Question 1: What curve shape is most ideal for gradient descent?

Question 2: What is the behavior of the ideal condition number in a gradient descent algorithm?

Question 3: What benefits does the Newton method pose over standard gradient descent?

Answer Question 1: Strongly convex

Answer Question 2: The condition number adjusts how far the gradient descent moves the algorithm. A small condition number descends slowly, requiring many iterations but remains accurate and won't overshoot by much. A larger condition number descends quite fast, requiring less iterations, but risks overshooting continuously as it lacks the fine granularity.

Answer Question 3: The newton method is far more accurate with faster descents due to the fact it uses second derivatives, however, it requires more preemptive computation to find that Hessian Matrix