Batch Runtime for One Optimizer Pass

Legend:
- Full Parameter
- LoRA
- Adapter
- BitFit
- Out of Memory

GPT-2

Sequence Length 512: 8.23x, 46.64x, 104.96x

Sequence Length 1024: 7.99x, 44.52x, 106.27x

GPT-2 XL

Sequence Length 512: 30.67x, 101.15x, 359.03x

Sequence Length 1024: OOM

Y-axis: Time (ms)

X-axis: Sequence Length