**Exercise 4.17.2** Consider the quadratic function $Q(x) = \frac{1}{2}\mathbf{x}^T A\mathbf{x} - b\mathbf{x}$, with $A$ nonsingular square matrix of order $n$.

($a$) Find the gradient $\nabla Q$;

($b$) Write the gradient descent iteration;

($c$) Find the Hessian $H_Q$;

($d$) Write the iteration given by Newton's formula and compute its limit.

(a) $\nabla Q = AX - b$

(b) $X^{n+1} = X^n - \eta \nabla Q(X^n)$

$= X^n - \eta(AX^n - b)$

$= (I - \eta A)X^n + \eta b$

(c) $H_Q = \frac{1}{2}A$

(d) $X^{n+1} = X^n - H_Q^{-1}(X^n)\nabla Q(X^n)$

$= X^n - \frac{1}{2}A^{-1}(AX^n - b)$

$= X^n - \frac{1}{2}X^n + \frac{1}{2}A^{-1}b$

$= \frac{1}{2}X^n + \frac{1}{2}A^{-1}b$

$X^{n+1} = \frac{1}{2}X^n + \frac{1}{2}A^{-1}b$

$= \frac{1}{2}\left(\frac{1}{2}X^{n-1} + \frac{1}{2}A^{-1}b\right) + \frac{1}{2}A^{-1}b$

$= \frac{1}{2}\cdot\frac{1}{2}X^{n-1} + \frac{1}{2}\cdot\frac{1}{2}A^{-1}b + \frac{1}{2}A^{-1}b$

$= \frac{1}{2}\cdot\frac{1}{2}\left(\frac{1}{2}X^{n-2} + \frac{1}{2}A^{-1}b\right) + \frac{1}{2}\cdot\frac{1}{2}A^{-1}b + \frac{1}{2}A^{-1}b$

$$= \frac{1}{2^3} X^{n-2} + \frac{1}{2^3} A^{-1}b + \frac{1}{2^2} A^{-1}b + \frac{1}{2} A^{-1}b$$

$$\vdots$$

$$= \frac{1}{2^{n+1}} X^0 + A^{-1}b \left( \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \cdots + \frac{1}{2^{n+1}} \right)$$

$$\lim_{n \to \infty} X^{n+1} = 0 + A^{-1}b \cdot \frac{\frac{1}{2}}{1 - \frac{1}{2}}$$

$$= \boxed{A^{-1}b}$$

**Exercise 5.10.4** Explain the equivalence between the linear regression algorithm and the learning of a linear neuron.

linear regression algorithm

$$\hat{\beta} = \arg\min_{\beta} Q(\beta)$$

$$Q(\beta) = (Y - X\beta)^T (Y - X\beta)$$

$$= Y^T Y - 2\beta^T X^T Y + \beta^T X^T X \beta$$

$$\frac{\partial}{\partial \beta} Q(\beta) = -2X^T Y + 2X^T X \beta = 0$$

$$\therefore \hat{\beta} = (X^T X)^{-1} X^T Y$$

learning of a linear neuron

$$Y = \sum_{j}^{n} w_j X_j = w^T X = X^T w$$

$$w^* = \arg\min_{w} E[(Z-Y)^2]$$

$$E[(Z-Y)^2] = E[Z^2 - 2ZY + Y^2]$$

$$= E[Z^2 - 2ZX^T w + (w^T X)(X^T w)]$$

$$= E[Z^2] - 2E[ZX^T]w + w^T E[XX^T]w$$

$$= c - 2b^T w + w^T A w$$

$$(b = E[ZX],\ c = E[Z^2],\ A = E[XX^T])$$

$$\nabla_w E[(Z-Y)^2] = 2Aw - 2b = 0$$

$$\therefore w_* = A^{-1} b$$

* 동일한 Error function을 minimize 시키는 방법으로

$\hat{\beta}$, $w_*$를 구하므로 ⇒ equivalent.