Optimization 1 - 098311

Winter 2021 - HW 4

Ido Czerninski 312544596, Asaf Gendler 301727715

November 19, 2020

Problem 1:

a)

solution 1:

since $f(x)$ has a L Lipschitz continuous gradient, it is twice continuously differentiable and from the equivalence we saw in the lecture:

$$\left\|\nabla^2 f(x)\right\| \leq L$$

in addition from the linear approximation theorem we know that, there exists $\xi \in [x, y]$ such that:

$$f(y) = f(x) + \nabla f(x)^T (y-x) + \frac{1}{2}(y-x)^T \nabla^2 f(\xi)(y-x)$$

from the Cauchy–Schwarz inequality:

$$\left|(y-x)^T \nabla^2 f(\xi)(y-x)\right| \leq \|(y-x)\| \left\|\nabla^2 f(\xi)(y-x)\right\|$$

and from the induced norm inequality:

$$\|(y-x)\| \left\|\nabla^2 f(\xi)(y-x)\right\| \leq \|(y-x)\| \left\|\nabla^2 f(\xi)\right\| \|(y-x)\| = \|(y-x)\|^2 \left\|\nabla^2 f(\xi)\right\| \leq L \|(y-x)\|^2$$

thus:

$$\left|(y-x)^T \nabla^2 f(\xi)(y-x)\right| \leq L \|(y-x)\|^2$$

$$(y-x)^T \nabla^2 f(\xi)(y-x) \leq L \|(y-x)\|^2$$

putting it back in the linear approximation theorem:

$$f(y) = f(x) + \nabla f(x)^T (y-x) + \frac{1}{2}(y-x)^T \nabla^2 f(\xi)(y-x) \leq f(y) = f(x) + \nabla f(x)^T (y-x) + \frac{1}{2} L \|(y-x)\|^2$$

$$\boxed{f(y) \leq f(x) + \nabla f(x)^T (y-x) + \frac{1}{2} L \|(y-x)\|^2}$$

$\blacksquare$

solution 2:

we know that $f(x)$ has a L Lipschitz continuous gradient, thus:

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

multiply both sides by the positive constant $\|x - y\|$:

$$\|\nabla f(x) - \nabla f(y)\| \|x - y\| \leq L \|x - y\|^2$$

using the Cauchy–Schwarz inequality:

$$\left|(\nabla f(x) - \nabla f(y))^T (x - y)\right| \leq \|\nabla f(x) - \nabla f(y)\| \|x - y\|$$

thus:

$$\left|(\nabla f(x) - \nabla f(y))^T (x - y)\right| \leq L \|x - y\|^2$$

$$(\nabla f(x) - \nabla f(y))^T (x - y) \leq L \|x - y\|^2$$

define the function:

$$g(t) = f(x + t(y - x))$$

and notice the next properties:

$$g(0) = f(x)$$

$$g(1) = f(y)$$

$$g'(t) = \nabla f(x + t(y - x))^T (y - x)$$

$$g'(0) = \nabla f(x)^T (y - x)$$

$$g(1) - g(0) = \int_0^1 g'(t)\, dt \longrightarrow g(1) = g(0) + \int_0^1 g'(t)\, dt$$

because of the inequality we proved, than:

$$\begin{aligned} g'(t) - g'(0) &= \nabla f(x + t(y - x))^T (y - x) - \nabla f(x)^T (y - x) = \\ &= (\nabla f(x + t(y - x)) - \nabla f(x))^T (y - x) = \\ &= \frac{1}{t} (\nabla f(x + t(y - x)) - \nabla f(x))^T (x + t(y - x) - x) \\ &\leq \frac{1}{t} L \|x + t(y - x) - x\|^2 = \frac{1}{t} L \|t(y - x)\|^2 = tL \|(y - x)\|^2 \end{aligned}$$

thus:

$$g'(t) \le g'(0) + tL\left\|(y-x)\right\|^2$$

putting it all together:

$$g(1) = g(0) + \int_0^1 g'(t)\,dt \le g(0) + \int_0^1 g'(0) + tL\left\|(y-x)\right\|^2 dt =$$

