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

**Part 2: (1)**

Question:  
In this question we extend the convergence of stationary point result delivered in class from gradient descent to stochastic gradient descent.  
Let be a twice continuously differentiable and -smooth function that attains its global minimum  
. Suppose we run stochastic gradient descent over :

Where is a (possibly time varying) learning rate, and stands for a time-independent noise with zero mean () and variance (). Let .  
Assume , and derive an upper bound on the number of iterations needed for reaching an expected  
-stationary point, i.e. on that will ensure:

Proof:

Let be a twice continuously differentiable and -smooth function that attains its global minimum  
.  
Reminder: In class we proved the following lemma:  
Lemma:  
Let be twice continuously differentiable and -smooth. Then:

So, for any :

We’ll take expected value over :  
Denote- , so

is independent of :

Plugging in the SGD step definition

We achieve:

, because , and linearity of expected value with inner product.

Now, let’s focus on :

Here, once again, . And by ’s distribution’s definition.

So we have:

Plugging in the result:

Plugging in :

Assume that for steps we didn’t achieve an expected ()-stationary point.  
I.e.  for all .

For any random variable : .  
So, for steps :

So:

This inequality holds for all . So:

Since is ’s global minimum,

So we have

Where , so:

So, for any an expected -stationary point will be reached within a number of steps at most