Convergence of Gradient Descent after Homeomorphic Transformation

Vatsal Patel[1]
BS Mathematics

Abstract

Gradient Descent belongs to the class of optimization algorithms which is used to minimize a function. It is by far the most popular optimization strategy used in machine learning & deep learning. It is applied on a function obtained as a result of the error while training the model. Hence used to minimize the error. To approach this minimum at a faster rate is an important aspect of reducing standard computation time. To do the same we explore Homeomorphic Topological Spaces related to our function represented here as a topological space & perform gradient descent.

Keywords: Gradient Descent, Convex Optimization, Homeomorphic Transformation, Convergence, Homeomorphic Topological Space.

1.Introduction

Gradient Descent is usually performed on a Convex Surface. The same surface can be idealized as a Topological Space defined under Euclidean topology. One Can perform gradient descent on it and note down the steps taken to reach its minimum. As we have the understanding of Topological spaces, we can define a map in such a way that the surface will become more convex. Then we will possibly reduce the steps of gradient descent that were there in the original surface after mapping. Such a mapping must be homeomorphic in order to be one-one & onto, hence the original topological space was mapped to a more convex homeomorphic space. In the first part of the paper in Section 2 we have discussed a few terminologies we would be working with that includes gradient descent algorithm, homeomorphic topological space and inverse function theorem. Section 3 represents the theoretical approach in brief on how to reduce the steps of convergence & also verify how it works. While in Section 4 we will apply the theoretical approach and see how it goes with different surfaces and maps. Section 5 includes results and observations deduced from the examples. Then we have a short discussion on selection of the homeomorphic map in Section 6 based on the observations and then finally concluding with a small hypothesis in Section 7.

[1] E-Mail - vatsal.patel@xaviers.edu.in (Vatsal Patel)

2. Preliminaries

2.1. Gradient Descent Algorithm

Gradient descent algorithm is an iterative algorithm to optimize a differentiable function. To optimize is to minimize it, thus it will find the local minima of the function. The idea is to iteratively move in the direction opposite to the gradient of that function at the current point. Thus it will move in the direction towards the minima.

$x_{k+1} = x_k - \alpha\,(\nabla f)_{x_k}$ represent our gradient descent algorithm for a multivariable function f , $x \in R^n$ and $\alpha \in R$. α usually manages the step size of the descent(1-5).

2.2. Homeomorphic Topological Space.

A continuous function between topological spaces that has a continuous inverse function is called a Homeomorphism. They are the isomorphisms in the category of topological spaces, that is they are the mappings that preserve all the topological properties of a given space.

Two spaces with a homeomorphism between them are called homeomorphic and from a topological point of view are the same.

Definition of Homeomorphism:

Let (X,τ) and (Y,τ') be two topological spaces. A function $f : X \rightarrow Y$ between two topological spaces is homeomorphism if it has the following properties :

1) f is a bijection (one-one and onto)

2) f is continuous

3) the inverse function $f^{-1}$is continuous.

If such a function exists then X and Y are Homeomorphic Topological Spaces(6-12).

2.3. Inverse Function Theorem.

Theorem 2.3.1. Let U be an open set in R^n , and let $f : U \rightarrow R^n$ be continuously differentiable. Suppose that $x_0 \in U$ and $Df(x_0)$ is invertible i.e $det\, J_f \neq 0$.[2] Then there exists a smaller neighbourhood $V \ni x_0$ such that f is a homeomorphism onto its image. Furthermore, V may be taken small enough so that $f^{-1}$is

[2] det is short form of determinant & J_f represents for the Jacobian of function f

also continuously differentiable, with its derivative satisfying $D(f^{-1})_y = (Df)^{-1}{}_{f^{-1}(y)}$. Moreover, if $\underline{f}$ is of class $\underline{C^k}$, (k ∈ N ∪ {∞}), then so is $\underline{f^{-1}}$ [13-17].

3. Convergence of gradient descent post homeomorphic transformation

3.1 Methodology to approach the minimum faster.

Let $X \subseteq R^3, X \neq \emptyset$ be the surface on which we are performing gradient descent. Let d be the Euclidean metric on X. Define a Euclidean topology τ on X under euclidean metric d. Then (X,τ) is a Topological Space.

We assume that the surface X is represented as $X = \{(x,y,z)|\ x \in (a,b), y \in (c\ ,d), z = f(x,y)\}$ where $a,b,c,d\ \in R.$

Let there be some $Y \subseteq R^3, Y \neq\ \emptyset$.

Define a map $\Psi: X \longrightarrow Y$ as $\Psi(x,y,z)\ = (u(x,y,z), v(x,y,z), w(x,y,z))$, where $u,v,w : R^3 \rightarrow R$.

The idea here about the selection of the map Ψ should be such that we have an improved convergence on surface Y. It will be discussed further in section 3.

Check if $det\, J_\Psi\ \neq 0$, If true then according to Theorem 2.3.1 clearly Ψ is a homeomorphism.

So considering Ψ , surface Y will be represented as $Y = \{(m,n,o)|\ m \in (p,q), n \in (r\ ,s), o = g(m,n)\}$ where $p,q,r,s \in R$.

Define a Euclidean topology τ' on Y under euclidean metric d. Then (Y,τ') is a Topological Space. As a result (X,τ) and (Y,τ') are homeomorphic topological spaces perhaps the same.

Now to find Ψ^{-1} , we need to find some $\Omega(u,v,w)\ =\ (\ x(u,v,w)\ , y(u,v,w), z(u,v,w)\ )$ then $\Psi^{-1} =\ \Omega$ .This can be done once we have a map defined.

Perform gradient descent on surface Y and use the inverse map Ψ^{-1} on the last point obtained from the algorithm. Hence one can reach the minimum of surface X faster using this method.

3.2 To verify if the convergence is improved

To perform a gradient descent on a surface we usually choose a random point for the start but to compare if the convergence has improved post homeomorphic map we need to have some consistency.

For the same initially choose a random point $P_0 = (a,b,c)$ on the surface X. Perform gradient descent and note down the number of iterations to converge. Now map P_0 on surface Y using the map i.e $R_0 = \Psi(a,b,c)$ and now perform gradient descent starting off from point R_0 on surface Y, note down the number of iterations required to converge.

