Motivation

Population Risk (g) \rightarrow \text{Strongly Morse} \rightarrow \text{Empirical Risk (f)}

What if the Hessian is degenerate?

Assumptions

1. In $\mathcal{T} \triangleq \{x \in B(l) : \|\text{grad} \ g(x)\|_2 < \epsilon\}$, $|\lambda_{\text{min}}(\text{hess} \ g(x))| > \eta$
2. Gradient proximity:
   \[
   \sup_{x \in B(l)} \|\text{grad} \ f(x) - \text{grad} \ g(x)\|_2 \leq \frac{\eta}{2}
   \]
3. Hessian proximity:
   \[
   \sup_{x \in B(l)} \|\text{hess} \ f(x) - \text{hess} \ g(x)\|_2 \leq \frac{\eta}{2}
   \]

Main Theorem

Denote $f$ and $g$ as the empirical risk and the population risk. Let $\mathcal{D}$ be a maximal connected and compact subset of $\mathcal{B}$. With the above assumptions, we have

(a) $\mathcal{D}$ contains at most one local minimum of $g$. If $g$ has $K(K = 0, 1)$ local minima in $\mathcal{D}$, then $f$ also has $K$ local minima in $\mathcal{D}$.
(b) If $g$ has strict saddles in $\mathcal{D}$, then $f$ has any saddle points in $\mathcal{D}$, they must be strict saddle points.

Local Minima Distance

Corollary

Let $\{\hat{x}_i\}_{i=1}^{L_{\text{opt}}}$ and $\{x_i\}_{i=1}^{L_{\text{opt}}}$ denote the local minima of the empirical risk and its population risk. Denote $L_{\text{opt}} = \sup_{x \in B(l)} \|\nabla^3 g(x)\|$ with $\nabla^3 g(x)$ being the third-order Riemannian derivative of $g(x)$ with respect to $x \in \mathbb{R}^N$. Define $\epsilon_{\text{opt}} \triangleq c \sqrt{\text{opt}^{-1} \text{N} \log(M)} \leq \epsilon$. Then, we have

\[
\|\hat{x}_i - x_i\|_2 \leq \frac{4M}{\eta}, \quad \text{if} \ M \geq c(\eta/L_{\text{opt}}^2 \epsilon_{\text{opt}}^2)^{-1}N \log(M)
\]

Matrix Sensing

- Population risk: $(U \in \mathbb{R}^{N \times k}, X \in \mathbb{R}^{N \times N}$ is PSD with rank $r)$
  \[
  f(U) \triangleq \frac{1}{2} \|AUU^\top - X\|^2_F
  \]
  \[
  g(U) = E[f(U)] = \frac{1}{2} \|UU^\top - X\|^2
  \]

Lemma 1

- Assump 1 is true by setting
  \[
  \epsilon \leq \min\{\frac{1}{80}, \frac{1}{40c^{-1}}\lambda_{\text{min}}^2\}, \eta = 0.06\lambda_k
  \]
- Assumps 2&3 are true if we construct the symmetric operator $A$ with another operator $B$ $(A_m = \frac{1}{2}(B_m + B_m^\top))$ satisfying the RIP:
  \[
  \lambda_{\text{min}}(\text{hess} \ g(x)) \geq \sqrt{\frac{1}{2} \|\nabla^2 g(x)[U] | U = \frac{1}{2} \|\nabla^2 g(x)[U] | U \}
  \]

Note that the RIP holds w.h.p if $M \geq C(r + k)N/\epsilon_{\text{opt}}^2$.

Phase Retrieval

- Population risk: $(y_m = |(a_m, x^*)|^2, 1 \leq m \leq M, x^* \in \mathbb{R}^N)$
  \[
  f(x) = \frac{1}{2M} \sum_{m=1}^{M} (|a_m, x|^2 - y_m)^2
  \]

Lemma 2

- Assump 1 is true by setting
  \[
  \epsilon \leq 0.3963 \|x^*\|^2_2, \eta = 0.22 \|x^*\|^2_2
  \]
- Assumps 2&3 hold w.h.p if $M \geq CN^2$ and $a_m \in \mathbb{R}^N$ is a Gaussian random vector with entries following $\mathcal{N}(0, 1)$.

Phase retrieval:

- $f(x) = \frac{1}{2M} \sum_{m=1}^{M} (|a_m, x|^2 - y_m)^2$
- $y_m = |(a_m, x^*)|^2, 1 \leq m \leq M$
- $N = 2, x^* = [1 - 1]^\top$

Conclusions

We study the problem of establishing a correspondence between the critical points of the empirical risk and its population risk \textit{without} the strongly Morse assumption.

Acknowledgement

This work was supported by NSF grant CCF-1704204, and the DARPA Lagrange Program under ONR/SPAWAR contract N660011824020.