Exercises 4.17

Exercise 4.17.1

Let $f(x_1, x_2) = e^{x_1}\sin(x_2)$, with $(x_1, x_2) \in (0,1) \times (0, \frac{\pi}{2})$.

a. Show that f is a harmonic function;

b. Find $\|\nabla f\|$;

c. Show that the equation $\nabla f = 0$ does not have any solutions;

d. Find the maxima and minima for the function f.

Exercise 4.17.2

Consider the quadratic function $Q(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T A\mathbf{x} - b\mathbf{x}$, with A nonsingular square matrix of order n.

a. Find the gradient $\|\nabla Q\|$;

b. Write the gradient descent iteration;

c. Find the Hessian H_Q;

d. Write the iteration by Newton's formula and compute its limit.

Exercise 4.17.3

Let A be a nonsingular square matrix of order n and $b \in \mathbb{R}^n$ a given vector. Consider the linear system $A\mathbf{x} = b$. The solution can be approximated using the following steps:

a. Associate the cost function $C(\mathbf{x}) = \frac{1}{2}\|A\mathbf{x} - \mathbf{b}\|^2$. Find its gradient, $\nabla C(\mathbf{x})$, and Hessian $H_C(\mathbf{x})$;

b. Write the gradient descent algorithm iteration which converges to the system solution $\mathbf{x}$ with the inital value $\mathbf{x}^0 = 0$;

c. Write Newton's iteration which converges to the system solution $\mathbf{x}$ with the initial value $\mathbf{x}^0 = 0$.

Exercise 4.17.4

a. Let $(a_n)_n$ be a sequence with $a_0 > 0$ satisfying the inequality
$a_{n+1} \leq \mu a_n + K, \ \forall n \geq 1$, with $\mu \in (0,1)$ and $K > 0$. Show that the sequence $(a_n)_n$ is bounded from above.

b. Consider the momentum method equations (4.4.16) – (4.4.17), and asume that the function f has a bounded gradient $\|\nabla f\| \leq M$. Show that the sequence of velocities, $(v^n)_n$ is bounded.

Exercise 4.17.5

a. Let f and g two integrable functions. Verify that

$$\int (f \star g)(x)dx = \int f(x)dx \int g(x)dx$$

b. Show that $\|f \star g\| \leq \|f\|_1 \|g\|_1$

c. Let $f_\sigma := f \star G_\sigma$ where $G_\sigma = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}}$. Prove that $\|f_\sigma\|_1 \leq \|f\|_1$ for $\sigma > 0$

Exercise 4.17.6

Show that the convolution of two Gaussians is also a Gaussian:

$$G_{\sigma_1} \star G_{\sigma_2} = G_\sigma, \text{ where } \sigma = \sqrt{\sigma_1^2 + \sigma_2^2}.$$

Exercise 4.17.7

Show that the if n have the sum equal to s,

$$\sigma_1 + \ldots + \sigma_n = s,$$

then the numbers for which the sum of their squares, $\sum_{j=1}^{n} \sigma_j^2$, its minimum occurs for the case when all the numbers are equal to $\frac{s}{n}$.

SOLUTIONS

Exercise 4.17.1 (a)

By definition a function is harmonic when satisfies the condition $\nabla^2 f = 0$. Let's corroborate this is indeed fullfilied by the function $f(x_1, x_2) = e^{x_1}\sin(x_2)$. For $\frac{\partial^2}{\partial x_1^2} f$ and $\frac{\partial^2}{\partial x_2^2} f$ we have:

$$\frac{\partial^2}{\partial x_1^2} e^{x_1}\sin(x_2) = \sin(x_2)\frac{\partial^2}{\partial x_1^2} e^{x_1} = e^{x_1}\sin(x_2)$$
$$\frac{\partial^2}{\partial x_2^2} e^{x_1}\sin(x_2) = e^{x_1}\frac{\partial^2}{\partial x_2^2}\sin(x_2) = -e^{x_1}\sin(x_2)$$