$$= g(0) + g'(0) + L\left\|(y-x)\right\|^2 \left[\frac{t^2}{2}\right]_0^1 = g(0) + g'(0) + \frac{1}{2}L\left\|(y-x)\right\|^2$$

now going back from g to f:

$$f(y) = f(x) + \nabla f(x)^T (y-x) + \frac{1}{2}L\left\|(y-x)\right\|^2$$

■

b)

let's look at:

$$g_k(x) = f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{t_k}\left\|x - x_k\right\|_2^2 =$$

$$= f(x_k) + \nabla f(x_k)^T x - \nabla f(x_k)^T x_k + \frac{1}{t_k}\left(x^T x - 2x_k^T x + x_k^T x_k\right) =$$

$$= x^T \underbrace{\left(\frac{1}{t_k}I\right)}_{A} x + 2\underbrace{\left(\frac{1}{2}\nabla f(x_k) - \frac{1}{t_k}x_k\right)^T}_{b} x + \underbrace{f(x_k) - \nabla f(x_k)^T x_k + \frac{1}{t_k}x_k^T x_k}_{c}$$

this is a quadratic function of x , and in addition $A = \frac{1}{t_k}I$ is positive definite $(t_k > 0)$, hence:

$$x = -A^{-1}b = -t_k I\left(\frac{1}{2}\nabla f(x_k) - \frac{1}{t_k}x_k\right) = x_k - \frac{1}{2}t_k \nabla f(x_k)$$

is a strict global minimum of $g_k(x)$

therefore:

$$x_{k+1} = \arg\min_{x \in R^n} \left\{g_k(x)\right\} = x_k - \frac{1}{2}t_k \nabla f(x_k)$$

c)

for:

$$0 < t_k \le \frac{2}{L}$$

$$\Downarrow$$

$$\frac{1}{t_k} \geq \frac{L}{2}$$

Therefore:

$$\begin{aligned} g_k(x) &= f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{t_k} \|x - x_k\|_2^2 \\ &\geq f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{L}{2} \|x_k - x\|_2^2 \end{aligned}$$

using the inequality we proved in section a:

$$f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{L}{2} \|x_k - x\|_2^2 \geq f(x)$$

hence for the interval $I = \left(0, \frac{2}{L}\right] \subset R_{++}$, for $t_k \in I$ we have:

$$f(x) \leq g_k(x), \quad \forall x \in R^n$$

d)

using the inequality from section c in the point x_{k+1} we get:

$$g_k(x_{k+1}) \geq f(x_{k+1})$$

$$f(x_k) + \nabla f(x_k)^T (x_{k+1} - x_k) + \frac{1}{t_k} \|x_{k+1} - x_k\|_2^2 \geq f(x_{k+1})$$

plugging $x_{k+1} = x_k - \frac{1}{2} t_k \nabla f(x_k)$:

$$f(x_k) - \frac{1}{2} t_k \|\nabla f(x_k)\|_2^2 + \frac{1}{4} t_k \|\nabla f(x_k)\|_2^2 \geq f(x_{k+1})$$

$$f(x_k) - f(x_{k+1}) \geq \underbrace{\frac{1}{4} t_k}_{M>0} \|\nabla f(x_k)\|_2^2$$

thus for $\nabla f(x_k) \neq 0$:

$$f(x_k) \geq \underbrace{M}_{>0} \underbrace{\|\nabla f(x_k)\|_2^2}_{>0} + f(x_{k+1}) > f(x_{k+1})$$

$$f(x_k) > f(x_{k+1})$$

Problem 2:

a)

Let $x_0 \in \mathbb{R}^n$ be the initial guess for the newton method.

We know:

$$x_1 = x_0 + d_0$$

such that:

$$\nabla^2 f(x_0)\, d_0 = -\nabla f(x_0)$$

in our case:

$$\nabla f(x_0) = 2Ax_0 + 2b$$
$$\nabla^2 f(x_0) = 2A$$

hence d_0 holds:

$$2Ad_0 = -2Ax_0 - 2b$$

$A \succ 0$ thus invertible. By multiplying both side in A^{-1} we get:

$$2d_0 = -2x_0 - 2A^{-1}b$$
$$d_0 = -x_0 - A^{-1}b$$

hence:

$$x_1 = x_0 + d_0 = x_0 - x_0 - A^{-1}b$$
$$= -A^{-1}b \underbrace{=}_{*} \underset{x\in\mathbb{R}^n}{\operatorname{argmin}} f(x)$$

The minimum is attained after one iteration.

