Characterization of Convex Objective Functions and Optimal Expected Convergence Rates of SGD

Phuong Ha Nguyen

Marten van Dijk¹, Lam M. Nguyen² and Dzung T. Phan²

1. Secure Computation Laboratory, ECE, University of Connecticut
2. IBM Research, Thomas J. Watson Research Center

International Conference on Machine Learning (ICML)
Long Beach, California, 2019
Problem Setting

- Solve

\[
\min_{w \in \mathbb{R}^d} \{ F(w) = E_\xi [f(w; \xi)] \}
\]
Problem Setting

- **Solve**
  \[
  \min_{w \in \mathbb{R}^d} \{ F(w) = E_\xi [f(w; \xi)] \}
  \]

- **Assumptions**
  - **Convex:**
    \[
    f(w; \xi) - f(w'; \xi) \geq \langle \nabla f(w'; \xi), (w - w') \rangle
    \]
  - **Smooth:**
    \[
    ||\nabla f(w; \xi) - \nabla f(w'; \xi)|| \leq L ||w - w'||
    \]
Problem Setting

- **Solve**
  \[ \min_{w \in \mathbb{R}^d} \{ F(w) = E_\xi [ f(w; \xi) ] \} \]

- **Assumptions**
  - Convex:
    \[ f(w; \xi) - f(w'; \xi) \geq \langle \nabla f(w'; \xi), (w - w') \rangle \]
  - Smooth:
    \[ ||\nabla f(w; \xi) - \nabla f(w'; \xi)|| \leq L ||w - w'|| \]

- **Find a** \( w_t \) **close to**
  \[ W^* = \{ w_* \in \mathbb{R}^d : \forall w \in \mathbb{R}^d, F(w) \geq F(w_*) \} \]
Problem Setting

- Solve
  \[ \min_{w \in \mathbb{R}^d} \{ F(w) = E_\xi [f(w; \xi)] \} \]

- Assumptions
  - Convex:
    \[ f(w; \xi) - f(w'; \xi) \geq \langle \nabla f(w'; \xi), (w - w') \rangle \]
  - Smooth:
    \[ ||\nabla f(w; \xi) - \nabla f(w'; \xi)|| \leq L||w - w'|| \]

- Find a \( w_t \) close to
  \[ W^* = \{ w* \in \mathbb{R}^d : \forall w \in \mathbb{R}^d, F(w) \geq F(w*) \} \]

---

Stochastic Gradient Descend (SGD):

Initialize: \( w_0 \)

Iterate:
\[ \text{for } t = 0, 1, 2, \ldots, \text{ do} \]
  - Choose \( \eta_t > 0 \)
  - Generate random \( \xi_t \)
  - Compute \( \nabla f(w_t; \xi_t) \)
  - Update \( w_{t+1} = w_t - \eta_t \nabla f(w_t; \xi_t) \)
\[ \text{end for} \]
Problem Setting

- **Solve**
  \[
  \min_{w \in \mathbb{R}^d} \{ F(w) = E_{\xi} [f(w; \xi)] \} 
  \]

- **Assumptions**
  - **Convex:**
    \[
    f(w; \xi) - f(w'; \xi) \geq \langle \nabla f(w'; \xi), (w - w') \rangle
    \]
  - **Smooth:**
    \[
    \|\nabla f(w; \xi) - \nabla f(w'; \xi)\| \leq L \|w - w'\|
    \]

- **Find a** \(w_t\) close to
  \[
  W^* = \{ w_* \in \mathbb{R}^d : \forall w \in \mathbb{R}^d, F(w) \geq F(w_*) \}
  \]

- **Problem: Characterize Expected Convergence Rates**
  \[
  E \left[ \inf_{w_* \in W^*} \|w_t - w_*\|^2 \right] \text{ and } E[F(w_t) - F(w_*)]
  \]

---

**Stochastic Gradient Descend (SGD):**

**Initialize:** \(w_0\)

**Iterate:**

\[
\text{for } t = 0, 1, 2, \ldots, \text{ do}
\]

- Choose \(\eta_t > 0\)
- Generate random \(\xi_t\)
- Compute \(\nabla f(w_t; \xi_t)\)
- Update \(w_{t+1} = w_t - \eta_t \nabla f(w_t; \xi_t)\)

**end for**
Beyond convex and strongly convex functions

Plain Convex

\[ F(w) - F(w_*) \geq 0 \]

Strongly Convex

\[ F(w) - F(w_*) \geq \frac{\mu}{2} ||w - w_*||^2 \]
$\omega$-Convexity

Plain Convex

$F(w) - F(w_*) \geq 0$

$\omega$-Convex

$\omega (F(w) - F(w_*)) \geq \inf_{w_* \in W_*} ||w - w_*||^2,$
$\omega' > 0, \omega'' < 0,$

Strongly Convex

$F(w) - F(w_*) \geq \frac{\mu}{2} ||w - w_*||^2$
**ω-Convexity with curvature $h \in [0,1]$**

Plain Convex

$F(w) - F(w_*) \geq 0$

$\omega$ - Convex

$\omega (F(w) - F(w_*)) \geq \inf_{w_* \in W, \omega' > 0, \omega'' < 0, ||w - w_*||^2}$

Strongly Convex

$F(w) - F(w_*) \geq \frac{\mu}{2} ||w - w_*||^2$

$(F(w) - F(w_*))^h \geq \alpha \inf_{w_* \in W, ||w - w_*||^2}$

$h = 0 \quad h \in (0,1) \quad h = 1$
\( \omega \)-Convexity with curvature \( h \in [0,1] \)

Plain Convex

\[
F(w) - F(w_*) \geq 0
\]

\( \omega \) - Convex

\[
\omega(F(w) - F(w_*)) \geq \inf_{w_* \in W^*} ||w - w_*||^2,
\]
\[
\omega' > 0, \omega'' < 0,
\]

\( \omega \) - Convex

\[
(F(w) - F(w_*))^h \geq \alpha \inf_{w_* \in W^*} ||w - w_*||^2
\]

Strongly Convex

\[
F(w) - F(w_*) \geq \frac{\mu}{2}||w - w_*||^2
\]

\( h = 0 \longrightarrow h \in (0,1) \longrightarrow h = 1 \)

HEB (Holderian Error Bound): \( (F(w) - F(w_*))^h \geq \alpha \inf_{w_* \in W^*} ||w - w_*||^2 \), where \( h \in (0,2] \).

