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

Phuong Ha Nguyen[1]

Marten van Dijk[1], Lam M. Nguyen[2] and Dzung T. Phan[2]

Marten Lam P. Ha Dzung

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**

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

Problem Setting

- **Solve**

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

- **Assumptions**
  - **Convex:**

$$f(w;\xi) - f(w';\xi) \geq \langle \nabla f(w';\xi), (w - w') \rangle$$

  - **Smooth:**

$$||\nabla f(w;\xi) - \nabla f(w';\xi)|| \leq L||w - w'||$$

UCONN

Problem Setting

- **Solve**

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

- **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**

$$W^* = \{w_* \in R^d : \forall_{w \in R^d}, F(w) \geq F(w_*)\}$$

UCONN

Problem Setting

- **Solve**

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

- **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**

$$W^* = \{w_* \in R^d : \forall_{w \in R^d}, F(w) \geq F(w_*)\}$$

Stochastic Gradient Descend (SGD):

Initialize: w_0
Iterate:
for $t = 0, 1, 2, \dots,$ **do**
  Choose $\eta_t > 0$
  Generate random ξ_t
  Compute $\nabla f(w_t;\xi_t)$
  Update $w_{t+1} = w_t - \eta_t \nabla f(w_t;\xi_t)$
end for

Problem Setting

- **Solve**

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

- **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**

$$W^* = \{w_* \in R^d : \forall_{w \in R^d}, F(w) \geq F(w_*)\}$$

- **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:
for $t = 0, 1, 2, \ldots,$ **do**
  Choose $\eta_t > 0$
  Generate random ξ_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

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

ω-Convexity

Plain Convex

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

ω − Convex

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

$$\omega' > 0, \omega'' < 0,$$

Strongly Convex

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

ω-Convexity with curvature $h \in [0,1]$

Plain Convex

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

ω − Convex

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

$$\omega' > 0, \omega'' < 0,$$

Strongly Convex

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

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

$h = 0$ ⟷ $h \in (0,1)$ ⟷ $h = 1$

ω-Convexity with curvature $h \in [0,1]$

Plain Convex

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

ω − Convex

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

Strongly Convex

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

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

$h = 0$ ⟷ $h \in (0,1)$ ⟷ $h = 1$

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

HEB and ω-convexity are not subclasses of one another but they do intersection for $h \in (0,1]$.
[Bolte, J., Nguyen, T. P., Peypouquet, J., and Suter, B. W. From error bounds to the complexity of first order descent methods for convex functions. Mathematical Programming, 165(2):471–507, Oct 2017]

UCONN

Close to optimal stepsize

Plain Convex

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

$h = 0$

ω − Convex

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

Strongly Convex

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

$h = 1$

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

$h \in (0,1)$

SGD

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

Close to optimal stepsize

UCONN

Convergence Rate of SGD

Plain Convex

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

$h = 0$

ω − Convex

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

Strongly Convex

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

$h = 1$

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

$h \in (0,1)$

SGD

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

Close to optimal stepsize

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

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

UCONN

Convergence Rate of SGD

Plain Convex

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

$h = 0$

ω − Convex

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

Strongly Convex

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

$h = 1$

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

$h \in (0,1)$

$$E\left[\inf_{w_* \in W^*} ||w_t - w_*||^2\right] = O\big(t^{-h/(2-h)}\big)$$ [Useless,0] ⟷ [Useful,1]

$0 \leftarrow \quad h \quad \rightarrow 1$

$$\frac{1}{t}\sum_{i=t+1}^{2t} E[F(w_i) - F(w_*)] = O(t^{-1/(2-h)})$$ [Useful,0] ⟷ [Useful,1]

UCONN

Convergence Rate of SGD

Plain Convex

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

$h = 0$

ω – Convex

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

Strongly Convex

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

$h = 1$

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

$h \in (0,1)$

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

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

h= ½

$$F(w) = H(w) + \lambda G(w), H(w) - convex$$

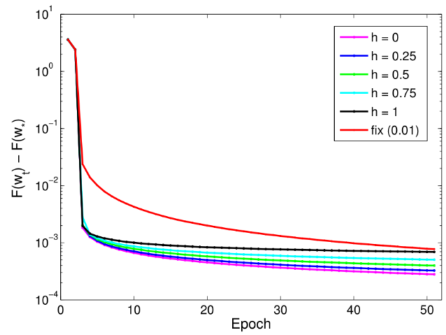
$$G(w) = \sum_{i=1}^{d} [e^{w_i} + e^{-w_i} - 2 - w_i^2]$$

UCONN

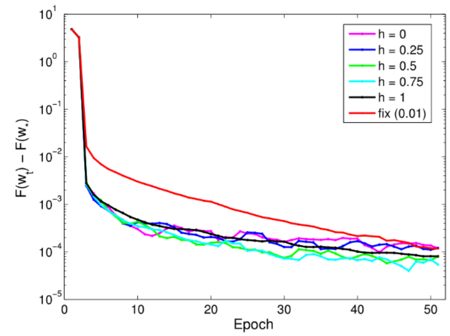
Experiment

Curvature 0 (convex)

$$f_i(w) = \log(1 + \exp(-y_i x_i^T w))$$



(a)



(b)

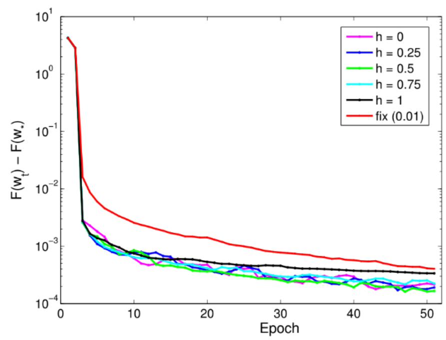
Curvature unknown

$$f_i^a(w) = f_i(w) + \lambda||w||$$

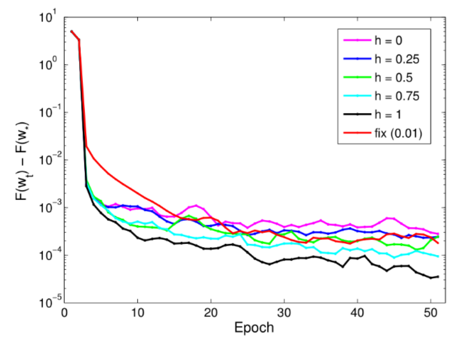
Curvature ½

$$f_i^a(w) = f_i(w) + \lambda G(w)$$

$$G(w) = \sum_{i=1}^{d} [e^{w_i} + e^{-w_i} - 2 \ - w_i^2]$$



(c)



(d)

Curvature 1 (strongly convex)

$$f_i^c(w) = f_i(w) + \frac{\lambda}{2}||w||^2$$

Conclusion

- ω-convexity notion: plain convex, strongly convex and something in between
- SGD with ω-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