Convex Optimization

Part 2: Gradient descent (more)

Namhoon Lee

POSTECH

26 Sep 2022

Admin

Attendance

- The university policy requires students to attend $> 3/4$ of a course to claim the credit. Please record your attendance using the app “포스텍 전자출결”.

Assignment 1

- due by this Friday

Strong convexity

f is strongly convex with parameter $\alpha > 0$ if, for all x, y and $t \in [0, 1]$,

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) - \frac{\alpha}{2} t(1-t)\|x - y\|_2^2$$

- A stronger version of convexity

For f being α-strongly convex is equivalent to the following:

$$f(x) - \frac{\alpha}{2}\|x\|_2^2 \quad \text{is convex.}$$

- For twice differentiable f this means $\nabla^2 f(x) \succeq \alpha I$.

A consequence of α-strong convexity

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\alpha}{2}\|y - x\|^2 \qquad \forall x, y$$

i.e., a quadratic lower bound on f.

Consequence of quadratic lower bound

Bound on suboptimality

$$\frac{\alpha}{2}\|x - x^*\|^2 \leq f(x) - f(x^*) \leq \frac{1}{2\alpha}\|\nabla f(x)\|_2^2$$

- The right-hand inequality is a.k.a. Polyak-Łojasiewicz (PL) inequality.

Proof.

The proof is done similarly as for smoothness. $\square$

Coercivity of gradient

Coercivity

If f is α-strongly convex, then

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \alpha \|x - y\|_2^2 \qquad \forall x, y$$

- a.k.a. strong monotonicity of ∇f.

Proof.

The proof follows by adding the quadratic lower bounds with x, y switched. $\square$

Extension of co-coercivity

Extension of co-coercivity

For f being α-strongly convex and β-smooth, the co-coercivity of gradient extends to

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \frac{\alpha\beta}{\alpha + \beta} \|x - y\|_2^2 + \frac{1}{\alpha + \beta} \|\nabla f(x) - \nabla f(y)\|_2^2 \qquad \forall x, y$$

Proof.

First, f being α-strongly convex means the following is convex.

$$g(x) = f(x) - \frac{\alpha}{2}\|x\|_2^2$$

Thus,

$$\begin{aligned}
0 &\leq \langle \nabla g(x) - \nabla g(y), x - y\rangle && \text{(monotonicity of } g\text{)} \\
&= \langle \nabla f(x) - \nabla f(y), x - y\rangle - \alpha\|x - y\|_2^2 && \text{(def. of } g\text{)} \\
&\leq (\beta - \alpha)\|x - y\|_2^2 && \text{(}\beta\text{-smoothness)}
\end{aligned}$$

which shows that g is $(\beta - \alpha)$-smooth (from the first and last lines).
Then writing out co-coercivity of ∇g (and rearranging terms) will finish the proof. □

Convergence for smooth and strongly convex functions

Theorem

For β-smooth and α-strongly convex functions, gradient descent with the step size $\eta = 2/(\alpha+\beta)$ after T iterations satisfies

$$f(x_{T+1}) - f^* \le \rho^T \frac{\beta R^2}{2}$$

where $\rho = \left(\frac{\kappa-1}{\kappa+1}\right)^2$ *with* $\kappa = \beta/\alpha$ *and* $R = \|x_1 - x^*\|_2$.

- This achieves the linear convergence rate of $\mathcal{O}(\rho^t)$.
- The number of iterations to reach ϵ-accuracy is $\mathcal{O}(\log(1/\epsilon))$.
- Big κ leads to slow convergence.

Proof.

For GD with step size $\eta = 2/(\alpha + \beta)$ we can write

$$\begin{aligned}
\|x_{t+1} - x^*\|_2^2 &= \|x_t - \frac{2}{\alpha + \beta}\nabla f(x_t) - x^*\|_2^2 \\
&= \|x_t - x^*\|_2^2 - \frac{4}{\alpha + \beta}\langle \nabla f(x_t), x_t - x^* \rangle + \left(\frac{2}{\alpha + \beta}\right)^2 \|\nabla f(x_t)\|_2^2 \\
&\leq \left(\frac{\alpha - \beta}{\alpha + \beta}\right)^2 \|x_t - x^*\|_2^2
\end{aligned}$$

where κ the last inquality follows from the extension of co-coercivity. Expanding on t

$$\|x_{t+1} - x^*\|_2^2 \leq \rho^t \|x_1 - x^*\|_2^2$$

where $\rho = \left(\frac{\kappa-1}{\kappa+1}\right)^2$ with $\kappa = \beta/\alpha$. Further using the suboptimality bound we derived previously will finish the proof. $\square$

Any questions?