Math 164 HW3

Shilpa Bojjireddy

October 25, 2024

---

1. Perform two iterations leading to the minimization of

$$f(x) = x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_1^2 + x_2^2 + 3 \quad (1)$$

using the steepest descent method with the starting point $x^{(0)} = 0$. Also determine an optimal solution analytically.

If we solve for a minimum analytically, we get:

$$\nabla f(x) = \begin{bmatrix} x_1 + 1 \\ 2x_2 + \frac{1}{2} \end{bmatrix} = 0$$

$$x_1 = -1$$

$$x_2 = -\frac{1}{4}$$

The Steepest Descent Method is

$$x^{k+1} := x^k - \alpha_k \nabla f(x^k) \tag{1}$$

$$\alpha_k = argmin_{\alpha \geq 0} f(x^k - \alpha \nabla f(x^k)) \tag{2}$$

$$\nabla f(x) = \begin{bmatrix} 1 + x_1 \\ \frac{1}{2} + 2x_2 \end{bmatrix}, \nabla f(x_0) = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix}, x_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

First we will use the first derivative test to compute α_0 with (2)
Define $g(\alpha)$ as:

$$g(\alpha) = -\alpha - 0.25\alpha + 0.75\alpha^2 + 3$$

Minimize $g(\alpha)$:

$$\alpha_0 = \arg\min_{\alpha \geq 0} g(\alpha)$$

Differentiate $g(\alpha)$ and set to zero:

$$g'(\alpha) = \frac{3}{2}\alpha - \frac{5}{4} = 0 \implies \alpha_0 = \frac{5}{6}$$

First iteration:

$$x^1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \frac{5}{6} \begin{bmatrix} 1 \\ 0.5 \end{bmatrix} = \begin{bmatrix} -\frac{5}{6} \\ -\frac{5}{12} \end{bmatrix}$$

To find x_2, we first determine

$$\nabla f(x^{(1)}) = \begin{bmatrix} \frac{1}{6} \\ -\frac{2}{6} \end{bmatrix}$$

After some calculations, we find

$$\alpha_1 = argmin(f(x^1) - \alpha \nabla f(x^{(1)})) = \frac{5}{9}$$

Second Iteration:

$$x_2 = x_1 - \alpha_1 \nabla f(x^1) = \begin{bmatrix} -\frac{5}{6} \\ -\frac{5}{12} \end{bmatrix} - \frac{5}{9} \begin{bmatrix} \frac{1}{6} \\ -\frac{2}{6} \end{bmatrix} = \begin{bmatrix} -0.926 \\ -0.231 \end{bmatrix}$$

---

2. Consider the two sequences $\{x^{(k)}\}$ and $\{y^{(k)}\}$ defined iteratively as follows:

$$x^{(k+1)} = ax^{(k)} \quad (2) \tag{3}$$
$$y^{(k+1)} = \left(y^{(k)}\right)^b \quad (3) \tag{4}$$

where $a \in \mathbb{R}$, $b \in \mathbb{R}$, $0 < a < 1$, $b > 1$, $x^{(0)} \neq 0$, $y^{(0)} \neq 0$, and $|y^{(0)}| < 1$.

(a) Derive a formula for $x^{(k)}$ in terms of $x^{(0)}$ and a. Use this to deduce that $x^{(k)} \to 0$.

(b) Derive a formula for $y^{(k)}$ in terms of $y^{(0)}$ and b. Use this to deduce that $y^{(k)} \to 0$.

(c) Find the order of convergence of $\{x^{(k)}\}$ and the order of convergence of $\{y^{(k)}\}$.

(d) Calculate the smallest number of iterations k such that $|x^{(k)}| \leq c|x^{(0)}|$, where $0 < c < 1$.
Hint: The answer is in terms of a and c. You may use the notation $\lceil z \rceil$ to represent the smallest integer not smaller than z.

(e) Calculate the smallest number of iterations k such that $|y^{(k)}| \leq c|y^{(0)}|$, where $0 < c < 1$.

(f) Compare the answer of part (e) with that of part (d), focusing on the case where c is very small.