Now one can compare the between the number of steps on surface X to that on surface Y. Check out the results for the same with different surfaces in section 5.

4. Examples to showcase implementation of Section 3

4.1 Example using a simplest convex surface

Let surface X be a paraboloid surface, i.e say $X = \{(x, y, z)| \; x \in (-15,15), y \in (-15\;,15), z = x^2 + y^2\}$.Here $f(x, y) = x^2 + y^2$.

Consider the homeomorphic map $\Psi(x, y, z) \; = (x, y, \alpha z)$, where $\alpha \in R, \alpha > 0$. Then we have $det \, det J_\Psi \; = \alpha \neq 0$.

Thus (X, τ) and (Y, τ') are homeomorphic topological spaces. Where we have just scaled the z component of X by a factor of α. Here $Y = \{(m, n, o)| \; m \in (-15,15), n \in (-15\;,15), o = g(m, n)\}$ where clearly we have know that $Y = \Psi(X)$ and so $Y = \{(x, y, \alpha z)| \; x \in (-15,15), y \in (-15\;,15), \alpha z = \; \alpha(x^2 + y^2)\}$, Thus $m = x, n = y$ & $o \; = \alpha$z. So we have $o = g(x, y) = \; \alpha \, f(x, y)$.

Note that the 3^{rd} component mapped in Ψ preserves the initial representation of surface X, that is we have $g(x, y) = \; \alpha \, f(x, y)$.

One can now computationally perform the example mentioned above with different values of α. Check out Table 1 & Section 5 for interpretations & results.

Now to find Ψ^{-1} , as we already know that $\Psi(x, y, z) \; = \big(u(x, y, z), v(x, y, z), w(x, y, z)\big)$ where $u, v, w : R^3 \rightarrow R$.(We need to find some $\Omega(u, v, w) \; = \; (\; x(u, v, w) \; , y(u, v, w), z(u, v, w) \;)$ then $\Psi^{-1} = \; \Omega$).

For our case we have $u(x, y, z) = x$, $v(x, y, z) = y$, $w(x, y, z) = \alpha z$.

$u(x, y, z) \; = \; x \quad \Rightarrow \quad x \; = \; u \quad , \therefore x(u, v, w) \; = \; u$.

$v(x, y, z) \; = \; y \quad \Rightarrow \quad y \; = \; v \quad , \therefore \; y(u, v, w) \; = \; v$

$w(x, y, z) = \; \alpha z \quad \Rightarrow \quad z \; = \frac{w}{\alpha}, \therefore \; z(u, v, w) = \frac{w}{\alpha}$

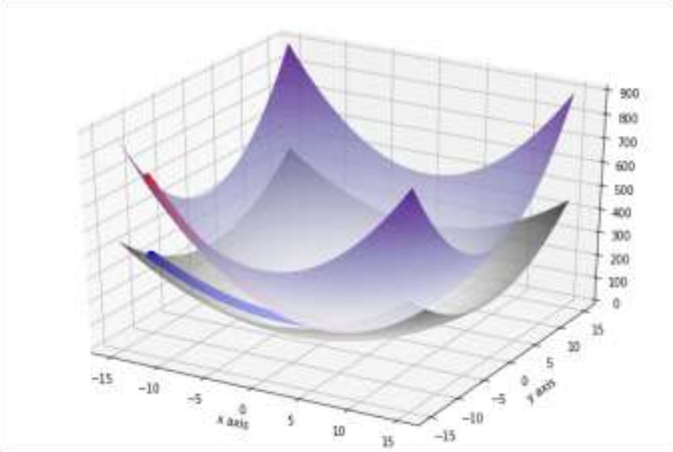
So $\Omega \, (u, v, z) = \; \big(x(u, v, w), y(u, v, w), z(u, v, w)\big) = \big(u, v, \frac{w}{\alpha}\big) \quad \therefore \quad \Psi^{-1}(x, y, z) \; = (x \; , y \; , \frac{z}{\alpha})$.

Clearly Ψ^{-1} is a continuous function. One can computationally verify using the inverse map on the final point obtained after performing gradient descent on surface Y by comparing it with the final point obtained after gradient descent on surface X[3]. If both these points are almost equivalent and at the same time the number of steps taken on surface Y are less than that on surface X then we can easily say that our homeomorphic map is working. Thus it is proved that the homeomorphic map can be implemented to reduce the steps of convergence and hence reduce computation time. Reach out to Table 4 for the same.

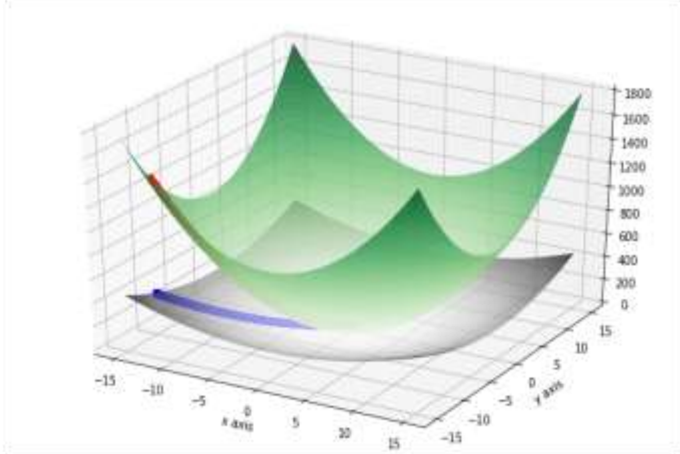
Table 1

[3] Using the terminologies Surface X and Space X for the same structure as mentioned in the theory in Section 3.