$(*)$Since $A \succ 0$ we know that the minimum of the quadratic form is attained exactly at the resulted point.

b)

We assume t can be any **constant** in $\mathbb{R}_{++}$:

Let's find a condition on x_0 in which the optimum point is attained after only one step:

The update rule of the gradient descend algorithm is given by:

$$x_{min} = x_0 - t\nabla f\left(x_0\right)$$

$$\begin{aligned} -A^{-1}b &= x_0 - t\left(2Ax_0 + 2b\right) \\ &= \left(I - 2tA\right)x_0 - 2tb \end{aligned}$$

$$\left(I - 2tA\right)x_0 = \left(2tI - A^{-1}\right)b$$

$$-\left(2tI - A^{-1}\right)Ax_0 = \left(2tI - A^{-1}\right)b$$

$$\underbrace{\left(2tI - A^{-1}\right)}_{B}\left(Ax_0 + b\right) = 0$$

First, we see that if B is invertible $\boxed{x_0 = A^{-1}b = x_{min}}$ is the only point that will achieve optimality after one iteration (by multiplying by the inverse in both sides).

Otherwise, we know that $\mathrm{Ker}\left(B\right) \neq \{0\}$ and the resulted equality hold if and only if:

$$Ax_0 + b \in \mathrm{Ker}\left(B\right)$$

Let v be any eigen vector (not necessarily normalized) with a zero eigen value of B:

$$Ax_0 + b = v$$

$$\boxed{x_0 = A^{-1}\left(v - b\right)}$$

Problem 3:

a)

$$f(x,y) = f_1(x,y)^2 + f_2(x,y)^2$$

$$f_1(x,y) = -13 + x + ((5-y)y - 2)y$$
$$f_2(x,y) = -29 + x + ((y+1)y - 14)y$$

First, for convenient, let's find the first and second derivative of f_1, f_2:

$$\frac{\partial}{\partial x} f_1 = 1$$

$$\begin{aligned}\frac{\partial}{\partial y} f_1 &= y(5-2y) + ((5-y)y - 2) \\ &= -2y^2 + 5y + 5y - y^2 - 2 \\ &= -3y^2 + 10y - 2\end{aligned}$$

$$\frac{\partial^2}{\partial x^2} f_1 = 0$$
$$\frac{\partial^2}{\partial y^2} f_1 = -6y + 10$$
$$\frac{\partial^2}{\partial xy} f_1 = 0$$
$$\frac{\partial}{\partial x} f_2 = 1$$

$$\begin{aligned}\frac{\partial}{\partial y} f_2 &= (2y+1)y + ((y+1)y - 14) \\ &= 2y^2 + y + y^2 + y - 14 \\ &= 3y^2 + 2y - 14\end{aligned}$$

$$\frac{\partial^2}{\partial x^2} f_2 = 0$$

$$\frac{\partial^2}{\partial y^2} f_2 = 6y + 2$$

$$\frac{\partial^2}{\partial xy} f_2 = 0$$

Now, let's find the stationary point by checking when the gradient is equal to zero:

$$\begin{aligned}
\frac{\partial}{\partial x} f(x,y) &= 2 \cdot f_1(x,y) \cdot \frac{\partial}{\partial x} f_1(x,y) + 2 \cdot f_2(x,y) \cdot \frac{\partial}{\partial x} f_2(x,y) \\
&= 2\left(f_1(x,y) + f_2(x,y)\right) = 0 \\
&\Rightarrow f_2(x,y) = -f_1(x,y)
\end{aligned}$$

