Speed of Convergence

Exercise 35 (Gradient descent)**.**
The exercises 35 and 36 are aimed at proving that the gradient algorithm for minimizing f, where f is convex and differentiable has convergence rate $O(1/k)$ in general (where k is the number of iterations) and $O((\frac{Q-1}{Q})^k)$ when f is strongly convex (where Q is called the *condition number*).
Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function whose gradient is L-Lipschitz continuous *i.e.* $\|\nabla f(y) - \nabla f(x)\| \leq L\|y - x\|$ for all x, y.

1. Prove that for all x, y, $\langle \nabla f(y) - \nabla f(x), y - x\rangle \leq L\|y - x\|^2$.
2. Set $\varphi(t) = f(x + t(y - x))$ for all $t \in [0, 1]$. Prove that
$$f(y) - f(x) - \langle \nabla f(x), y - x\rangle = \varphi(1) - \varphi(0) - \varphi'(0)\,.$$
3. Deduce that
$$f(y) - f(x) - \langle \nabla f(x), y - x\rangle = \int_0^1 \langle \nabla f(x + t(y - x)) - \nabla f(x), y - x\rangle dt$$
4. Using the first question, conclude that
$$f(y) \leq f(x) + \langle \nabla f(x), y - x\rangle + \frac{L}{2}\|y - x\|^2\,.$$

We consider the gradient algorithm *i.e.*, the sequence (x_k) defined by $x_{k+1} = x_k - \gamma\nabla f(x_k)$ where $\gamma > 0$ is a constant step size.

5. Show that
$$x_{k+1} = \arg\min_{y\in\mathbb{R}} \langle \nabla f(x_k), y - x_k\rangle + \frac{1}{2\gamma}\|x_k - y\|^2.$$
6. Prove that for all $z \in \mathbb{R}^n$,
$$\langle \nabla f(x_k), x_{k+1}-x_k\rangle+\frac{1}{2\gamma}\|x_k-x_{k+1}\|^2 = \langle \nabla f(x_k), z-x_k\rangle+\frac{1}{2\gamma}\|x_k-z\|^2-\frac{1}{2\gamma}\|x_{k+1}-z\|^2. \tag{7}$$
7. Deduce that $f(x_{k+1}) \leq f(x_k) - \dfrac{1}{\gamma}(1 - \dfrac{\gamma L}{2})\|x_{k+1} - x_k\|^2$.
8. Provide a condition on γ which ensures that when $x_{k+1} \neq x_k$, $f(x_{k+1}) < f(x_k)$.

From now on, we set $\gamma = \frac{1}{L}$.

9. Using (7), show that for all $z \in \mathbb{R}^n$,
$$f(x_{k+1}) \leq f(x_k) + \langle \nabla f(x_k), z - x_k\rangle + \frac{L}{2}\|x_k - z\|^2 - \frac{L}{2}\|x_{k+1} - z\|^2\,. \tag{8}$$

We assume from now on that f is convex and admits (at least) one minimizer $x^\star$.

10. Show that
$$f(x_{k+1}) \le f(x^\star) + \frac{L}{2}\|x_k - x^\star\|^2 - \frac{L}{2}\|x_{k+1} - x^\star\|^2 .$$

11. Deduce that for all $k \ge 1$,
$$\sum_{i=1}^{k} f(x_i) \le k f(x^\star) + \frac{L}{2}\|x_0 - x^\star\|^2 .$$

12. Show that
$$f(x_k) - f(x^\star) \le \frac{L\|x_0 - x^\star\|^2}{2k} .$$

Exercise 36 (Gradient descent strongly convex functions)**.**
We assume from now on that f is μ-strongly convex. Thus, for any x, y,
$$f(y) \ge f(x) + \langle \nabla f(x), y - x\rangle + \frac{\mu}{2}\|y - x\|^2 .$$

1. Using Eq. (8), prove that
$$f(x_{k+1}) \le f(x^\star) + \frac{L-\mu}{2}\|x_k - x^\star\|^2 - \frac{L}{2}\|x_{k+1} - x^\star\|^2 .$$

