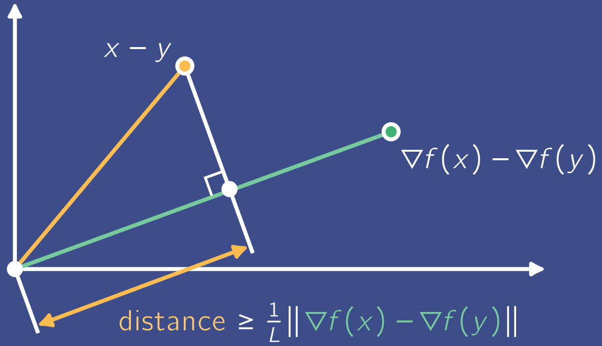
Baillon-Haddad Theorem

a bridge from functions to monotone operators



Howard Heaton

Typal Academy

Overview

Smoothness bounds gradients and aligns them with the path of descent

Baillon-Haddad Theorem

$$(\text{convex}) + (L\text{-smooth}) \implies \left(\tfrac{1}{L}\text{-cocoercive}\right)$$

Significance

- Bridges convex analysis and monotone operator theory
- Simplifies convergence proofs for popular methods (*e.g.* gradient descent)

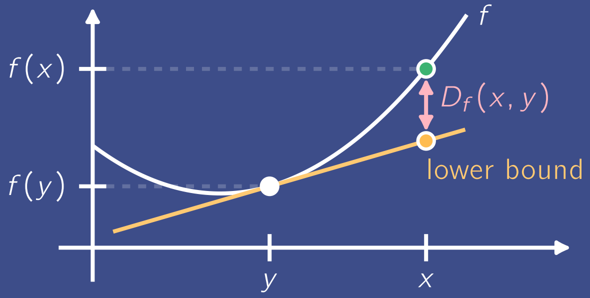
Convexity and Bregman Divergences

For a differentiable function $f\colon \mathbb{R}^n \to \mathbb{R}$, the Bregman divergence D_f is

$$D_f(x, y) = f(x) - f(y) - (x - y) \cdot \nabla f(y).$$

Convexity of f ensures the Bregman divergence is always nonnegative,

i.e. the linear estimate from y to x never exceeds function value at x.



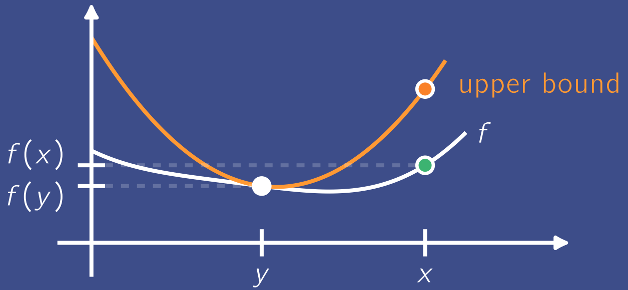
L-smooth Functions

A differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is L-smooth provided ∇f is L-Lipschitz, *i.e.*

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \quad \text{for all } x \text{ and } y.$$

Descent Lemma If f is L-smooth, then

$$f(x) \leq f(y) + (x - y) \cdot \nabla f(y) + \frac{L}{2}\|x - y\|^2, \quad \text{for all } x \text{ and } y.$$



Operators

Each operator herein is a function mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$.

An operator Q is nonexpansive provided it is 1-Lipschitz, *i.e.*

$$\|Q(x) - Q(y)\| \leq \|x - y\|, \quad \text{for all } x \text{ and } y.$$

An operator T is averaged is there is $\alpha \in (0,1)$ and nonexpansive Q such that

$$T(x) = (1-\alpha)x + \alpha Q(x) = x + \alpha\big(Q(x) - x\big), \quad \text{for all } x \text{ and } y.$$

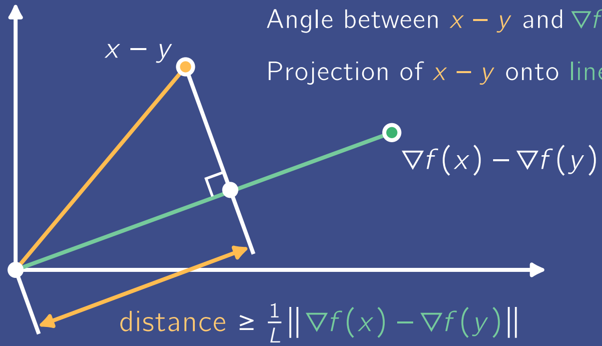
An operator C is β-cocoercive provided

$$(x-y)\cdot\big(C(x) - C(y)\big) \geq \beta\|C(x) - C(y)\|^2, \quad \text{for all } x \text{ and } y.$$

Baillon-Haddad Theorem

If $f : \mathbb{R}^n \to \mathbb{R}$ is convex and L-smooth, then the gradient ∇f is $\frac{1}{L}$-cocoercive, *i.e.*

$$(x - y) \cdot \Big(\nabla f(x) - \nabla f(y) \Big) \geq \frac{1}{L} \| \nabla f(x) - \nabla f(y) \|^2, \quad \text{for all } x \text{ and } y.$$



Angle between $x - y$ and $\nabla f(x) - \nabla f(y)$ is acute

Projection of $x - y$ onto line has length $\geq \frac{1}{L} \|\nabla f(x) - \nabla f(y)\|$

