightarrow \tilde{f}(1-\alpha) = f(\alpha x_1 + (1-\alpha)x_2)$$

$$\Rightarrow f(\alpha x_1 + (1-\alpha)x_2) = \tilde{f}(1-\alpha)$$

$$= \tilde{f}(\alpha \cdot 0 + (1-\alpha)\cdot 1)$$

$$\leq \alpha \tilde{f}(0) + (1-\alpha)\tilde{f}(1)$$

$$\boxed{f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2)},$$

$\Rightarrow f$ is convex if $\tilde{f}$ is convex

$\Rightarrow$ $f: \mathbb{R}^n \to \mathbb{R}$ is convex if and only if $\tilde{f}(t) = f(x+tv)$ is convex.

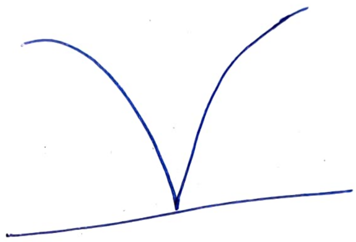
⇒ If we restrict the lines only to the co-ordinate axes,

Then this condition does not holds ie.,

$f(x)$ can be non convex even if $\tilde{f}(t) = f(x+te_i)$ is convex.

⇒ An example of such a function can be the quasiconvex function given in the slides

ie, $f(x) = \sqrt{|x|}$



⇒ In this function $\tilde{f}(t)=f(x+te_i)$ is convex but function itself is not convex.

⇒ $\tilde{f}(t)= f(x+te_i)$ is convex because from any point, ~~if w~~ the line from it parallel to coordinates axis ~~∈ Domain~~ but the two points which are not belong to the set but the two points which are not considered joined by lines parallel to coordinate axis

like

(x,y) joined by line which proves that $\sqrt{|x|}$ itself is not convex, since

$f(\theta x+(1-\theta)y) > \theta f(x)+(1-\theta)f(y)$

(2)

Given $f(x) = \sum_{i=1}^{k} |a_i^T x + b_i|$

let $f_i(x) = |a_i^T x + b_i|$

Then $f(x) = \sum_{i=1}^{k} f_i(x)$

$\Rightarrow$ Proving each $f_i(x)$ is convex:

$\Rightarrow$ ~~$f_i(x)$~~ Every (input) domain $(f_i(x)) = R^n$

$\Rightarrow$ Input set is convex

and $f_i(\alpha x_1 + (1-\alpha)x_2) = |a_i^T(\alpha x_1 + (1-\alpha)x_2) + b_i|$

$$= |\alpha a_i^T x_1 + (1-\alpha) a_i^T x_2 + \alpha b_i + (1-\alpha) b_i|$$

$$= |\alpha(a_i^T x_1 + b_i) + (1-\alpha)(a_i^T x_2 + b_i)|$$

$$\leq |\alpha(a_i^T x_1 + b_i)| + |(1-\alpha)(a_i^T x_2 + b_i)|$$

$[\because |x_1 + x_2| \leq |x_1| + |x_2|]$

$$\leq \alpha|a_i^T x_1 + b_i| + (1-\alpha)|a_i^T x_2 + b_i|$$

$(\because \alpha \geq 0 \;\&\; \alpha \leq 1)$

$\Rightarrow f_i(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f_i(x_1) + (1-\alpha) f_i(x_2)$

$\Rightarrow$ Each $f_i(x)$ is convex.

$\Rightarrow f(x) = f_1(x) + f_2(x) + \dots + f_k(x)$

Again $\mathrm{Dom}(f) \in R^n \Rightarrow$ Input set is convex and $f(\alpha x_1 + (1-\alpha) x_2)$

$$= f_1(\alpha x_1 + (1-\alpha) x_2) + \dots + f_K(\alpha x_1 + (1-\alpha) x_2)$$

$$\leq \alpha f_1(x_1) + (1-\alpha) f_1(x_2) + \dots + \alpha f_K(x_1) + (1-\alpha) f_K(x_2)$$

$$\leq \alpha [f_1(x_1) + f_2(x_1) + \dots + f_K(x_1)] + (1-\alpha) [f_1(x_2) + f_2(x_2) + \dots + f_K(x_2)]$$

$$\leq \alpha f(x_1) + (1-\alpha) f(x_2)$$

$\Rightarrow$ $f(x)$ is also convex

[$\because$ Sum of convex functions is convex as proved above].