2. Define $\Delta_{k+1} = f(x_{k+1}) - f(x^\star) + \frac{L}{2}\|x_{k+1} - x^\star\|^2$. Show that
$$\Delta_{k+1} \le \left(1 - \frac{\mu}{L}\right)\Delta_k .$$

3. Conclude that
$$\begin{aligned} f(x_k) - f(x^\star) &\le \left(1 - \frac{\mu}{L}\right)^k \Delta_0 \\ \|x_k - x^\star\|^2 &\le \left(1 - \frac{\mu}{L}\right)^k \frac{2\Delta_0}{L} . \end{aligned}$$

4. The ratio $Q = L/\mu$ is called the *condition number* of f. Discuss the influence of Q on the convergence rate.

Exercise 37 (ADMM)**.**
We consider the ADMM

$$
\begin{aligned}
x_{k+1} &\in \arg\min_x f(x) + \langle \phi_k, Mx\rangle + \frac{\rho}{2}\|Mx - z_k\|^2 \\
z_{k+1} &= \arg\min_z g(z) - \langle \phi_k, z\rangle + \frac{\rho}{2}\|Mx_{k+1} - z\|^2 \\
\phi_{k+1} &= \phi_k + \rho(Mx_{k+1} - z_{k+1})
\end{aligned}
$$

for the resolution of the optimization problem

$$\min_x f(x) + g(Mx)$$

where $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ are convex functions, $M : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map and there exists a solution x^* to the problem.

1. Denote $\varphi : x \mapsto F(x) + \frac{\rho}{2}\|Mx - z_0\|^2$, where F is convex and let $p \in \arg\min \varphi$.
   Show that $(x \mapsto \varphi(x) - \frac{\rho}{2}\|Mx - Mp\|^2)$ is convex and that
   $$p \in \arg\min_x \varphi(x) - \frac{\rho}{2}\|Mx - Mp\|^2.$$

2. Deduce that for all x,
   $$F(p) + \frac{\rho}{2}\|Mp - z_0\|^2 \le F(x) + \frac{\rho}{2}\|Mx - z_0\|^2 - \frac{\rho}{2}\|Mx - Mp\|^2$$

3. Show that for all $x \in \mathbb{R}^n$, $z \in \mathbb{R}^m$, $\phi \in \mathbb{R}^m$,

$$f(x_{k+1}) + \langle \phi_k, Mx_{k+1}\rangle + \frac{\rho}{2}\|Mx_{k+1} - z_k\|^2 \le f(x) + \langle \phi_k, Mx\rangle + \frac{\rho}{2}\|Mx - z_k\|^2 - \frac{\rho}{2}\|Mx - Mx_{k+1}\|$$

$$g(z_{k+1}) - \langle \phi_k, z_{k+1}\rangle + \frac{\rho}{2}\|Mx_{k+1} - z_{k+1}\|^2 \le g(z) - \langle \phi_k, z\rangle + \frac{\rho}{2}\|Mx_{k+1} - z\|^2 - \frac{\rho}{2}\|z - z_{k+1}\|^2$$

$$-\langle \phi_{k+1}, Mx_{k+1} - z_{k+1}\rangle + \frac{1}{2\rho}\|\phi_{k+1} - \phi_k\|^2 = -\langle \phi, Mx_{k+1} - z_{k+1}\rangle + \frac{1}{2\rho}\|\phi - \phi_k\|^2 - \frac{1}{2\rho}\|\phi - \phi_{k+1}\|^2$$

4. Denote the Lagrangian function of the problem as
   $$L(x, z, \phi) = f(x) + g(z) + \langle \phi, Mx - z\rangle.$$
   Show that for all $x \in \mathbb{R}^n$, $z \in \mathbb{R}^m$, $\phi \in \mathbb{R}^m$,
   $$
   \begin{aligned}
   L(x_{k+1}, z_{k+1}, \lambda) - L(x, z, \phi_{k+1}) \le{}& \frac{1}{2\rho}\Big[\|\phi_k - \phi\|^2 - \|\phi_{k+1} - \phi\|^2 - \|\phi_k - \phi_{k+1}\|^2\Big] \\
   &+ \frac{\rho}{2}\Big[\|z_k - Mx\|^2 - \|z_{k+1} - Mx\|^2 - \|Mx_{k+1} - z_k\|^2 + \|Mx_{k+1} - z_{k+1}\|^2\Big]
   \end{aligned}
   $$

