

A Proof of Proposition 1

Proposition 1 (Link between MixUp and gradient Lipschitz continuity). *The property $P1 \Rightarrow P2$.*

Proof. For all x and x' in \mathbb{R}^d we have

$$\begin{aligned} & f(\lambda x + (1 - \lambda)x') \\ &= f(x') + \lambda \int_0^1 \langle \nabla f(\lambda tx + (1 - \lambda)t x'), x - x' \rangle dt \\ &= f(x') + \lambda [f(x) - f(x')] + \lambda \left[\int_0^1 \langle \nabla f(\lambda tx + (1 - \lambda)t x'), x - x' \rangle dt - (f(x) - f(x')) \right] \end{aligned}$$

Therefore

$$\begin{aligned} & |f(\lambda x + (1 - \lambda)x') - (\lambda f(x) + (1 - \lambda)f(x'))| \\ &= \lambda \left| \int_0^1 \langle \nabla f(\lambda tx + (1 - \lambda)t x') - \nabla f(tx + (1 - t)x'), x - x' \rangle dt \right| \\ &\leq \lambda \int_0^1 |\langle \nabla f(\lambda tx + (1 - \lambda)t x') - \nabla f(tx + (1 - t)x'), x - x' \rangle| dt \\ &\leq \lambda \int_0^1 \|\nabla f(\lambda tx + (1 - \lambda)t x') - \nabla f(tx + (1 - t)x')\| \|x - x'\| dt \\ &\leq \lambda \int_0^1 (1 - \lambda)t \kappa \|x - x'\|^2 dt \\ &= \frac{\lambda(1 - \lambda)\kappa}{2} \|x - x'\|^2, \end{aligned}$$

where the second inequality follows Cauchy-Schwarz inequality and the third inequality is the property $P1$. \square