reading list

Karimi, H., Nutini, J. and Schmidt, M., 2016, September. Linear convergence of gradient and proximal-gradient methods under the polyak-łojasiewicz condition. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases
http://www.optimization-online.org/DB_FILE/2016/08/5590.pdf

Soudry, D. and Carmon, Y., 2016. No bad local minima: Data independent training error guarantees for multilayer neural networks. arXiv preprint arXiv:1605.08361.
https://arxiv.org/pdf/1605.08361

Du, S.S., Zhai, X., Poczos, B. and Singh, A., 2018. Gradient descent provably optimizes over-parameterized neural networks. ICLR 2019
https://arxiv.org/pdf/1810.02054

Allen-Zhu, Z., Li, Y. and Song, Z., 2019, May. A convergence theory for deep learning via over-parameterization. In International Conference on Machine Learning (pp. 242-252). PMLR.
http://proceedings.mlr.press/v97/allen-zhu19a/allen-zhu19a.pdf

Du, S., Lee, J., Li, H., Wang, L. and Zhai, X., 2019, May. Gradient descent finds global minima of deep neural networks. In International conference on machine learning (pp. 1675-1685). PMLR.
http://proceedings.mlr.press/v97/du19c/du19c.pdf

Liu, C., Zhu, L. and Belkin, M., 2022. Loss landscapes and optimization in over-parameterized non-linear systems and neural networks. Applied and Computational Harmonic Analysis.
Vancouver
https://arxiv.org/abs/2003.00307