The Ky Fan Norms and Beyond: Dual Norms and Combinations for Matrix Optimization

Alexey Kravatskiy   KRAVTSKII.AIU@PHYSTECH.EDU
Moscow Institute of Physics and Technology (MIPT)

Ivan Kozyrev   KOZYREV.IN@PHYSTECH.EDU
Moscow Institute of Physics and Technology (MIPT)
Marchuk Institute of Numerical Mathematics

Alexander Vinogradov   VINOGRADOV.AM@PHYSTECH.EDU
Moscow Institute of Physics and Technology (MIPT)

Nikolai Kozlov   KOZLOV.NA@PHYSTECH.EDU
Moscow Institute of Physics and Technology (MIPT)

Daniil Merkulov   DANIIL.MERKULOV@PHYSTECH.EDU
Moscow Institute of Physics and Technology (MIPT)
Skoltech, HSE, AI4Science

Ivan Oseledets   I.OSELEDETS@SKOLTECH.RU
AIRI, Skoltech
Marchuk Institute of Numerical Mathematics

Abstract

In this article, we explore the use of various matrix norms for optimizing functions of weight matrices, a crucial problem in training large language models. Moving beyond the spectral norm that underlies the Muon update, we leverage the duals to the Ky Fan k-norms to introduce a family of Muon-like algorithms we name *Fanions*, which happen to be similar to Dion. Then working with the duals of convex combinations of the Ky Fan k-norms and the Frobenius norm or the l_∞ norm, we construct the families of *F-Fanions* and *S-Fanions* respectively. Their most prominent members are *F-Muon* and *S-Muon*. We complement our theoretical analysis with an extensive empirical study of the algorithms across a wide range of tasks and settings, from which it follows that F-Muon and S-Muon are always on par with Muon, while on fine-tuning of NanoGPT and synthetic linear least squares they are even better than vanilla Muon optimizer.