

TRANSFORMER - Gradient Alignment vs. Hessian Operator Norm

