
Submission and Formatting Instructions for International Conference on Machine Learning (ICML 2024)

Anonymous Authors¹

Abstract

This document provides a basic paper template and submission guidelines. Abstracts must be a single paragraph, ideally between 4–6 sentences long. Gross violations will trigger corrections at the camera-ready phase.

1. Electronic Submission

References

¹Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

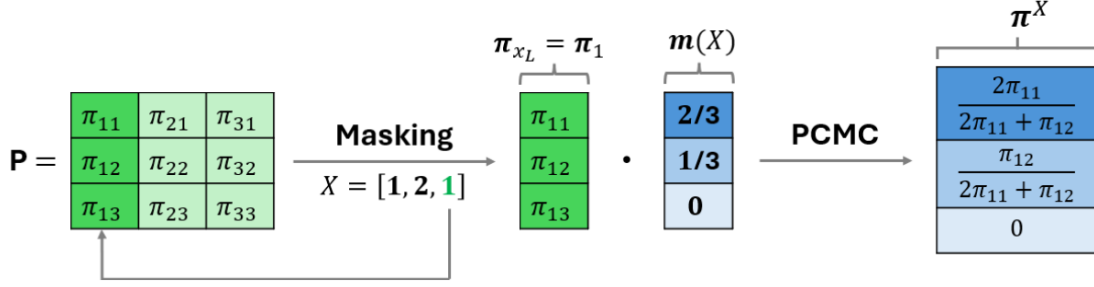


Figure 1. Demonstration of Definition 2.1. We provide an example where the vocabulary size $K = 3$ and the input prompt $X = [1, 2, 1]$, which results in a frequency vector $m(X)$. P represents the transition matrix of the base Markov chain.

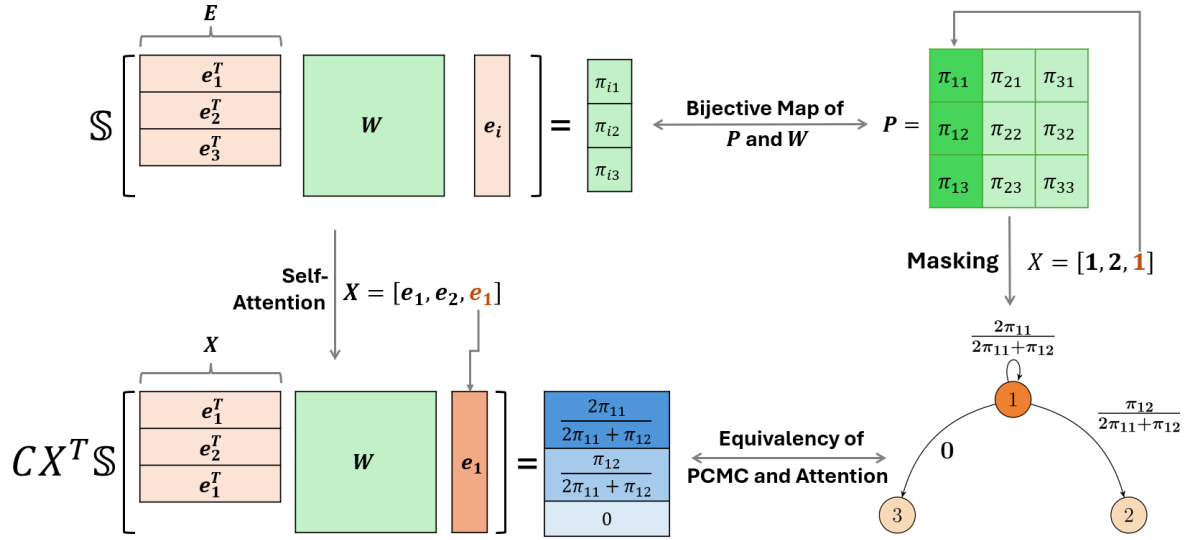


Figure 2. Illustration of the Equivalency between the Attention and PCMC models. We provide an example where the vocabulary size $K = 3$ and the input prompt is $X = [1, 2, 1]$. The upper figure represents how the token probabilities $S(EW e_i)$ can be mapped to a base transition matrix P . The left-lower figure demonstrates the output of the self-attention given an input prompt X . The right-lower figure derives PCMC transitions from this P given the same prompt. The resulting next token probabilities are the same for both of the models. The masking operation is demonstrated in a more detailed way in Figure 1.

A. Additional Figures