Robust Online Convex Optimization in the Presence of Outliers

Tim van Erven
Sarah Sachs
Wouter M. Koolen
Wojciech Kotłowski

Extreme Outliers Can Break Learning

Robustifying Any OCO Algorithm

1. Any OCO ALG with regret bound \( R_T(G) \) if gradients have length at most \( G \)
2. Top-k Filter: simple strategy to filter out large gradients
ALG must be able to adapt to gradient length \( G \)

**Theorem** (At most \( k \) outliers). **On linear losses, ALG + Top-k Filter achieves**

\[
R_T(u, S) \leq B_T(2G(S)) + 4DG(S)(k + 1) \quad \text{for any } S : T - |S| \leq k.
\]

Feed ALG gradients \( \leq 2G(S) \)

Consequences

<table>
<thead>
<tr>
<th>Losses</th>
<th>Minimax Robust Regret</th>
</tr>
</thead>
<tbody>
<tr>
<td>General convex</td>
<td>( O(\sqrt{T} + k) )</td>
</tr>
<tr>
<td>General convex + i.i.d.</td>
<td>&quot;</td>
</tr>
<tr>
<td>Strongly convex</td>
<td>( O(\ln(T) + k) )</td>
</tr>
</tbody>
</table>

Efficient Filtering Approach

Top-k Filter:
- Maintain list \( L_t \) of \( k + 1 \) largest gradient lengths seen so far
- Filter round if \( ||g_t|| > 2 \text{ min } L_t \); otherwise pass to ALG

Main Ideas:
1. Never pass ALG gradients \( > 2G(S) \):
   - \( L_t \) contains at least 1 inlier, because at most \( k \) outliers
   - Hence \( \text{min } L_t \leq G(S) \)
2. Overhead for filtering is \( O(k) \):
   - Every filtered round is also added to \( L_t \)
   - Therefore \( \text{min } L_t \) (at least) doubles every \( k + 1 \) filtered rounds
   - Hence last \( k + 1 \) filtered rounds dominate

Quantile Outliers

Which extra assumptions allow sublinear dependence on number of outliers \( k \)?

- \( ||g_t|| \leq L \| X_t \| \) for i.i.d. \( X_t \) (e.g. hinge loss, logistic loss)
- Inliers \( S_p \) are rounds s.t. \( \| X_t \| \) less than \( p \)-quantile \( X_p \)

**Theorem** (Sublinear Outlier Overhead). Suppose ALG has regret bound \( R_T(X) \), concave in \( T \), if non-filtered \( X_t \) have length at most \( X \). Then ALG + \( p \)-Quantile Filter achieves

\[
\mathbb{E} \left[ \max_{u \in W} R_T(u, S_p) \right] \leq B_{pT}(X_p) + O \left( LD_X \sqrt{p(1 - p)T \ln T + \ln(T)^2} \right).
\]

Robust regret: measure regret only on (unknown) inlier rounds

**Price of Robustness = Overhead over usual regret rate**:
- At most \( k \) adversarial outliers: \( O(k) \)
- \( p \)-Quantile outliers: \( O(\sqrt{T(1 - p)T \ln T + \ln(T)^2}) \)

Summary

Tim is looking for a postdoc, starting anytime in 2022. Please get in touch if you want to come to Amsterdam!