098311 Optimization 1 Spring 2018
HW 4

Chen Tessler 305052680
Orr Krupnik 302629027

May 9, 2018

Problem 1.

1. An ellipse is defined by $E(x, \Sigma, r) = \left\{y \in \mathbb{R}^n : (y-x)^T \Sigma^{-1} (y-x) \leq r^2\right\}$ for any given PD matrix Σ. For a given set of points $y_1, ..., y_m \in \mathbb{R}^n$ the best ellipsoid fit is defined as the solution to this problem:

$$\min_{x,r} \sum_{i=1}^{n} \left((x-y_i)^T \Sigma^{-1} (x-y_i) - r^2\right)^2$$

   Find an equivalent LS problem.

2. Matlab question - see code.

Solution

1. Since Σ is PD, it is invertible and Σ^{-1} is also PD. Therefore, we can define the matrix $S \in \mathbb{R}^{n \times n}$ such that $\Sigma^{-1} = S^T S$. We can now write:

$$\min_{x,r} \sum_{i=1}^{m} \left((x-y_i)^T \Sigma^{-1} (x-y_i) - r^2\right)^2 = \min_{x,r} \sum_{i=1}^{m} \left((x-y_i)^T S^T S (x-y_i) - r^2\right)^2 =$$

$$= \min_{x,r} \sum_{i=1}^{m} \left(||S(x-y_i)||_2^2 - r^2\right)^2 = \min_{x,r} \sum_{i=1}^{m} \left(||Sx||_2^2 + ||Sy_i||_2^2 - 2y_i^T S^T S x - r^2\right)^2 =$$

$$= \min_{x,r} \sum_{i=1}^{m} \left(-2y_i^T \Sigma^{-1} x + ||Sy_i||_2^2 + ||Sx||_2^2 - r^2\right)^2 (*)$$

   We now define: $x_{n+1} = ||Sx||_2^2 - r^2$, and redefine $\tilde{x} = (x, x_{n+1})^T$. We also define $z_i = (-2y_i^T \Sigma^{-1}, 1)^T$ for each y_i. In this way, $(*)$ turns into:

$$(*) = \min_{\tilde{x}} \sum_{i=1}^{m} \left(z_i^T \tilde{x} + ||Sy_i||_2^2\right)^2$$

   Finally, defining A as the $m \times (n+1)$ matrix with z_i^T along its rows, and $b \in \mathbb{R}^m$ as $b = (||Sy_1||_2^2, ..., ||Sy_m||_2^2)^T$, the above sum can be written as $\min_{\tilde{x}} ||A\tilde{x} - b||_2^2$, which is exactly a linear least squares formulation.

2. See full solution in Matlab code. The results are: $x = \begin{pmatrix} -0.5361 \\ -0.5663 \end{pmatrix}$, $r = 0.3009$.

Problem 2.
Consider the minimization problem

$$\min\{x^T Q x : x \in \mathbb{R}^2\}$$

where Q is a positive definite 2×2 matrix. Suppose that we use the diagonal matrix:

$$D = \begin{pmatrix} Q_{1,1}^{-1} & 0 \\ 0 & Q_{2,2}^{-1} \end{pmatrix}$$

Show that the above scaling matrix improves the condition number of Q in the sense that

$$\kappa(D^{1/2} Q D^{1/2}) \leq \kappa(Q) \ .$$

Solution

First, let us calculate the eigenvalues of Q (by determinant and trace). Note Q is PD, thus symmetric and therefore $Q_{1,2} = Q_{2,1}$.

$$\begin{aligned}
\lambda_{max}(Q) + \lambda_{min}(Q) &= \mathrm{tr}(Q) = Q_{1,1} + Q_{2,2} \\
\lambda_{max}(Q) \cdot \lambda_{min}(Q) &= \det(Q) = Q_{1,1}Q_{2,2} - Q_{1,2}^2 \\
&\Rightarrow \lambda(Q)(Q_{1,1} + Q_{2,2} - \lambda(Q)) = Q_{1,1}Q_{2,2} - Q_{1,2}^2 \\
&\Rightarrow \lambda(Q)^2 - (Q_{1,1} + Q_{2,2})\lambda(Q) + Q_{1,1}Q_{2,2} - Q_{1,2}^2 = 0 \\
&\Rightarrow \lambda_{max,min}(Q) = \frac{Q_{1,1} + Q_{2,2} \pm \sqrt{(Q_{1,1} + Q_{2,2})^2 - 4Q_{1,1}Q_{2,2} + 4Q_{1,2}^2}}{2}
\end{aligned}$$