HEB and \( \omega \)-convexity are not subclasses of one another but they do intersect for \( h \in (0,1] \).

Close to optimal stepsize

Plain Convex

\[ F(w) - F(w_*) \geq 0 \]

\( h = 0 \)

\[ \omega \in \text{Convex} \]

\[ \omega(F(w) - F(w_*)) \geq \inf_{w_* \in W^*} \|w - w_*\|^2, \]

\( \omega' > 0, \omega'' < 0, \)

\[ (F(w) - F(w_*))^h \geq \alpha \inf_{w_* \in W^*} \|w - w_*\|^2 \]

\( h \in (0,1) \)

Strongly Convex

\[ F(w) - F(w_*) \geq \frac{\mu}{2} \|w - w_*\|^2 \]

\( h = 1 \)

SGD

\[ \eta_t = \frac{c}{(t+\Delta)^{1/(2-h)}} \]

Close to optimal stepsize
Convergence Rate of SGD

Plain Convex

\[ F(w) - F(w_*) \geq 0 \]

\[ h = 0 \]

\[ (F(w) - F(w_*))^h \geq \alpha \inf_{w_* \in W} ||w - w_*||^2 \]

\[ h \in (0,1) \]

SGD

\[ \eta_t = \frac{c}{(t+\Delta)^{1/(2-h)}} \]

Close to optimal stepsize

\[ E \left[ \inf_{w_* \in W} ||w_t - w_*||^2 \right] = O(t^{-h/(2-h)}) \]

\[ \frac{1}{t} \sum_{i=t+1}^{2t} E[F(w_i) - F(w_*)] = O(t^{-1/(2-h)}) \]
Convergence Rate of SGD

Plain Convex
$F(w) - F(w_*) \geq 0$
$h = 0$

$\omega \rightarrow \text{Convex}$
$\omega (F(w) - F(w_*)) \geq \inf_{w_* \in W^*} ||w - w_*||^2$
$\omega' > 0, \omega'' < 0,$

Strongly Convex
$F(w) - F(w_*) \geq \frac{\mu}{2} ||w - w_*||^2$
$h = 1$

$\omega \rightarrow \text{Strongly Convex}$
$(F(w) - F(w_*))^h \geq \alpha \inf_{w_* \in W^*} ||w - w_*||^2$
$h \in (0,1)$

$E \left[ \inf_{w_* \in W^*} ||w_t - w_*||^2 \right] = O(t^{-h/(2-h)})$

$E \left[ F(w_i) - F(w_*) \right] = O(t^{-1/(2-h)})$

$\frac{1}{t} \sum_{i=t+1}^{2t} E[F(w_i) - F(w_*)] = O(t^{-1/(2-h)})$

[Useless,0] $0 \leftarrow h \rightarrow 1$ [Useful,1] [Useful,1]
**Convergence Rate of SGD**

Plain Convex

\[ F(w) - F(w_*) \geq 0 \]

\[ h = 0 \]

\[ \omega - \text{Convex} \]

\[ \omega(F(w) - F(w_*)) \geq \inf_{w_* \in W_*} ||w - w_*||^2, \]

\[ \omega' > 0, \omega'' < 0, \]

\[ (F(w) - F(w_*))^h \geq \alpha \inf_{w_* \in W_*} ||w - w_*||^2 \]

\[ h \in (0,1) \]

Strongly Convex

\[ F(w) - F(w_*) \geq \frac{\mu}{2} ||w - w_*||^2 \]

\[ h = 1 \]

\[ E \left[ \inf_{w_* \in W_*} ||w_t - w_*||^2 \right] = O(t^{-h/(2-h)}) \]

\[ \frac{1}{t} \sum_{i=t+1}^{2t} E[F(w_i) - F(w_*)] = O(t^{-1/(2-h)}) \]

\[ h = \frac{1}{2} \]

\[ F(w) = H(w) + \lambda G(w), H(w) - \text{convex} \]

\[ G(w) = \sum_{i=1}^{d} [e^{w_i} + e^{-w_i} - 2 - w_i^2] \]
Curvature 0 (convex)
\[ f_i(w) = \log(1 + \exp(-y_i x_i^T w)) \]

Curvature \( \frac{1}{2} \)
\[ f^{a}_i(w) = f_i(w) + \lambda G(w) \]
\begin{align*}
G(w) = \sum_{i=1}^{d} [e^{w_i} + e^{-w_i} - 2 - w_i^2]
\end{align*}

Curvature unknown
\[ f^{a}_i(w) = f_i(w) + \lambda ||w|| \]

Curvature 1 (strongly convex)
\[ f^{c}_i(w) = f_i(w) + \frac{\lambda}{2} ||w||^2 \]
Conclusion

- $\omega$-convexity notion: plain convex, strongly convex and something in between
- SGD with $\omega$-convex objective functions

Thank you for your attention! 😊

https://arxiv.org/abs/1810.04100

Poster Number: #193 – Pacific Ballroom. – 06:30—09:00PM – 06/11