Solutions to exercises combining matrix algebra with multivariate calculus

1. Consider the following function $\phi : \mathbb{R}^2 \to \mathbb{R}$ where μ and Ω are:

$$\phi(z) = \phi(x, y) = \frac{1}{2\pi}(\det\ \Omega)^{-\frac{1}{2}} e^{-\frac{1}{2}(z-\mu)^t \Omega (z-\mu)}$$

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \text{and} \quad \Omega = \begin{bmatrix} \omega_x & \omega_{xy} \\ \omega_{xy} & \omega_y \end{bmatrix}.$$

This function is in fact the probability density function of the bivariate Gaussian distribution.

(a) Compute the partial derivatives and Jacobian matrix of ϕ at the points $(0, 0)$ and μ, respectively.

The first order partial derivatives of ϕ are given by:

$$\frac{\partial \phi}{\partial x} = \phi(x, y) \left(-\omega_x(x - \mu_1) - \omega_{xy}(y - \mu_2)\right)$$
$$\frac{\partial \phi}{\partial y} = \phi(x, y) \left(-\omega_y(y - \mu_2) - \omega_{xy}(x - \mu_1)\right)$$

The second order partial derivatives of ϕ are given by:

$$\frac{\partial^2 \phi}{\partial x^2} = \phi(x, y) \left(\left(-\omega_x(x - \mu_1) - \omega_{xy}(y - \mu_2)\right)^2 - \omega_x\right)$$
$$\frac{\partial^2 \phi}{\partial y^2} = \phi(x, y) \left(\left(-\omega_y(y - \mu_2) - \omega_{xy}(x - \mu_1)\right)^2 - \omega_y\right)$$
$$\frac{\partial^2 \phi}{\partial x \partial y} = \phi(x, y) \left(\left(-\omega_x(x - \mu_1) - \omega_{xy}(y - \mu_2)\right)\left(-\omega_y(y - \mu_2) - \omega_{xy}(x - \mu_1)\right) - \omega_{xy}\right)$$

(b) Compute the Hessian matrix of ϕ at μ, say $H\phi(\mu)$.

The Hessian matrix at μ is given by:

$$H\phi(\mu) = \frac{1}{2\pi}(\det\ \Omega)^{-\frac{1}{2}} \begin{bmatrix} -\omega_x & -\omega_{xy} \\ -\omega_{xy} & -\omega_y \end{bmatrix} = -\frac{1}{2\pi}(\det\ \Omega)^{-\frac{1}{2}}\Omega$$

(c) Compute the eigenvalues of $H\phi(\mu)$.

The eigenvalues λ_1, λ_2 of $H\phi(\mu)$ can be found by noting that

i. $\lambda_1 + \lambda_2 = \text{trace} H\phi(\mu)$ and $\lambda_1\lambda_2 = \det H\phi(\mu)$.

ii. If λ is an eigenvalue of A then $\alpha\lambda$ is an eigenvalue of αA for any scalar α.

Thus, the eigenvalues of $H\phi(\mu)$ are given by:

$$\lambda_1 = -\frac{1}{2\pi}(\det\ \Omega)^{-\frac{1}{2}} \left(\frac{\omega_x + \omega_y + \sqrt{(\omega_x - \omega_y)^2 + 4\omega_{xy}^2}}{2}\right)$$

$$\lambda_2 = -\frac{1}{2\pi}(\det\ \Omega)^{-\frac{1}{2}} \left(\frac{\omega_x + \omega_y - \sqrt{(\omega_x - \omega_y)^2 + 4\omega_{xy}^2}}{2}\right)$$

5. Study the critical points of the "residual sum of squares" function $\psi : \mathbb{R}^n \to \mathbb{R}$ defined as

$$\psi(x) = (Ax - b)^t(Ax - b)$$

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$.

