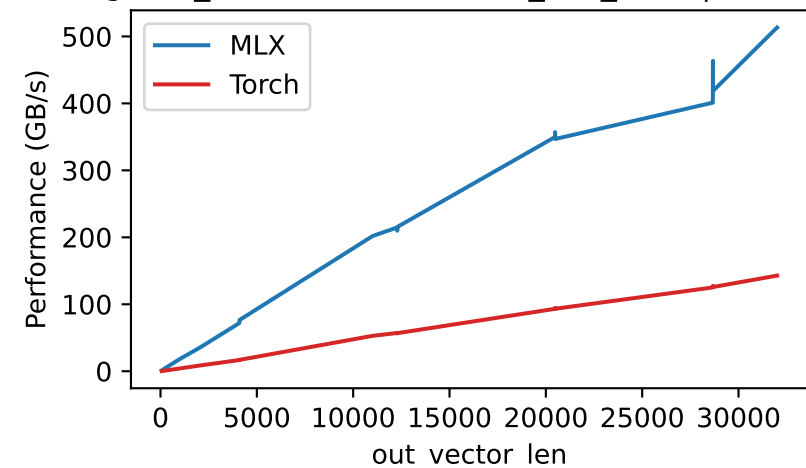
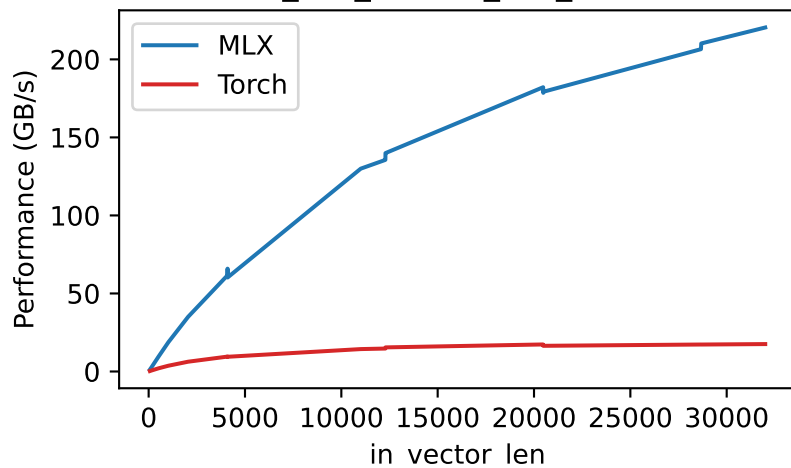


Apple M2 Ultra: float16 gemv_t

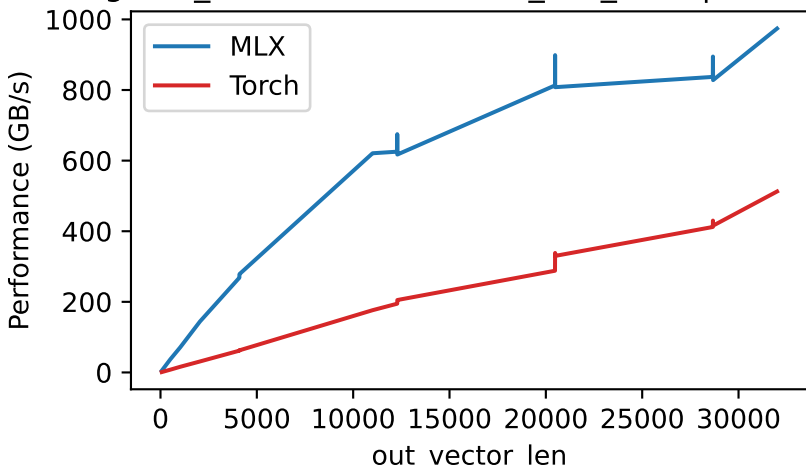
gemv_t ([1, 128] [128, out_vec_len]) | float16



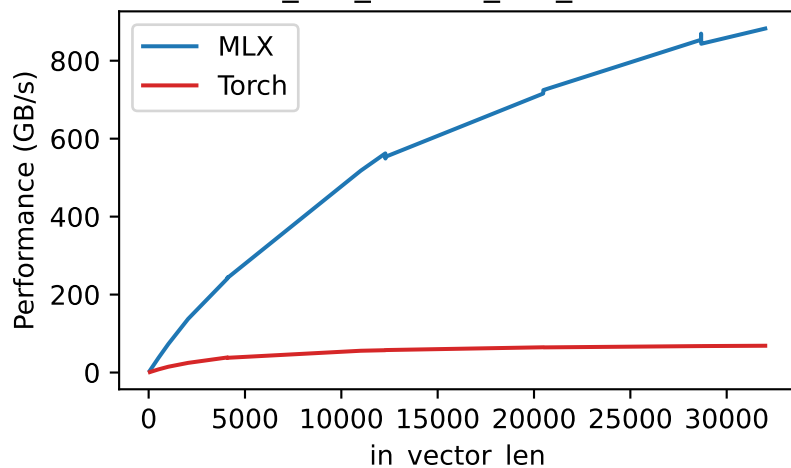
gemv_t ([1, in_vec_len] [in_vec_len, 128]) | float16