However $f(x)$ cannot be effectively minimized by Gradient Descent and Newton's method.

$\rightarrow$ The function $f(x) = \sum_{i=1}^{K} |a_i^T x + b_i|$ is not differentiable when, ~~at~~ $x =$ solution of $a_i^T x + b = 0$ (for each $i = 1 \dots K$)

ie., if $a_i^T x + b = 0$, then $f_i(x)$ is not differentiable and hence $f(x)$ is not differentiable.

$\Rightarrow$ When ~~x~~ minimizing using gradient descent we might not be able to take the exact step $x^+ = x + t\Delta x$ since Δx may not be possible to calculate for some

x and thus we might be approximating Δx here by some value which is not as effective.

Thus, gradient descent won't effectively minimize $f(x)$.

→ Similarly, the second derivative for each $f_i(x)$ doesn't exist or vanishes and thus the second derivative for $f(x)$ vanishes.

Thus, Newton's Method is also not able to effectively minimize this $f(x)$.

[$\because \Delta x = -H(x)^{-1} g(x)$] and $H(x)$ vanishes here]

For the smoothed function

$$\hat{f}(x) = \sum_{i=1}^{K} \sqrt{(a_i^T x + b_i)^2 + \epsilon}$$

consider each $\hat{f}_i(x) = \sqrt{(a_i^T x + b_i)^2 + \epsilon}$

Then $\hat{f}(x) = \sum_{i=1}^{K} \hat{f}_i(x)$

⇒ For $\hat{f}_i(x)$, $x \in R^n$ ⇒ Domain is convex.

& $\hat{f}_i(x) = \sqrt{(a_i^T x + b_i)^2 + \epsilon}$

(can be written in terms of 2-norm)

$$\hat{f}_i(x) = \sqrt{(a_i^T x + b_i)^2 + (\sqrt{\epsilon})^2} = \left\| \begin{matrix} a_i^T x + b_i \\ \sqrt{\epsilon} \end{matrix} \right\|_2$$

Then $\hat{f}_i(\alpha x_1 + (1-\alpha)x_2) = \left\| \dfrac{a_i^T(\alpha x_1 + (1-\alpha) x_2) + b_i}{\sqrt{E}} \right\|_2$

$$= \left\| \frac{\alpha a_i^T x_1 + (1-\alpha) a_i^T x_2 + \alpha b_i + (1-\alpha) b_i}{\alpha\sqrt{E} + (1-\alpha)\sqrt{E}} \right\|_2$$

$$= \left\| \frac{\alpha(a_i^T x_1 + b_i) + (1-\alpha)(a_i^T x_2 + b_i)}{\alpha\sqrt{E} + (1-\alpha)\sqrt{E}} \right\|_2$$

$$\leq \left\| \frac{\alpha(a_i^T x + b_i)}{\alpha\sqrt{E}} \right\|_2 + \left\| \frac{(1-\alpha)(a_i^T x_2 + b_i)}{(1-\alpha)\sqrt{E}} \right\|_2$$

(following $\|X+Y\| \leq \|X\| + \|Y\|$ and $\|cX\| = |c|\,\|X\|$)

$$\cancel{\leq |\alpha| \left\| \frac{a_i x_1 + b_i}{\sqrt{E}} \right\| + |(1-\alpha)| \left\| \quad \right\|}$$

$$\leq |\alpha| \left\| \frac{(a_i^T x_1 + b_i)}{\sqrt{E}} \right\|_2 + |(1-\alpha)| \left\| \frac{(a_i^T x_2 + b_i)}{\sqrt{E}} \right\|_2$$

$$\boxed{\Rightarrow \hat{f}_i(\alpha x_1 + (1-\alpha)x_2) \leq \alpha \hat{f}_i(x_1) + (1-\alpha)\hat{f}_i(x_2)}$$

($\because$ Both α, $(1-\alpha) \geq 0$)

$\Rightarrow \hat{f}_i(x)$ is convex

$\Rightarrow \hat{f}(x) = \sum_{i=1}^{k} \hat{f}_i(x)$ is also convex since sum of convex functions is convex as proved previously.

