Method

Scalable Influence Functions

Algorithm 1 Arnoldi

1: procedure ARNOLDI(v, n)
2:   1. Build orthonormal basis for the Krylov subspace \( K_n \).
3:      \( \tilde{K}_n = \text{Span}(v, Hv, H^2v, \ldots, H^{n-1}v) \)
4:   2. Diagonalize \( K_n \) to its top eigenvalues.
5:      \( \text{Discard the last row of } A \) and the last \( v \).
6:   3. Compute \( A \)'s eigenvalues \( \lambda_j \) and eigenvectors \( w_j \).
7:   4. Set \( (\lambda_j, w_j) \) to the corresponding eigenvectors.
8:      \( \text{Set } (\tilde{\lambda}_j, \tilde{w}_j) \) to the projection onto the spans \( [c_j]_n \) basis.
9:   5. Return \( (\tilde{\lambda}_j, \tilde{w}_j) \).
10: end procedure

The procedure additionally builds an orthonormal basis for the Krylov subspace so that the diagonalization of the restriction \( H \) to \( K_n \) yields an approximation of the largest (in absolute value) eigenvalues of \( H \) and of the corresponding eigenvectors.

The matrix \( H \) gets replaced with now diagonal \( \tilde{H} \) simplifying the matrix appearing in the definition of \( T_H \) and dispensing with the expensive LISSA [2] procedure.

Results

MNIST Benchmark

Corrupted MNIST – Small Model (Figure 1 & 2)

- [2] considers only a 10% subset of data and a CNN with 3k parameters.
- We evaluate ability to recall synthetically mislabeled examples [4].
- IFs outperform gradient methods and RandProj does not capture correct eigenvalues.

Corrupted MNIST – Large Model (Table 1)

- CNN with 800k parameters.
- LISSA is prohibitively slow; Arnoldi & RandProj scale well; no advantage in using curvature information.