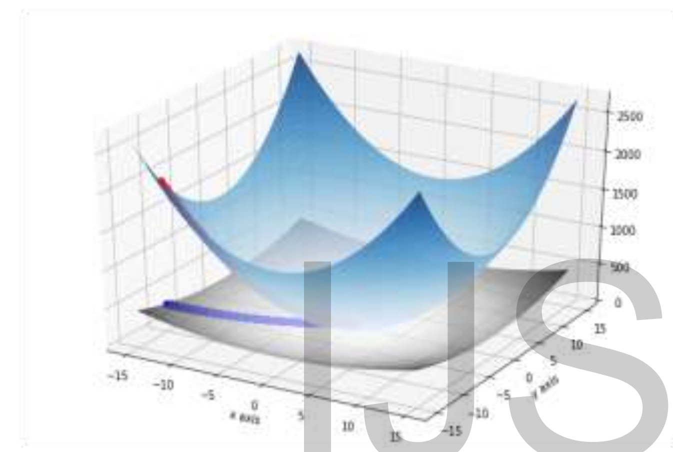
1. $\Psi(x, y, z) = (x, y, 2z)$



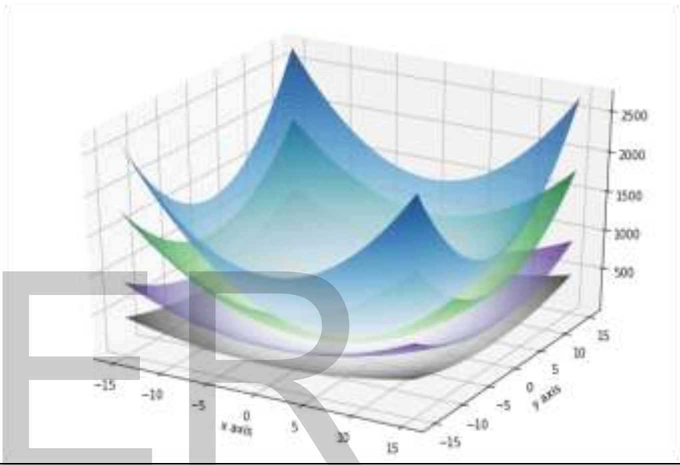
2. $\Psi(x, y, z) = (x, y, 4z)$



3. $\Psi(x, y, z) = (x, y\ 6z)$



All presented together.



The Greyish-White surface represent our Space X while the coloured surface represent for different values of α for Space Y. One can clearly see that after implementation of homeomorphic maps we are increasing the convexity of the Surface X and the same thing is reducing the number of steps taken to reach the minimum.

4.2 Example to reduce steps on a plane surface with non negative bounds

Let surface X be a Plane surface, i.e say $X = \{(x, y, z)|\ x \in (0,10), y \in (0,10), z = x + y\}$.Here $f(x, y) = x + y$.

Consider the homeomorphic map $\Psi(x, y, z) = (x, y, z^\beta)$, where $\beta \in R, \beta > 0$. Then we have $det\ det J_\Psi = \beta z^{\beta-1}$. For $det\ det J_\Psi$ to be non zero we have already taken care a our plane is having x and y in the domain of open interval 0 to 10 and as a result z = x + y can never be 0, $\therefore det\ det J_\Psi \neq 0$.

Thus (X, τ) and (Y, τ') are homeomorphic topological spaces. Here $Y = \{(m, n, o)|\ m \in (0,10), n \in (0\ ,10), o = g(m, n)\}$ where clearly we have know that $Y = \Psi(X)$ and so $Y = \{(x, y, z^\beta)|\ x \in (0,10), y \in (0\ ,10), z^\beta = (x + y)^\beta\}$, Thus $m = x, n = y\ \&\ o = z^\beta$. So we have $o = g(x, y) = (f(x, y))^\beta$.

Note that the 3^{rd} component mapped in Ψ preserves the initial representation of surface X, that is we have $g(x, y) = (f(x, y))^\beta$.

One can now computationally perform the example mentioned above with different values of β. Check out Table 2 & 4 for interpretations. For $0 < \beta < 1$ we did not improve the convergence, while it is improved for $\beta > 1$.

Now to find Ψ^{-1} , as we already know that $\Psi(x,y,z) = \big(u(x,y,z), v(x,y,z), w(x,y,z)\big)$ where $u, v, w : R^3 \to R$.

For our case we have $u(x,y,z) = x$, $v(x,y,z) = y$, $w(x,y,z) = z^{\beta}$.

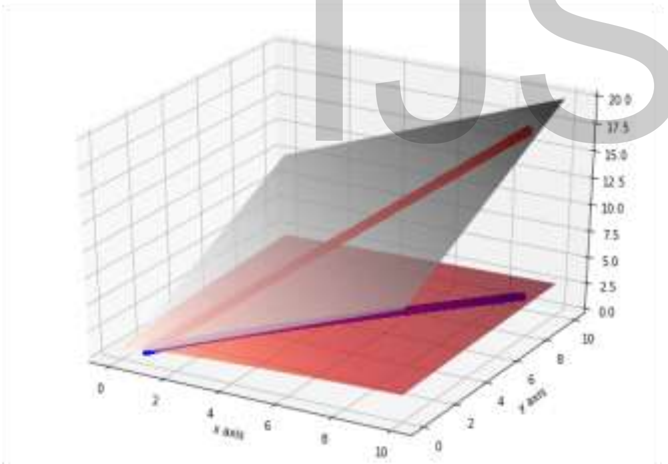
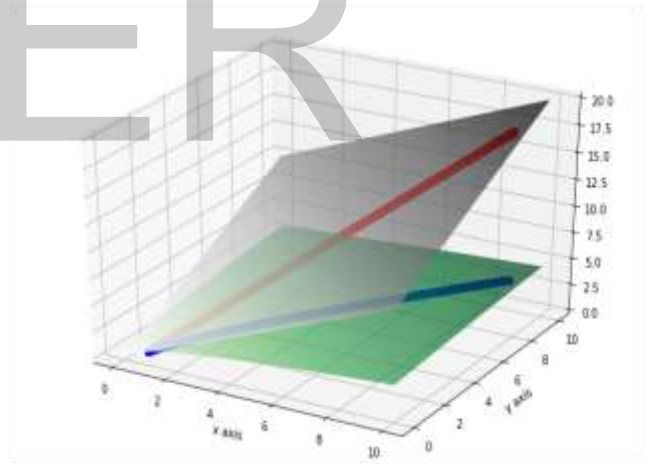
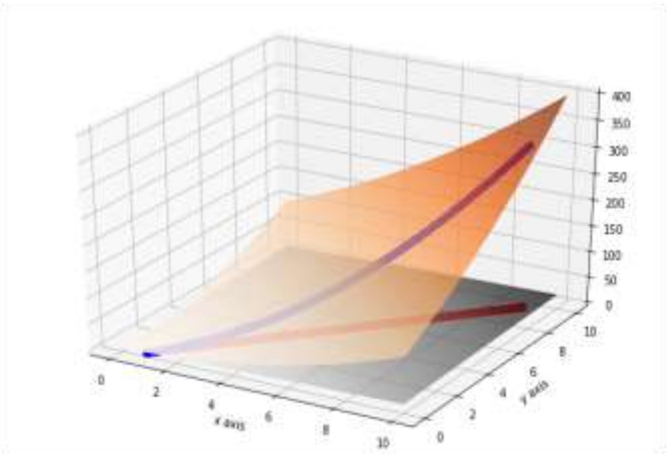
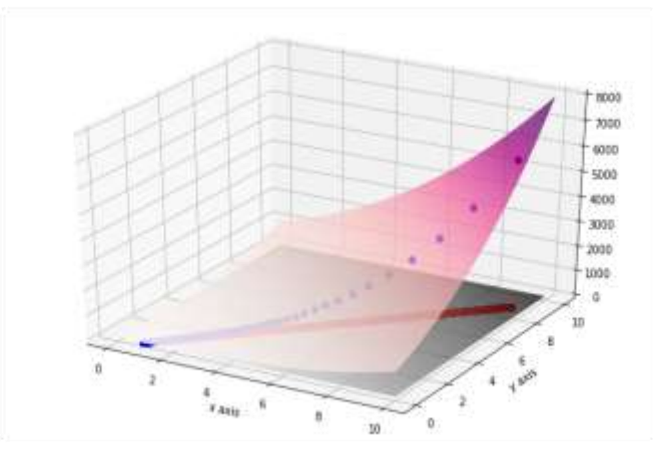
$u(x,y,z) = x \quad \Rightarrow \quad x = u \quad , \therefore x(u,v,w) = u$.

