

連続最適化アルゴリズムの探索方向ノイズと 深層学習モデルの汎化性能の関係

佐藤 尚樹

明治大学大学院 理工学研究科 情報科学専攻

(指導教員：飯塚 秀明)

要旨

近年の人工知能の急速な技術革新は、深層学習をはじめとする機械学習手法の進歩によるものである。一般に深層学習モデルの訓練には、確率的勾配降下法(SGD)やモーメンタム法(SGD with momentum)をはじめとする連続最適化アルゴリズムが使用され、それらの収束性と安定性については、理論と応用の両側面から非常に活発に研究されてきた。しかし、未だに理論と応用の間には大きな乖離があり、実験的には観察されているが理論的には説明されていない現象が多く存在する。例えば、SGD with momentum は特にバッチサイズが大きい時に SGD よりも汎化性能が優れることが観察されているが、それがなぜかは明らかになっていない。このような、いつ、どのような条件で、なぜ訓練が成功するのかという当然の疑問には、アルゴリズムの収束性と安定性に関する既存の理論だけでは答えることができない。

本稿は、最も基本的な連続最適化アルゴリズムである SGD と SGD with momentum に注目し、それらが暗黙のうちに抱えている確率的なノイズに着目する。まず、このノイズが目的関数を平滑化しているとみなせること、その平滑化の度合いは訓練に使われるアルゴリズムの学習率やバッチサイズ等のパラメータで定まることを示す。次に、アルゴリズムごとの平滑化の度合いを推定することで、確率的ノイズの大きさと深層学習モデルの汎化性能との関係について理論的に解析する。得られた理論は畳み込みニューラルネットワークベースの深層学習モデルによる画像分類タスクの数値実験によって検証する。最後に、これらの結果によって上述したいくつかの未解明の現象の説明が可能となることを示す。