NUMERICAL OPTIMIZATION

Sheet 5: Gradient descent

~~EXERCISE ONE~~ At the lecture, we have had the following chain of arguments. First, we have concluded that for ~~gradient descent~~, assuming the exact line search, the exact formula for the step x_{k+1} for a convex quadratic problem is

$$x_{k+1} = x_k - \left(\frac{\nabla f_k^T \nabla f_k}{\nabla f_k^T Q \nabla f_k} \right) \nabla f_k. \tag{1}$$

Then, we introduced the error norm $\|x\|_Q^2 = x^T Q x$ and claimed that it captures the difference between the current function value and the minimum:

$$\frac{1}{2} \|x - x^*\|_Q^2 = f(x) - f(x^*). \tag{2}$$

With those, we can finally define the exercise: using the steps above to show an exact formula for the error norm difference:

$$\|x_{k+1} - x^*\|_Q^2 = \left\{ 1 - \frac{\left(\nabla f_k^T \nabla f_k\right)^2}{\left(\nabla f_k^T Q \nabla f_k\right)\left(\nabla f_k^T Q^{-1} \nabla f_k\right)} \right\} \|x_k - x^*\|_Q^2. \tag{3}$$

First, use (1) to show that

$$\|x_k - x^*\|_Q^2 - \|x_{k+1} - x^*\|_Q^2 = 2\alpha_k \nabla f_k^T Q (x_k - x^*) - \alpha_k^2 \nabla f_k^T Q \nabla f_k,$$

Second, use the fact that $\nabla f_k = Q(x_k - x^*)$ to obtain

$$\|x_k - x^*\|_Q^2 - \|x_{k+1} - x^*\|_Q^2 = \frac{2\left(\nabla f_k^T \nabla f_k\right)^2}{\left(\nabla f_k^T Q \nabla f_k\right)} - \frac{\left(\nabla f_k^T \nabla f_k\right)^2}{\left(\nabla f_k^T Q \nabla f_k\right)}$$

and

$$\|x_k - x^*\|_Q^2 = \nabla f_k^T Q^{-1} \nabla f_k.$$

EXERCISE TWO From the lecture: a function $f : \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz-continuous if there exists a constant L such that

$$\forall x, y : \|f(x) - f(y)\| \leq L \cdot \|x - y\|.$$

Which of the following functions are Lipschitz-continuous and which have a Lipschitz-continuous gradient?

1. $f(x) = x \log x$,
2. $f(x) = x^3$,
3. $f(\vec{x}) = \|x\|_1$,
4. $f(\vec{x}) = \frac{1}{2} \cdot \|Ax - b\|_2^2$,
5. $f(\vec{x}) = \log \sum_{i=1}^n e^{x_i}$.

EXERCISE THREE Let us run through a very illustrative example for gradient descent. Suppose we are in $\mathbb{R}^2$ and we are given some constant γ. We minimize the quadratic function $f(x) = \frac{1}{2}(x_1^2 + \gamma x_2^2)$ and start at the point $x_0 = (\gamma, 1)$.

1. What is the local minimum of this function?
2. What are the eigenvalues of its gradient?
3. Suppose we use gradient descent with exact line search – so after computing the gradient, we find the minimum value on this line. What is the coordinate of x_1?
4. Do one more step to compute x_2, and create a general formula for x_k (it is quite nice).

EXERCISE FOUR The *condition number* λ_n/λ_1, the largest eigenvalue of the Hessian divided by the smallest one, has a huge impact on the running time of gradient descent. There is a second way of defining the condition number, and that is as follows:

Consider a positive semidefinite matrix A. Let M be the smallest non-negative number such that $MI - A$ is positive semidefinite, and let m be the largest non-negative number such that $A - mI$ is positive semidefinite. Then, M/m, if it is well-defined, is the condition number of A.

Your tasks are:

1. First, show that if A is a matrix with $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ its eigenvalues, then if we add c times the identity matrix, getting $A + cI$, then the eigenvalues of $A + cI$ are $\lambda_1 + c, \lambda_2 + c, \ldots, \lambda_n + c$. This sounds daunting, but it is not. Start from the fact that the set of eigenvalues A are the roots of the polynomial $\det(A - xI)$, where $x = (x_1, \ldots, x_n)$ is a vector of variables.
2. Then, conclude that $M \approx \lambda_n$ and $m \approx \lambda_1$. In this claim we write $\approx$ instead of $=$ to allow you to have some small ε extra term in your arguments.

EXERCISE FIVE Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function. Assume that y is a local minimum of f along every line that passes trough y, that is for each $h \in \mathbb{R}^n$

$$g(s) = f(y + sh)$$

has local minimum at $s = 0$. Show that $f'(y) = 0$. Check that $f(x_1, x_2) = (x_1 - x_2^2)(x_1 - 2x_2^2)$ satisfies the condition above with $y = (0, 0)$, but $(0, 0)$ is not a local minimum of f.

EXERCISE SIX Consider the problem of minimizing the quadratic function $x^T Ax + b^T x + c$. Show that if A is not positive definite, then the function is unbounded from below. Show that if optimality condition $2Ax = b$ is not solvable, then the function is unbounded from below.