$v(x,y,z) = y \quad \Rightarrow \quad y = v \quad , \therefore y(u,v,w) = v$

$w(x,y,z) = z^{\beta} \quad \Rightarrow \quad z = w^{\frac{1}{\beta}}, \quad \therefore z(u,v,w) = w^{\frac{1}{\beta}}$

So $\Omega\,(u,v,z) = \big(x(u,v,w), y(u,v,w), z(u,v,w)\big) = \left(u, v, w^{\frac{1}{\beta}}\right) \quad \therefore \Psi^{-1}(x,y,z) = (x\,, y\,, z^{\frac{1}{\beta}})$. Clearly Ψ^{-1} is a continuous function. Using the inverse map we can verify if it converges to the same point on surface X by comparing the inverse of the final point on surface/space Y .

Table 2

4. $\Psi(x,y,z) = (x,y,z^{\frac{1}{3}})$ 	5. $\Psi(x,y,z) = (x,y,z^{\frac{1}{2}})$ 
6. $\Psi(x,y,z) = (x,y,z^{2})$ 	7. $\Psi(x,y,z) = (x,y,z^{3})$ 

The Greyish-White surface represent our Space X while the coloured surface represent for different values of $\beta$for Space Y. The implementation of a homeomorphic map on such plane surfaces is increasing the curvature of the inclined plane and hence reducing the steps taken to descend to its minima i.e point with lowest value on this surface.

4.3 Example to generalize the idea of selection of homeomorphic map

Let surface X be a Convex surface, i.e say $X = \{(x, y, z) |\, x \in (-5,2), y \in (-5,5), z = \sin \sin x\ \}$.Here $f(x, y) = \sin \sin x$.

Consider the homeomorphic map $\Psi(x, y, z) = (x, y, z^{\beta} + \alpha z + \gamma)$, where $\alpha, \beta, \gamma \in R, \alpha > 0, \beta > 1$. Then we have $\det det J_{\Psi} = \beta z^{\beta-1} + \alpha$. Check for the condition when $\beta z^{\beta-1} + \alpha = 0$, i.e $z = \sqrt[\beta-1]{-\frac{\alpha}{\beta}}$.

Depending on what z is we should check if the domain of x and y results in z=0. If it doesn't then we can conclude that $\det det J_{\Psi} \neq 0$.

In our case $z = \sin \sin x$, so $x = z$, $\therefore x = (\sqrt[\beta-1]{-\frac{\alpha}{\beta}})$. Hence try different values of α, β such a way that the value of x for which $\det det J_{\Psi} \neq 0$ doesn't lie in the domain of x.

Thus (X, τ) and (Y, τ') are homeomorphic topological spaces. Where we have just scaled the z component of X by a factor of α. Here $Y = \{(m, n, o) |\, m \in (-5,2), n \in (-5\ ,5), o = g(m, n)\}$ where clearly we have know that $Y = \Psi(X)$ and so $Y = \{(x, y, z^{\beta} + \alpha z + \gamma) |\, x \in (0,10), y \in (0\ ,10), z^{\beta} + \alpha z + \gamma = (\sin \sin x\ )^{\beta} + \alpha \sin \sin x + \gamma\}$, Thus $m = x, n = y$ & $o = z^{\beta} + \alpha z + \gamma$. So we have $o = g(x, y) = \left(f(x, y)\right)^{\beta} + \alpha f(x, y) + \gamma$.

Note that the 3^{rd} component mapped in Ψ preserves the initial representation of surface X, that is we have $g(x, y) = \left(f(x, y)\right)^{\beta} + \alpha f(x, y) + \gamma$.

One can now computationally perform the example mentioned above with different values of α & β. Check out Table 3 & 4 for interpretations. For $0 < \beta < 1$ we did not improve the convergence, while it is improved for $\beta > 1$.

Now to find Ψ^{-1} , as we already know that $\Psi(x, y, z) = \left(u(x, y, z), v(x, y, z), w(x, y, z)\right)$ where $u, v, w : R^3 \to R$.

For our case we have $u(x, y, z) = x$, $v(x, y, z) = y$, $w(x, y, z) = z^{\beta} + \alpha z + \gamma$.

$u(x, y, z) = x \quad \Rightarrow \quad x = u \quad , \therefore x(u, v, w) = u$.

