Assignment 2

Botond Kovacs

January 12, 2024

1 (6 points)

$$min_x f(x) = \frac{1}{2m}\sum_{i=1}^{m}(x-u_i)^2$$

(a)

In part c) we prove that that f is a convex function and its gradient is $\frac{1}{m}\sum_{i=1}^{m}(x-u_i) = x - \frac{\sum_{i=1}^{m} u_i}{m}$. Hence, the function has an extreme value at $x = \frac{\sum_{i=1}^{m} u_i}{m}$, which is a global minimum, because the second derivative is 1 and the function is convex, so all local minimums all global minimums. Thus, the solution of the problem is $x = \frac{\sum_{i=1}^{m} u_i}{m}$.

(b)

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\|$$

$$\|x - \frac{\sum_{i=1}^{m} u_i}{m} - (y - \frac{\sum_{i=1}^{m} u_i}{m})\| \leq L\|x-y\|$$

$$\|x-y\| \leq L\|x-y\|$$

$$1 \leq L$$

So the Lipschitz constant of ∇f is 1.

(c)

For a twice differentiable $f : \mathbb{R}^n \to \mathbb{R}$, the following holds:

- f is μ-strongly convex if and only if $\nabla^2 f(x) \succeq \mu I$ for all x.

It is trivial that f is twice differentiable. Its gradient is $x - \frac{\sum_{i=1}^{m} u_i}{m}$, so its Hessian is 1, hence the problem is 1-strongly convex.

(d)

We know that for L-smooth functions a reasonable of the stepsize is $\frac{1}{L^2}$. From b) we know that our f is L-smooth, so let α be 1.

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

$$x_{k+1} = x_k - (x_k - \frac{\sum_{i=1}^{m} u_i}{m})$$

$$x_{k+1} = \frac{\sum_{i=1}^{m} u_i}{m}$$

So regardless of what x_0 is starting from x_1 we already arrive at the optimal solution: $\frac{\sum_{i=1}^{m} u_i}{m}$.

(e)

$$f_i(x) := \frac{1}{2}(x - u_i)^2$$
$$\nabla f_i(x) = x - u_i$$

- $x_0 = 0$
- $x_1 = x_0 - \alpha_0 \nabla f_1(x_0) = 0 - 1 * (0 - u_1) = u_1$
- $x_2 = x_1 - \alpha_1 \nabla f_2(x_1) = u_1 - \frac{1}{2} * (u_1 - u_2) = \frac{u_1+u_2}{2}$
- $x_3 = x_2 - \alpha_2 \nabla f_3(x_2) = \frac{u_1+u_2}{2} - \frac{1}{3} * (\frac{u_1+u_2}{2} - u_3) = \frac{u_1+u_2+u_3}{3}$
- $x_k = x_{k-1} - \alpha_{k-1} \nabla f_k(x_{k-1}) = \frac{u_1+...+u_{k-1}}{k-1} - \frac{1}{k} * (\frac{u_1+...+u_{k-1}}{k-1} - u_k) = \frac{u_1+...+u_k}{k}$

So the kth iterate is always the average of $u_1, ..., u_k$ where for an index $j > m$, $u_j = u_{j-m}$. This also means that the algorithm arrives at the solution in every mth iterate, but converges around it otherwise.

2 (1 point)

$$f_k(x) := f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2} \langle \nabla^2 f(x_k)(x - x_k), x - x_k \rangle$$

For the minimization of $f_k(x)$ we need its gradient to be 0:

$$\nabla f_k(x) = 0$$
$$\nabla f(x_k) + \nabla^2 f(x_k)(x - x_k) = 0$$

Knowing that $\nabla^2 f(x_k)$ is positive definit from this we obtain $x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} f(x_k)$, which is exactly the Newton method applied to $\min_x f(x)$.

