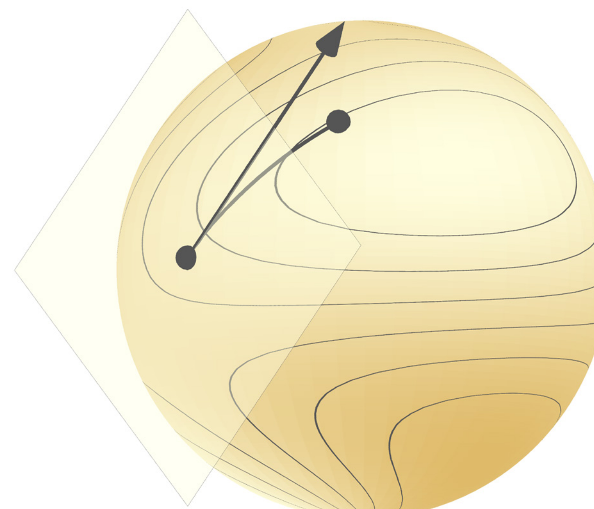
213

Linear convergence with Polyak–Łojasiewicz



Spring 2023

Optimization on manifolds, MATH 512 @ EPFL

Instructor: Nicolas Boumal

A comfortable assumption to make

Let $f : \mathcal{M} \to \mathbf{R}$ be differentiable on a Riemannian manifold $\mathcal{M}$.

Def.: We say f satisfies the Polyak–Łojasiewicz condition with constant $\mu > 0$ on a set $S \subseteq \mathcal{M}$ if

$$f(x) - f^* \leq \frac{1}{2\mu} \|\mathrm{grad} f(x)\|_x^2 \qquad \text{for all } \; x \in S,$$

where $f^* = \inf_{x \in S} f(x)$.

In words: within S, the gradient norm^2 bounds the optimality gap.

PŁ + sufficient decrease ⇒ linear cvgce

Let $f : \mathcal{M} \to \mathbf{R}$ be differentiable on a Riemannian manifold $\mathcal{M}$.
Consider a sequence of points $x_0, x_1, x_2, \ldots$ on $\mathcal{M}$.

Theorem. Assume the following hold for all k:

1. Decrease: $f(x_k) - f(x_{k+1}) \geq \frac{1}{2L} \|\mathrm{grad} f(x_k)\|_{x_k}^2$
2. PŁ: $f(x_k) - f^* \leq \frac{1}{2\mu} \|\mathrm{grad} f(x_k)\|_{x_k}^2$

Then, $f(x_k) - f^* \leq \left(1 - \frac{1}{L/\mu}\right)^k (f(x_0) - f^*) \qquad \forall k$

Pf :

$$
\begin{aligned}
f(x_{k+1}) - f^* &= f(x_{k+1}) - f(x_k) + f(x_k) - f^* \\
&\leq -\frac{1}{2L} \|\operatorname{grad} f(x_k)\|_{x_k}^2 + f(x_k) - f^* \\
&\leq -\frac{\cancel{2}\mu}{\cancel{2}L}\left(f(x_k) - f^*\right) + f(x_k) - f^* \\
&= \left(1 - \frac{\mu}{L}\right)\left(f(x_k) - f^*\right).
\end{aligned}
$$

PŁ holds for geodesically strongly convex f

Fact/Def.: A function $f: \mathcal{M} \to \mathbf{R}$ is geodesically μ-strongly convex on a geodesically convex set S if

$$\operatorname{Hess} f(x) \succcurlyeq \mu \operatorname{Id} \sim \langle v, \operatorname{Hess} f(x)[v] \rangle_x \geq \mu \|v\|_x^2$$

for all $x \in S$.

Fact: Under that definition, for all $x, x^* \in S$, we have:

$$f(x) - f(x^*) \leq \frac{1}{2\mu} \|\operatorname{grad} f(x)\|_x^2$$

can be replaced with $\inf_{x^* \in S} f(x^*) \doteq f^*$.

Since $x, x^* \in S$ which is g-convex, there exists a geodesic segment $c : [0,1] \to \mathcal{M}$ s.t. $c(0) = x$, $c(1) = x^*$, $c(t) \in S$ $\forall t \in [0,1]$

let $g = f \circ c$;

$$f(x^*) = g(1) = g(0) + \int_0^1 g'(t)\,dt = g(0) + \int_0^1 g'(0) + \int_0^t g''(\tau)\,d\tau\,dt$$

$$g'(t) = Df(c(t))[c'(t)] = \langle \operatorname{grad} f(c(t)), c'(t) \rangle_{c(t)}$$

$$g''(t) = \left\langle \frac{D}{dt} \operatorname{grad} f(c(t)), c'(t) \right\rangle_{c(t)} + \underbrace{\cancel{\left\langle \operatorname{grad} f(c(t)), \frac{D}{dt} c'(t) \right\rangle_{c(t)}}}_{0}$$

$$= \langle \operatorname{Hess} f(c(t))[c'(t)], c'(t) \rangle_{c(t)}$$

$$\geq \mu \|c'(t)\|^2_{c(t)} \quad \forall t \in [0,1]$$ since f is μ-strongly g-convex in S.

$$\frac{d}{dt}\|c'(t)\|^2_{c(t)} = \frac{d}{dt}\langle c'(t), c'(t)\rangle_{c(t)} = 0$$

$$= \mu\|c'(0)\|_x^2 \doteq \mu\|v\|_x^2 \quad \text{w/} \quad v = c'(0).$$

$$\Rightarrow f(x^*) \geq f(x) + \langle \operatorname{grad} f(x), v\rangle_x + \int_0^1\int_0^t \mu\|v\|_x^2 \, d\tau \, dt$$

$$= f(x) + \langle \operatorname{grad} f(x), v\rangle_x + \frac{\mu}{2}\|v\|_x^2$$

$$\geq \inf_{u \in T_x\mathcal{M}} \left[f(x) + \langle \operatorname{grad} f(x), u\rangle_x + \frac{\mu}{2}\|u\|_x^2 \right]$$

$u = -\frac{1}{\mu}\operatorname{grad} f(x)$

$$= f(x) - \frac{1}{2\mu}\|\operatorname{grad} f(x)\|_x^2$$

To learn more about the PŁ condition, see for example arxiv.org/abs/2303.00096.

Application: intrinsic means

Given m points $x_1, \dots, x_m$ on a Riemannian manifold (say, a sphere), what is a good notion of mean or average of those points?

The Fréchet mean is any minimizer of:

$$f(x) = \frac{1}{2m} \sum_{i=1}^{m} \mathrm{dist}(x, x_i)^2$$

It's fun to explore how the above applies to this f.