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.*

Answers:

1. Let $S = \mathrm{lev}_{f(x^{(0)})} f$. By strong convexity it is a bounded set. f being C^2, its Hessian is continuous and thus bounded on S, where all $x^{(k)}$ lives.

2. For any t such that $y := x^{(k)} - t\nabla f(x^{(k)}) \in S$, there exists $z \in [x^{(k)}, y]$, such that

$$\begin{aligned}
f(y) =& f(x^{(k)}) - \nabla f(x^{(k)})^\top (y - x^{(k)}) \\
&+ \frac{1}{2}(y - x^{(k)})^\top \nabla^2 f(z)(y - x^{(k)})
\end{aligned}$$

replacing y by its value, and using the upper bound on $\nabla^2 f(z)$ yields the result.

3. Use $t = 1/M$ in the upperbound of the previous question. Note that this choice of t minimizes said upper bound.

4. We have, by strong convexity,

$$\begin{aligned}
f(y) \geq& f(x) + \nabla f(x)^\top (y - x) + \frac{m}{2}\|y - x\|^2 \\
\geq& f(x) + \nabla f(x)^\top (\tilde{y} - x) + \frac{m}{2}\|\tilde{y} - x\|^2 \\
=& f(x) - \frac{1}{2m}\|\nabla f(x)\|^2
\end{aligned}$$

where $\tilde{y}$ minizes the right hand side. We then apply to $y = x^\sharp$.

5. By the previous question we have $\|\nabla f(x^{(k)})\|^2 \geq 2m(f(x^{(k)}) - v^\sharp)$. Question 3 then yields

$$f(x^{(k+1)}) \leq f(x^{(k)}) - m/M(f(x^{(k)}) - v^\sharp)$$

substracting $v^\sharp$ on each sides yields the result.

6. Recursively we get $f(x^{(k)}) - v^\sharp \leq c^k(f(x^{(0)}) - v^\sharp)$, with $c = 1 - m/M$. In particular, for any $\varepsilon > 0$, we have $f(x^{(k)}) - v^\sharp \leq \varepsilon$ after at most $\frac{\ln(\varepsilon) - \ln(f(x^{(0)} - v^\sharp))}{\ln(c)}$ iterations.

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*

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

Answers:

1. We have $\nabla^2 f(x) \preceq MI$, thus $(y-x)^\top (MI - \nabla^2 f(x))(y-x) \geq 0$.
2. Using Taylor remainder theorem we have the existence of $z \in [x, y]$ such that

$$f(y) = f(x) + \nabla f(x)^\top (y-x) + \frac{1}{2}(y-x)^\top \nabla^2 f(z)(y-x)$$

3. We obtain

$$f(x^{(k+1)}) \leq f(x^{(k)}) - (1 - \frac{Mt}{2})\|\nabla f(x^{(k)})\|_2^2$$

and with the condition on t we have $1 - Mt/2 \geq 1/2$.

4. We use the convexity inequality to get $f(x^{(k)}) \leq f(x^\sharp) + \nabla f(x^{(k)})^\top (x^{(k)} - x^\sharp)$ and the result of the previous question.
5. We have

$$\begin{aligned}
&\frac{1}{2t}(\|x^{(k)} - x^\sharp\|^2 - \|x^{(k+1)} - x^\sharp\|^2) \\
&\quad = \frac{1}{2t}(\|x^{(k)} - x^\sharp\|^2 - \|x^{(k)} - x^\sharp - t\nabla f(x^{(k)})\|^2) \\
&\quad = \nabla f(x^{(k)})^\top (x^{(k)} - x^\sharp) - \frac{t}{2}\|\nabla f(x^{(k)})\|^2
\end{aligned}$$

Thus,

$$\begin{aligned}
f(x^{(k+1)}) &\leq v^\sharp + \nabla f(x^{(k)})^\top (x^{(k)} - x^\sharp) - \frac{t}{2}\|\nabla f(x^{(k)})\|^2 \\
&= v^\sharp + \frac{1}{2t}(\|x^{(k)} - x^\sharp\|^2 - \|x^{(k+1)} - x^\sharp\|^2)
\end{aligned}$$

6. Sum the previous inequality
7. As $f(x^{(i)}) - v^\sharp$ is non-increasing we have that the last term of the sum is lower than the mean of the sum.

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*

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

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).*

Answers:

1. f is finite convex and thus continuous on X which is compact, yielding the existence and finiteness of v.

   f is subdifferentiable, thus we have the existence of $g^{(k)}$, and an optimal solution to $\mathcal{P}^{(k)}$ exists as the solution of a bounded linear programm.

2. We have seen that on any compact Kincluded in the domain of a convex function f, f is L-Lipschitz. Here $\text{dom}(f) = \mathbb{R}^n$, so on the compact $K = X + B(0, \varepsilon)$ f is L-Lipschitz, and on X any subgradient g is of norm lower than L.

3. $f(x_k) \to v$ iff $\forall \varepsilon > 0, \exists N_\varepsilon \in \mathbb{N}, \quad k \geq K \implies f(x_k) \leq v + \varepsilon$. Hence $K_\varepsilon \subset [N_\varepsilon]$.

4. By subgradient inequality $f(y) \geq f(x^{(k)}) + \left\langle g^{(k)}, y - x^{(k)} \right\rangle$. Thus, for all k, $v \geq v^{(k)}$. Further, note that $v^{(k)} = f(x^{(k)})$, hence using $k = k_1$, and $y = x^{(k_2)}$ we get

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

5. As $k_2 \in K_\varepsilon$, we have $f(x^{(k_2)}) = v^{(k_2)} > v + \varepsilon \geq f(x^{(k_1)}) + \left\langle g^{(k_1)}, x^{(k_2)} - x^{(k_1)} \right\rangle + \varepsilon$ by the previous question.

6. We have

$$\varepsilon < |f(x^{(k_2)}) - f(x^{(k_1)})| + |\left\langle g^{(k_1)}, x^{(k_2)} - x^{(k_1)} \right\rangle| \leq 2L\|x^{(k_2)} - x^{(k_1)}$$

   by Cauchy-Schwartz and question 2.

7. If $f(x^{(k)}) \not\to v$, then there exists $\varepsilon > 0$ such that $(x^{(k)})_{k \in K_\varepsilon})$ is not finite. As X is compact we can exctract a converging subsequence, that is $x^{(\sigma(k)}$ such that $x^{(\sigma(k))} \to x^\star$ and $\sigma(k) \in K_\varepsilon$, which is in contradiction with the result of 6.