$v(x, y, z) = y \quad \Rightarrow \quad y = v \quad , \therefore y(u, v, w) = v$

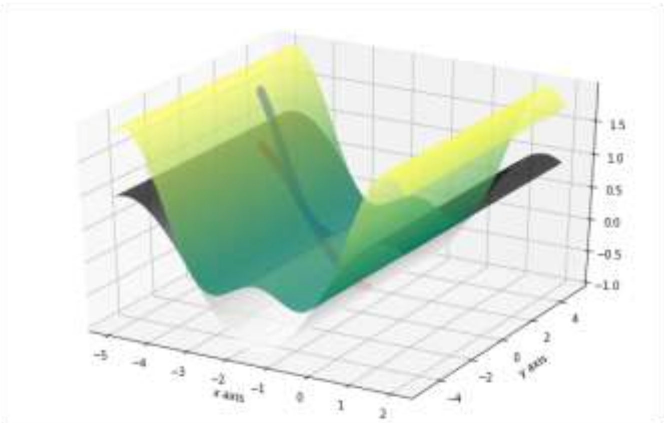
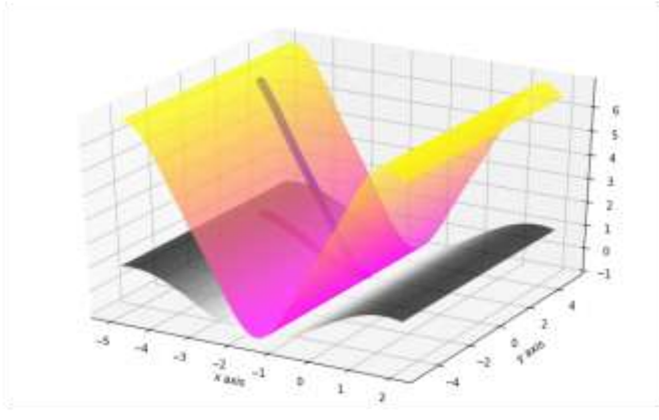
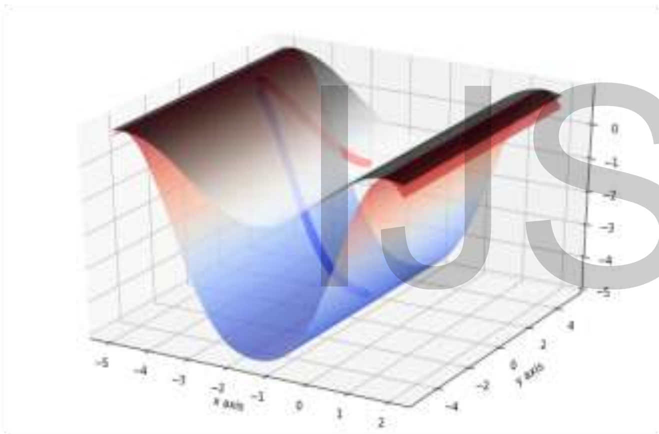
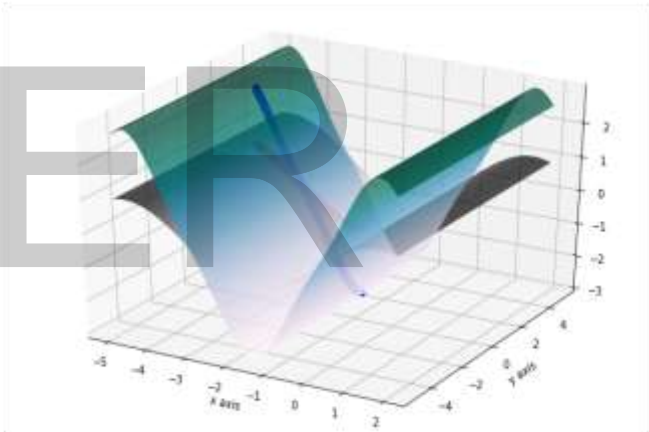
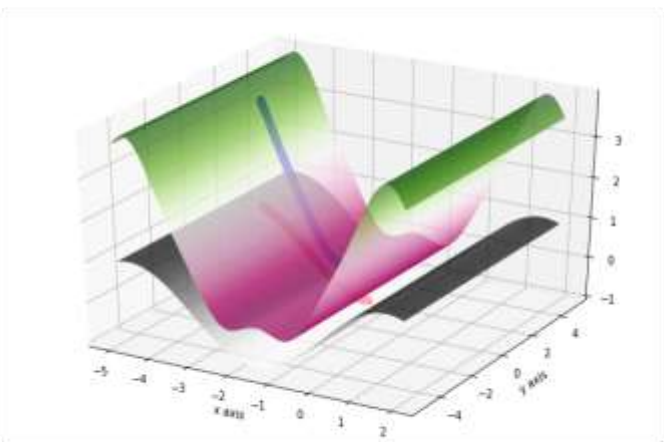
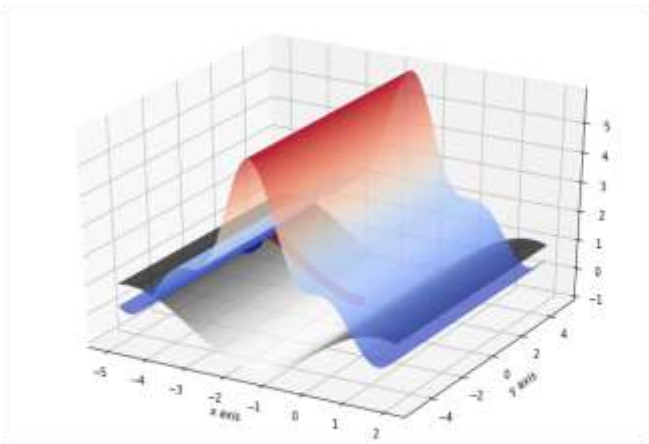
$w(x, y, z) = z^{\beta} + \alpha z + \gamma \quad \Rightarrow \quad z = \frac{w-\gamma-z^{\beta}}{\alpha}$, as surface X is basically , $z = \sin \sin x$ & $x = u$

$\Rightarrow z = \frac{w-\gamma-u)^{\beta}}{\alpha} \qquad \therefore z(u, v, w) = \frac{w-\gamma-u)^{\beta}}{\alpha}$