(a)
$$x^{(1)} = ax^{(0)} \implies x^{(2)} = a(ax^{(0)}) \implies x^{(k)} = a^k x^{(0)}$$

where $a \in \mathbb{R}$, $0 < a < 1$. Thus, we can conclude:

$$x^{(0)} \cdot \lim_{k \to \infty} a^k = 0$$

(b)

$$y^{(1)} = (y^{(0)})^b$$
$$y^{(2)} = (y^{(1)})^b = ((y^{(0)})^b)^b = (y^{(0)})^{b^2}$$
$$y^{(3)} = (y^{(2)})^b = (y^{(0)})^{b^3}$$
$$y^{(k)} = (y^{(0)})^{b^k}$$

$b \in \mathbb{R}$, $|y^{(0)}| < 1$. So, we can conclude

$$\lim_{k \to \infty} (y^{(0)})^{b^k} = 0$$

(c) According to the definition of order of convergence, p is given by:

$$\lim_{k \to \infty} \frac{\|x^{(k+1)} - x^*\|}{\|x^{(k)} - x^*\|^p} = L, \quad \text{where } 0 < L < \infty.$$

Substituting $x^* = 0$:

$$\lim_{k \to \infty} \frac{|a^{k+1} x^{(0)}|}{(|a^k x^{(0)}|)^p} = \lim_{k \to \infty} \frac{a^{k+1} |x^{(0)}|}{a^{kp} |x^{(0)}|^p}.$$

Simplifying,

$$\lim_{k \to \infty} \frac{a^{k+1}}{a^{kp}} = \lim_{k \to \infty} a^{k(1-p)} \cdot a.$$

For the limit to be finite and non-zero, $1 - p = 0$, implying $p = 1$. So, $\{x^{(k)}\}$ converges *linearly* with order $p = 1$.

According to the definition of order of convergence, p is given by:

$$\lim_{k \to \infty} \frac{\|y^{(k+1)} - y^*\|}{\|y^{(k)} - y^*\|^p} = L, \quad \text{where } 0 < L < \infty.$$

Substituting $y^* = 0$:

$$\lim_{k \to \infty} \frac{\left|(y^{(0)})^{b^{k+1}}\right|}{\left|(y^{(0)})^{b^k}\right|^p} = \lim_{k \to \infty} \frac{(y^{(0)})^{b^{k+1}}}{(y^{(0)})^{p \, b^k}}.$$

Simplifying,

$$\lim_{k \to \infty} (y^{(0)})^{b^{k+1} - p \, b^k}.$$

For all $p > 0$

$$b^{k+1} - p \, b^k \to \infty \quad \text{as} \quad k \to \infty.$$

And the limit goes to 0, indicating the order of convergence is ∞.

(d)

$$|x^{(k)}| \le c|x^{(0)}| \implies |a^k x^{(0)}| \le c|x^{(0)}|$$
$$a^k x^{(0)} \le c x^{(0)}$$
$$k \ln a \le \ln c, \ln a < 0$$

3

$$k \geq \left\lceil \frac{\ln c}{\ln a} \right\rceil$$

(e)

$$|y^{(k)}| \leq c|y^{(0)}|$$

$$\Rightarrow \left|(y^{(0)})^{b^k}\right| \leq c\,|y^{(0)}|$$

$$\left|(y^{(0)})^{b^k-1}\right| \leq c$$

$$(b^k - 1)\ln|y^{(0)}| \leq \ln(c)$$

$$b^k \geq 1 + \frac{\ln(c)}{\ln|y^{(0)}|}$$

$$k \geq \left\lceil \log_b\left(1 + \frac{\ln(c)}{\ln|y^{(0)}|}\right) \right\rceil$$

(f) When c is very small, the $y^{(k)}$ sequence converges faster than the $x^{(k)}$ sequence, implying that the number of iterations needed for the $y^{(k)}$ sequence to reach a small value is less than that for the $x^{(k)}$ sequence.

---

3. This exercise explores a zero-finding algorithm. Suppose that we wish to solve the equation $h(\mathbf{x}) = 0$, where

$$h(\mathbf{x}) = \begin{bmatrix} 4 + 3x_1 + 2x_2 \\ 1 + 2x_1 + 3x_2 \end{bmatrix} \quad (4)$$

Consider using an algorithm of the form

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha h(\mathbf{x}^{(k)}),$$

where α is a scalar constant that does not depend on k.

(a) Find the solution of $h(\mathbf{x}) = 0$.

(b) Find the largest range of values of α such that the algorithm is globally convergent to the solution of $h(\mathbf{x}) = 0$.

(c) Assuming that α is outside the range of values in part (b), give an example of an initial condition $\mathbf{x}^{(0)}$ of the form $\begin{bmatrix} x_1 \\ 0 \end{bmatrix}^{\top}$ such that the algorithm is guaranteed not to satisfy the descent property.