From the latter follows $\nabla^2 f = \dfrac{\partial^2}{\partial x_1^2} f + \dfrac{\partial^2}{\partial x_2^2} f = 0$ i.e the function f is harmonic. $\square$

Exercise 4.17.1 (b)

By the pythagorean identity between the trigonometric functions sin and cos follows:

$$\|\nabla f\| = \sqrt{\nabla f \cdot \nabla f} = \sqrt{(e^{x_1}\cos(x_2))^2 + (e^{x_1}\sin x_2)^2} = e^{x_1}$$

Exercise 4.17.1 (c)

$$\nabla f = (e^{x_1}\cos(x_2), e^{x_1}\sin(x_2)) = 0 \iff e^{x_1} = 0$$

The equation $e^x = 0$ is known to not have a solution. Therefore, $\nabla f = 0$ is not solvable. $\square$

Exercise 4.17.1 (d)

Let's define the extension of f over the compact $K := [0,1] \times [0, \frac{\pi}{2}]$ with the same association rule as above. On this set f is also harmonic. Then, f reaches its minimun and maximum on the boundaries of the set K; note that both functions e^{x_1} and $\sin(x_2)$ are increasing, then the maxima is met at the point when both functions reach their maximum.

This means the maximum of the function f is met on the point $(1, \frac{\pi}{2})$ with a value of $f(1, \pi) = e$. Likewise, the minimum is reached at the point $(0,0)$ with a value of $f(0,0) = 0$. $\square$

Exercise 4.17.2 (a)

Computing the gradient we get: $\nabla Q = \frac{1}{2}(A + A^T)\mathbf{x} - b\mathbf{1}$. Then, expressing the norm in terms of the interior product we have:
$\|\nabla Q(\mathbf{x})\| = (\dfrac{1}{4}\mathbf{x}^T(A + A^T)^2\mathbf{x} - b\mathbf{1}(A + A^T)\mathbf{x} + b^2\mathbf{1}^T\mathbf{1})^{\frac{1}{2}}$

Exercise 4.17.2 (b)

The equations that describe the iterations made in the GDA[1] is the sequence of vectors $\{\mathbf{x}^n\}_{n\in\mathbb{N}}$ satisfying the following recursion: $\mathbf{x}^{n+1} = \mathbf{x}^n - \delta(\dfrac{1}{2}(A + A^T)\mathbf{x}^n - b\mathbf{1})$

[1] From now on, the learning rate δ in the GDA an its variants will be supposed constant, unless is stated otherwise.

Exercise 4.17.2 (c)

From excercise 4.17.2.a we know that $\nabla Q = \frac{1}{2}(A + A^T)\mathbf{x} + b\mathbf{1}$. Note that taking the derivative of a vectorial function is indeed the same as computing its Hessian-matrix. By using the afore mentioned observation, we get:

$H_Q(\mathbf{x}) = \frac{\partial}{\partial \mathbf{x}} \nabla Q = \frac{1}{2}(A + A^T)$.

Exercise 4.17.2 (d)

The sequence of iterations produced by Newton's method $\{\mathbf{x}^n\}_{n\in\mathbb{N}}$ are given with following recurrence relationship:

$$\begin{aligned} \mathbf{x}^{n+1} &= \mathbf{x}^n - (\frac{1}{2}(A + A^T))^{-1}(\frac{1}{2}(A + A^T)\mathbf{x}^n - b\mathbf{1}) \\ &= (\frac{1}{2}(A + A^T))^{-1}b\mathbf{1}. \end{aligned}$$

We note that the sequence is a constant. This in turn implies the limit is trivialy given by the expression: $\mathbf{x}^\star = (\frac{1}{2}(A + A^T))^{-1}b\mathbf{1}$.

