Assignment 7

1. General Comments.

- 5, 5.2 You claim (I hope I got it right): if $f(\mathbf{z}) = f(z_1, \ldots, z_m)$ is a quadratic function such that at a point $\mathbf{a}$ one has

$$\frac{\partial f}{\partial z_i}(\mathbf{a}) = 0, \text{and } \frac{\partial^2 f}{\partial z_i^2}(\mathbf{a}) > 0$$

then $\mathbf{a}$ is a local (global?) minimum of f.

Let's look at a simple example of a quadratic function $f(x, y) = x^2 - 3xy + y^2$. Clearly when $\mathbf{a} = 0$ one has

$$\frac{\partial f}{\partial x}(\mathbf{a}) = 0, \frac{\partial f}{\partial y}(\mathbf{a}) = 0, \text{and } \frac{\partial^2 f}{\partial x^2}(\mathbf{a}) = 2 > 0, \ \frac{\partial^2 f}{\partial y^2}(\mathbf{a}) = 2 > 0.$$

On the other hand

$$f(0, 0) = 0 > -1 = f(1, 1).$$

The point is that to reach a conclusion about max/min of $f(\mathbf{z})$ one needs a bit more information than just positivity of $\frac{\partial^2 f}{\partial z_i^2}(\mathbf{a})$.

- 5, 5.1 I would say the following:

(a) If $\log x = \log y$, then $x = y$.

(b) If $\log x = \frac{1}{m} \log(a_1 a_2 \ldots a_m)$, then $\log x = \log \left[(a_1 a_2 \ldots a_m)^{\frac{1}{m}}\right]$, and $x = (a_1 a_2 \ldots a_m)^{\frac{1}{m}}$.

$(a_1 a_2 \ldots a_m)^{\frac{1}{m}}$ is called the geometric mean of $\{a_1, \ldots, a_m\}$. The example shows that, in addition to the data set, the centroid depends of the distance like function.

Note that in the case of just two numbers a_1 and a_2 the geometric mean is $\sqrt{a_1 a_2}$, while the arithmetic mean is $\frac{a_1 + a_2}{2}$. It is not hard to show that

$$\sqrt{a_1 a_2} \leq \frac{a_1 + a_2}{2} \text{ (can you do this?)}$$

A more general question is whether this inequality holds for 3 (or more) numbers, that is whether

$$(a_1 a_2 \ldots a_m)^{\frac{1}{m}} \leq \frac{a_1 + \ldots + a_m}{m} \text{ for each set of } m \text{ positive numbers.}$$

We do not need these property of the geometric mean vs. the arithmetic mean for the Boley's paper, hence the two problems above are not "formally" assigned. However this is a useful property.

- 6.2 is fine, however it can be shorten. In addition to $\mathbf{x}^T \mathbf{v}' = 0$ we also know that $\mathbf{x}^T \mathbf{v} = 0$ (after all $\mathbf{v}$ is the projection of $\mathbf{c}$ onto a hyperplane perpendicular to $\mathbf{x}$). This means that the line four of the inequalities

$$|\mathbf{v} - \mathbf{v}'|^2 + |\mathbf{x}\mathbf{c}^T\mathbf{x}|^2 + 2\mathbf{x}^T\mathbf{v}\mathbf{c}^T\mathbf{x}$$

is, in fact,

$$|\mathbf{v} - \mathbf{v}'|^2 + |\mathbf{x}\mathbf{c}^T\mathbf{x}|^2,$$

and this completes the proof.

I can also suggest the following:

$$|\mathbf{c} - \mathbf{v}'|^2 = |\mathbf{c} - \mathbf{v} + \mathbf{v} - \mathbf{v}'|^2 = |\mathbf{c} - \mathbf{v}|^2 + 2(\mathbf{c} - \mathbf{v})^T(\mathbf{v} - \mathbf{v}') + |\mathbf{v} - \mathbf{v}'|^2.$$

Since the vector $\mathbf{c} - \mathbf{v}$ is perpendicular to both $\mathbf{v}$ and $\mathbf{v}'$ the dot product $(\mathbf{c} - \mathbf{v})^T(\mathbf{v} - \mathbf{v}') = 0$, hence

$$|\mathbf{c} - \mathbf{v}'|^2 = |\mathbf{c} - \mathbf{v}|^2 + |\mathbf{v} - \mathbf{v}'|^2 \geq |\mathbf{c} - \mathbf{v}|^2.$$

- 6.3 is fine. I would note the following: since $\mathbf{c} - \mathbf{w}$ is a constant vector one has

$$\sum_{i=1}^{m}(\mathbf{a}_i - \mathbf{c})^T(\mathbf{c} - \mathbf{w}) = \left[\sum_{i=1}^{m}(\mathbf{a}_i - \mathbf{c})^T\right](\mathbf{c} - \mathbf{w}) = \left[\sum_{i=1}^{m}\mathbf{a}_i - m\mathbf{c}\right](\mathbf{c} - \mathbf{w}) = [m\mathbf{c} - m\mathbf{c}]\,(\mathbf{c} - \mathbf{w}) = 0.$$

- 6.4 is fine.
- 6.5 the argument involving diagonalization is very nice. The result leads to two immediate questions:

Problem 7.1: True or False? All the eigenvalues λ_i of the matrix BB^T are non negative.

Problem 7.2: Let λ be the largest eigenvalue of BB^T. How to identify a unit eigenvector $\mathbf{v}$ of BB^T so that $BB^T\mathbf{v} = \lambda\mathbf{v}$.