Therefore,

$$\begin{aligned}
\kappa(Q) = \frac{\lambda_{max}(Q)}{\lambda_{min}(Q)} &= \frac{Q_{1,1} + Q_{2,2} + \sqrt{(Q_{1,1} + Q_{2,2})^2 - 4Q_{1,1}Q_{2,2} + 4Q_{1,2}^2}}{Q_{1,1} + Q_{2,2} - \sqrt{(Q_{1,1} + Q_{2,2})^2 - 4Q_{1,1}Q_{2,2} + 4Q_{1,2}^2}} = \\
&= \frac{Q_{1,1} + Q_{2,2} + \sqrt{(Q_{1,1} - Q_{2,2})^2 + 4Q_{1,2}^2}}{Q_{1,1} + Q_{2,2} - \sqrt{(Q_{1,1} - Q_{2,2})^2 + 4Q_{1,2}^2}} = \frac{1 + \sqrt{\frac{(Q_{1,1} - Q_{2,2})^2 + 4Q_{1,2}^2}{(Q_{1,1} + Q_{2,2})^2}}}{1 - \sqrt{\frac{(Q_{1,1} - Q_{2,2})^2 + 4Q_{1,2}^2}{(Q_{1,1} + Q_{2,2})^2}}}
\end{aligned}$$

Next, let us calculate $D^{1/2}QD^{1/2}$ and its eigenvalues (by determinant and trace):

$$Q^{-1} = \frac{1}{\det(Q)} \begin{pmatrix} Q_{2,2} & -Q_{1,2} \\ -Q_{1,2} & Q_{1,1} \end{pmatrix} \Rightarrow D = \frac{1}{\det(Q)} \begin{pmatrix} Q_{2,2} & 0 \\ 0 & Q_{1,1} \end{pmatrix}$$

$$\Rightarrow D^{1/2} = \frac{1}{\sqrt{\det(Q)}} \begin{pmatrix} \sqrt{Q_{2,2}} & 0 \\ 0 & \sqrt{Q_{1,1}} \end{pmatrix}$$

$$\Rightarrow D^{1/2}QD^{1/2} = \frac{1}{\det(Q)} \begin{pmatrix} Q_{1,1}Q_{2,2} & Q_{1,2}\sqrt{Q_{1,1}Q_{2,2}} \\ Q_{1,2}\sqrt{Q_{1,1}Q_{2,2}} & Q_{1,1}Q_{2,2} \end{pmatrix}$$

$$\lambda_{max}(D^{1/2}QD^{1/2}) + \lambda_{min}(D^{1/2}QD^{1/2}) = \mathrm{tr}(D^{1/2}QD^{1/2}) = 2\frac{Q_{1,1}Q_{2,2}}{\det(Q)}$$

$$\lambda_{max}(D^{1/2}QD^{1/2}) \cdot \lambda_{min}(D^{1/2}QD^{1/2}) = \det(D^{1/2}QD^{1/2}) = \frac{(Q_{1,1}Q_{2,2})^2 - Q_{1,2}^2 Q_{1,1}Q_{2,2}}{(\det(Q))^2} = \frac{Q_{1,1}Q_{2,2}}{\det(Q)}$$

$$\Rightarrow \lambda(D^{1/2}QD^{1/2})^2 - 2\frac{Q_{1,1}Q_{2,2}}{\det(Q)}\lambda(D^{1/2}QD^{1/2}) + \frac{Q_{1,1}Q_{2,2}}{\det(Q)} = 0$$

$$\Rightarrow \lambda_{min,max}(D^{1/2}QD^{1/2})^2 = \frac{2\frac{Q_{1,1}Q_{2,2}}{\det(Q)} \pm \sqrt{4\left(\frac{Q_{1,1}Q_{2,2}}{\det(Q)}\right)^2 - 4\frac{Q_{1,1}Q_{2,2}}{\det(Q)}}}{2} =$$