$$\begin{aligned}
& \boxed{f_2(x,y) = -f_1(x,y)} \\
\iff & -29 + x + ((y+1)y - 14)y = 13 - x - ((5-y)y - 2)y \\
\iff & -42 + 2x + y((y+1)y - 14 + (5-y)y - 2) = 0 \\
\iff & -42 + 2x + y(6y - 16) = 0 \\
\iff & 6y^2 - 16y + 2x - 42 = 0 \\
\iff & \boxed{x = -3y^2 + 8y + 21}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial y} f(x,y) &= 2 \cdot f_1(x,y) \cdot \frac{\partial}{\partial y} f_1(x,y) + 2 \cdot f_2(x,y) \cdot \frac{\partial}{\partial y} f_2(x,y) \\
&= 2\left[f_1(x,y) \cdot \left(-3y^2 + 10y - 2\right) + f_2(x,y) \cdot \left(3y^2 + 2y - 14\right)\right]
\end{aligned}$$

By using the condition we got on f_2 and in x:

$$
\begin{aligned}
&= 2\left[f_1(x,y)\cdot\left(-3y^2+10y-2\right)-f_1(x,y)\cdot\left(3y^2+2y-14\right)\right]\\
&= 2f_1(x,y)\left[-6y^2+8y+12\right]\\
&= -4f_1(x,y)\left(3y^2-4y-6\right)\\
&= -4f_1(x,y)\left(3y^2-4y-6\right)\\
&= -4\left(-13+x+((5-y)y-2)y\right)\left(3y^2-4y-6\right)\\
&= -4\left(-13+x+\left(5y-y^2-2\right)y\right)\left(3y^2-4y-6\right)\\
&= -4\left(-13+x+5y^2-y^3-2y\right)\left(3y^2-4y-6\right)\\
&= -4\left(-13-3y^2+8y+21+5y^2-y^3-2y\right)\left(3y^2-4y-6\right)\\
&= -4\left(-y^3+2y^2+6y+8\right)\left(3y^2-4y-6\right)\\
&= 4\left(y^3-2y^2-6y-8\right)\left(3y^2-4y-6\right)\\
&= 4\left(y^3-4y^2+2y^2-8y+2y-8\right)\left(3y^2-4y-6\right)\\
&= 4\left(y^2(y-4)+2y(y-4)+2(y-4)\right)\left(3y^2-4y-6\right)\\
&= 4(y-4)\left(y^2+2y+2\right)\left(3y^2-4y-6\right)\\
&= 4(y-4)\left(y^2+2y+2\right)\left[y-\left(\frac{2+\sqrt{22}}{3}\right)\right]\left[y-\left(\frac{2-\sqrt{22}}{3}\right)\right]
\end{aligned}
$$

$p(y)=y^2+2y+2>0, \quad \forall y\in R$ because:

$$\Delta = 4-8=-4<0$$

hence there are exactly three stationery points:

$$y_1=4, \quad y_2=\frac{2+\sqrt{22}}{3} \quad y_3=\frac{2-\sqrt{22}}{3}$$

using:

$$x=-3y^2+8y+21$$

$$
\begin{aligned}
x_1 &= 5\\
x_2 &= \frac{53+4\sqrt{22}}{3}\\
x_3 &= \frac{53-4\sqrt{22}}{3}
\end{aligned}
$$

hence the stationery points are:

$$(5,4), \quad \left(\frac{53+4\sqrt{22}}{3},\frac{2+\sqrt{22}}{3}\right), \quad \left(\frac{53-4\sqrt{22}}{3},\frac{2-\sqrt{22}}{3}\right)$$

$f(x,y)$ is twice continuous differentiable over R^2,and R^2 is an open set, therefore, all the optimum points must be stationery points, now we just need to classify the stationery points that we have found using the hessian:

$$\frac{\partial^2}{\partial x^2}f(x,y) = \frac{\partial}{\partial x}2\left(f_1(x,y)+f_2(x,y)\right) = 2(1+1) = 4$$

$$\begin{aligned}\frac{\partial^2}{\partial y\partial x}f(x,y) &= \frac{\partial}{\partial y}2\left(f_1(x,y)+f_2(x,y)\right) = 2\left(-3y^2+10y-2+3y^2+2y-14\right) = \\ &= 2(12y-16) = 8(3y-4)\end{aligned}$$