③ (a)

$$f(x) = \sum_{(i,j) \in E} \|x_i - x_j\|_2^p$$

→ The gradient for each term in the summation can be calculated as :

$$\frac{\partial \|x_i - x_j\|_2^p}{\partial x_i} = p \|x_i - x_j\|_2^{p-1} \cdot \frac{\cancel{2}(x_i - x_j)}{\cancel{2}\sqrt{(x_i-x_j)^T(x_i-x_j)}}$$

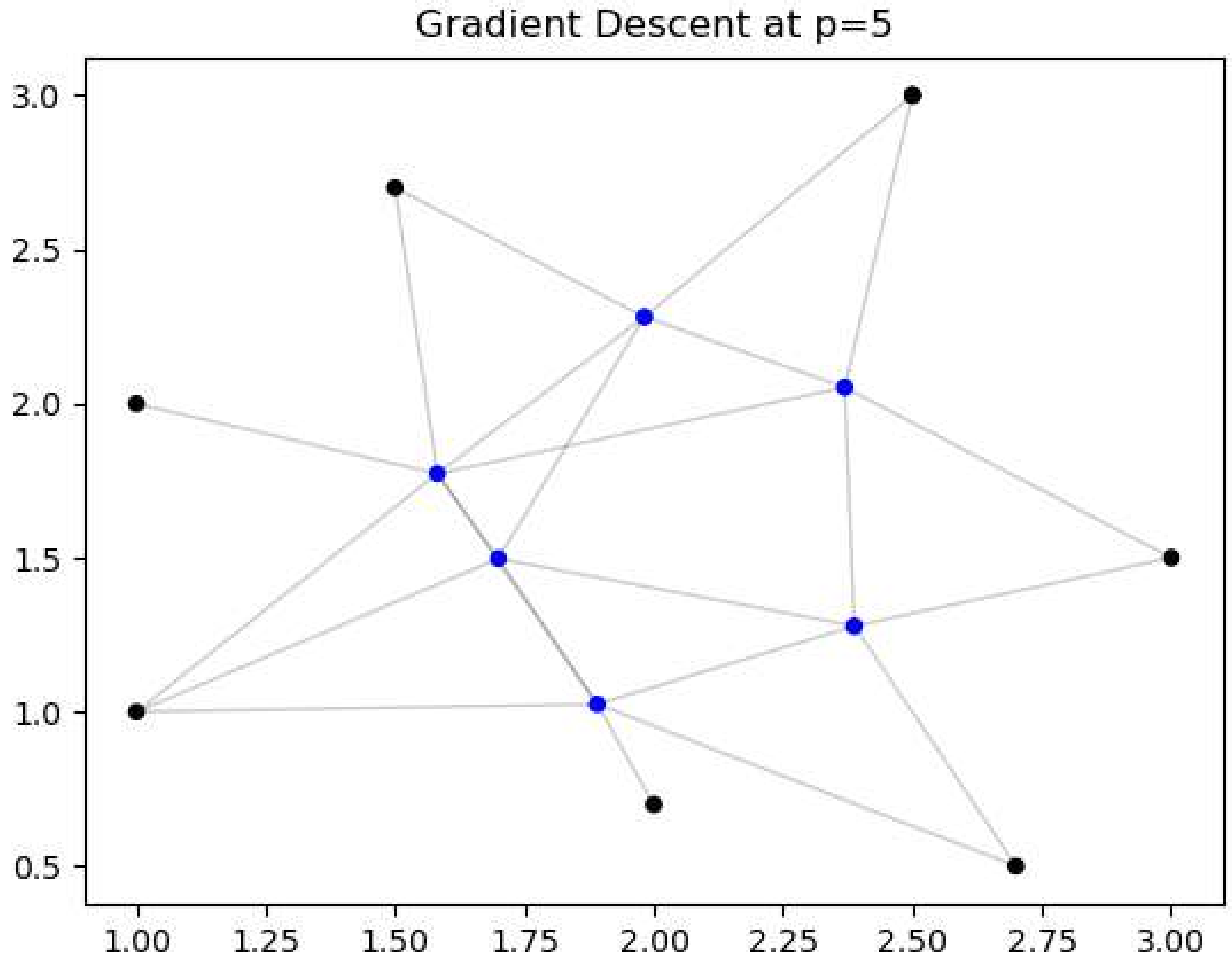
$$= p \|x_i - x_j\|_2^{p-2} \cdot (x_i - x_j) \quad \text{[Using chain rule)}$$

$$\frac{\partial(\|x_i - x_j\|_2^p)}{\partial x_j} = -p \|x_i - x_j\|_2^{p-2} (x_i - x_j)$$

⇒ The gradient of f can be calculated in the same way by adding the gradient of each term x_i, x_j to the gradient vector.

