Question 1: What are the benefits and disadvantages of using Newton's method vs gradient descent?

Question 2: Is gradient descent always guaranteed to converge to the global minima?

Question 3: What is the condition number of the problem, and what does it mean?

Answer Question 1: While Newton's method may converge faster, it requires creating a matrix of the second partial derivatives of the function. This operation is much more computationally expensive than gradient descent, which only uses the first order derivative.

Answer Question 2: No.

Answer Question 3: The condition number is defined as $\frac{L}{\alpha}$, and it measures the difficulty of a solution to the convex problem. A higher number will indicate a hard task, while a lower number indicates an easier one.