$$= \frac{Q_{1,1}Q_{2,2}}{\det(Q)} \pm \sqrt{\left(\frac{Q_{1,1}Q_{2,2}}{\det(Q)}\right)^2 - \frac{Q_{1,1}Q_{2,2}}{\det(Q)}} = \frac{Q_{1,1}Q_{2,2} \pm \sqrt{(Q_{1,1}Q_{2,2})^2 - Q_{1,1}Q_{2,2}\det(Q)}}{\det(Q)} =$$

$$= \frac{Q_{1,1}Q_{2,2} \pm \sqrt{(Q_{1,1}Q_{2,2})^2 - Q_{1,1}Q_{2,2}(Q_{1,1}Q_{2,2} - Q_{1,2}^2)}}{\det(Q)} =$$

$$= \frac{Q_{1,1}Q_{2,2} \pm \sqrt{Q_{1,1}Q_{2,2}Q_{1,2}^2}}{\det(Q)}$$

Thus,

$$\kappa(D^{1/2}QD^{1/2}) = \frac{\lambda_{max}(D^{1/2}QD^{1/2})}{\lambda_{min}(D^{1/2}QD^{1/2})} = \frac{Q_{1,1}Q_{2,2} + \sqrt{Q_{1,1}Q_{2,2}Q_{1,2}^2}}{Q_{1,1}Q_{2,2} - \sqrt{Q_{1,1}Q_{2,2}Q_{1,2}^2}} = \frac{1 + \sqrt{\frac{Q_{1,2}^2}{Q_{1,1}Q_{2,2}}}}{1 - \sqrt{\frac{Q_{1,2}^2}{Q_{1,1}Q_{2,2}}}}$$

Now, both sides of the inequality we are trying to prove are of the form $\frac{1+\sqrt{x}}{1-\sqrt{x}}$. If we can show the "$x$" for $\kappa(D^{1/2}QD^{1/2})$ is always smaller, we prove our inequality (since the nominator and denominator are both closer to 1, hence their ratio is smaller). It now remains to show, therefore, that $\frac{Q_{1,2}^2}{Q_{1,1}Q_{2,2}} \leq \frac{(Q_{1,1}-Q_{2,2})^2+4Q_{1,2}^2}{(Q_{1,1}+Q_{2,2})^2}$.

If $Q_{1,2} = 0$ the result holds trivially. If $Q_{1,1} = Q_{2,2} = Q_e$, we get equality. For all other cases, we can use the fact that $0 < \det(Q) = Q_{1,1}Q_{2,2} - Q_{1,2}^2$ (since Q is PD), and consequentially, the fact that $Q_{1,1} > 0, Q_{2,2} > 0$:

$$\begin{aligned}
Q_{1,2}^2 &< Q_{1,1}Q_{2,2} \\
Q_{1,2}^2(Q_{1,1} - Q_{2,2})^2 &< Q_{1,1}Q_{2,2}(Q_{1,1} - Q_{2,2})^2 \\
4Q_{1,2}^2Q_{1,1}Q_{2,2} + Q_{1,2}^2(Q_{1,1} - Q_{2,2})^2 &< 4Q_{1,2}^2Q_{1,1}Q_{2,2} + Q_{1,1}Q_{2,2}(Q_{1,1} - Q_{2,2})^2 \\
4Q_{1,2}^2Q_{1,1}Q_{2,2} + Q_{1,2}^2(Q_{1,1}^2 - 2Q_{1,1}Q_{2,2} + Q_{2,2}^2) &< 4Q_{1,2}^2Q_{1,1}Q_{2,2} + Q_{1,1}Q_{2,2}(Q_{1,1} - Q_{2,2})^2 \\
Q_{1,2}^2(Q_{1,1}^2 + 2Q_{1,1}Q_{2,2} + Q_{2,2}^2) &< Q_{1,1}Q_{2,2}(4Q_{1,2}^2 + (Q_{1,1} - Q_{2,2})^2) \\
\frac{Q_{1,2}^2}{Q_{1,1}Q_{2,2}} &< \frac{(Q_{1,1} - Q_{2,2})^2 + 4Q_{1,2}^2}{(Q_{1,1} + Q_{2,2})^2}
\end{aligned}$$

Combining with the edge cases above, we get $\frac{Q_{1,2}^2}{Q_{1,1}Q_{2,2}} \leq \frac{(Q_{1,1}-Q_{2,2})^2+4Q_{1,2}^2}{(Q_{1,1}+Q_{2,2})^2}$ and therefore $\kappa(D^{1/2}QD^{1/2}) \leq \kappa(Q)$.

