

**Question 1:** Between Gradient Descent and Newton's Method, which method of minimizing risk converges faster, and why?

**Question 2:** What does 'convergence at a linear rate' refer to in the reading?

**Question 3:** When discussing strongly convex problems, what does the 'condition number' refer to? Why is it important/significant?

**Answer Question 1:** Newton's method converges faster to the global minimum, since it is a second-order optimization method (uses the 2nd derivative of the given function  $f$ ). Compared to gradient descent which is a first-order method

**Answer Question 2:** When a method is described as 'converging at a linear rate', the method must be utilizing a constant step size, and converges exponentially quickly to the problem's minimum risk.

**Answer Question 3:** The condition number encodes how hard a strongly convex problem is to solve. This is also partially due to the fact that the number of iterations ( $T$ ) to be used, is dependent on this ratio. ( $k = L/\mu$ ), where  $k$  is the condition number. This relationship between  $T$  and  $k$  can lead to scaling issues later on.