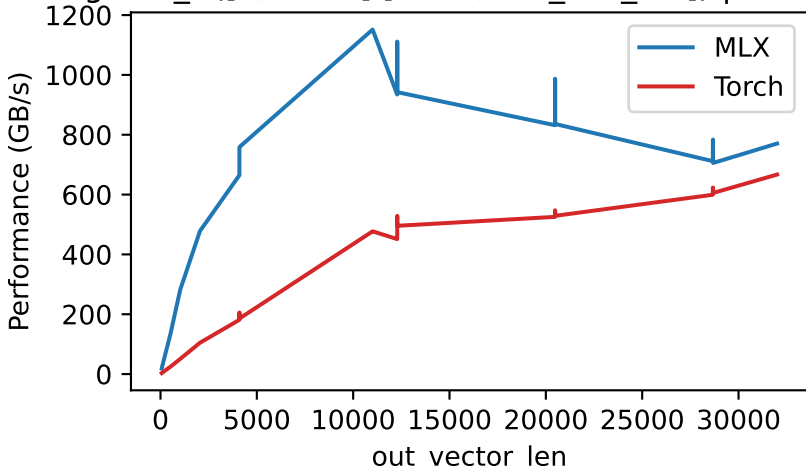
gemv_t ([1, 512] [512, out_vec_len]) | float16



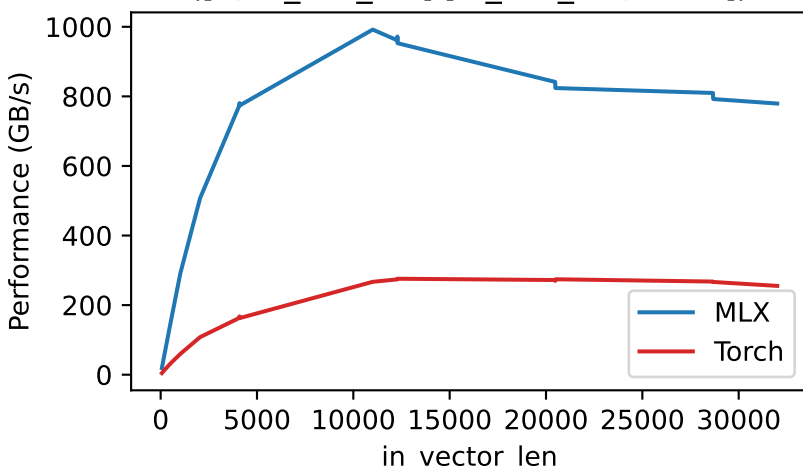
gemv_t ([1, in_vec_len] [in_vec_len, 512]) | float16



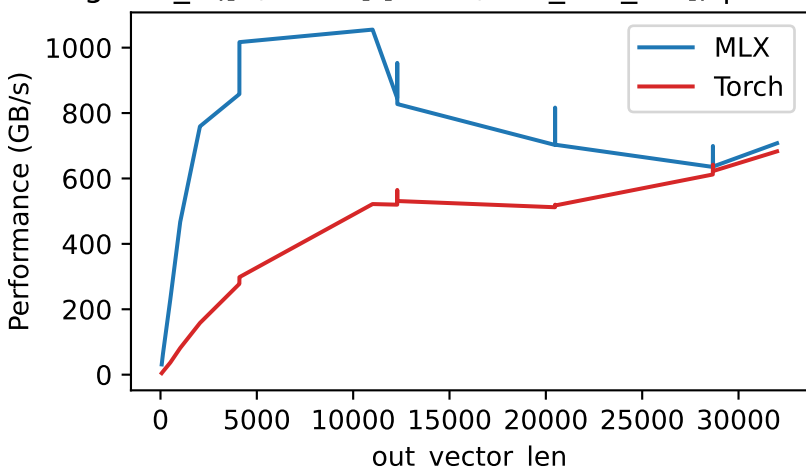
gemv_t ([1, 2048] [2048, out_vec_len]) | float16



gemv_t ([1, in_vec_len] [in_vec_len, 2048]) | float16



gemv_t ([1, 4096] [4096, out_vec_len]) | float16



gemv_t ([1, in_vec_len] [in_vec_len, 4096]) | float16