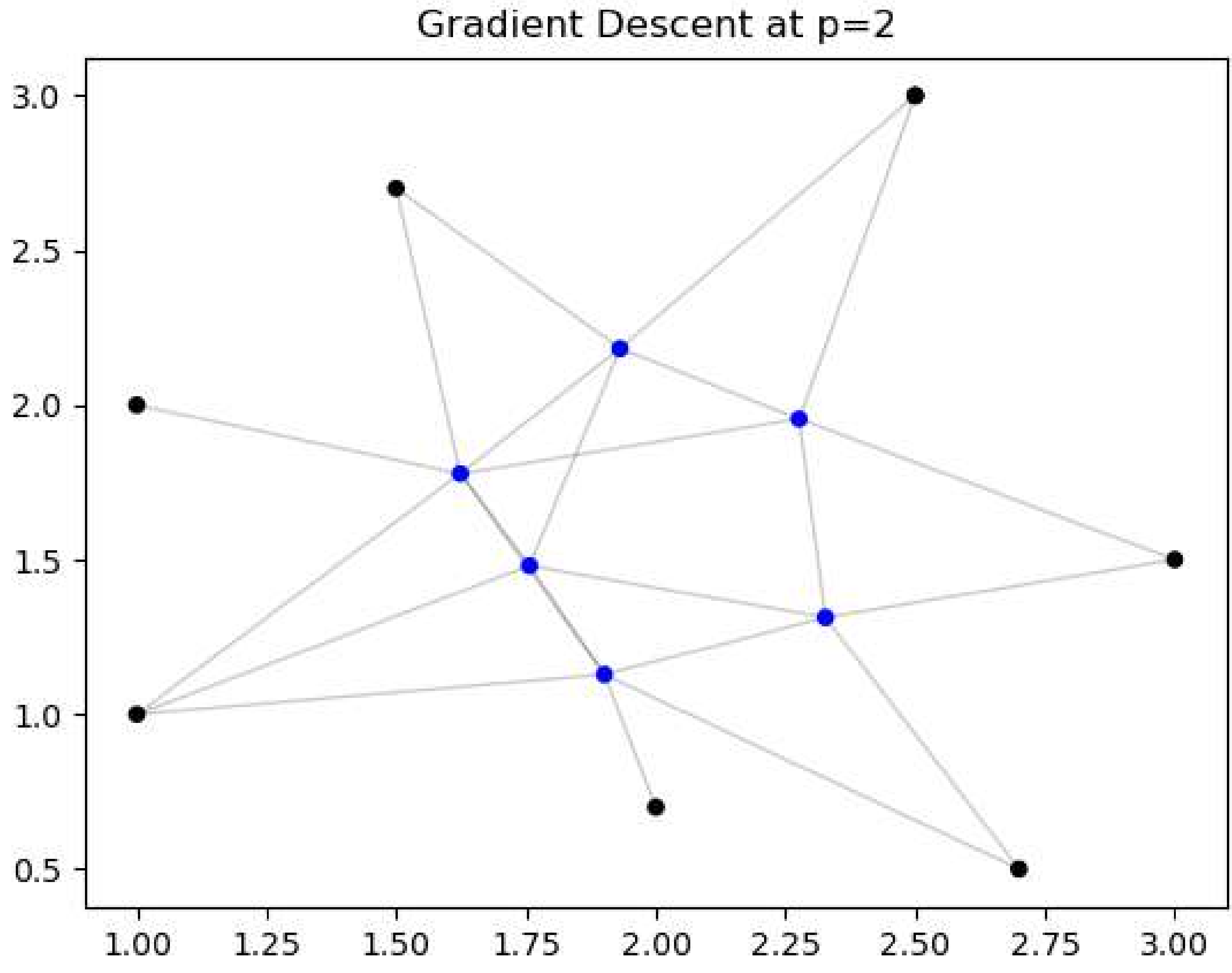
[Note that gradient $(x_i) = 0$ if $i <= m$]
(since they are the fixed nodes).

