Optimization for Machine Learning

TD 1 : Gradient Dynamics and Convexity

The exercises difficulty is indicated by the number of stars. Exercises marked with $\emptyset$ are beyond the scope of this class and aren't expected to be done or prepared.

Lyapunov energies for the gradient flow

Here we assume that $f \in \Gamma_0(H) \cap C^1(H)$, and that $u \in C^1([0, +\infty[; H)$ is a trajectory solution of

$$(\forall t > 0) \quad \dot{u}(t) + \nabla f(u(t)) = 0, \quad u(0) = u^0. \tag{1}$$

We will say that a function $\mathcal{E} : H \to \mathbb{R}$ is a Lyapunov energy *if $t \mapsto \mathcal{E}(u(t))$ is decreasing.*

(*) **Exercise 1** (Nonexpansive dynamic). Let $u, v : [0, +\infty[\to H$ be two solutions of the continuous Gradient Descent dynamic (1), starting from two distinct points $u(0)$ and $v(0)$. Show that this dynamic is nonexpansive:

$$(\forall t > 0)(\forall h \geq 0) \quad \|u(t+h) - v(t+h)\| \leq \|u(h) - v(h)\|.$$

To prove this, you can study the variations of $t \mapsto (1/2)\|u(t+h) - v(t+h)\|^2$.

Correction:

$$\begin{aligned}
\frac{d}{dt}(1/2)\|u(t+h) - v(t+h)\|^2 &= \langle \dot{u}(t+h) - \dot{v}(t+h), u(t+h) - v(t+h) \rangle \\
&= -\langle \nabla f(u(t+h)) - \nabla f(v(t+h)), u(t+h) - v(t+h) \rangle \\
&\leq 0 \quad \text{because } \nabla f \text{ is monotone.}
\end{aligned}$$

We therefore just proved that this function is decreasing, which means that at any t it is smaller than its value at $t = 0$. Which is what we wanted.

(*) **Exercise 2** (Getting closer to the solution). Let $x^* \in \operatorname{argmin} f \neq \emptyset$. Show that $\mathcal{E}(x) = (1/2)\|x - x^*\|^2$ is a Lyapunov energy.

Correction:

$$\begin{aligned}
(\mathcal{E} \circ x)'(t) &= \frac{d}{dt}(1/2)\|x(t) - x^*\|^2 = \langle \dot{x}(t), x(t) - x^* \rangle = \langle \nabla f(x(t)), x^* - x(t) \rangle \\
&\leq f(x^*) - f(x(t)) \quad \text{because } f \text{ is convex} \\
&\leq 0 \quad \text{because } f(x^*) = \inf f \leq f(x(t)).
\end{aligned}$$