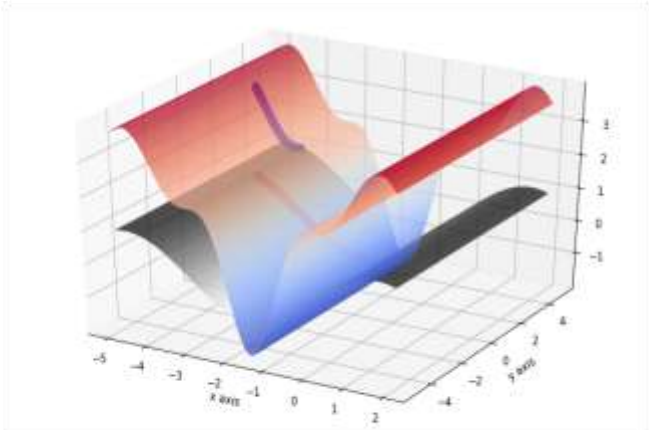
So $\Omega\ (u, v, z) = \left(x(u, v, w), y(u, v, w), z(u, v, w)\right) = \left(u, v, \frac{w-\gamma-u)^{\beta}}{\alpha}\right)$

$\therefore \Psi^{-1}(x, y, z) = (x\ , y\ , \frac{z-\gamma-x)^{\beta}}{\alpha})$. Clearly Ψ^{-1} is a continuous function. Thus use the same to verify the verify the inverse of the final point on surface Y that whether it is equivalent to that of final point on surface X

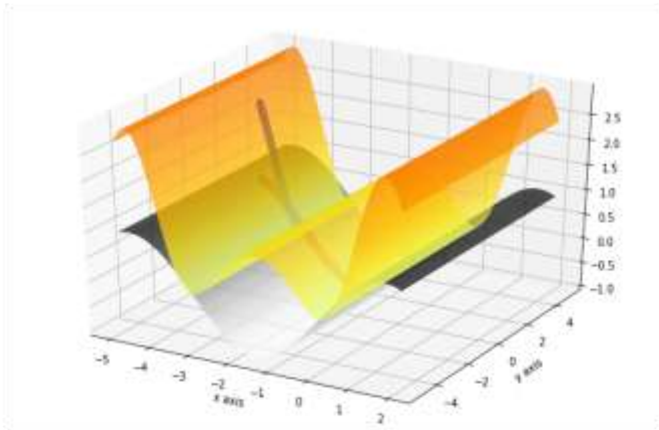
Table 3

8. $\Psi(x,y,z) = (x,y,z^2+z)$ 	9. $\Psi(x,y,z) = (x,y,z^2+3z+3)$ 
10. $\Psi(x,y,z) = (x,y,z^2+3z-3)$ 	11. $\Psi(x,y,z) = (x,y,z^3+2z)$ 
12. $\Psi(x,y,z) = (x,y,z^4+2z+1)$ 	13. $\Psi(x,y,z) = (x,y,-z^4+3z^3+2)$ 

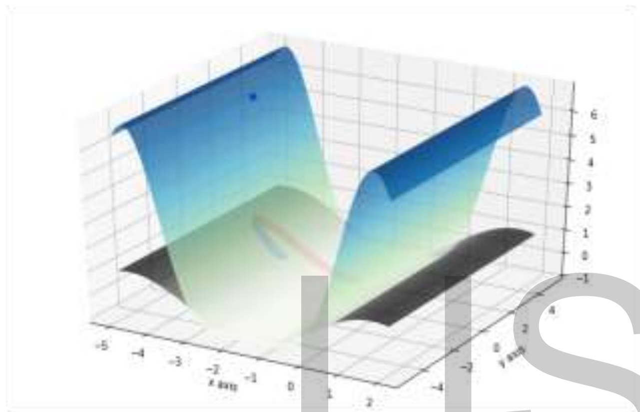
14. $\Psi(x,y,z) = (x,y,z^4 - 3z^3 + 2)$



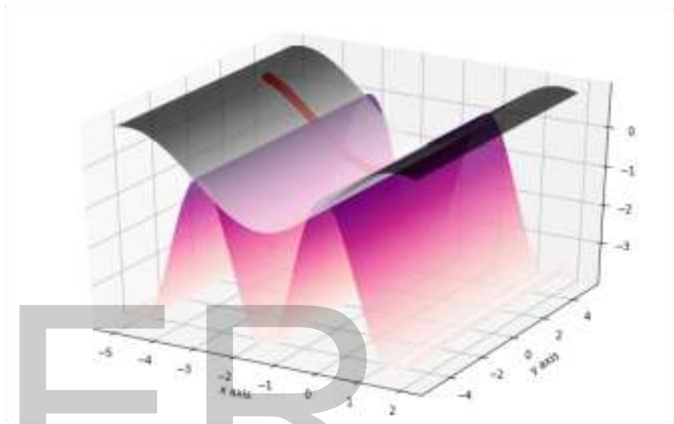
15. $\Psi(x,y,z) = (x,y,z^3 + z^2)$



16. $\Psi(x,y,z) = (x,y,z^3 + 2z^2 + 3z)$



17. $\Psi(x,y,z) = (x,y,z^3 - 4z^2 - z)$



The Greyish-White surface represents Space X while the coloured surface represents different homeomorphic maps for Space Y. It can be often noticed that few types of map result in two or more local minima at the same time it shows reduction in the number of steps , such surfaces are not useful to us. While at the same time we might also increase the convexity so much that the actual area of the minima will increase with the homeomorphic map and it will again converge with fewer steps , such maps are also of no use.Also the Homeomorphic map which works might change depending on the type of our surface X ,However one can always deduce what will work after a few trail runs.