(a) What are the partial derivatives of the function $f(x) = (w^t x - c)$, with $w \in \mathbb{R}^n$ and $c \in \mathbb{R}$
The partial derivatives of the function $f(x) = (w^t x - c)$ are given by:

$$\frac{\partial f}{\partial x_i} = w_i$$

Therefore, the gradient of f is:

$$\nabla f(x) = w$$

(b) What are the partial derivatives of the function $f(x) = (w^t x - c)^2$, with w, c are as above.
The partial derivatives of the function $f(x) = (w^t x - c)^2$ are given by:

$$\frac{\partial f}{\partial x_i} = 2(w^t x - c)w_i$$

Therefore, the gradient of f is:

$$\nabla f(x) = 2(w^t x - c)w$$

(c) Prove that the expression defining $\psi(x)$ can be written as:

$$(w_1^t x - b_1)^2 + (w_2^t x - b_2)^2 + ... + (w_m^t x - b_m)^2$$

where $w_1, ..., w_m \in \mathbb{R}^n$ are the row vectors of A. Taking into account the row vs column perspective of matrix multiplication, we have:

$$Ax = \begin{bmatrix} w_1^t x \\ w_2^t x \\ \vdots \\ w_m^t x \end{bmatrix}$$

Therefore:

$$(Ax - b)^t(Ax - b) = \begin{bmatrix} w_1^t x - b_1 & w_2^t x - b_2 & ... & w_m^t x - b_m \end{bmatrix} \begin{bmatrix} w_1^t x - b_1 \\ w_2^t x - b_2 \\ \vdots \\ w_m^t x - b_m \end{bmatrix}$$

$$= (w_1^t x - b_1)^2 + (w_2^t x - b_2)^2 + ... + (w_m^t x - b_m)^2$$

(d) In view of the previous remark, can you compute the partial derivatives of $\psi(x)$?
The first order partial derivatives of $\psi(x)$ are given by:

$$\frac{\partial \psi}{\partial x_j} = 2\sum_{i=1}^{m}(w_i^t x - b_i)w_{i,j}$$

Therefore, the gradient of ψ is:

$$\nabla \psi(x) = 2\sum_{i=1}^{m}(w_i^t x - b_i)w_i = 2A^t(Ax - b)$$

(e) Prove that the jacobian matrix is given by:

$$J\psi(x) = 2(Ax - b)^t A.$$

The Jacobian matrix is the transpose of the gradient, hence:

$$J\psi(x) = (\nabla\psi(x))^t = (2A^t(Ax - b))^t = 2(Ax - b)^t A$$

(f) Compute the critical point(s) of ψ.

The critical points of ψ are found by equating the gradient (or equivalently, the Jacobian matrix) to zero:

$$\nabla\psi(x) = 2A^t(Ax - b) = 0 \implies A^tAx = A^tb$$

If A is full-rank, the critical point is given by:

$$x^* = (A^tA)^{-1}A^tb$$

Note that this is precisely the least-squares solution to the overdetermined system $Ax = b$.

(g) Do the second partial derivatives of ψ (hence the Hessian matrix) differ between points? Why?

The second order partial derivatives of $\psi(x)$ are given by:

$$\frac{\partial^2\psi}{\partial x_j \partial x_k} = 2\sum_{i=1}^{m} w_{i,j}w_{i,k}$$

Therefore, the Hessian matrix of ψ is:

$$H\psi(x) = 2\sum_{i=1}^{m} w_i w_i^t = 2A^tA$$

The second partial derivatives do not depend on the point x because they are constant with respect to x.

(h) Compute the Hessian matrix of ψ, say H. Verify that $H\psi = 2A^tA$

As derived above, the Hessian matrix of ψ is:

$$H\psi(x) = 2A^tA$$

(i) Observe that if A is full-rank, the Hessian of ψ must be symmetric, positive-definite, meaning that all its eigenvalues $\lambda_1, ..., \lambda_n > 0$. What does this tell you about the classification of the critical point(s) of ψ?

This means that the critical point(s) of ψ are local minima.

(j) Provide an expression for the local quadratic approximation of ψ at the critical point(s). Try to come up with a compact, matrix-algebra expression.

The local quadratic approximation of ψ at the critical point x^* is given by:

$$\psi(x) \approx \psi(x^*) + \frac{1}{2}(x - x^*)^t H\psi(x^*)(x - x^*)$$

Since $\nabla\psi(x^*) = 0$, the linear term vanishes, resulting in:

$$\psi(x) \approx \psi(x^*) + (x - x^*)^t A^tA(x - x^*) = \psi(x^*) + \|A(x - x^*)\|^2$$

In fact, in this case the quadratic approximation is exact, since ψ is already a quadratic function.