We are grateful to the reviewers for their time and comments. For the reviewers’ convenience, we briefly state below the novel contributions of our work, as summarized by R4 (whom we thank for the expert summary):

“This paper shows that under mild conditions, SGD converges to a critical point of general non-convex functions and avoids all strict saddle points, with probability 1. It also presents a convergence rate analysis of SGD once it enters the neighborhood of a local minimum.”

In what follows, we address the reviewers’ comments in order, tagging the reviewers concerned in each as #RX.

#R1: From high probability to probability 1. The target probability threshold \( \zeta \) of Ge et al. is hard-coded in the algorithm’s step-size. Therefore, getting results for different probability thresholds (in order to apply Borel-Cantelli) would necessitate running different algorithms, destroying in this way the validity of the results of Ge et al.

#R2: On the rates of Jin et al [14]. First, as can be seen from (E.3) and (E.42), Thm. 4 gives the precise bound

\[
\mathbb{E}[f(X_n) - f(x^*) \mid X_1 \in \mathcal{U}_1] \leq \frac{3}{\sqrt{\kappa}} \frac{\sigma^2 + \kappa}{\sqrt{n}} + o\left(\frac{1}{n}\right).
\]

This is not possible for (at least) two reasons:

1. The statements for SGD in [14] and related papers are also asymptotic because they involve an unknown, probabilistic constant hidden in the \( O(\cdot) \) notation; see Theorem 3, Corollary 4 and Theorem 5 in [14], as well as the corresponding statements in the papers mentioned by R1.

2. The asymptotic value convergence rate of [14] and related papers is \( O(1/\sqrt{n}) \); by contrast, the value convergence guarantee that we provide is \( O(1/n) \). The reviewers are therefore incorrect in stating that our rates are similar to those of [14] and related works.

#R3: On the rates of escape. Deriving rates of escape that hold with probability 1 is a whole new paper in itself.

#R4: On the dependence on \( d \). Great question! The rate does not explicitly depend on \( d \), see (1) above.