5. Results & Observations

Table 4

Sr.no	$\Psi: X \longrightarrow Y \,.\, X,Y \subseteq R^3$ $\Psi(x,y,z) =$	*Steps in X*	*Final point in X*	*Steps in Y*	*Final point in Y*	$\Psi^{-1}: Y \longrightarrow X \,.\, X,Y \subseteq R^3$ $\Psi^{-1}(x,y,z) =$	*Inverse of Final point in Y*	*Map useful wrt X*
1	$(x,y,2z)$	20756	(-3.46e-18, -3.46e-18, 2.40e-35)	10367	(-3.46e-18, -3.45e-18, 4.79e-35)	$(x,y,\frac{z}{2})$	(-3.46e-18, -3.45e-18, 2.39e-35)	Yes
2	$(x,y,4z)$	20756	(-3.46e-18, -	5174	(-3.42e-18, -3.46e-18,	$(x,y,\frac{z}{4})$	(-3.42e-18, -3.46e-18,	Yes

			3.46e-18, 2.40e-35)		9.49e-35)		2.37e-35)	
3	$(x, y, 6z)$	20756	(-3.46e-18, -3.46e-18, 2.40e-35)	3442	(-3.43e-18, -3.40e-18, 1.40e-34)	$(x, y, \frac{z}{6})$	(-3.43e-18, -3.40e-18, 2.33e-35)	Yes
4	$(x, y, z^{\frac{1}{3}})$	8594	(0.9033, 0.00031, 0.90364)	111472	(0.9032, 0.00028, 0.96676)	(x, y, z^3)	(0.9032, 0.00028, 0.90357)	No
5	$(x, y, z^{\frac{1}{2}})$	8594	(0.90332, 0.00031, 0.90364)	50736	(0.90351, 0.00050, 0.95080)	(x, y, z^2)	(0.90351, 0.00050, 0.90402)	No
6	(x, y, z^2)	8594	(0.90332, 0.00031,0.90364)	746	(0.90463, 0.00162, 0.82130)	$(x, y, z^{\frac{1}{2}})$	(0.90463, 0.00162, 0.90625)	Yes
7	(x, y, z^3)	8594	(0.90332, 0.00031, 0.90364)	171	(0.90368, 0.00067, 0.73966)	$(x, y, z^{\frac{1}{3}})$	(0.90368, 0.00067, 0.90436)	Yes
8	$(x, y, z^2 + z)$	17183	(-1.57079, 1.85618, -0.9999)	5972	(-2.6179, 1.85618, -0.24999)	(x, y, zx) *Last component depends on type of surface X*	(-2.6179, 1.8561, -0.5000)	No
9	$(x, y, z^2 + 3z + 3)$	17183	(-1.57079, 1.85618, -0.99999)	3205	(-1.57079, 1.85618, -0.99999)	$(x, y, \frac{,zx - 3}{3})$	(-1.57079, 1.85618, -1.6666)	Yes
10	$(x, y, z^2 + 3z - 3)$	17183	(-1.57079, 1.85618, -0.99999)	13989	(-1.57079, 1.85618, -4.99999)	$(x, y, \frac{,zx + 3}{3})$	(-1.57079, 1.85618, -0.99999)	Yes
11	$(x, y, z^3 + 2z)$	17183	(-1.57079, 1.85618, -0.99999)	4025	(-1.57079, 1.85618, -2.9999)	$(x, y, \frac{zx}{2})$	(-1.57079, 1.85618, -0.66666)	Yes
12	$(x, y, z^4 + 2z + 1)$	17183	(-1.57079, 1.85618, -0.99999)	3406	(-2.22472, 1.85618, -0.19055)	$(x, y, \frac{zx - 1}{2})$	(-2.22472, 1.85618, -0.79370)	No
13	$(x, y, -z^4 + 3z^3 + 2)$	17183	(-1.57079, 1.85618, -0.99999)	13357	(-3.14159, 1.85618, 2.0)	$(x, y, \frac{z}{x+x+(\frac{2}{sinx})})$	(-3.14159, 1.856187, 2.45e-06)	No
14	$(x, y, z^4 - 3z^3 + 2)$	17183	(-1.57079, 1.85618, -0.9999)	2987	(-4.71238, 1.85618, 4.92e-14)	$(x, y, \frac{z}{x-x+(\frac{2}{sinx})})$	(-4.7123, 1.856187, 0.62359)	No
15	$(x, y, z^3 + z^2)$	17183	(-1.5707, 1.85618, -	2220	(-3.14159, 1.85618,	$(x, y, \frac{z}{x+3sinx})$	(-3.1415, 1.8561, -	No

			0.9999)		9.55e-13)		4.60e-07)	
16	$(x, y, z^3 + 2z^2 + 3z)$	17183	(-1.5707, 1.85618, -0.9999)	3584	(-3.52609, 1.85618, -0.56249)	$(x, y, \frac{zx - x}{3})$	(-3.52609, 1.85618, -0.29889)	No
17	$(x, y, z^3 - 4z^2 - z)$	17183	(-1.5707, 1.85618, -0.9999)	0	(-4.30027, 1.85618, -3.43495)	$(x, y, x - 4x - z)$	(-4.30027, 1.85618, 0.84598)	No

Map-1,2,3 represents section 4.1. It is quite clear from the observations that implementing a map that scales the Z-component is a useful approach as it reduces the steps by the scaling factor.

Map-4,5,6,7 represents section 4.2. It can be observed that for $0 < \beta < 1$ the steps of convergence are much more on surface Y than on X, while for $\beta > 1$ it seems much more applicable. One must keep in mind that they must only use these maps after understanding the domain of the surface, because the inverse of a squared z component will only give the positive values and we might have to put some filter in the inverse map for the same to avoid problems.

Map 8 to 17 represents the surface of Section 4.3. It gives us a much better idea about generalising the idea of selection of homeomorphic maps i.e the z component must be a function of z itself. However it still depends on the type of surface X, in our case it was z=sin(x) which had a specific restricted domain for x and y. But the same results would go along with any other similar case as well where the surface is convex in its domain of x and y added that the random start point for gradient descent must be around the vicinity of the minima as seen sometime it might converge to other minimas of the surface it it existed. The vivid observations definitely would let us hypothesize our selection of the homeomorphic map discussed ahead.

6. Discussion on selection of Homeomorphic map

Keeping in note that we begin with $\Psi(x, y, z) = (u(x, y, z), v(x, y, z), w(x, y, z))$, However we did not end up using a lot of different types of u(x,y,z) & v(x,y,z) because it will just change the domain of the surface post the homeomorphic transformation & depending on the same it will extend or contract in the euclidean plane.Still to explain how it would go assume homeomorphic map of the form $\Psi(x, y, z) = (x^{\mu} + \lambda x^{\kappa} + \delta, y^{\xi} + \rho y^{\varkappa} + \eta, z^{\beta} + \alpha z^{\sigma} + \gamma)$, where $\mu, \xi, \lambda, \delta, \rho, \kappa, \varkappa, \eta, \alpha, \beta, \sigma, \gamma \in R$. Then we can deduce the following:

1) μ, ξ will contribute to increasing the domain of x & y respectively by a large factor and as a result the surface may extend or reduce in the plane by large amounts.
2) When μ, ξ=0 & $\kappa, \varkappa = 1$, then λ, ρ will be responsible for scaling the domain of x & y respectively and as a result of that the surface will extend or reduce.