Exercise 4.17.3 (a)

Expressing C as a quadratic form and simplifying we get: $C(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T A^T A\mathbf{x} - \mathbf{b}^T A\mathbf{x} + \frac{1}{2}\mathbf{b}^T\mathbf{b}$. On the other hand, the fact that the matrix $A^T A$ is symmetric, yields:

- $\nabla C(\mathbf{x})$:
$$\begin{aligned} \frac{\partial}{\partial \mathbf{x}} C(\mathbf{x}) &= \frac{\partial}{\partial \mathbf{x}} \left[\frac{1}{2}\mathbf{x}^T A^T A\mathbf{x} - \mathbf{b}^T A\mathbf{x} + \frac{1}{2}\mathbf{b}^T\mathbf{b}\right] \\ &= A^T A\mathbf{x} - \mathbf{b}^T A \end{aligned}$$

- $H_C(\mathbf{x})$:
$$\begin{aligned} \frac{\partial^2}{\partial^2 \mathbf{x}} C(\mathbf{x}) &= \frac{\partial}{\partial \mathbf{x}} \left[A^T A\mathbf{x} - \mathbf{b}^T A\right] \\ &= A^T A \end{aligned}$$

Exercise 4.17.3 (b)

Using exercise 4.17.3.a is clear that the iteration in the GDA method is: $\mathbf{x}^{n+1} = (Id - \delta A^T A)\mathbf{x}^n + \delta\mathbf{b}^T A$. Iterating the latter equations gives:

$$\begin{aligned} \mathbf{x}^{n+1} &= (Id - \delta A^T A)\mathbf{x}^n + \delta\mathbf{b}^T A \\ &\vdots \\ \mathbf{x}^{n+1} &= (Id - \delta A^T A)^n\mathbf{x}^0 + \delta\sum_{k=0}^{n-1}(Id - \delta A^T A)^n\mathbf{b}^T A \end{aligned}$$

Note that the eigenvalues of are all positive, and furthermore in absolute value less than one (The proof of this can be found in the Apendix G). Implying the sequence $\sum_{k=0}^{n-1}(Id - \delta A^T A)^n$ converges, which in turn implies $(Id - \delta A^T A)^n \to \mathbb{O}$ and that the limit $\mathbf{x}^\star$ is well definied and has the value: $\mathbf{x}^\star = (A^T A)^{-1}\mathbf{b}^T A$.

Exercise 4.17.3 (c)

The sequence of Newton's method iterations is given by:

$$\begin{aligned}\mathbf{x}^{n+1} &= \mathbf{x}^n - (A^T A)^{-1}(A^T A\mathbf{x}^n - \mathbf{b}^T A) \\ &= \mathbf{x}^n - \mathbf{x}^n + (A^T A)^{-1}\mathbf{b}^T A \implies \mathbf{x}^\star = \lim_{n\to\infty} \mathbf{x}^n = (A^T A)^{-1}\mathbf{b}^T A.\end{aligned}$$

Exercise 4.17.4 (a)

The sequence $\{a_n\}_{n\in\mathbb{N}}$ satisfies the inequality $a_{n+1} \leq \mu a_n + K$, $\forall n \geq 1$. Then, iterating this inequality we have:

$$\begin{aligned}a_{n+1} &\leq \mu a_n + K \\ &\leq \mu^2 a_{n-1} + (1+\mu)K \leq \ldots \leq \mu^n a_0 + K\sum_{j=0}^{n-1}\mu^j\end{aligned}$$

By hypothesis $a_0 > 0$, $\mu \in (0,1)$ and $K > 0$. Implying the sequence of partials sums $\{\sum_{j=0}^{n-1}\mu^j\}_{\{n\,:\,n\geq 1\}}$ is convergent. Thus the sequence $\{a_n\}_{n\in\mathbb{N}}$ is bounded by above. $\square$

Exercise 4.17.4 (b)