$$\begin{aligned}\frac{\partial^2}{\partial y^2}f(x,y) &= \frac{\partial}{\partial y}2\left[f_1(x,y)\cdot\frac{\partial}{\partial y}f_1(x,y)+f_2(x,y)\cdot\frac{\partial}{\partial y}f_2(x,y)\right] = \\ &= 2\left(\left(\frac{\partial}{\partial y}f_1(x,y)\right)^2+f_1(x,y)\frac{\partial^2}{\partial y^2}f_1(x,y)+\left(\frac{\partial}{\partial y}f_2(x,y)\right)^2+f_2(x,y)\frac{\partial^2}{\partial y^2}f_2(x,y)\right) = \\ &= 2\left(\left(-3y^2+10y-2\right)^2+f_1(x,y)(-6y+10)+\left(3y^2+2y-14\right)^2+f_2(x,y)(6y+2)\right)\end{aligned}$$

we saw that at the stationery points $f_1(x,y) = f_2(x,y) = -(y-4)\left(y^2+2y+2\right)$

$$\frac{\partial^2}{\partial y^2}f(x,y) = 2\left(\left(-3y^2+10y-2\right)^2-(y-4)\left(y^2+2y+2\right)(-12y+8)+\left(3y^2+2y-14\right)^2\right)$$

$$\nabla^2 f(x,y) = \begin{pmatrix} 4 & 8(3y-4) \\ 8(3y-4) & 2\left(\left(-3y^2+10y-2\right)^2-(y-4)\left(y^2+2y+2\right)(-12y+8)+\left(3y^2+2y-14\right)^2\right)\end{pmatrix}$$

$$\nabla^2 f(5,4) = \begin{pmatrix} 4 & 64 \\ 64 & 3728\end{pmatrix}$$

$$Tr\left(\nabla^2 f(5,4)\right) = 3732 > 0$$

$$det\left(\nabla^2 f(5,4)\right) = 10.816 > 0$$

it's a $2X2$ matrix, hence:

$$\nabla^2 f(5,4) \succ 0$$

Therefore $(5,4)$ is a strict local minimum

$$\nabla^2 f\left(\frac{53+4\sqrt{22}}{3},\frac{2+\sqrt{22}}{3}\right)=\begin{pmatrix} 4 & -16+8\sqrt{22} \\ -16+8\sqrt{22} & -643.52 \end{pmatrix}$$

$$Tr\left(\nabla^2 f\left(\frac{53+4\sqrt{22}}{3},\frac{2+\sqrt{22}}{3}\right)\right)=-639.52<0$$

$$det\left(\nabla^2 f\left(\frac{53+4\sqrt{22}}{3},\frac{2+\sqrt{22}}{3}\right)\right)=-3037.33<0$$

it's a $2X2$ matrix, hence $\nabla^2 f\left(\frac{53+4\sqrt{22}}{3},\frac{2+\sqrt{22}}{3}\right)$ is indefinite.

Therefore $\left(\frac{53+4\sqrt{22}}{3},\frac{2+\sqrt{22}}{3}\right)$ is a saddle.

$$\nabla^2 f\left(\frac{53-4\sqrt{22}}{3},\frac{2-\sqrt{22}}{3}\right)=\begin{pmatrix} 4 & -16-8\sqrt{22} \\ -16-8\sqrt{22} & 901.89 \end{pmatrix}$$

$$Tr\left(\nabla^2 f\left(\frac{53-4\sqrt{22}}{3},\frac{2-\sqrt{22}}{3}\right)\right)=905.89>0$$

$$det\left(\nabla^2 f\left(\frac{53-4\sqrt{22}}{3},\frac{2-\sqrt{22}}{3}\right)\right)=742.81>0$$

it's a $2X2$ matrix, hence:

$$\nabla^2 f\left(\frac{53-4\sqrt{22}}{3},\frac{2-\sqrt{22}}{3}\right)\succ 0$$

Therefore $\left(\frac{53-4\sqrt{22}}{3},\frac{2-\sqrt{22}}{3}\right)$ is a strict local minimum.

in addition:

$$f(x,y)=f_1(x,y)^2+f_2(x,y)^2\geq 0$$

and equality hold if and only if $f_1(x,y)=f_2(x,y)=0$.

we saw that:

$$f_1(5,4)=f_2(5,4)=0$$

hence $(5,4)$ is a global minimizer of $f(x,y)$.