Problem 3. Let $A \in \mathbb{R}^{m\times n}$ with a_i being its i-th row, and let $y \in \{0,1\}^m$ be a binary vector. Consider the logistic regression problem defined by

$$\min_{x\in\mathbb{R}^n} \left\{ f(x) = -\sum_{i=1}^{m} \left[y_i \log h_i(x) + (1-y_i)\log(1-h_i(x)) \right] \right\}$$

where $h_i : \mathbb{R}^n \to \mathbb{R}$ is given by:

$$h_i(x) = \frac{\exp(a_i^T x)}{1+\exp(a_i^T x)}$$

The gradient and hessian of f are given by

$$\begin{aligned}
\nabla f(x) &= -\sum_{i=1}^{m} \left[y_i(1-h_i(x)) - (1-y_i)h_i(x) \right] a_i^T \\
\nabla^2 f(x) &= \mathbf{A}^T\mathbf{W}(x)\mathbf{A}
\end{aligned}$$

where $\mathbf{W}$ is a diagonal matrix with $\mathbf{W}_{i,i} = h_i(x)(1-h_i(x))$ for any $i = 1,2,...,n$.

1. Show that for any $i = 1,...,m$

$$\nabla h_i(x) = h_i(x)(1-h_i(x))a_i^T$$

   Use this result to prove that the gradient of f is given by:

$$\nabla f(x) = -\sum_{i=1}^{m} \left[y_i(1-h_i(x)) - (1-y_i)h_i(x) \right] a_i^T$$

2. Matlab section, see in HW description.

Solution

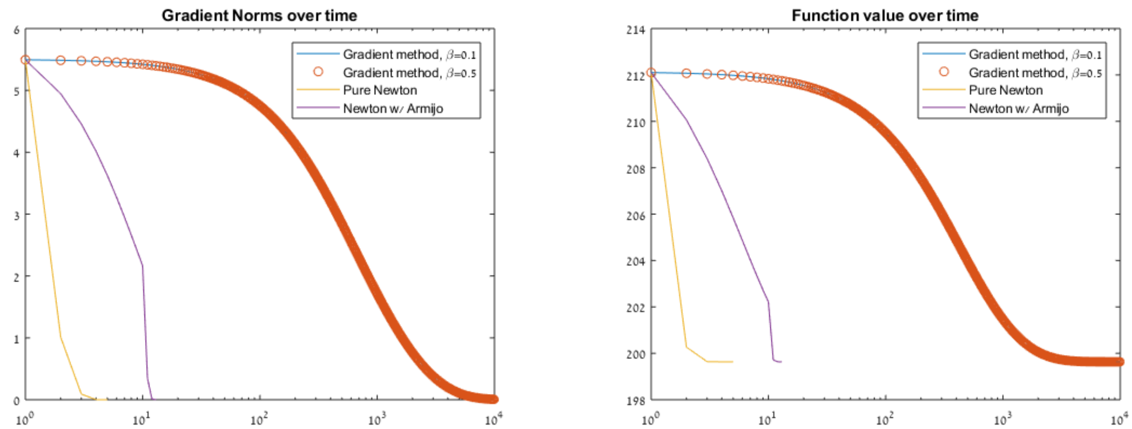

Figure 1: Gradient Norms and Function Values.

