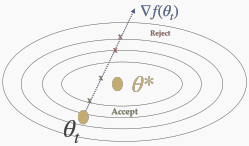
Armijo Line-search Can Make (Stochastic) Gradient Descent Provably Faster (Vaswani, Babanezhad, ICML'25. https://arxiv.org/abs/2503.00229)

- Performance of GD is highly sensitive to its step-size ("learning-rate").
- Armijo line-search is a classic method to set the step-size for GD.
- For smooth functions, in practice, GD-LS typically converges faster than constant step-size GD.

× Theoretically, in the worst case, GD-LS can not be faster than constant step-size GD.



This paper: Identifies a class of non-uniform smooth objective functions including convex losses (e.g. logistic regression, multiclass classification with cross-entropy loss) and non-convex losses (e.g. softmax policy gradient for RL, generalized linear models) for which GD-LS can

✓ be provably faster than constant step-size GD (e.g. exponentially faster for logistic regression),

✓ match the fast convergence of specialized algorithms (e.g. match natural policy gradient for policy optimization in RL),

✓ do so with minimal hyper-parameter tuning.