Motivated by solving the stochastic composite optimization problem by Empirical Minimization:

\[ w^*_n = \arg \min_{w_n} \left\{ P(w) + E_{z \sim P}[f(w, z)] + R(w) \right\} \]

Study the convergence of the empirical minimizer of (1)

\[ \hat{w} = \arg \min_{w} \left\{ P_n(w) + \frac{1}{n} \sum_{i=1}^{n} f(w, z_i) + R(w) \right\} \]

where \( z_1, \ldots, z_n \) are i.i.d samples from \( P \).

Goal: to establish the fast convergence rate of the empirical minimizer in terms of \( P(\hat{w}) - P(w^*_n) \).

Main Results

The three recent studies [1, 2, 3] focus on establishing fast rates in terms of risk minimization without a regularizer

\[ \min_{w} F(w) + E_{z \sim P}[f(w, z)] \]


\[ \hat{w} = \arg \min_{w} \left\{ F_n(w) + \frac{1}{n} \sum_{i=1}^{n} f(w, z_i) + \frac{1}{n} g(w) \right\} \]

Mehta [2] targeted on the original risk minimization as (3)

Table 1: Difference of fast rates between our work and the related works [1, 2, 3]

<table>
<thead>
<tr>
<th>Related Work</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \min_{w} F(w) )</td>
<td>[ 1, 2, 3 ]</td>
</tr>
<tr>
<td>( \min_{w} F(w) + R(w) )</td>
<td>[ 1, 2, 3 ]</td>
</tr>
<tr>
<td>( F(w) )</td>
<td>[ 1, 2, 3 ]</td>
</tr>
<tr>
<td>( w^*_n )</td>
<td>[ 1, 2, 3 ]</td>
</tr>
<tr>
<td>( \text{`s-convex regularizer} )</td>
<td>[ 1 ]</td>
</tr>
<tr>
<td>( \text{High probability} )</td>
<td>[2]</td>
</tr>
</tbody>
</table>

Our result is more general.

Theoretical Results

Assumption 1:

- \( W \) is a closed and bounded convex set, i.e., there exists \( R \) such that \( \| w \|_2 \leq R \) for all \( w \in W \).
- \( f(w, z) \) is a \( \mathcal{G} \)-Lipschitz continuous, \( L \)-smooth and \( \beta \)-exp-concave function of \( w \in W \) for any \( z \in Z \).
- \( R(w) \) is a convex function.

Theorem 1

For the stochastic composite minimization problem (1), we consider the empirical minimizer \( \hat{w} \) by solving (2). Under Assumption 1, with probability at least \( 1 - \delta \), we have

\[ P(\hat{w}) - P(w^*_n) \leq O\left( \frac{\log n}{n} + \frac{\log(1/\delta)}{n\sigma} \right). \]

Remark 1: When \( R(w) = 0 \), directly obtain a fast rate with high probability of the empirical risk minimizer for the exp-concave risk minimization.

Remark 2: Linear dependence on dimensionality \( d \) is unavoidable [4].

Theorem 2

For the risk minimization problem (3), we consider the regularized empirical risk minimizer \( \hat{w} \) by solving (4). Under Assumption 1 (i), (ii), and that \( g(x) \) is bounded over \( \mathcal{W} \) such that \( \sup_{w \in \mathcal{W}}|g(w) - g(w')| \leq B \), with probability at least \( 1 - \delta \), we have

\[ F(\hat{w}) - \min_{w \in \mathcal{W}} F(w) \leq O\left( \frac{\log n}{n} + \frac{\log(1/\delta)}{n\sigma} \right). \]

Remark 1: Address the open problem raised in [1] about high probability bound for strongly regularized empirical risk minimizer.

Remark 2: Extend the fast rate to any regularized empirical risk minimizer as long as the regularizer is convex.

Analysis Technique

- Step 1: using the convexity of \( P(w) \), the optimality condition of \( \hat{w} \), and Cauchy-Schwarz inequality:

\[ P(\hat{w}) - P(w^*_n) \leq \| G(\hat{w}, w^*_n) \|_H \| \hat{w} - w^*_n \|_H + \| \Delta_0(\hat{w}) \|_H \| \hat{w} - w^*_n \|_H \]

where \( G(\hat{w}, w^*_n) = \nabla P(\hat{w}) - \nabla P(w^*_n) \), \( G_n(\hat{w}, w^*_n) = \nabla P_n(\hat{w}) - \nabla P_n(w^*_n) \), \( \Delta_0(\hat{w}) = \nabla P(\hat{w}) - \nabla P_n(\hat{w}) \), \( H \) is local norm.

- Step 2: using concentration inequality [5], union bound and covering number of \( \mathcal{W} \), with probability at least \( 1 - \delta \),

\[ \| G(\hat{w}, w^*_n) \|_H \leq O\left( \frac{\log(2\delta)}{n} \right) \]

- Step 3: using Young’s inequality and do some algebra:

\[ P(\hat{w}) - P(w^*_n) \leq O\left( \frac{\log n}{n} + \frac{\log(1/\delta)}{n\sigma} \right). \]

Table 1: Difference of fast rates between our work and the related works [1, 2, 3]

<table>
<thead>
<tr>
<th>Related Work</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{`s-convex regularizer} )</td>
<td>[ 1 ]</td>
</tr>
<tr>
<td>( \text{High probability} )</td>
<td>[2]</td>
</tr>
</tbody>
</table>