1. Using the chain rule, and denoting $y(x) = a_i^T x$, we have:

$$\nabla h_i(x) = \frac{\partial}{\partial y}\frac{\exp(y)}{1+\exp(y)}\nabla y(x) = \frac{\exp(y)(1+\exp(y)) - \exp(2y)}{(1+\exp(y))^2}a_i^T = \frac{\exp(y)}{(1+\exp(y))^2}a_i^T =$$
$$= \frac{\exp(y)}{1+\exp(y)}\frac{1}{1+\exp(y)}a_i^T = \frac{\exp(y)}{1+\exp(y)}\left(1 - \frac{1}{1+\exp(y)}\right)a_i^T = h_i(x)(1-h_i(x))a_i^T$$

Now, again using the chain rule, we have:

$$\nabla f(x) = -\sum_{i=1}^{m}\left[y_i\frac{1}{h_i(x)}\nabla h_i(x) - (1-y_i)\frac{1}{(1-h_i(x))}\nabla h_i(x)\right] =$$
$$= -\sum_{i=1}^{m}\left[y_i\frac{1}{h_i(x)}h_i(x)(1-h_i(x))a_i^T - (1-y_i)\frac{1}{(1-h_i(x))}h_i(x)(1-h_i(x))a_i^T\right] =$$
$$= -\sum_{i=1}^{m}\left[y_i(1-h_i(x))a_i^T - (1-y_i)h_i(x)a_i^T\right]$$

2. Results for the various methods are shown in Figure 1

(a) Grad Method with $\beta = 0.1$

(b) Grad Method with $\beta = 0.5$

(c) Pure Newton

(d) Newton with Backtracking

Method	(a)	(b)	(c)	(d)
Avg. Iteration Time	1.47e-4	1.46e-4	4.95e-4	5.86e-4
Total Time	1.471	1.455	2.47e-3	8.8e-3
Iterations	1e4	1e4	5	15

Under the given parameters, we notice that the pure Newton method converges not only in less iterations but also in less time. Additionally, the backtracking parameters need to be selected carefully, for instance in (d), backtracking is shown to reduce the rate of convergence as opposed to the pure Newton approach. Additionally, the backtracking step size selection does not assist either gradient method simulation - both run their course to the maximal number of iterations. However, when using different parameters or on a different problem, not only might the backtracking approach converge faster, but in certain cases the pure Newton method has been shown to diverge - a combination between backtracking and the Newton method overcomes this issue.

Our conclusion is: Which algorithm will perform best is problem dependent, and many times 'ill-selected' hyper-parameters (such as those selected for backtracking in this experiment) can have a major effect on the convergence rate.

Problem 4.
Consider the Fermat-Weber problem. Given a set of m anchor points $a_1, ..., a_m \in \mathbb{R}^n$ and m weights $w_1, ..., w_m > 0$, find a solution to

$$\min_{x\in\mathbb{R}^n} \left\{ f(x) = \sum_{i=1}^{n} w_i ||x - a_i||_2 \right\}$$

Let

$$p \in argmin\{f(a_i) : i = 1, ..., m\}$$

Suppose that

$$|| \sum_{i\neq p} w_i \frac{a_p - a_i}{||a_p - a_i||} || > w_p$$

1. Show that there exists a direction $d \in \mathbb{R}^n$ such that $f'(a_p; d) < 0$.
2. Show that there exists $x_0 \in \mathbb{R}^n$ satisfying $f(x_0) < min\{f(a_1), ..., f(a_p)\}$. Explain how to compute such a vector.

Solution

1. We assume at all points a_i are unique. In case of identical points (e.g. a_k, a_r) we can define a new weighting scheme by omitting the repetitions (e.g. $\hat{w}_k = (w_k + w_r), \hat{w}_r = 0$).

   First, we define an alternative function, $g(x) \triangleq \sum_{i\neq p} w_i ||x - a_i||_2$. Notice that the gradient of $g(x)$ is defined at a_p and is equal to:

$$\nabla g_i(a_p) = w_i \frac{a_p - a_i}{||a_p - a_i||}$$

   Such that

$$||\nabla g(a_p)|| = \sum_{i\neq p} ||w_i \frac{a_p - a_i}{||a_p - a_i||}|| \geq || \sum_{i\neq p} w_i \frac{a_p - a_i}{||a_p - a_i||} || > w_p$$

Where the first inequality holds due to the triangle inequality and the last holds by the assumption provided in the question.