If the function f satisfies $\|\nabla f\| \leq M$. Then, the set of equations describying the Momentum GDA (4.4.16-4.4.17) in norm satisfy the following inequalities:

$$\begin{cases}\|\mathbf{x}^{n+1}\| \leq \|\mathbf{x}^n\| + \|\mathbf{v}^{n+1}\| \\ \\ \|\mathbf{v}^{n+1}\| \leq \mu\|\mathbf{v}^n\| + \eta M\end{cases}$$

An application of the above exercise with $a_n = \|\mathbf{v}^n\|$, $K = \eta M$ yields the desired result. $\square$

Exercise 4.17.5 (a)

Without loss of generality let's assume that the convolution product $f \star g$ is being integrated over the whole real line. Then, by the definition of the convolution product we get:

$$\begin{aligned}\int_{\mathbb{R}}(f\star g)(x)dx &= \int_{\mathbb{R}}\int_{\mathbb{R}} f(u)g(u-x)dudx = \int_{\mathbb{R}}\int_{\mathbb{R}} f(v)g(t)dvdt \quad \text{(applying the substitution: } u = v, u - x = t) \\ &= \int_{\mathbb{R}} f(v)dv\int_{\mathbb{R}} g(t)dt\end{aligned}$$

Where the last equality follows from Fubbini's theorem. $\square$

Exercise 4.17.5 (b)

An application of the integral triangle inequality to the above identity, yields:

$$\left|\int_{\mathbb{R}}(f\star g)(x)dx\right| \leq \int_{\mathbb{R}}|(f\star g)(x)|\,dx = \|(f\star g)(x)\|_1 = \int_{\mathbb{R}}\int_{\mathbb{R}}|f(x)|\,|g(u-x)|\,dudx$$
$$\leq \int_{\mathbb{R}}|f(x)|\,dx\int_{\mathbb{R}}|g(x)|\,dx = \|f(x)\|_1\|g(x)\|_1$$

□

Exercise 4.17.5 (c)

Noting that G_σ is a density, then $\|G_\sigma(x)\|_1 = 1$. Using the inequality proven in excercise 4.17.5.b we get:

$$\|f_\sigma(x)\|_1 \leq \|f(x)\|_1\|G_\sigma(x)\|_1 \leq \|f(x)\|_1$$

□

Exercise 4.17.6

This is in esence what was proven in Exercise 3.15.10.b

Exercise 4.17.7

Let $\boldsymbol{\sigma} := (\sigma_1, \ldots, \sigma_n)$. One has to optimize the function $f(\boldsymbol{\sigma}) = \boldsymbol{\sigma}^T\boldsymbol{\sigma}$ constrained to $\sigma_1 + \ldots + \sigma_n = s$. This is solved by Lagrange's multipliers method. The Lagrangian function to optimize is then:
$\mathcal{L}(\boldsymbol{\sigma}, \lambda) = \boldsymbol{\sigma}^T\boldsymbol{\sigma} - \lambda(\displaystyle\sum_{j=0}^{n}\sigma_j - s)$.

Note the constraining condition can be abbrebiated to $\|\boldsymbol{\sigma}\|_1 = s$; computing the Lagrangian gradient and equating it to zero, we get:

$$\nabla_{\sigma,\lambda}\mathcal{L} = 0 \Longleftrightarrow \begin{cases} \sigma = -\lambda\cdot\mathbf{1} \\ \|\boldsymbol{\sigma}\|_1 - s = 0. \end{cases}$$

From the first and second equations follows that $\sigma_k = \frac{s}{n}$, $\forall k \in \{1, \ldots, n\}$. On the other hand, the Bordered Hessian has the form:

$$H_{\mathcal{L}}(\boldsymbol{\sigma}, \lambda) = \begin{bmatrix} 0 & \boldsymbol{\sigma} \\ \boldsymbol{\sigma}^T & Id \end{bmatrix}$$

Which is obvously positively definite. Thus, the application of Lagrange's multipliers method gives a contrained minimum.

□