MAP569 Machine Learning II

PC4 : Convergence rates in optimization

1 Convergence rates for Projected Gradient Descent

We consider a convex function $f : \mathbb{R}^d \to \mathbb{R}$, a closed convex set $\mathcal{C} \subset \mathbb{R}^d$ and the optimisation problem

$$\min_{x \in \mathcal{C}} f(x). \tag{1}$$

We denote by $\pi_{\mathcal{C}} x = \operatorname{argmin}_{u \in \mathcal{C}} \|x - u\|^2$ the projection of x onto the convex set $\mathcal{C}$. To solve (1), we can apply the Projected Gradient Descent algorithm (with $\eta > 0$) :

For $k = 1, \ldots, K-1$,

$$y_{k+1} = x_k - \eta \nabla f(x_k) \ ,$$
$$x_{k+1} = \pi_{\mathcal{C}} y_{k+1} \ ,$$

Return $\quad f(x_K)$.

1.1 Basic facts

1. Let $u \in \mathcal{C}$ and $0 < t < 1$. Why do we have $\|z - (tu + (1-t)\pi_{\mathcal{C}} z)\|^2 \geq \|z - \pi_{\mathcal{C}} z\|^2$?
2. Investigating this inequality for t small, prove that

$$\langle u - \pi_{\mathcal{C}} z, z - \pi_{\mathcal{C}} z \rangle \leq 0 \quad \text{and} \quad \|\pi_{\mathcal{C}} z - z\|^2 + \|u - \pi_{\mathcal{C}} z\|^2 \leq \|u - z\|^2.$$

3. Assume that f differentiable and convex. For any $x, h \in \mathbb{R}^d$ and $t \in [0,1]$, we set $F(t) = f(x + th)$. Prove that $F(1) - F(0) \geq F'(0)$. Conclude that $f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle$ for all x, y.

1.2 Rate for Lipschitz convex functions

We assume here that $\mathcal{C} \subset B(x_1, R)$. Let x^* be a minimizer of (1) and define $\bar{x}_K = (x_1 + \ldots + x_K)/K$. In this section, we will prove that if $\|\nabla f(x)\| \leq L$ for all $x \in \mathcal{C}$, and $\eta = R/(L\sqrt{K})$, then

$$f(\bar{x}_K) - f(x^*) \leq \frac{LR}{\sqrt{K}} \ .$$

1. Using question 1.1.3, prove that

$$f(x_k) - f(x^*) \leq \frac{1}{\eta} \langle x_k - y_{k+1}, x_k - x^* \rangle = \frac{\eta}{2} \|\nabla f(x_k)\|^2 + \frac{1}{2\eta} \left(\|x_k - x^*\|^2 - \|y_{k+1} - x^*\|^2 \right).$$

2. Using question 1.1.2, prove that

$$\frac{1}{K} \sum_{k=1}^{K} f(x_k) - f(x^*) \leq \frac{\eta L^2}{2} + \frac{\|x_1 - x_*\|^2}{2\eta K} \ .$$

3. Conclude.

1.3 Rate for strongly convex functions

When the function f is strongly convex, then the PGD converges much faster. In the following, we assume that f is α-strongly convex :

$$f(y) - f(x) \leq \langle \nabla f(y), y - x \rangle - \frac{\alpha}{2} \|y - x\|^2 . \tag{2}$$

We also assume that ∇f is β-Lipschitz. We will prove that, for $\eta = 1/\beta$,

$$\|x_{K+1} - x^*\|^2 \leq \|x_1 - x^*\|^2 \, e^{-\rho K} \ ,$$

with $\rho = \alpha/\beta$.
Define $g(x) = \beta \left(x - \pi_{\mathcal{C}}(x - \frac{1}{\beta} \nabla f(x)) \right)$. The key of the proof is the inequality :

$$\forall (x, y) \in \mathcal{C}^2 \ , \ \ f\left(\pi_{\mathcal{C}}(x - \frac{1}{\beta} \nabla f(x)) \right) - f(y) \leq \langle g(x), x - y \rangle - \frac{1}{2\beta} \|g(x)\|^2 - \frac{\alpha}{2} \|x - y\|^2 \ , \tag{3}$$

which evaluates the progress made by one step of the PGD algorithm.

1. Assume first that (3) holds. Prove the following (in)equalities :

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &= \|x_k - x^*\|^2 - \frac{2}{\beta} \langle g(x_k), x_k - x^* \rangle + \frac{1}{\beta^2} \|g(x_k)\|^2 \ , \\ &\leq (1 - \rho) \|x_k - x^*\|^2 \leq e^{-\rho k} \|x_1 - x^*\|^2 . \end{aligned}$$

2. It remains to prove (3). With the mean value theorem, prove that

$$f(y) - f(x) = \int_0^1 \langle \nabla f(x + t(y - x)), y - x \rangle \, dt \leq \langle \nabla f(x), y - x \rangle + \frac{\beta}{2} \|y - x\|^2 \ . \tag{4}$$

3. Set $x^+ = \pi_{\mathcal{C}}(x - \frac{1}{\beta} \nabla f(x))$. Using (2) and (4), check that

$$f(x^+) - f(y) \leq \langle \nabla f(x), x^+ - x \rangle + \frac{\beta}{2} \|x^+ - x\|^2 + \langle \nabla f(x), x - y \rangle - \frac{\alpha}{2} \|y - x\|^2 \ .$$

4. With question 1.1.2, prove that $\langle \nabla f(x), x^+ - y \rangle \leq \langle g(x), x^+ - y \rangle$ for all $y \in \mathcal{C}$.
5. Conclude that

$$\begin{aligned} f(x^+) - f(y) &\leq \langle g(x), x^+ - y \rangle + \frac{1}{2\beta} \|g(x)\|^2 - \frac{\alpha}{2} \|y - x\|^2 \ , \\ &= \langle g(x), x - y \rangle - \frac{1}{2\beta} \|g(x)\|^2 - \frac{\alpha}{2} \|y - x\|^2 . \end{aligned}$$