We now follow the definition of the directional derivative of f at the point a_p:

$$\begin{aligned}
f'(a_p; d) &= \lim_{\beta \to 0} \frac{f(a_p + \beta d) - f(a_p)}{\beta} \\
&= \lim_{\beta \to 0} \frac{w_p||a_p + \beta d - a_p|| + \sum_{i \neq p} w_i(||a_p + \beta d - a_i|| - ||a_p - a_i||}{\beta} \\
&= \lim_{\beta \to 0} \frac{w_p \beta ||d|| + \sum_{i \neq p} w_i(||a_p + \beta d - a_i|| - ||a_p - a_i||}{\beta} \\
&\overset{(a)}{=} w_p||d|| + \lim_{\beta \to 0} \frac{\sum_{i \neq p} w_i(||a_p + \beta d - a_i|| - ||a_p - a_i||}{\beta} \\
&\overset{(b)}{=} w_p||d|| + \nabla g(a_p)^T d \\
&\overset{(c)}{=} w_p||\nabla g(a_p)|| - ||\nabla g(a_p)||^2 \\
&= ||\nabla g(a_p)||(w_p - ||\nabla g(a_p)||) \\
&\overset{(d)}{<} 0
\end{aligned}$$

(a): As $\frac{w_p \beta ||d||}{\beta} \equiv w_p||d||$, it can be taken out of the limit.
(b): From the definition of the function $g(x)$, the identical problem without the anchor a_p.
(c): By selecting a vector $d = -\nabla g(a_p)$.
(d): As $||\nabla g(a_p)|| > w_p > 0$ and $w_p - ||\nabla g(a_p)|| < 0$ by definition.

2. By definition we have that $p = argmin_i\{f(a_1), ..., f(a_p), ..., f(a_n)\}$. As such $f(a_p) \leq f(a_i)\ \forall i$.

   Following the linear approximation theorem, we define $\hat{x}$ a unit norm vector and $\beta \in \mathbb{R}_{++}$. The linear approximation theorem tells us that:

$$f(a_p + \beta \hat{x}) = f(a_p) + f'(a_p; \hat{x})\beta$$

   yet, in Q4.1 we showed that when $\beta \hat{x} = -||\nabla g(a_p)||$, $f'(a_p; \hat{x}) < 0$.

$$f(a_p - \beta \frac{\nabla g(a_p)}{||\nabla g(a_p)||}) = f(a_p) + f'(a_p; -\frac{\nabla g(a_p)}{||\nabla g(a_p)||})\beta < f(a_p)$$

   hence $\exists \beta > 0$ such that $f(a_p - \beta \frac{\nabla g(a_p)}{||\nabla g(a_p)||} < f(a_p) \leq f(a_i)\ \forall i$.

   We can find such a vector using the following steps:

   (a) Find $p = argmin_i f(a_i)$.

   (b) Compute $\nabla g(a_p)$ as defined above (w.r.t p found in step 1).

   (c) Perform a gradient descent step in the direction of $-\nabla g(a_p)$ using exact line search, denote the point found as x^*.

(d) Return x^* as found in step 3, this point is ensured to satisfy the requirement.

Problem 5. In the "source localization problem" we are given m locations of sensors $a_1, ..., a_m \in \mathbb{R}^n$ and approximate distances between the sensors and an unknown "source" located at $x \in \mathbb{R}^n$:

$$d_i \approx ||x - a_i||.$$

The problem is to find and estimate x given the locations $a_1, ..., a_m$ and the approximate distances $d_1, ..., d_m$. A particular formulation of the source localization problem consists of minimizing the following objective function:

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) = \sum_{i=1}^{m} (||x - a_i|| - d_i)^2 \right\}$$

We will denote the set of sensors by $\mathcal{A}$.

1. Show that the optimality condition $\nabla f(x) = 0 (x \notin \mathcal{A})$ is the same as

$$x = \frac{1}{m} \left[\sum_{i=1}^{m} a_i + \sum_{i=1}^{m} d_i \frac{x - a_i}{||x - a_i||} \right]$$

2. Show that the corresponding fixed point method

$$x^{k+1} = \frac{1}{m} \left[\sum_{i=1}^{m} a_i + \sum_{i=1}^{m} d_i \frac{x^k - a_i}{||x^k - a_i||} \right]$$

is a gradient descent method, assuming that $x^k \notin \mathcal{A}$ for all $k \geq 0$. What is the step-size?