5. Denote $\bar{x}_k = \frac{1}{k}\sum_{i=1}^k x_i$, $\bar{z}_k = \frac{1}{k}\sum_{i=1}^k z_i$ and $\bar{\phi}_k = \frac{1}{k}\sum_{i=1}^k \phi_i$.
   Show that for all $\phi \in \mathbb{R}^m$ and for all solution x^*, we have
$$L(\bar{x}_k, \bar{z}_k, \phi) - L(x, z, \bar{\phi}_k) \leq \frac{\rho}{2k}\|z_0 - Mx^*\|^2 + \frac{1}{2\rho k}\|\phi - \phi_0\|^2$$

6. For $\beta > 0$, denote
$$S_\beta(\bar{x}, \bar{z}, \phi_0) = \sup_\phi f(\bar{x}) + g(\bar{z}) + \langle \phi, M\bar{x} - \bar{z}\rangle - f(x^*) - g(Mx^*) - \frac{\beta}{2}\|\phi_0 - \phi\|^2$$
   Show that
$$S_\beta(\bar{x}, \bar{z}, \phi_0) = f(\bar{x}) + g(\bar{z}) - f(x^*) - g(Mx^*) + \langle \phi_0, M\bar{x} - \bar{z}\rangle + \frac{1}{2\beta}\|M\bar{x} - \bar{z}\|^2$$
   and that
$$S_{\frac{1}{\rho k}}(\bar{x}_k, \bar{z}_k, \phi_0) \leq \frac{\rho}{2k}\|z_0 - Mx^*\|^2.$$

7. Show that for all $\bar{x}, \bar{z}, \phi_0, \beta$, and for a saddle point (x^*, ϕ^*),
   (a) $f(\bar{x}) + g(\bar{z}) - f(x^*) - g(Mx^*) \geq -\langle \phi^*, M\bar{x} - \bar{z}\rangle \geq -\|\phi^*\|.\|M\bar{x} - \bar{z}\|$
   (b) $f(\bar{x}) + g(\bar{z}) - f(x^*) - g(Mx^*) = S_\beta(\bar{x}, \bar{z}, \phi_0) - \langle \phi_0, M\bar{x} - \bar{z}\rangle - \frac{1}{2\beta}\|M\bar{x} - \bar{z}\|^2$
   $\leq S_\beta(\bar{x}, \bar{z}, \phi_0) + \|\phi_0\|.\|M\bar{x} - \bar{z}\|$
   (c) $\|M\bar{x} - \bar{z}\| \leq \beta\Big(\|\phi_0 - \phi^*\| + \big(\|\phi_0 - \phi^*\|^2 + \frac{2}{\beta}S_\beta(\bar{x}, \bar{z}, \phi_0)\big)^{1/2}\Big)$

8. Show that the ergodic sequence $(\bar{x}_k, \bar{z}_k, \bar{\phi}_k)$ has a convergence speed as $O(1/k)$ in function value $f(\bar{x}_k) + g(\bar{z}_k)$ and feasibility.

9. Assume that g is Lipschitz continuous. Show that $f(\bar{x}_k) + g(M\bar{x}_k)$ converges to $f(x^*) + g(Mx^*)$ as $O(1/k)$.

In this exercise, we showed the convergence speed for the ergodic mean of the iterates. We shall refer to “Damek Davis and Wotao Yin, (2016), Convergence rate analysis of several splitting schemes, in Splitting Methods in Communication, Imaging, Science, and Engineering (pp. 115-163), Springer International Publishing” for the non-ergodic speed of the ADMM.