$f(x,y)$ doesn't have a global maximum since it is not bounded, for example in the direction $(t,0)$:

$$f(t,0)=f_1(t,0)^2+f_2(t,0)^2=(-13+t)^2+(-29+t)^2\xrightarrow{t\to\infty}\infty$$

to summarize:

$$\begin{cases} (5,4) & \text{global minimizer} \\ \frac{53-4\sqrt{22}}{3}, \frac{2-\sqrt{22}}{3} & \text{strict local minimum} \\ \frac{53+4\sqrt{22}}{3}, \frac{2+\sqrt{22}}{3} & \text{saddle} \end{cases}$$

b)

```
Gradiant_Backtracking:
x0 = (-50,7) solution: (5,4) iterations: 2252
x0 = (20,7) solution: (5,4) iterations: 2447
x0 = (20,-18) solution: (11.4128,-0.8968) iterations: 2472
x0 = (5,-10) solution: (5,4) iterations: 2123

Gradiant_Newton_Backtracking:
x0 = (-50,7) solution: (5,4) iterations: 8
x0 = (20,7) solution: (5,4) iterations: 8
x0 = (20,-18) solution: (11.4128,-0.89681) iterations: 16
x0 = (5,-10) solution: (11.4128,-0.89681) iterations: 13

Guass_Newton_Backtracking:
x0 = (-50,7) solution: (5,4) iterations: 29
x0 = (20,7) solution: (5,4) iterations: 29
x0 = (20,-18) not converged
x0 = (5,-10) not converged
>>
```

Figure 1:

We can see in general, that if the algorithms converge, then they converge to stationary points as expected. Gradient method with backtracking and gradient Newton hybrid method with backtracking converged to a local minimum one time and two time respectively. The rest was the global minimum. It can be seen that Gradient method with backtracking was the slowest while gradient Newton hybrid method with backtracking was the fastest. As shown in class, it is no surprising since Newton method has a quadratic convergence rate when near the optimum. Another thing we see is that the Gauss Newton method with backtracking did not converge in two occasions. Remember that in Gauss Newton we assume that the matrix $J(x,y)^T J(x,y)$ is invertible.

Let's see when is the jacobian non invertible and thus $J(x,y)^T J(x,y)$ non invertible:

$$J(x,y) = \begin{pmatrix} \frac{\partial}{\partial x} f_1(x,y) & \frac{\partial}{\partial y} f_1(x,y) \\ \frac{\partial}{\partial x} f_2(x,y) & \frac{\partial}{\partial y} f_2(x,y) \end{pmatrix} = \begin{pmatrix} 1 & -3y^2+10y-2 \\ 1 & 3y^2+2y-14 \end{pmatrix}$$

$$\begin{aligned} det(J(x,y)) &= 3y^2+2y-14-\left(-3y^2+10y-2\right) = \\ &= 3y^2+2y-14+3y^2-10y+2 = \\ &= 6y^2-8y-12 = \\ &= 2\left(3y^2-4y-6\right) = 0 \end{aligned}$$

$$y_{1,2} = \frac{4 \pm \sqrt{16+72}}{6} = \frac{2 \pm \sqrt{22}}{3}$$

so at every point with a y coordinate of $\frac{2\pm\sqrt{22}}{3}$ the Jacobian matrix becomes non invertible, thus if the algorithm reached such a point during the iterations it will not converge. Interesting, both the stationery points that we have found (that are not the global minimum) have such y values, so it is very likely the algorithm converged to their area, and that the jacobian became very ill conditioned and thus the algorithm did not converge.