⇒ Implemented Gradient Descent with line search and the graphs are as follows :


Gradient Descent at p=5
3.0
2.5
2.0
1.5
1.0
0.5
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00

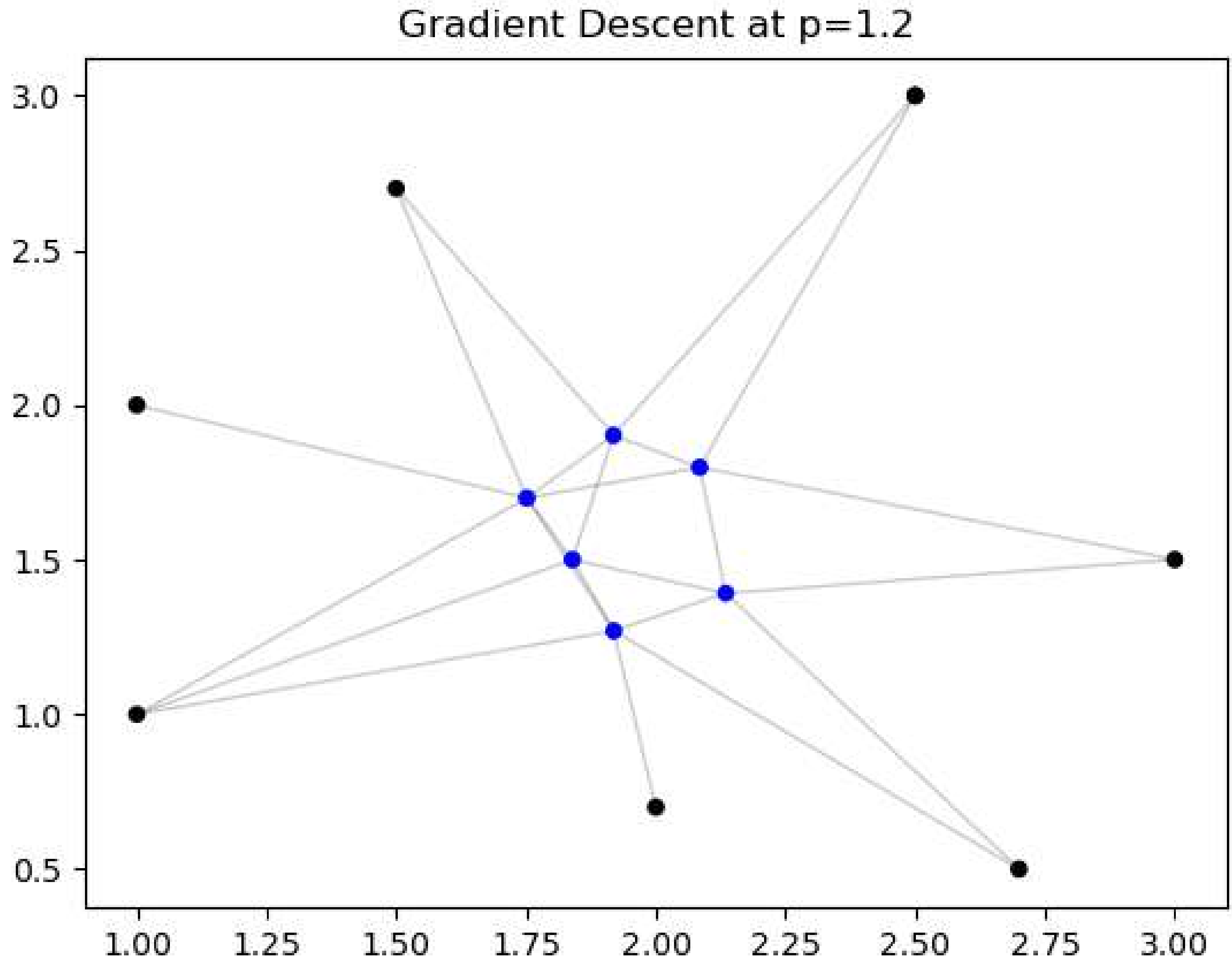
Gradient Descent (p=5)