3 (3 points)

If f is differentiable and μ-strongly convex for its minimizer $x*$ and every x it holds that

$$\|\nabla f(x)\| \geq \mu \|x - x*\|$$

We know from class that for this class of functions it holds that

$$f(x^*) \geq f(x) + < \nabla f(x), x^* - x > + \frac{\mu}{2} \|x - x^*\|^2$$
$$\geq f(x) - \|\nabla f(x)\| \|x^* - x\| + \frac{\mu}{2} \|x - x^*\|^2$$
$$\geq f(x^*) + \frac{\mu}{2} \|x - x^*\|^2 - \|\nabla f(x)\| \|x^* - x\| + \frac{\mu}{2} \|x - x^*\|^2$$

The second row came from the Cauchy-Schwarz inequality and the third row came from the same inequality also used in the first row after switching x and x_0:

$$f(x) \geq f(x^*) + < \nabla f(x^*), x - x^* > + \frac{\mu}{2} \|x - x^*\|^2$$
$$f(x) \geq f(x^*) + \frac{\mu}{2} \|x - x^*\|^2$$

as $\nabla f(x^*) = 0$, because x^* is a minimizer. And this gives us

$$0 \geq -\|\nabla f(x)\| \|x^* - x\| + \mu \|x - x^*\|^2$$
$$\|\nabla f(x)\| \|x^* - x\| \geq \mu \|x - x^*\|^2$$
$$\|\nabla f(x)\| \geq \mu \|x - x^*\|$$

The last inequality is valid, if $\|x^* - x\|$ is not 0, but if it is, then the inequality is true with equality.

4 (2 points)

We want to apply $x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$.

- $f(x) = \sqrt{1+x^2}$
- $\nabla f(x) = \frac{x}{\sqrt{1+x^2}}$
- $\nabla^2 f(x) = \frac{1}{(1+x^2)^{3/2}}$
- $[\nabla^2 f(x)]^{-1} = (1+x^2)^{3/2}$

Hence $x_{k+1} = x_k - x_k(1+x_k^2)$. It is trivial that the minimizer is 0, but 3 different things can happen based on x_0. If its absolute value is 1, then x_k is jumping between 1 and -1. If its absolute value is larger than 1, then the absolute value of x_k is converging to infinity. If its absolute value is less than 1, then it is converging to the solution. Examples:

- if $x_0 = 1$
- $x_1 = x_0 - x_0(1+x_0^2) = 1 - 1 * 2 = -1$
- $x_2 = x_1 - x_1(1+x_1^2) = -1 - (-1) * 2 = 1$

- if $x_0 = 2$
- $x_1 = x_0 - x_0(1+x_0^2) = 2 - 2 * 5 = -8$
- $x_2 = x_1 - x_1(1+x_1^2) = -8 - (-8) * 65 = 512$
- $x_3 = x_2 - x_2(1+x_2^2) = 512 - 512 * (1+512^2) = 241785163922925834941235 2$

- if $x_0 = 0.5$
- $x_1 = x_0 - x_0(1+x_0^2) = 0.5 - 0.5 * 1.25 = -0.125$
- $x_2 = x_1 - x_1(1+x_1^2) = -0.125 - (-0.125) * (1+0.125^2) = 0.001953125$

5 (2 points)

$$3x^2y + y^2 = 1$$
$$x^4 + xy^3 = 1$$

When solving an equation(system), the Newton method takes the following form:

$$x_{k+1} = x_k - J^{-1}(x_k)F(x_k)$$

$$F(x,y) = \begin{bmatrix} 3x^2y + y^2 - 1 \\ x^4 + xy^3 - 1 \end{bmatrix}$$

$$J(x,y) = \begin{bmatrix} 6xy & 3x^2 + 2y \\ 4x^3 + y^3 & 3xy^2 \end{bmatrix}$$

Let $(x_0, y_0) = (1,1)$. The first iteration is given by:

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \nabla F(1,1)^{-1} F(1,1) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 6 & 5 \\ 5 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{11}{7} \\ -\frac{2}{7} \end{bmatrix}$$

The second iteration:

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} \frac{11}{7} \\ -\frac{2}{7} \end{bmatrix} - \nabla F(\frac{11}{7}, -\frac{2}{7})^{-1} F(\frac{11}{7}, -\frac{2}{7}) = \begin{bmatrix} 1.23626365 \\ 0.02614397 \end{bmatrix}$$

$$\begin{bmatrix} x_3 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1.05944398 \\ 0.22318294 \end{bmatrix}$$

$$\begin{bmatrix} x_4 \\ y_4 \end{bmatrix} = \begin{bmatrix} 1.00015183 \\ 0.29733566 \end{bmatrix}$$

$$\begin{bmatrix} x_5 \\ y_5 \end{bmatrix} = \begin{bmatrix} 0.99288389 \\ 0.30631382 \end{bmatrix}$$

$$\begin{bmatrix} x_5 \\ y_5 \end{bmatrix} = \begin{bmatrix} 0.99272621 \\ 0.30646792 \end{bmatrix}$$

$$\begin{bmatrix} x_5 \\ y_5 \end{bmatrix} = \begin{bmatrix} 0.99272489 \\ 0.30646863 \end{bmatrix}$$

Here the method starts converging. With different x_0 we can obtain 2 more solutions: $x = -1.00669, y = 0.299443$ and $x = -0.428033, y = -1.311892$.

