Exercise 17.3

We have $f(x) = \frac{1}{2}\left((h_1(x))^2 + \dots + (h_m(x))^2\right)$,

and so

$$\frac{\partial f}{\partial x_i}(x) = h_1(x)\frac{\partial h_1}{\partial x_i}(x) + \dots + h_m(x)\frac{\partial h_m}{\partial x_i}(x) = (h(x))^T \begin{bmatrix} \frac{\partial h_1}{\partial x_i}(x) \\ \vdots \\ \frac{\partial h_m}{\partial x_i}(x) \end{bmatrix}.$$

Thus

$$\nabla f(x) = \left[\frac{\partial f}{\partial x_1}(x) \;\middle|\; \cdots \;\middle|\; \frac{\partial f}{\partial x_n}(x)\right]$$

$$= \left[(h(x))^T \begin{bmatrix} \frac{\partial h_1}{\partial x_1}(x) \\ \vdots \\ \frac{\partial h_m}{\partial x_1}(x) \end{bmatrix} \;\middle|\; \cdots \;\middle|\; (h(x))^T \begin{bmatrix} \frac{\partial h_1}{\partial x_n}(x) \\ \vdots \\ \frac{\partial h_m}{\partial x_n}(x) \end{bmatrix} \right]$$

$$= (h(x))^T \begin{bmatrix} \frac{\partial h_1}{\partial x_1}(x) & \cdots & \frac{\partial h_1}{\partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial h_m}{\partial x_1}(x) & \cdots & \frac{\partial h_m}{\partial x_n}(x) \end{bmatrix}$$

$$= (h(x))^T (\nabla h(x)).$$

Also, $[F(x)]_{ij} = \frac{\partial^2 f}{\partial x_j \partial x_i}(x) = \sum_{k=1}^{m} h_k(x) \frac{\partial^2 h_k}{\partial x_j \partial x_i}(x) + \sum_{k=1}^{m} \frac{\partial h_k}{\partial x_j}(x) \frac{\partial h_k}{\partial x_i}(x)$.

The (i,j)th entry of $(\nabla h(x))^T \nabla h(x)$ is

$$\sum_{k=1}^{m} \left[(\nabla h(x))^T\right]_{ik} \left[\nabla h(x)\right]_{kj} = \sum_{k=1}^{m} \left[\nabla h(x)\right]_{ki} \left[\nabla h(x)\right]_{kj}$$

$$= \sum_{k=1}^{m} \frac{\partial h_k}{\partial x_i}(x) \frac{\partial h_k}{\partial x_j}(x).$$

The (i,j)th entry of $\sum_{k=1}^{m} h_k(x) H_k(x)$ is

$$\sum_{k=1}^{m} h_k(x) \left[H_k(x)\right]_{ij} = \sum_{k=1}^{m} h_k(x) \frac{\partial^2 h_k}{\partial x_j \partial x_i}(x)$$

Hence

$$F(x) = (\nabla h(x))^T \nabla h(x) + \sum_{k=1}^{m} h_k(x) H_k(x).$$

Exercise 17.4

With $h(x) := \begin{bmatrix} h_1(x) \\ h_2(x) \\ h_3(x) \\ h_4(x) \end{bmatrix} = \begin{bmatrix} x_1^2 - x_2 - \delta_1 \\ x_1^2 + x_2 - \delta_2 \\ x_2^2 - x_1 - \delta_3 \\ x_2^2 + x_1 - \delta_4 \end{bmatrix}$, we have $f(x) = \frac{1}{2} h(x)^T h(x) \geq 0$ for all $x \in \mathbb{R}^2$

(1) If all $\delta_i = 0$, then $h(\hat{x}) = 0 \in \mathbb{R}^4$ and $f(\hat{x}) = 0$.

So $\forall x \in \mathbb{R}^2$, $f(x) \geq 0 = f(\hat{x})$.

Hence $\hat{x}$ is a global minimizer of f.

(2) We have

$$\nabla h(x) = \begin{bmatrix} 2x_1 & -1 \\ 2x_1 & 1 \\ -1 & 2x_2 \\ 1 & 2x_2 \end{bmatrix}.$$

With $x^{(0)} := 0$, we have $h(x^{(0)}) = \begin{bmatrix} 0.1 \\ -0.1 \\ 0.2 \\ -0.2 \end{bmatrix}$,

$$\nabla h(x^{(0)}) = \begin{bmatrix} 0 & -1 \\ 0 & 1 \\ -1 & 0 \\ 1 & 0 \end{bmatrix}.$$

Hence

$$(\nabla h(x^{(0)}))^T \nabla h(x^{(0)}) = \begin{bmatrix} 0 & 0 & -1 & 1 \\ -1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 0 & 1 \\ -1 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix},$$ and

$$(\nabla h(x^{(0)}))^T h(x^{(0)}) = \begin{bmatrix} 0 & 0 & -1 & 1 \\ -1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.1 \\ -0.1 \\ 0.2 \\ -0.2 \end{bmatrix} = \begin{bmatrix} -0.4 \\ -0.2 \end{bmatrix}$$

Thus the system $(\nabla h(x^{(0)}))^T \nabla h(x^{(0)}) d = -(\nabla h(x^{(0)}))^T h(x^{(0)})$ becomes $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} d = -\begin{bmatrix} -0.4 \\ -0.2 \end{bmatrix}$, and so $d = \begin{bmatrix} 0.2 \\ 0.1 \end{bmatrix}$.

Hence

$$x^{(1)} = x^{(0)} + d = 0 + d = d = \begin{bmatrix} 0.2 \\ 0.1 \end{bmatrix}.$$

The gradient of the objective function at $x^{(1)}$ is

$$\nabla f(x^{(1)}) = (h(x^{(1)}))^T \, \nabla h(x^{(1)}).$$

We have

$$h_1(x^{(1)}) = 0.04 - 0.1 + 0.1 = 0.04,$$
$$h_2(x^{(1)}) = 0.04 + 0.1 - 0.1 = 0.04,$$
$$h_3(x^{(1)}) = 0.01 - 0.2 + 0.2 = 0.01,$$
$$h_4(x^{(1)}) = 0.01 + 0.2 - 0.2 = 0.01,$$

and so

$$h(x^{(1)}) = \begin{bmatrix} 0.04 \\ 0.04 \\ 0.01 \\ 0.01 \end{bmatrix}.$$

Also,

$$\nabla h(x^{(1)}) = \begin{bmatrix} 2(0.2) & -1 \\ 2(0.2) & 1 \\ -1 & 2(0.1) \\ 1 & 2(0.1) \end{bmatrix} = \begin{bmatrix} 0.4 & -1 \\ 0.4 & 1 \\ -1 & 0.2 \\ 1 & 0.2 \end{bmatrix}.$$

Hence

$$\nabla f(x^{(1)}) = (h(x^{(1)}))^T \, \nabla h(x^{(1)})$$
$$= \begin{bmatrix} 0.4 & 0.4 & -1 & 1 \\ -1 & 1 & 0.2 & 0.2 \end{bmatrix} \begin{bmatrix} 0.04 \\ 0.04 \\ 0.01 \\ 0.01 \end{bmatrix}$$
$$= \begin{bmatrix} 0.032 \\ 0.004 \end{bmatrix} \neq 0.$$

Since the gradient $\nabla f(x^{(1)}) \neq 0$, $x^{(1)}$ cannot be a local minimizer of f.