Iteration Number	Gradient Norm	Function Value
0	2797.2461880005344	3975.7081832991157
1	573.8654949362551	523.8810085212804
2	209.75512984760002	152.49929886461447
3	112.25885095256761	72.22548258752673
4	66.20163768882858	35.674936661217785
5	27.63132720644318	20.6796323436583
6	23.847228787627206	13.865522815462384
7	12.536042972608914	10.494713664563376
8	9.357037431439641	8.923561367790375
9	4.5663658593071	8.179678314434067
10	3.9551426623889645	7.954191489661324
11	1.5708204239381278	7.812895134195343
12	1.31856243177595	7.784374537911032
13	0.5216593516873984	7.767378744396812
14	0.4161873119029447	7.7635045599733
15	0.257803879377294	7.761837374091349
16	0.1598680579434653	7.761179069859618
17	0.09966442762012276	7.760918011800768
18	0.06243328000974904	7.760812155883202
19	0.039402922964885505	7.760768440782698
20	0.025086989719048113	7.760749909074177
21	0.02194980016200586	7.7607428359898245
22	0.01384791915900659	7.760738477308732
23	0.008816886904342385	7.760736582761471
24	0.005675763733405506	7.7607357342401855
25	0.0037011723472191015	7.7607353428885535
26	0.00324644964232779	7.76073517334239
27	0.0020817663374463062	7.760735068944664
28	0.001351483194817036	7.760735020710833
29	0.0012144520181874322	7.760735000974398
30	0.0007750430708589154	7.760734987324004
31	0.000500367620960676	7.760734981075489
32	0.00032736153119606795	7.76073497813209
33	0.0002903010133733095	7.7607349768515155
34	0.00018637822818233147	7.7607349760257165
35	0.0001211684194116532	7.760734975641361
36	0.00010909800698575286	7.760734975483276
37	6.967703953919038e-05	7.760734975372981


Gradient Descent at p=2
3.0
2.5
2.0
1.5
1.0
0.5
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00

Gradient Descent (p=2)

Iteration Number	Gradient Norm	Function Value
0	41.63535686435429	90.69427951304912
1	22.107975564611102	33.58844388990197
2	8.458330877030024	17.016243703357485
3	5.918605192616898	14.56975659001894
4	2.7394851300685414	13.38234738168802
5	1.9094445355695198	13.093742045682099
6	0.9117513173147596	12.963179212781306
7	0.6247506633054593	12.927568697400433
8	0.4459249191928512	12.914847779349175
9	0.2058981668845021	12.908279380616456
10	0.14476577995714196	12.906732773747779
11	0.06836057711783801	12.906002690797251
12	0.04725694066892458	12.90581245022543
13	0.022875546477257237	12.90572998098572
14	0.015523794770788357	12.905706346012899
15	0.01099721672247473	12.905698018543884
16	0.005135478925786571	12.905693915155886
17	0.0035800353442157983	12.90569289633236
18	0.0017118943304249472	12.905692436437196
19	0.00117235197503996	12.905692310443662
20	0.0008364069765330435	12.905692265475997
21	0.00038648316653820474	12.905692242327962
22	0.00027158640868598234	12.905692236858021
23	0.0001283507384082397	12.90569223428358
24	8.867390193004231e-05	12.905692233610504


Gradient Descent at p=1.2
3.0
2.5
2.0
1.5
1.0
0.5
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00

Gradient Descent (p=1.2)

Iteration Number	Gradient Norm	Function Value
0	11.208240058911064	64.33623036000427
1	13.182079027850897	43.995899972045315
2	8.7428775943619	24.213167492450392
3	4.711456849447196	18.52993256004202
4	4.50179751034553	17.06558173092728
5	2.5650851025558086	16.095120784853307
6	1.7726236658517505	15.705392792341254
7	1.4612495085854595	15.541628072603716
8	1.2806281993912456	15.443102071290458
9	0.8563120002638069	15.375630909555412
10	0.807298473001589	15.340781907647184
11	0.5751415069468772	15.315218069594073
12	0.42012016684216935	15.299512674327445
13	0.4102699212255736	15.290538414085974
14	0.30393274291576805	15.283666255369674
15	0.2309364596744388	15.279441758963323
16	0.2328516744293254	15.276930528235978
17	0.17609184538669978	15.274945882011071
18	0.13407651398366224	15.273661927030979
19	0.10338480597230898	15.272837578073007
20	0.10551312676402584	15.272346615060405
21	0.0804789296706539	15.271949006464752
22	0.06171090323098455	15.271690890033863
23	0.04771733455335079	15.271523330230679
24	0.049347768720995944	15.271426875413926
25	0.03768733579981557	15.271344396788372
26	0.02891910547773627	15.271291025074976
27	0.02233652331056084	15.271256336787902
28	0.01735102220332262	15.271233491779805
29	0.01784134634708597	15.27121995327948
30	0.013670093353557348	15.271208750926727
31	0.01053266864436091	15.271201489757317
32	0.008159302777574503	15.271196728603071
33	0.006359083728560445	15.271193584844589
34	0.00650297488285792	15.271191672088777
35	0.004995860124131566	15.271190137491155
36	0.0038577959384004335	15.271189136541906
37	0.0029960033857848946	15.271188478326478
38	0.0031063591369356307	15.271188097574644
39	0.002379398142151332	15.271187770522298