6 (5 points)

$$p(x) = \sum_{i=0}^{m} a_i x^i$$

$$p'(x) = \sum_{i=1}^{m} i \cdot a_i x^{i-1}$$

The Newton-method in this case:

$$x_{k+1} = x_k - \frac{p(x_k)}{p'(x_k)} = x_k - \frac{\sum_{i=0}^{m} a_i x_k^i}{\sum_{i=1}^{m} i \cdot a_i x_k^{i-1}}$$

Because $x_k > u_n$, there are two possibilities:

- $p(x_k) \geq 0$ for all k and p is convex here, thus $p'(x_k) \geq 0$ also stands
- $p(x_k) \leq 0$ for all k and p is concave here, thus $p'(x_k) \leq 0$ also stands

In each case $\frac{p(x_k)}{p'(x_k)} \geq 0$, meaning that the series x_k is decreasing, hence converging to the closest solution less than or equal to x_k, which is exactly u_n.

7 (3 points)

- 1) $\min_x f(x)$
- 2) $\min_x g(f(x))$

First we show that 1) and 2) are equivalent. $f(x)$ is convex so $\min_x f(x)$ exists. Let's suppose that x^* is a solution of 1), but not a solution of 2). In this case there exists x' such that $g(f(x')) < g(f(x^*))$. As g is an increasing function it is equivalent as saying $f(x') < f(x^*)$, which is a contradiction, because $x*$ is a solution of 1). Now let's suppose that x^* is a solution of 2), but not of 1). In this case for all $x g(f(x^*)) \leq g(f(x))$. As g is an increasing function it is equivalent to stating that $f(x^*) < f(x)$ for all x, which is a contradiction, since x^* is not a solution of 1). Hence 1) and 2) are equivalent.

For 1) the Newton method looks usual:

$$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$$

For 2):

$$x_{k+1} = x_k - [g''(f(x_k))(\nabla f(x_k))^T \nabla f(x_k) + g'(f(x_k)) \nabla^2 f(x_k)]^{-1} \nabla g(f(x_k))$$

Since applying the chain and the product rules

$$\nabla g(f(x)) = g'(f(x)) \nabla f(x)$$

and

$$\nabla^2 g(f(x)) = g''(f(x))(\nabla f(x))^T \nabla f(x) + g'(f(x)) \nabla^2 f(x)$$

8 (3 points)

$$\min_x \|x\|^2 \text{ subject to } x_1 + ... + x_n = 1$$

The objective function is convex, thus the problem must have a solution. We form Lagrangian:

$$L(x, \lambda) = \|x\|^2 + \lambda(x_1 + ... + x_n - 1)$$

Critical points of L can be found from

$$0 = 2x_1 + \lambda$$

$$.$$

$$.$$

$$.$$

$$0 = 2x_n + \lambda$$

$$0 = x_1 + ... + x_n - 1$$

This implies that $x_1 = ... = x_n = \frac{-\lambda}{2}$, furthermore $\frac{-\lambda * n}{2} = 1$. And by solving it, we get $x_1 = ... = x_n = \frac{1}{n}$. This point is a global minimizer. This point satisfies the constraint, and the objective function's value is $\frac{1}{n}$.

9 (3 points)

$$\min_x \|x - u\|^2 \text{ subject to } Ax = o$$

The objective function is convex, thus the problem must have a solution. We denote our Lagrange multipliers with the vector λ. We form Lagrangian:

$$L(x, \lambda_1, ..., \lambda_m) = \|x - u\|^2 + \lambda^T Ax$$

Critical points of L can be found from

$$0 = 2(x - u) + A^T \lambda$$

and

$$0 = Ax$$

From the first equation we get $x = u - \frac{1}{2}A^T\lambda$, combining it with the second one we get $0 = A(u - \frac{1}{2}A^T\lambda)$. This gives us $\lambda = 2(AA^T)^{-1}Au$ and finally $x = u - A^T(AA^T)^{-1}Au$. This is our global minimizer.