1 Exercises on unconstrained optimization with equalities

Exercise 6. We may see this by finding a contraexample. Consider the one-dimensional real convex function $f(x) = x^2 - x$, with $f'(x) = 2x - 1$, and which clearly has a unique minimum at $x^* = \frac{1}{2}$. Recall the Wolfe conditions:

$$\begin{cases} \nabla f(x_k + \alpha_k p_k)^T p_k \geq c_2 (\nabla f(x_k))^T p_k \\ f(x_k + \alpha_k p_k) < f(x_k) + c_1 \alpha_k (\nabla f(x_k))^T p_k \end{cases}$$

Note that for the one dimensional case we consider $p_k = 1$ and therefore we may allow α_k to be negative, thus, the Wolfe conditions for this function are given by the following expressions:

$$\begin{cases} f'(x_k + \alpha_k) \geq c_2 f'(x_k) \\ f(x_k + \alpha_k) < f(x_k) + c_1 \alpha_k f'(x_k) \end{cases}$$

Consider therefore the line-search method with step α_k, and for instance, initial point $x_0 = 0$, thus note that the Wolfe conditions bring us to the following conditions:

$$2\alpha_0 - 1 \geq -c_2 \iff \alpha_0 \geq \frac{1 - c_2}{2}$$
$$\alpha_0^2 - \alpha_0 \leq -c_1 \alpha_0 \iff \alpha_0 \leq 1 - c_1$$

Now consider for example $c_1 = \frac{3}{4} > \frac{1}{4} = c_2$, thus we have

$$\alpha_0 \leq \frac{1}{4} \text{ and } \alpha_0 \geq \frac{3}{8},$$

contradicting the Wolfe conditions. Hence, we see that we need $0 < c_1 < c_2 < 1$ to be sure that there exists a step size α_k that satisfies the Wolfe conditions.

Exercise 11. Consider the function $f(x, y) = (x + y^2)^2$ at $x_0 = (1, 0)$. Note that $p_k^T = (-1, 1)$ is a descent direction, since $\nabla f(x, y) = (2(x + y^2), 2(x + y^2)2y)$, and then:

$$p_k^T \nabla f(x_0) = (-1, 1) \begin{pmatrix} 2 \\ 0 \end{pmatrix} = -2 < 0$$

Consider now the problem of finding the minimizers of

$$min_{\alpha \in \mathbb{R}^+} f(\mathbf{x} + \alpha \mathbf{p})$$

where we consider $\mathbf{x} = \mathbf{x_0} = \mathbf{(1, 0)^T}$ and its descent direction $\mathbf{p^T} = (\mathbf{-1, 1})$. Hence the corresponding condition $\nabla g(\alpha^*) = 0$ where $g(\alpha) = f(\mathbf{x} + \alpha \mathbf{p})$ corresponds to the directional derivative throughout the descent direction: $\nabla f(\mathbf{x} + \alpha \mathbf{p})\mathbf{p^T} = \mathbf{0} = \frac{\partial \mathbf{f}}{\partial \alpha}(\mathbf{1} - \alpha, \alpha)$, this is:

$$0 = \frac{\partial f}{\partial \alpha}(1 - \alpha, \alpha) = \frac{\partial}{\partial \alpha}(1 - \alpha + \alpha^2)^2 = 2(1 - \alpha + \alpha^2)(2\alpha - 1)00 \iff \begin{cases} \alpha = \frac{1}{2} \\ \alpha = \frac{1 \pm \sqrt{1-4}}{2} \notin \mathbb{R} \end{cases}$$

Thus, we have $\hat{\alpha} = \frac{1}{2}$, which clearly is a minimum because $\frac{\partial^2 f}{\partial \alpha^2}(1 - \alpha, \alpha) = 4\alpha(3\alpha - 1)$, and therefore $\frac{\partial^2 f}{\partial \alpha^2}(1 - \hat{\alpha}, \hat{\alpha}) = \frac{1}{2} > 0$. Note that the computation of the corresponding directional derivative is equivalent to the calculating $\nabla f(\mathbf{x} + \alpha \mathbf{p})\mathbf{p^T} = \mathbf{0}$.