40	0.0018314456755543266	15.271187558437415
41	0.0014173100563097114	15.27118741964422
42	0.0011030813379144695	15.271187328012378
43	0.0011369068238284074	15.271187273615729
44	0.00087238775721952	15.27118722834051
45	0.0006728423907852711	15.271187198869063
46	0.0005218038395608405	15.271187179503926
47	0.0004070503481822989	15.271187166675075
48	0.0004167928955001624	15.271187158858082
49	0.00032044075134533254	15.27118715255593
50	0.00024766770395969736	15.27118714843588
51	0.00019251623420110992	15.27118714571822
52	0.000199642825824025	15.27118714413839
53	0.0001530414484370621	15.271187142781294
54	0.00011791038760146187	15.271187141898825
55	9.134538776732743e-05	15.27118714131942

(3)

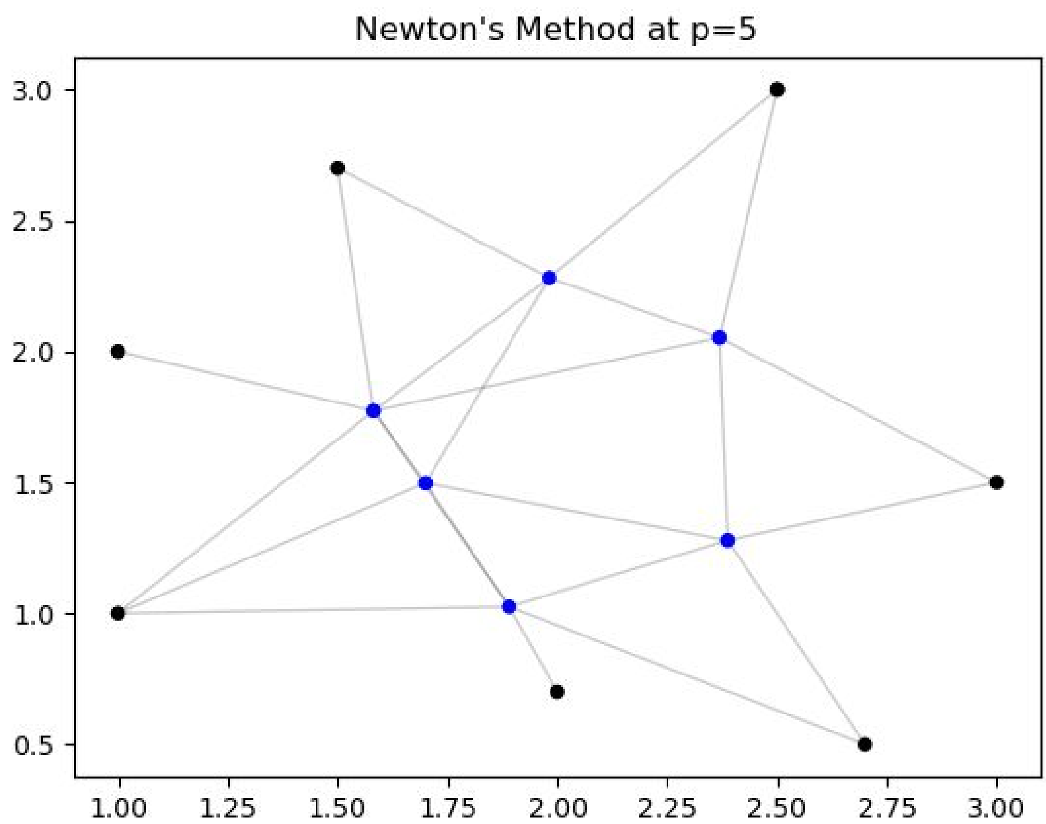
(b)

⇒ Found Hessian using the centered differences.