Problem 4:

define:

$$f\left(x\right)=\sum_{i=1}^{m}\left(\underbrace{||x-a_{i}||_{2}}_{=g_{i}(x)}-d_{i}\right)^{2}$$

$$g_{i}\left(x\right)=||x-a_{i}||_{2}=\sqrt{\sum_{j=1}^{n}\left(x_{j}-a_{ij}\right)^{2}}$$

for $x\notin\mathcal{A}=\{a_{1},a_{2},...,a_{m}\}$:

$$\begin{aligned}\frac{\partial g_{i}}{\partial x_{k}}&=\frac{1}{2g_{i}\left(x\right)}\cdot2\left(x_{k}-a_{ik}\right)\\&=\frac{x_{k}-a_{ik}}{g_{i}\left(x\right)}\end{aligned}$$

a)

Let's compute the gradient of f:

$$\begin{aligned}\frac{\partial f}{\partial x_{k}}&=2\sum_{i=1}^{m}\left(g_{i}\left(x\right)-d_{i}\right)\frac{\partial g_{i}}{\partial x_{k}}\\&=2\sum_{i=1}^{m}\left(g_{i}\left(x\right)-d_{i}\right)\frac{x_{k}-a_{ik}}{g_{i}\left(x\right)}\\&=2\sum_{i=1}^{m}\left(||x-a_{i}||_{2}-d_{i}\right)\frac{x_{k}-a_{ik}}{||x-a_{i}||_{2}}\\&=\end{aligned}$$

$$\begin{aligned}\nabla f\left(x\right)&=2\sum_{i=1}^{m}\left(||x-a_{i}||_{2}-d_{i}\right)\frac{\left(x-a_{i}\right)}{||x-a_{i}||_{2}}\\&=2\sum_{i=1}^{m}\left(x-a_{i}\right)-2\sum_{i=1}^{m}d_{i}\frac{\left(x-a_{i}\right)}{||x-a_{i}||_{2}}\underbrace{=}_{\text{demand}}0\\\Rightarrow&\sum_{i=1}^{m}\left(x-a_{i}\right)-\sum_{i=1}^{m}d_{i}\frac{\left(x-a_{i}\right)}{||x-a_{i}||_{2}}=0\\\Rightarrow&mx-\sum_{i=1}^{m}a_{i}-\sum_{i=1}^{m}d_{i}\frac{\left(x-a_{i}\right)}{||x-a_{i}||_{2}}=0\\\Rightarrow&\boxed{x=\frac{1}{m}\left(\sum_{i=1}^{m}a_{i}+\sum_{i=1}^{m}d_{i}\frac{\left(x-a_{i}\right)}{||x-a_{i}||_{2}}\right)}\end{aligned}$$

b)

First we know:

$$\begin{aligned}\nabla f\left(x_k\right) &= \sum_{i=1}^{m}\left(x_k - a_i\right) - \sum_{i=1}^{m} d_i \frac{\left(x_k - a_i\right)}{\|x - a_i\|_2} \\ &= m x_k - \sum_{i=1}^{m} a_i - \sum_{i=1}^{m} d_i \frac{\left(x_k - a_i\right)}{\|x_k - a_i\|_2}\end{aligned}$$

$$\begin{aligned}x_{k+1} &= \frac{1}{m}\left(\sum_{i=1}^{m} a_i + \sum_{i=1}^{m} d_i \frac{\left(x_k - a_i\right)}{\|x_k - a_i\|_2}\right) \\ &= \frac{1}{m}\left(m x_k - m x_k + \sum_{i=1}^{m} a_i + \sum_{i=1}^{m} d_i \frac{\left(x_k - a_i\right)}{\|x_k - a_i\|_2}\right) \\ &= x_k + \frac{1}{m}\left(-m x_k + \sum_{i=1}^{m} a_i + \sum_{i=1}^{m} d_i \frac{\left(x_k - a_i\right)}{\|x_k - a_i\|_2}\right) \\ &= x_k - \frac{1}{m}\left(m x_k - \sum_{i=1}^{m} a_i - \sum_{i=1}^{m} d_i \frac{\left(x_k - a_i\right)}{\|x_k - a_i\|_2}\right) \\ &= x_k - \frac{1}{m}\nabla f\left(x\right)\end{aligned}$$

we can conclude that the fixed point method is equivalent to the gradient decent method with constant size $t = \frac{1}{m}$.