Empirical Risk Minimization

Without loss of generality, we restrict our attention to (1) if \( r(w) \) is a Lipschitz continuous function.

1. ERM for Lipschitz continuous random functions
   Assume \( f(w, z) \) is a G-Lipschitz continuous function w.r.t. \( w \) for any \( z \in Z \). If \( r(w) \) is present, it can be absorbed into \( f(w, z) \). It is notable that we do not assume \( f(w, z) \) is convex in terms of \( w \) or any \( z \).

   \[
   P(\bar{w}) - P_{z} \leq O\left(\frac{d \log n + \log(1/\delta)}{n}\right)^{1/2},
   \]
   where \( a = \delta(\log(32/2\alpha e^{2})) + \log(1/\delta))\) is a constant.

2. ERM for non-negative, Lipschitz continuous and smooth convex random functions
   Besides the Lipschitz continuity, we further assume \( f(w, z) \) is a non-negative and L-smooth convex function w.r.t. \( w \) for any \( z \in Z \). It is notable that we do not assume \( (r(w)) \) is smooth.

   \[
   P(\bar{w}) - P_{z} \leq O\left(\frac{d \log n + \log(1/\delta)}{n}\right)^{1/2} + \frac{(\log(32/2\alpha e^{2})) + \log(1/\delta))\) is a constant.
   \]

Efficient SA for Lipschitz Continuous Random Functions

**Algorithm 1** SSG(\(w_1, \gamma, T, W)\)

1. Require: \( w_1 \in W \), \( \gamma > 0 \) and \( T \)
2. Ensure: \( w_T \)
3. for \( k = 1, \ldots, T \) do
   1. \( w_{k+1} = B(w_k, \gamma g) \)
4. end for
5. return \( w_T \)

**Algorithm 2** ASA(\(w_1, \gamma, n_0, R_0)\)

1. Set \( R_0 = 2R_0, w_0 = w, m = \frac{1}{2} \log \frac{2R_0}{\epsilon} \)\) - 1, \( n_0 = \lfloor m \rfloor \)
2. for \( k = 1, \ldots, n \) do
   1. \( \gamma_k = \gamma \) and \( R_k = R_{k-1}/2 \)
   2. \( w_0 = \text{SSG}(w_0, \gamma_k, n_0, W \cap B(w_0, R_{k-1})) \)
3. end for
4. return \( w_m \)

ASA for G-Lipschitz continuous random functions. Suppose \( w_m - w^*_w \leq R_0 \), where \( w^*_w \) is the closest optimal solution to \( w \). Define \( a = \max(\alpha(e^{G^2}(R_0/2^2))) \).

For \( n \geq 100 \) and any \( \delta \in (0, 1) \), with probability at least \( 1 - \delta \), we have

\[
P(\bar{w}_m) - P_{z} \leq O\left(\frac{d \log n + \log(\log(n/\delta))}{n}\right)^{1/2}.
\]