Exercise 16. Consider the real function $f(x) = x^2 + e^x - 3$. Note that clearly it is a continuous differentiable function, since it is a sum of continuous differentiable functions. It is then directly to see that it is a $\mathcal{C}^2$ functions, with the corresponding derivatives:

$$\frac{df}{dx}(x) = f'(x) = 2x + e^x$$
$$\frac{d^2 f}{dx^2}(x) = f''(x) = 2 + e^x$$

Note that $f''(x) > 0$, $\forall x \in \mathbb{R}$, then $f''(x)$ does not have any zero. Hence, the continuous function $f'(x)$ has no change in its monotony, this is, it either increases or decreases during its whole domain. Thus, if it has any zero it just has one, it may just cross the x-axis once. Notice now that $f'(x) > 0$, $\forall x \geq 0$, for instance $f'(0) = e^0 = 1 > 0$. Meanwhile, we have $f'(-1/2) = -1 + e^{\frac{1}{2}} < 0$. Therefore, we have $f'(0) \cdot f'(-1/2) < 0$ and by applying the Bolzano theorem we have that there exists a zero of the function $f'(x)$ in $(-1/2, 0)$. Since we have argued that it has only one zero, its unique root is $y_1 \in (-1/2, 0)$. This means that we are certain that $f'(x) \neq 0$ outside this interval.
Considering a similar approach, observe that $f(-1/2) = \frac{1}{4} + e^{-\frac{1}{2}} - 3 \approx -2.14 < 0$, while $f(-2) = 4 + e^{-2} - 3 > 0$. Then applying again Bolzano's theorem we have that there exists a zero $x_1 \in (-2, -1/2)$ of the function $f(x)$. Recalling the fact that $f'(x) \neq 0$ in this interval, we have as a result of the Rolle's theorem that this is the unique zero in the interval.
Similarly, notice that $f(0) = 1 - 3 < 0$, and $f(1) = 1 + e - 3 > 0$, and we can argue as aforementioned to see that there is a unique zero in this interval $x_2 \in (0, 1)$.

We have found two zeros of the function f, x_1, x_2. However, we know that $f'(x)$ has only one zero, and consequently the function f has just one change of monotony, and thus in order for the function to cross again the x- axis it may need to have an additional change of monotony, contradicting the fact that f' has a unique zero. Hence, the function f has only two solutions, located at the intervals $x_1 \in (-2, -1/2)$ and $x_2 \in (0, 1)$.

Consider the corresponding Newton method, this is the algorithmic approach given by the succession

$$x_{k+1} = x_k - \alpha_k \frac{f(x_k)}{f'(x_k)}$$

with $\alpha_k = 0.1$ fixed, and a threshold fixed of 10^{-3} of difference between iterations as a stop criterion. Thus, considering for each case the middle point of the interval considered the Newton method is implemented, obtaining the results shown in the Table 1.

Initial point	Solution	Iterations
-1.25	-1.67723537	4
0.5	0.83449305	4

Table 1: Computation of zeros of f using the Newton method.

Recall now the theoretical result studied in class, that states that the Newton method as described above convergences quadratically when considering the starting point close enough to the initial point. This can be seen directly by noting that all the zeros found are simple, since the unique point where the derivative vanishes is not in the intervals where the zeros are allocated.
Considering the iterates obtained by fixing the initials points as described above, $x_1^0 = -1.25, x_2^0 = 0.5$, we can also prove the quadratically convergence of this method experimentally. By applying the Newton method as aforementioned we obtain the results shown in 2 and 3.

Iterates	ε_n	$\frac{\lvert\varepsilon_{n-1}\rvert}{\lvert\varepsilon_n\rvert}$	$\frac{\lvert\varepsilon_{n-1}\rvert}{\lvert\varepsilon_n\rvert^2}$
-1.250000	0.427233	0.217112	0.508181
-1.769990	0.092757	0.029946	0.322842
-1.680010	0.002778	0.000957	0.344484
-1.677235	0.000003	0.000000	0.000000

Table 2: Computation of the errors, $\varepsilon_n = |x_n - x^*|$, considering the exact analytical solution of $\hat{x}_1 = -1.67723271$ considering the starting point of $x_1^0 = -1.25$.

Iterates	ε_n	$\frac{\lvert\varepsilon_{n-1}\rvert}{\lvert\varepsilon_n\rvert}$	$\frac{\lvert\varepsilon_{n-1}\rvert}{\lvert\varepsilon_n\rvert^2}$
0.500000	0.334487	0.243031	0.726578
0.915778	0.081291	0.041604	0.511788
0.837869	0.003382	0.001827	0.540337
0.834493	0.000006	0.000000	0.000000

Table 3: Computation of the errors, $\varepsilon_n = |x_n - x^*|$, considering the exact analytical solution of $\hat{x}_1 = 0.83448687$ considering the starting point of $x_1^0 = 0.5$.

Note that we clearly get a quadratic convergence of the method in this case, since the error for each iteration decreases quadratically in comparison to the previous one.

Note finally, that in case of not having the exact analytic solution (which may be the case for most of the practical cases) the study on the convergence can be conducted using the relative errors $\varepsilon_n = |x_n - x_{n-1}|$ instead, and in this case, the results are the same as the obtained using the exact solution.