Università della Svizzera italiana | Institute of Computing CI

Optimization Methods **Spring 2024**

Student: Birindelli Leonardo **Discussed with:** Morales Jeferson

Midterm Project

Due date: Sunday, 12 May 2024, 11:59 PM

Exercise 1

Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by $f(x) = \frac{1}{2}x^T Ax - b^T x$ with A symmetric positive definite. Let x_m be the minimizer of the function f. Le v be an eigenvector of A, and let λ be the associated eigenvalue. Suppose that we use Steepest Descent (SD) method to minimize f and the starting point for the SD algorithm is $x_0 = x_m + v$.

1. Prove that the gradient at x_0 is $\nabla f(x_0) = \lambda v$.

   To begin with the proof, it is required to remark the gradient formula of the quadratic form function which is the following:

   $$\nabla f(x) = Ax - b$$

   Then, considering the starting point $x_0 = x_m + v$ it is possible to write the gradient at that position as :

   $$\begin{aligned} \nabla f(x_0) &= Ax_0 - b && (1) \\ &= A(x_m + v) - b && (2) \\ &= Ax_m + Av - b && (3) \end{aligned}$$

   We know from the hypothesys that x_m is the minimizer of the function f, thus, it means that its gradient must be equal to zero according to the first necessary condition for a minimizer:

   $$\begin{aligned} \nabla f(x_m) &= 0 \\ Ax_m - b &= 0 \\ Ax_m &= b \end{aligned}$$

   Substituting Ax_m inside the expression (3), we obtain that :

   $$\nabla f(x_0) = b - Av - b = Av$$

   In addiction, we know beforehand the definition of eigenvalues and eigenvectors which affirms that :

$$Av = \lambda v \tag{4}$$

Hence it is possible to write that :

$$\nabla f(x_0) = \lambda v$$

Consequently, we have proved that $\nabla f(x_0) = \lambda v$.

2. How many iterations does the SD method take to minimize the function f if we use the optimal step length? Show the computations behind your reasoning.

   First of all, we know that the Steepest Descent method uses the negative value of the gradient of function at the given point as the step direction to reach its minimizer. Therefore, that direction, starting from x_0, is equal to the following value as we proved previously:

$$p = -\nabla f(x_0) = -\lambda v$$

   Now, we can show that the subsequent iteration x_1 is computed as following:

$$\begin{align} x_1 &= x_0 - \alpha \lambda v \tag{5} \\ &= x_m + v - \alpha \lambda v \qquad \text{where } x_0 = x_m + v \tag{6} \end{align}$$

   Secondly, we want to know the number of iterations of the gradient method in the case in which it is used the optimal step lenght, thus, we can write that:

$$\alpha_{opt} = \frac{\nabla f(x_i)^T \nabla f(x_i)}{\nabla f(x_i)^T A \nabla f(x_i)} \tag{7}$$

   Hence, the optimal step size of the starting point x_0 is:

$$\begin{align} \alpha_{opt} &= \frac{\nabla f(x_0)^T \nabla f(x_0)}{\nabla f(x_0)^T A \nabla f(x_0)} \tag{8} \\ &= \frac{\cancel{\lambda^2} v^T v}{\cancel{\lambda^2} v^T A v} \tag{9} \\ &= \frac{v^T v}{v^T \lambda v} \qquad Av = \lambda v \ (4) \tag{10} \\ &= \frac{1}{\lambda} \frac{\cancel{v^T v}}{\cancel{v^T v}} \tag{11} \\ &= \frac{1}{\lambda} \tag{12} \end{align}$$

   Knowing that l'optimal step size is equal to $\frac{1}{\lambda}$ we can semplify the expression 6, thus we achieve that:

$$\begin{align} x_1 &= x_m + v - \frac{1}{\cancel{\lambda}} \cancel{\lambda} v \tag{13} \\ x_1 &= x_m + v - v \tag{14} \\ x_1 &= x_m \tag{15} \end{align}$$

Then, knowing that x_m is the function minimizer as initial hypothesis, its gradient is equal to zero according to the first necessary condition for minimizer point and given this information we can write the next equation:

$$\nabla f(x_m) = \nabla f(x_1) = 0 \tag{16}$$

In conclusion, it is possible to notice that the Steepest Descent method converges in only 1 iteration.
In particular, the single iteration that brings the iterative method to converge, directly depends on the fact that we are using the optimal step lenght in the process, $\alpha_{opt} = \frac{1}{\lambda}$. This information lets the steepest method to find the best step size in order to get closer towards the minimizer of the function. On the other hand, if $\alpha \neq \frac{1}{\lambda}$, the method could have required at least 1 iteration concluding that:

$$x_1 \neq x_m \wedge \nabla f(x_m) \neq \nabla f(x_1) \Longrightarrow f(x_1) > f(x_m) \wedge \nabla f(x_1) > \nabla f(x_m) \tag{17}$$