(a)

$$3x_1 + 2x_2 = -4 \tag{5}$$
$$2x_1 + 3x_2 = -1 \tag{6}$$
$$x_1 = -2, x_2 = 1 \tag{7}$$

(b) By Theorem 8.3, For the fixed step size gradient algorithm, $x^{(k)} \to x^*$ for any $x^{(0)}$ if and only if

$$0 < \alpha < \frac{2}{\lambda_{max}(Q)}$$

The fixed step size gradient algorithm is

$$x^{(k+1)} = x^{(k)} - \alpha g^{(k)}, g^{(k)} = \nabla f(x^{(k)})$$

Note that since $g(k)$ is the gradient of $f(x^{(k)})$, $g(k)$ is linear.
$h(x)$ is linear, thus we can apply theorem 8.3, with $g(x) = h(x)$
Applying Theorem 8.3 to h(x), we find that

$$Q = J_h(x) = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$$

$$\lambda_{max} = 5, \lambda_{min} = 2$$

$$0 < \alpha < \frac{2}{5} = 0.4$$

(c) If the step size is too large, let's say that $\alpha = 1$, an example of an initial condition $x^{(0)}$ where the descent property is not satisfied is $x^{(0)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $f(x^{(0)}) = [4, 1]^T$.

$$x^{(1)} = x^{(0)} - \alpha h(x^{(0)}) \tag{8}$$
$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} -4 \\ -1 \end{bmatrix} \tag{9}$$
$$f(x^{(1)}) = [-10, -10]^T \tag{10}$$

Comparing the norms of $f(x^{(0)})$ and $f(x^{(1)})$, we can conclude that the descent property is not satisfied.

4. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x) = \frac{3}{2}\left(x_1^2 + x_2^2\right) + (1 + a)x_1x_2 - (x_1 + x_2) + b \quad (5)$$

where a and b are some unknown real-valued parameters.

(a) Write the function f in the usual multivariable quadratic form.

(b) Find the largest set of values of a and b such that the unique global minimizer of f exists, and write down the minimizer (in terms of the parameters a and b).

(c) Consider the following algorithm:

$$x^{(k+1)} = x^{(k)} - \frac{2}{5}\nabla f\left(x^{(k)}\right)$$

Find the largest set of values of a and b for which this algorithm converges to the global minimizer of f for any initial point $x^{(0)}$.

(a) The usual quadratic form is

$$f(x) = \frac{1}{2}x^T Q x - b^T x + c$$

$$\nabla f(x) = \begin{bmatrix} 3x_1 + (1 + a)x_2 \\ 3x_2 + (1 + a)x_1 \end{bmatrix}$$

$$\nabla f(x) = 2Qx \implies 2Q = \begin{bmatrix} 3 & (1 + a) \\ (1 + a) & 3 \end{bmatrix}$$

$$b = [1, 1]^T$$

The function f in quadratic form is

$$f(x) = \frac{1}{2}x^T \begin{bmatrix} 3 & (1 + a) \\ (1 + a) & 3 \end{bmatrix} x - \begin{bmatrix} 1 & 1 \end{bmatrix} x + b$$

(b) Recall that for the unique global minimizer of f to exist, Q must be positive definite, ensuring all the curvature is upwards.
To check positive definiteness of Q, we will check that the leading principle minors are all positive.

$$\nabla_1 = 3 > 0$$
$$\nabla_2 = det(Q) = 9 - (1 + a)^2$$
$$9 - (1 + a)^2 > 0 \to -4 < a < 2$$

We conclude that $-4 < a < 2$ and $b \in \mathbb{R}$
The minimizer $x^* = Q^{-1}b$

$$x^* = \frac{1}{9-(1+a)^2}\begin{bmatrix} 3 & -(1+a) \\ -(1+a) & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$x^* = \frac{1}{-(a+4)(a-2)}\begin{bmatrix} 2-a \\ 2-a \end{bmatrix}$$

$$x^* = \frac{1}{a+4}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

(c) Using Theroem 8.3, we know that for the algorithm to converge globally,

$$0 < \alpha < \frac{2}{\lambda_{max}(Q)}, \lambda_1 = 4+a, \lambda_2 = 2-a, \alpha = \frac{2}{5}$$

$$\lambda_{max}(Q) < \frac{2}{\alpha} = 5$$

$$\lambda_{max}(Q) = max(4+a, 2-a) \rightarrow -3 < a < 1$$

So, $-3 < a < 1$ and $b \in \mathbb{R}$

---

5. Let $f : \mathbb{R} \to \mathbb{R}$, $f \in C^3$, with first derivative f', second derivative f'', and unique minimizer x^*. Consider a fixed-step-size gradient algorithm:

$$x^{(k+1)} = x^{(k)} - \alpha f'\left(x^{(k)}\right).$$

Suppose that $f''(x^*) \neq 0$ and $\alpha = \frac{1}{f''(x^*)}$. Assuming that the algorithm converges to x^*, show that the order of convergence is at least 2.