3) The δ, η will only contribute to the shift of the surface's location. So one can select $\mu, \xi, \lambda, \delta, \rho, \eta$ as per they want to manipulate the domain.
4) For improving the convergence one can go with the values as $\mu, \xi, \delta, \eta = 0$ and $\kappa, \varkappa, \lambda, \rho = 1$ and focus more on selection of $\alpha, \beta, \sigma, \gamma$ added that we must have $\alpha > 0, \beta > 1$ as per our observations.
5) However selection of $\alpha, \beta, \sigma, \gamma$ still depends on the type of surface X, but until it has a convexity and it is expected to have only one minima in the domain then one can expect that scaling the z component with $\alpha > 0$ or increasing the power of z component with $\beta > 1$ will effectively improve the convergence.
6) One should keep in mind that the starting point for gradient descent should be in the vicinity of the minima.Else there is a possibility that it will converge to another local minima post homeomorphic tr

More over one can generalize that the mapping of z component in the homeomorphic map should be such that it is represented as a function of z i.e $F(z)$ or a polynomial of z i.e $P_n(z)$ for an improved convergence.One can also get a slight glance that if the function or the polynomial of z has a term of even order then there are high chances that the homeomorphic space would have more than one minima. Also if one is concerned about calculation of large numbers resulting from such mapping they can normalize or min-max scaling(i.e scale values between 0 to 1) the values of z of surface X and then take the homeomorphic map. Normalizing doesn’t change the distribution of the data hence the structure of the surface X will be preserved, it's like a pre-processing step one can take before taking the homeomorphic transformation.

To note that we also have neglected the case for surfaces with more than one minima as our original surface on which the gradient descent is performed. Taking homeomorphic transformation for surfaces with more than one local minima would increase the chance of the gradient descent algorithm to descend in either of those minimas and it is also a tough choice to select the random point in such a scenario.

7. Conclusion

Gradient descent iteratively converges to the local minima of a given surface. The same surface can be idealised as topological space under euclidean topology. By defining a homeomorphic map one can find another topological space which is identically the same. We can perform gradient descent on this new surface, at the same time reducing the number of steps taken to reach its local minima.So what one can do is directly perform gradient descent on the surface obtained from homeomorphic transformation (which ideally is supposed to be more convex) and than use the inverse map on the final point obtained in the descent which would map back to the minima of our original surface.Thus one can use the same method to iteratively converge to the local minima with less number of steps, hence saving the standard computation time.

Along the same lines we have also explored the selection of homeomorphic maps which would result in reduction of steps of gradient descent. We can conclude that it heavily relies on the type of primary surface & based on that we can increase it's convexity by taking a homeomorphic map such a way that the z-component of the map should be mapped to a function of z i.e $F(z)$ or either to a Polynomial of z i.e $P_n(z)$. Also we can conclude that this function or the polynomial of z having an odd order(power) would definitely result in a homeomorphic surface having less number of steps required to converge to its minima.

8. References

[1] Gradient Descent Algorithm . https://en.wikipedia.org/wiki/Gradient_descent

[2] Linear Algebra and Its Applications by Gilbert Strang.

[3] Convex Optimization: Algorithms and Complexity by Sébastien Bubeck

[4] Lecture on gradient descent by Prof. Gilbert Strang: https://ocw.mit.edu/courses/mathematics/18-065-matrix-methods-in-data-analysis-signal-processing-and-machine-learning-spring-2018/video-lectures/lecture-22-gradient-descent-downhill-to-a-minimum/

[5] Ruder, Sebastian. "An overview of gradient descent optimization algorithms." ArXiv abs/1609.04747 (2016): n. pag.

[6] Introduction to Topology by Bert Mendelson

[7] Topological spaces by Gerard Buskes

[8] Martin N.M., Pollard S. (1996) Homeomorphisms. In: Closure Spaces and Logic. Mathematics and Its Applications, vol 369. Springer, Boston, MA. https://doi.org/10.1007/978-1-4757-2506-3_5

[9] Xu, K., Chen, H., Liu, S., Chen, P., Weng, T., Hong, M., & Lin, X. (2019). Topology Attack and Defense for Graph Neural Networks: An Optimization Perspective. ArXiv, abs/1906.04214.

[10]Zaman, B., Ramos, L.M., & Beferull-Lozano, B. (2020). Online Joint Topology Identification and Signal Estimation with Inexact Proximal Online Gradient Descent. arXiv: Signal Processing.

[11] Homeomorphism . https://en.wikipedia.org/wiki/Homeomorphism

[12] Allendoerfer, Carl B. (1974). "Theorems about Differentiable Functions". Calculus of Several Variables and Differentiable Manifolds. New York: Macmillan. pp. 54–88.

[13] Baxandall, Peter; Liebeck, Hans (1986). "The Inverse Function Theorem". Vector Calculus. New York: Oxford University Press. pp. 214–225.

[14] Protter, Murray H.; Morrey, Charles B., Jr. (1985). "Transformations and Jacobians". Intermediate Calculus (Second ed.). New York: Springer. pp. 412–420.

[15] "Derivative of Inverse Functions". Math Vault. 2016-02-28. Retrieved 2019-07-26.

[16] Jaffe, Ethan. MIT OCW "Inverse Function Theorem" (PDF).

[17] Inverse Function Theorem. https://en.wikipedia.org/wiki/Inverse_function_theorem

Acknowledgments

I would like to thank my family and friends for the love & support they have given me. I would also like to thank the department of mathematics at St. Xavier's College Mumbai for giving me an opportunity to do a project which led me to develop this theory.

Supplementary Content

I have attached a supplementary code file which can recreate the same experiment. However one should keep in mind that as I have used a random function to select the starting point for gradient descent, it would lead to different results. https://github.com/Science1804/Convergence-of-Gradient-Descent-after-Homeomorphic-Transformation