The observations are as follows:

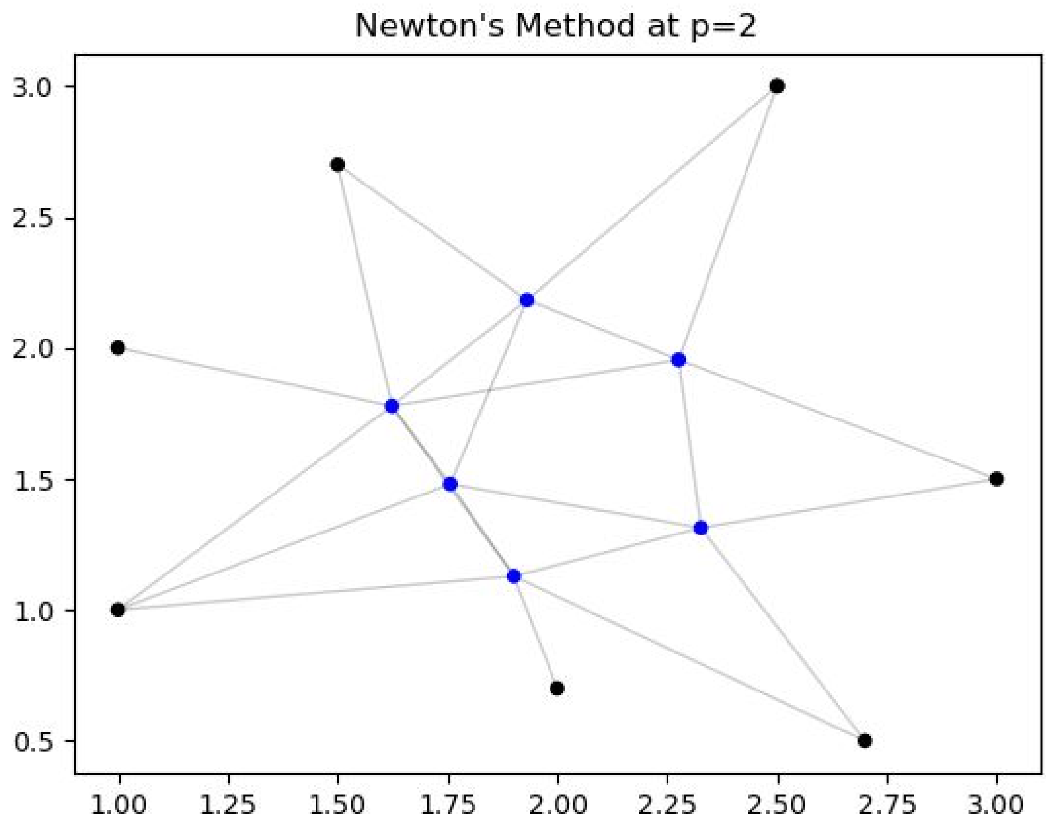
Newton's Method (p=5)

Iteration Number	Gradient Norm	Function Value
0	26909.489786636193	57568.74227426395
1	8559.146747615312	13946.831260245905
2	2730.9682193980702	3422.6520784888053
3	874.9823217815109	859.0687789913425
4	281.65920500685985	224.45297729007828
5	91.04009385323262	62.38015754035007
6	29.40844451314321	19.482083122884383
7	8.559556369859566	9.21252136893465
8	1.5806111588187577	7.8244291124991765
9	0.05519006587402884	7.760820834871707
10	6.185404719357273e-05	7.760734975371669



Newton's Method (p=2)

Iteration Number	Gradient Norm	Function Value
0	63.091468639277934	364.5543039392473
1	1.9981385694135693e-10	12.905692233186029



<u>Newton's Method (p=1.2)</u>

Iteration Number	Gradient Norm	Function Value
0	12.356630813692846	91.11234152982074
1	10.324180448892717	37.51444721050905
2	6.397603346365576	16.731341506661565
3	1.6918019307396321	15.446959126392866
4	0.274104689544287	15.277121629470772
5	0.014512908327198006	15.271196334055555
6	4.435685254581143e-05	15.27118714021955