Proof – Baillon-Haddad Theorem

Let x and y be given, and set $z = x - \frac{1}{L}(\nabla f(x) - \nabla f(y))$. Since f is L-smooth,

$$0 \leq f(x) - f(z) + (z - x) \cdot \nabla f(x) + \frac{L}{2}\|z - x\|^2 = -D_f(z, x) + \frac{L}{2}\|z - x\|^2.$$

Algebraic manipulations reveal

$$D_f(z, x) = D_f(z, y) - D_f(x, y) + (x - z) \cdot \Big(\nabla f(x) - \nabla f(y)\Big).$$

These two equations together imply

$$\begin{aligned} D_f(x, y) &\geq D_f(z, y) + (x - z) \cdot \Big(\nabla f(x) - \nabla f(y)\Big) - \frac{L}{2}\|z - x\|^2 \\ &\geq (x - z) \cdot \Big(\nabla f(x) - \nabla f(y)\Big) - \frac{L}{2}\|z - x\|^2, \end{aligned}$$

where the final inequality holds since convexity of f implies $D_f(z, y) \geq 0$.

Proof – Baillon-Haddad Theorem

Substituting for our choice of z yields

$$D_f(x, y) \geq (x - z) \cdot \Big(\nabla f(x) - \nabla f(y)\Big) - \frac{L}{2}\|z - x\|^2$$
$$= \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2.$$

Analogous argument shows

$$D_f(y, x) \geq \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2.$$

Adding these inequalities yields

$$\frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2 \leq D_f(x, y) + D_f(y, x)$$
$$= (x - y) \cdot \Big(\nabla f(x) - \nabla f(y)\Big).$$

∎

Gradient Descent Convergence

If $\alpha \in (0, 2/L)$ and f is convex, L-smooth and has a minimizer, then the iteration

$$x^{k+1} = x^k - \alpha \nabla f(x^k)$$

produces $\{x^k\}$ converging to a minimizer of f and $\|\nabla f(x^k)\|^2 = \mathcal{O}(1/k)$.

Proof Set $T(x) = x - \alpha \nabla f(x)$ so that $x^{k+1} = T(x^k)$. It suffices to show[†]

1) T is averaged, and

2) minimizers of f coincide with fixed points of T.

Since f is convex and L-smooth, ∇f is $\frac{1}{L}$-cocoercive.

[†]See previous lecture on Krasnosel'skiĭ-Mann iteration.

Gradient Descent with $2/L \implies$ Nonexpansive

Set $Q(x) = x - \frac{2}{L}\nabla f(x)$. The $\frac{1}{L}$-cocoercivity of ∇f implies, for all x and y,

$$\begin{aligned}
&\|Q(x) - Q(y)\|^2 \\
&= \|x - y\|^2 - \frac{4}{L}(x-y)^\top\big(\nabla f(x) - \nabla f(y)\big) + \frac{4}{L^2}\|\nabla f(x) - \nabla f(y)\|^2 \\
&= \|x - y\|^2 - \frac{4}{L}\underbrace{\left[(x-y)^\top\big(\nabla f(x) - \nabla f(y)\big) - \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2\right]}_{\geq 0} \\
&\leq \|x - y\|^2.
\end{aligned}$$

Taking square roots yields $\|Q(x) - Q(y)\| \leq \|x - y\|$, and so Q is nonexpansive.

Update is Averaged and Fixed Points are Minimizers

Setting $\theta = \alpha L/2$ yields $\theta \in (0, 1)$ and

$$T(x) = x - \frac{2\theta}{L}\nabla f(x) = (1-\theta)x + \theta Q(x) \quad \Longrightarrow \quad T \text{ is averaged.}$$

For convex and differentiable f,

$$\begin{aligned} x^\star \text{ minimizes } f \iff\ & 0 = \nabla f(x^\star) \\ \iff\ & 0 = -\alpha \nabla f(x^\star) \\ \iff\ & x^\star = x^\star - \alpha \nabla f(x^\star) \\ \iff\ & x^\star = T(x^\star). \end{aligned}$$

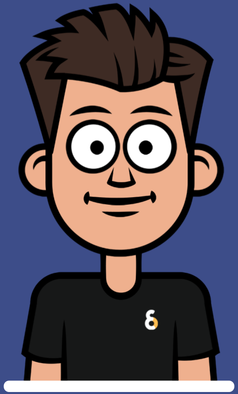
■

Summary

- Together L-smoothness and convexity of f yield $\frac{1}{L}$-cocoercivity of ∇f
- Baillon-Haddad theorem links smooth functions to monotone operator theory
- Enables simple algorithm design and analysis

Found this useful?

- ➕ Follow for more
- ♻️ Share it with friends
- 📝 Pre-order my online course



Howard Heaton

References

- Baillon, Haddad. *Quelques proprieétés des opérateurs angle-bornés et n-cycliquement monotones*, 1977.
- Bauschke, Combettes. *The Baillon-Haddad Theorem Revisited*, 2010.
- Beck. *First-Order Methods in Optimization*, 2017.
- Byrne. *On A Generalized Baillon–Haddad Theorem for Convex Functions on Hilbert Space*, 2014.
- Ryu, Yin. *Large-Scale Convex Optimization: Algorithms & Analysis via Monotone Operators*, 2022.