Exercises: Gradient algorithms

Exercise 1 (A quadratic example in $\mathbb{R}^2$). *Consider, for $\gamma > 0$, $f(x) = \frac{1}{2}(x_1^2 + \gamma x_2^2)$. We apply the gradient descent method with optimal step, starting at $x^{(0)} = (\gamma, 1)$.*

1. *Show that f is m-convex with M-Lipschitz gradient. Find the tightest m and M constants.*

2. *Show that*
$$x^{(k)} = \left(\gamma\Big(\frac{\gamma-1}{\gamma+1}\Big)^k, \Big(-\frac{\gamma-1}{\gamma+1}\Big)^k\right)$$
and
$$f(x^{(k)}) = \frac{\gamma(\gamma+1)}{2}\Big(\frac{\gamma-1}{\gamma+1}\Big)^{2k} f(x^{(0)})$$

3. *Show that, on this example, the convergence is exactly linear, that is $f(x^{(k)}) - v^\sharp$ is a geometric series. Give its reason. Compare with the theoretical bound.*

4. *When is this algorithm fast and slow ?*

Exercise 2 (Strongly convex - optimal step). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a m-convex $\mathcal{C}^2$ function. Define, for given $x^{(0)}$,*
$$\begin{aligned}
&\tilde{f}_k : t \mapsto f(x^{(k)} - t\nabla f(x^{(k)}))\\
&t^{(k)} = \arg\min_{t\in\mathbb{R}} \tilde{f}_k(t)\\
&x^{(k+1)} = x^{(k)} - t^{(k)}\nabla f(x^{(k)})
\end{aligned}$$

1. *Show that there exists $M \geq m$ such that $mI \preceq \nabla^2 f(x^{(k)}) \preceq MI$*

2. *Show that, for any interesting t (to be defined) we have*
$$\tilde{f}_k(t) \leq f(x^{(k)}) - t\|\nabla f(x^{(k)})\|_2^2 + \frac{Mt^2}{2}\|\nabla f(x^{(k)})\|_2^2$$

3. *Show that,*
$$f(x^{(k+1)}) \leq f(x^{(k)}) - \frac{1}{2M}\|\nabla f(x^{(k)})\|_2^2$$

4. *Show that $f(x^\sharp) \geq f(x) - \frac{1}{2m}\|\nabla f(x)\|_2^2$*

5. *Show that*
$$f(x^{(k+1)}) - f(x^\sharp) \leq (1 - \frac{m}{M})[f(x^{(k)}) - f(x^\sharp)]$$

6. *Show that the algorithm converges, and give its convergence speed.*

Exercise 3 (Strictly convex case). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^2, convex function with M-Lipschitz gradient such that $f(x^\sharp) = \inf f$. We define, for given $x^{(0)}$.*
$$x^{(k+1)} = x^{(k)} - t\nabla f(x^{(k)})$$
with $t \leq 1/M$.

1. *Show that, for all x and y $(y-x)^\top \nabla^2 f(z)(y-x) \leq M\|y-x\|_2^2$*

2. *Show that*
$$f(y) \leq f(x) + \nabla f(x)^\top (y-x) + \frac{M}{2}\|y-x\|^2$$

3. *Show that $f(x^{(k+1)}) \leq f(x^{(k)}) - \frac{t}{2}\|\nabla f(x^{(k)})\|^2$*

4. *Show that*
$$f(x^{(k+1)}) \leq v^\sharp + \nabla f(x^{(k)})^\top (x^{(k)} - x^\sharp) - \frac{t}{2}\|\nabla f(x^{(k)})\|^2$$

5. *Deduce that*
$$f(x^{(k+1)}) \leq v^\sharp + \frac{1}{2t}(\|x^{(k)} - x^\sharp\|^2 - \|x^{(k+1)} - x^\sharp\|^2)$$

6. *Show that*
$$\sum_{i=1}^{k} f(x^{(i)}) - v^\sharp \leq \frac{1}{2t}\|x^{(0)} - x^\sharp\|^2$$

7. *Conclude that*

$$f(x^{(k)}) - v^\sharp \leq \frac{1}{2kt}\|x^{(0)} - x^\sharp\|^2.$$

Exercise 4 (Kelley's convergence)**.** *We are going to prove that, if $f : \mathbb{R}^n \to \mathbb{R}$ is convex, and X a non-empty polytope (bounded polyhedron) then Kelley's cutting plane algorithm is converging.*
Consider $x_1 \in X$. We consider a sequence of points $(x^{(k)})_{k\in\mathbb{N}}$ such that $x^{(k+1)}$ is an optimal solution to

$$(\mathcal{P}^{(k)}) \qquad \underline{v}^{(k+1)} = \underset{x\in X}{\text{Min}} \quad z \qquad s.t. \quad f(x^{(\kappa)}) + \left\langle g^{(\kappa)}, x - x^{(\kappa)} \right\rangle \leq z \quad \forall \kappa \in [k]$$

where $g^{(k)} \in \partial f(x^{(k)})$.
Denote $v = \min_{x\in X} f(x)$.

1. *Show that v exists and is finite, and that there exists a sequence $x^{(k)}$.*

2. *Show that there exists L such that, for all k_1 and k_2, we have $\|f(x^{(k_1)}) - f(x^{(k_2)})\| \leq L\|x^{(k_1)} - x^{(k_2)}\|$, and $\|g^{(k)}\| \leq L$.*

3. *Let $K_\varepsilon = \{k \in \mathbb{N} \mid f(x^{(k)}) > v + \varepsilon\}$ be the set of index such that $x^{(k)}$ is not an ε-optimal solution. Show that $f(x_k) \to v$ if and only if K_ε is finite for all $\varepsilon > 0$*

4. *Consider $k_1, k_2 \in K_\varepsilon$, such that $k_2 > k_1$. Show that*

   $$f(x^{(k_1)}) + \left\langle g^{(k_1)}, x^{(k_2)} - x^{(k_1)} \right\rangle \leq \underline{v}^{(k_2)} \leq v$$

5. *Show that $\varepsilon + f(x^{(k_1)}) + \left\langle g^{(k_1)}, x^{(k_2)} - x^{(k_1)} \right\rangle < f(x^{(k_2)})$*

6. *Show that $\varepsilon < 2L\|x^{(k_2)} - x^{(k_1)}\|$.*

7. *Prove that $f(x^{(k)}) \to v$.*

8. *(Optional - hard) Find a complexity bound for the method (that is a number of iteration N_ε after which you are sure to have obtained a ε-optimal solution).*