**Foundations of Deep Learning – Homework Assignment #3**Adi Album & Tomer Epshtein

**Part 2: (3)**

Question:

Let be a twice continuously differentiable convex loss, overparameterized by a depth linear neural network with hidden widths :

Prove the following claims:

1. If we restrict the domain of to , where , then we obtain a smooth (Lipschitz gradient) function.
2. If we do not restrict the domain of , the function is smooth if and only if at least one of the following conditions hold:
   1. is constant
   2. is affine and the depth is .

Proof (A):

Let .

Denote by for and matrix otherwise.

Let’s take a look at the expression:

We saw in class:

For , let’s look at

Yielding

We’ll add and subtract terms…

So:

Denote , so

We will bound each separately.

Propositions:

1. , and :  
   Notice that if we denote we obtain:
2. , and :

We will use these propositions to complete our proof. We’ll prove them later.  
Note:

1. is a twice continuously differentiable convex function, is compact and convex and therefore is Lipschitz gradient. I.e there exists , such that :

(This is true by claim in course’s recap notes on convexity and Lipschitzness)

1. is a twice continuously differentiable function, is compact,  
   therefor is bounded on . I.e. there exists such that :

Bound on :

* By proposition 1:
* because where (from proposition 2),  
  so, by note (ii)
* because where   
  (from proposition 2), so

Bringing it together:

Bound on :

* because where   
  (from proposition 2), so
* by note (i):  
  Now, by proposition 1:
* because where   
  (from proposition 2), so

Bringing it together:

Bound on :

* because where   
  (from proposition 2), so
* because where (from proposition 2),  
  so, by note (ii)
* By proposition 1:

Bringing it together:

So:

Overall:

So is a smooth -Lipschitz-gradient function with .

We will now prove the two remaining propositions:

1. , and :  
   Notice that if we denote we obtain:
2. , and :

Proof of proposition 1:

Proof by induction:

* :  
  Immediate.
* :  
  Let :

Overall:

Proof of proposition 2:

Proof by induction

* Immediate.
* Let :

So

Proof (B):

Assume that is Lipschitz smooth on , I.e.

Let , so:

So,

We’ll add and subtract terms:

Overall:

First, we assume . We would like to prove is affine.  
So . Choose

* Define a series of matrices :  
  Where are arbitrary matrices (const w.r.t )
* Define a series of matrices   
  Where is the identity matrix.

For every , inequality holds for   
So:



So:

, but is constant w.r.t   
so   
 is affine.

Now, we assume . We would like to prove is constant.

. Choose

* Define a series of matrices :  
  Where is the identity matrix
* Define a series of matrices   
  Where is an arbitrary matrix
* For all , define a series of matrices :

For every , inequality holds for   
So:



So,

, but is constant w.r.t so   
 is const.

It remains to prove the opposite direction:

1. is constant is smooth
2. is affine and is smooth

Proof:

2. is affine

In either case . This will be the only assumption we need.

We saw in “Proof (A)”:

So in particular, selecting :

is -smooth on .