A Limit Theory of Foundation Models: A Mathematical Approach to Understanding Emergent Intelligence and Scaling Laws

Zongben Xu

（**Xi'an Jiaotong University- School of Mathematics and Statistics, & Xi'an Academy of Big Data and Artificial Intelligence**）

Abstract

Emergent intelligence has played a major role in the modern AI, manifesting as previously unobserved capabilities enabled by scaling model parameters. While existing studies primarily rely on empirical observations to characterize these phenomena, a rigorous theoretical framework remains underexplored. This study attempts to develop a mathematical approach to formulate emergent intelligence from the perspective of limit theory. Specifically, we introduce a performance function $\mathcal{E}(N,P,K)$, depedent on data size N, model size P and training steps K, to quantify intelligence behavior. We posit that intelligence emerges as a transition from finite to effectively infinite knowledge, and thus the emergent intelligence occurs only when the limit

$$\mathcal{E}(\infty,\infty,\infty)=\lim_{N,P,K\to\infty}\mathcal{E}(N,P,K)$$

exists, and whenever it exists, the emergent abilities corresponds to the limiting behavior. We define a parameter-limit infinite-dimensional system, called as the limit architecture, to embody the emergent abilities of foundation models, and we show that emergent intelligence rationally corresponds to the learning behavior of this limit architecture. By employing the tools from nonlinear Lipschitz operator theory, particularly the characteristic number $\mathrm{Lip}(T)$ of Lipschitz operator, the Lipschitz dual operator theory and the spectral property of compact operators, we prove that the necessary and sufficient conditions for existence of the limit architecture are (i) $\mathrm{Lip}(T_i)\le 1$ for all basic functional blocks T_i, $i>K_0$ for some integer K_0; and (ii) there exists a generalized projection operator T such that $\|T_i-T\|\le\epsilon_i$, where $\sum_{i=1}^{\infty}\epsilon_i<\infty$. Furthermore, we cast the scaling laws of foundation models as the convergence rates of $\mathcal{E}(N,P,K)$ to its asymptotic value $\mathcal{E}(\infty,\infty,\infty)$. Under general conditions, we establish the scaling laws of foundation models by leveraging the Lipschitz operator theory and empirical process theory.

The established theories show that: (i) the emergent intelligence of foundation models appears subject to the existence of the limit architecture,

and it is governed by three key factors-the number of training steps, the size of the training data, and the model size, all of which are in turn crucially determined by properties of the constituting functional blocks T_i; (ii) the emergent intelligence occurs only at the critical case when $\mathrm{Lip}(T_i) = 1$. This is a finding consistent with those empirically observed by Vock and Meisel (2025) and Müller et al. (2025); (iii) the emergent intelligence is essentially determined by an infinite-dimensional system, yet can be effectively approximated in practice through a finite-dimensional architecture, offering solid theoretical evidence for the practical applicability of foundation models in general, and the universal weight subspace hypothesis proposed by Kaushik et al. (2025) in particular. (iv) The established scaling laws shed some lights on the question posed by Simon et al. (2026): "can we develop a theory of scaling laws that both explains why power laws arise and predicts their exponents a priori". We apply the established theories to the foundation models with GPT-like architectures, showing that the GPT-2-like foundation models do exhibit intelligent emergence, whereas GPT-1-like models do not, and for GPT-2 like models, their scaling law with respect to model size follows a power-law. We provide a series of experiments to support the rationality and correctness of the established theories.