Solution

1. We start by calculating the gradient of the objective function, and comparing to 0:

$$\begin{aligned}
0 = \nabla f(x) = \sum_{i=1}^{m} 2(||x - a_i|| - d_i) \frac{x - a_i}{||x - a_i||} &= \sum_{i=1}^{m} 2(x - a_i - d_i \frac{x - a_i}{||x - a_i||}) \\
&= 2mx - 2 \sum_{i=1}^{m} (a_i + d_i \frac{x - a_i}{||x - a_i||} \\
&\Rightarrow x = \frac{1}{m} \left[\sum_{i=1}^{m} a_i + \sum_{i=1}^{m} d_i \frac{x - a_i}{||x - a_i||} \right]
\end{aligned}$$

2.

$$x^{k+1} = \frac{1}{m}\left[\sum_{i=1}^{m} a_i + \sum_{i=1}^{m} d_i \frac{x^k - a_i}{||x^k - a_i||}\right] = x^k - x^k + \frac{1}{m}\left[\sum_{i=1}^{m} a_i + \sum_{i=1}^{m} d_i \frac{x^k - a_i}{||x^k - a_i||}\right] =$$

$$= x^k - \frac{1}{m}\left[m x^k - \sum_{i=1}^{m} a_i - \sum_{i=1}^{m} d_i \frac{x^k - a_i}{||x^k - a_i||}\right] = x^k - \frac{1}{m}\left[\sum_{i=1}^{m} x^k - a_i - d_i \frac{x^k - a_i}{||x^k - a_i||}\right] =$$

$$= x^k - \frac{1}{2m}\left[\sum_{i=1}^{m} 2(||x^k - a_i|| - d_i)\frac{x^k - a_i}{||x^k - a_i||}\right] = x^k - \frac{1}{2m}\nabla f(x^k)$$

This shows the corresponding fixed point method is a gradient method, with step size $\frac{1}{2m}$.

Problem 6. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by:

$$f(x,y) = \begin{cases} |\arctan(\frac{y}{x})| \cdot \sqrt{x^2+y^2} & x, y \neq 0 \\ x^2 & y = 0 \\ y^2 & x = 0 \end{cases}$$

Prove that for any k:

$$\lim_{|x|\to\infty} f(x, kx) = \infty$$

but f is not coercive.

Solution For $k = 0$, we have $f(x, y) = x^2$ and therefore the limit holds trivially. For any other k, we can define:

$$g(x) = f(x, kx) = \begin{cases} |\arctan(k)| \cdot \sqrt{2}|x| & x, y \neq 0 \\ 0 & x = 0 \end{cases}$$

For this case we also have $\lim_{|x|\to\infty} f(x, kx) = \lim_{|x|\to\infty} g(x) = \infty$.
However, note the path $\varphi(t) = (t\cos(\frac{1}{t}), t\sin(\frac{1}{t}))$. Taking $t \to \infty$, we have $||(x, y)^T|| = \sqrt{t^2\cos^2(\frac{1}{t}) + t^2\sin^2(\frac{1}{t})} = |t| \to \infty$. However:

$$\lim_{t\to\infty} f(\varphi(t)) = \lim_{t\to\infty} f(t\cos(\frac{1}{t}), t\sin(\frac{1}{t})) = \lim_{t\to\infty} \begin{cases} |\frac{1}{t}| \cdot |t| & t\cos(\frac{1}{t}), t\sin(\frac{1}{t}) \neq 0 \\ t^2\cos^2(\frac{1}{t}) & t\sin(\frac{1}{t}) = 0 \\ t^2\sin^2(\frac{1}{t}) & t\cos(\frac{1}{t}) = 0 \end{cases} =$$

$$= \begin{cases} 1 & t\cos(\frac{1}{t}), t\sin(\frac{1}{t}) \neq 0 \\ t^2 & t\sin(\frac{1}{t}) = 0 \\ t^2 & t\cos(\frac{1}{t}) = 0 \end{cases}$$

Since both $t\sin(\frac{1}{t})$ and $t\cos(\frac{1}{t})$ do not go to 0 as $t \to \infty$, the limit above is equal 1, which gives us a path along which $||x|| \to \infty$ but $\lim_{||x||\to\infty} = 1$, and therefore f is not coercive.