Journal of Machine Learning **ISSN**: 2790-2048(e), 2790-203X(p)

Prompt Engineering Through the Lens of Optimal Control

Yifan Luo * [1], Yiming Tang † [1], Chengfeng Shen ‡ [1], Zhenan Zhou § [2], and Bin Dong ¶ [2,3,4]

[1] School of Mathematical Science, Peking University, Beijing, China.
[2] Beijing International Center for Mathematical Research, Peking University, Beijing, China.
[3] Center for Machine Learning Research, Peking University, Beijing, China.
[4] National Biomedical Imaging Center, Peking University, Beijing, China.

Abstract. Prompt engineering (PE) has emerged as a critical technique for guiding large language models (LLMs) in solving intricate tasks. Its importance is highlighted by its potential to significantly enhance the efficiency and effectiveness of human-machine interaction. As tasks grow increasingly complex, recent advanced PE methods have extended beyond the limitations of single-round interactions to embrace multi-round interactions, which allows for a deeper and more nuanced engagement with LLMs. In this paper, we propose an optimal control framework tailored for multi-round interactions with LLMs. This framework provides a unified mathematical structure that not only systematizes the existing PE methods but also sets the stage for rigorous analytical improvements. Furthermore, we extend this framework to include PE via ensemble methods and multi-agent collaboration, thereby enlarging the scope of applicability. By adopting an optimal control perspective, we offer fresh insights into existing PE methods and highlight theoretical challenges that warrant future research. Besides, our work lays a foundation for the development of more effective and interpretable PE methods.

Keywords:
Large language models,
Prompt engineering,
Optimal control.

Article Info.:
Volume: 2
Number: 4
Pages: 241 - 258
Date: December/2023
doi.org/10.4208/jml.231023

Article History:
Received: 23/10/2023
Accepted: 06/11/2023

Communicated by:
Zhi-Qin John Xu

1 Introduction

Prompt engineering (PE) first emerged in the field of large language models (LLMs) in 2020, as researchers realized that well-designed prompts could significantly enhance the capabilities of LLMs without additional model training [1, 4, 25, 29]. The development of PE can be contextualized within the larger scope of natural language programming [27,28] – an increasingly prevalent paradigm that allows for the manipulation of computational systems through natural language, thus offering a more intuitive alternative to traditional programming languages. Much like the transition from machine language to higher-level

*luoyf@pku.edu.cn
†1900010740@pku.edu.cn
‡2201110049@pku.edu.cn
§zhennan@bicmr.pku.edu.cn
¶Corresponding author. dongbin@math.pku.edu.cn

languages like C marked a significant leap in expressive power and ease of use, prompt engineering – or natural language programming in a broader sense – represents a further evolutionary leap, making it easier than ever to instruct machines in performing complex tasks. When implemented properly, PE can yield dramatic performance improvements, particularly in the context of advanced LLMs such as GPT-4 and Claude. In these sophisticated models, the gap between well-engineered and poorly conceived prompts can be stark, reinforcing the critical role of effective PE in leveraging the full potential of LLMs.

Initially, the focus of PE was on single-round prompting, a mechanism suited for relatively straightforward tasks. However, as the need for more complex problem-solving through natural language programming became evident, the field saw a shift towards more intricate forms of engagement, such as multi-round and even multi-agent interactions with LLMs [2, 41, 44, 45]. This evolution in PE bears a striking resemblance to the historical trajectory of optimal control theory [18], which itself originated from the need for point-to-point trajectory optimization and later expanded its scope to accommodate dynamic systems with feedback mechanisms.

The growing complexity of multi-round PE interactions presents significant challenges. Traditional PE approaches [4, 19, 39] often rely on heuristic or empirical methods that, while effective in specific scenarios, lack a systematic foundation amenable to rigorous analysis. This highlights the pressing need for a unified mathematical framework that can serve as a descriptive foundation and facilitate optimization of multi-round PE dynamics.

The primary aim of this paper is to introduce a novel optimal control framework tailored for multi-round interactions with LLMs. Unlike previous works with limited theoretical scopes [3,8,24,33], our approach offers a comprehensive mathematical structure for the systematic design, analysis, and optimization of PE methods, broadening its applicability to include ensemble and multi-agent strategies.

Adopting an optimal control perspective holds the promise of evolving PE along a trajectory similar to that of optimal control theory itself. Initial methodologies in PE mainly focused on single-round prompts, comparable to point-to-point trajectory optimization problems [14, 35, 42]. As optimal control theory incorporated feedback mechanisms for handling complex systems, our framework is designed to accommodate both single-round and multi-round interactions. This shift aims to offer a coherent understanding of the dynamics governing these intricate exchanges and to foster innovative applications transcending current limitations.

To realize these objectives, our methodology employs optimal control to conceptualize multi-round LLM interactions. While acknowledging existing gaps in mathematical rigor due to poorly understood metrics in discrete language spaces, the framework aims to serve as a unified lens for qualitatively evaluating existing PE techniques. Thus, it lays the groundwork for potential improvements in PE by providing an intuitive, structurally coherent approach to model extended dialogic interactions.

Contributions of this paper are summarized as follows:

1. We introduce a novel optimal control formulation that unifies a wide range of existing methods under a single mathematical framework. This provides a rigorous foundation for analyzing and improving prompt design.