To show that the order of convergence is at least 2, I will show that the error at each iteration is proportional to the square of the previous error, demonstrating that the error is decreasing quadratically with each iteration.
The error at the k'th iteration is defined as

$$e^{(k)} = x^{(k)} - x^* \tag{11}$$

Let's first approximate $f'(x^{(k)})$ near x^* using Taylor's approximation

$$f'(x^{(k)}) = f'(x^*) + f''(x^*)(x^{(k)} - x^*) + \frac{1}{2}f(\xi)(x^{(k)} - x^*)^2 \tag{12}$$

Let $\frac{1}{2}f(\xi) = c$, and $f'(x^*) = 0$, since it is the minimizer. After plugging these values in to (12) and substituting (12) into the gradient algorithm, we have

$$x^{(k+1)} = x^{(k)} - \frac{f''(x^*)(x^{(k)} - x^*) + c(x^{(k)} - x^*)^2}{f''(x^*)}$$

Simplifying further, we find that

$$x^{(k+1)} - x^* = c(x^{(k)} - x^*)^2$$

Error at the $k + 1$th iteration is $e^{(k+1)} = x^{(k+1)} - x^*$. We can substitute this and (11) into the above equation to get

$$e^{(k+1)} = c(e^{(k)})^2$$

We have shown that the error at the k+1th iteration is proportional to error at the kth iteration, which implies that the given gradient algorithm has at least second-order convergence.

---

6. Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by

$$f(x) = \frac{1}{2}x^\top Q x - x^\top b,$$

where $b \in \mathbb{R}^n$ and Q is a real symmetric positive definite $n \times n$ matrix. Suppose that we apply the steepest descent method to this function, with $x^{(0)} \neq Q^{-1}b$. Show that the method converges in one step, that is, $x^{(1)} = Q^{-1}b$, if and only if $x^{(0)}$ is chosen such that

$$g^{(0)} := Q x^{(0)} - b$$

is an eigenvector of Q.

$(\rightarrow)$:
First we will show that if the method converges in one step

$$x^{(1)} = x^* = Q^{-1}b$$

Then, $g^{(0)} = Q x^{(0)} - b$ is an eigenvector of Q. The steepest descent method is

$$Q^{-1}b = x^{(1)} = x^{(0)} - \alpha g^{(0)} = x^{(0)} - \alpha(Q x^{(0)} - b)$$

Simplifying this, we find

$$Q^{-1}b = x^{(0)} - \alpha g^{(0)}$$
$$b = Q x^{(0)} - \alpha Q g^{(0)}$$
$$\alpha Q g^{(0)} = Q x^{(0)} - b$$
$$Q g^{(0)} = \frac{1}{\alpha}Q x^{(0)} - b = \frac{1}{\alpha}g^{(0)}$$

We can let $\lambda = \alpha$, then since $Q g^{(0)} = \lambda g^{(0)}$, we can conclude that $g^{(0)}$ is an eigenvector of Q
$(\leftarrow)$:
For the backwards direction, we will show if $x^{(0)}$ is chosen such that $g^{(0)} = Q x^{(0)} - b$ is an eigenvector of Q, then the steepest descent method converges in one step.
The gradient at the initial point $x^{(0)}$ is:

$$g^{(0)} = Q x^{(0)} - b$$

The update for the next iteration is given by:

$$x^{(1)} = x^{(0)} - \alpha_k g^{(0)}$$

Since $g^{(0)}$ is an eigenvector of Q with eigenvalue λ, we have:

$$Qg^{(0)} = \lambda g^{(0)}$$

Therefore, the gradient can be written as:

$$g^{(0)} = Q^{-1}\lambda g^{(0)}$$

The update for $x^{(1)}$ becomes:

$$x^{(1)} = x^{(0)} - \alpha_k \left(Q^{-1}\lambda g^{(0)}\right)$$

Substituting for $g^{(0)}$, we get:

$$x^{(1)} = x^{(0)} - \alpha_k \left(\lambda Q^{-1}(Qx^{(0)} - b)\right)$$

Simplifying further:

$$x^{(1)} = x^{(0)} - \lambda\alpha_k \left(x^{(0)} - Q^{-1}b\right)$$

We choose α_k such that $\lambda\alpha_k = 1$, which satisfies Theorem 8.3. We conclude that the method converges in one step:

$$x^{(1)} = Q^{-1}b = x^*$$