Norwegian University of Science and Technology
Department of Engineering Cybernetics

Exercise 0
Solution

Problem 1 (25 %) Definitions

a For a continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ the gradient, $\nabla f(\mathbf{x})$, is

$$\nabla f(\mathbf{x}) = \left[\frac{\partial f}{\partial \mathbf{x}}\right]^\top = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} \tag{1}$$

b For a continuously differentiable function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ the Jacobian, $\frac{\partial \mathbf{f}}{\partial \mathbf{x}}$, is

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \tag{2}$$

c Using the definition, we see that it will be a column vector of length n.
A more intuitive approach: $f(\mathbf{x})$ is a function of $n(=\text{length}(\mathbf{x}))$ variables. The gradient tells us how the function is changing with respect to (w.r.t.) infinitesimally small changes in the different variables.
That means that the gradient must be a vector of length n.

d Using the definition, we see that it will be a matrix of size $m \times n$.

Problem 2 (25 %) Linear

a First calculate:

$$\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}$$

Then find:

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial \mathbf{f}}{\partial x_1} & \frac{\partial \mathbf{f}}{\partial x_2} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

Simplify to matrix form:

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \mathbf{A}$$

This is the Jacobian of $\mathbf{f}(\mathbf{x})$. $\mathbf{f}(\mathbf{x})$ is a vector of two elements.

b Utilizing the answer in **a**):

$$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}$$

Problem 3 (25 %) Nonlinear

a To easily see the resulting dimension of the matrix multiplication, we can do the following. First, write the dimension of the matrices/vectors after each other:

$$\begin{matrix} \mathbf{x}^T & \mathbf{G} & \mathbf{y} \\ (1\times 2)\cdot & (2\times 3)\cdot & (3\times 1) \end{matrix}$$

Second, to make sure there are no invalid multiplications: Wherever you see "$\ldots \times i)\cdot(j \times \ldots$", make sure $i = j$. If not: the multiplication cannot be done. Third, the resulting dimension will be the first and the last number. In our case, that will be: 1×1, which means $f(\mathbf{x}, \mathbf{y})$ is a scalar.

Note, this could also be used to, e.g., find the dimension of $\mathbf{x}^T\mathbf{G}$:
$(1 \times 2) \cdot (2 \times 3) \to (1 \times 3)$

Is $\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y})$ equal to $\dfrac{\partial f(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}}$? No, there is missing a transpose. Correct:
$\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y})^T = \dfrac{\partial f(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}}$

b First calculate:

$$\begin{aligned} f(\mathbf{x}, \mathbf{y}) &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \\ &= \begin{bmatrix} x_1 g_{11} + x_2 g_{21} & x_1 g_{12} + x_2 g_{22} & x_1 g_{13} + x_2 g_{23} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \\ &= x_1 y_1 g_{11} + x_2 y_1 g_{21} + x_1 y_2 g_{12} + x_2 y_2 g_{22} + x_1 y_3 g_{13} + x_2 y_3 g_{23} \end{aligned}$$

Then find:

$$\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}) = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} y_1 g_{11} + y_2 g_{12} + y_3 g_{13} \\ y_1 g_{21} + y_2 g_{22} + y_3 g_{23} \end{bmatrix}$$

Simplify to matrix form:

$$\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}) = \nabla_{\mathbf{x}}(\mathbf{x}^T \mathbf{G} \mathbf{y}) = \begin{bmatrix} y_1 g_{11} + y_2 g_{12} + y_3 g_{13} \\ y_1 g_{21} + y_2 g_{22} + y_3 g_{23} \end{bmatrix} = \mathbf{G}\mathbf{y}$$

c First calculate (taken from **a**)):

$$f(\mathbf{x}, \mathbf{y}) = x_1 y_1 g_{11} + x_2 y_1 g_{21} + x_1 y_2 g_{12} + x_2 y_2 g_{22} + x_1 y_3 g_{13} + x_2 y_3 g_{23}$$

Then find:

$$\nabla_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}) = \begin{bmatrix} \frac{\partial f}{\partial y_1} \\ \frac{\partial f}{\partial y_2} \\ \frac{\partial f}{\partial y_3} \end{bmatrix} = \begin{bmatrix} x_1 g_{11} + x_2 g_{21} \\ x_1 g_{12} + x_2 g_{22} \\ x_1 g_{13} + x_2 g_{23} \end{bmatrix}$$

Simplify to matrix form:

$$\nabla_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}) = \nabla_{\mathbf{y}}(\mathbf{x}^T \mathbf{G} \mathbf{y}) = \begin{bmatrix} x_1 g_{11} + x_2 g_{21} \\ x_1 g_{12} + x_2 g_{22} \\ x_1 g_{13} + x_2 g_{23} \end{bmatrix} = \mathbf{G}^T \mathbf{x}$$

d Here we must use the "product rule":

$$\begin{aligned} \nabla_{\mathbf{x}} f(\mathbf{x}) &= \nabla_{\mathbf{x}}(\mathbf{x}^T \mathbf{H} \mathbf{x}) \\ &= \underbrace{\mathbf{H}\mathbf{x}}_{\substack{\text{From differentiating} \\ \text{w.r.t. the first } \mathbf{x}. \\ \text{As we did in } \mathbf{b})}} + \underbrace{\mathbf{H}^T \mathbf{x}}_{\substack{\text{From differentiating} \\ \text{w.r.t. the last } \mathbf{x}. \\ \text{As we did in } \mathbf{c})}} \end{aligned}$$

If $\mathbf{H}$ is symmetric, then:

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = \mathbf{H}\mathbf{x} + \mathbf{H}^T \mathbf{x} = \mathbf{H}\mathbf{x} + \mathbf{H}\mathbf{x} = 2\mathbf{H}\mathbf{x}$$

Problem 4 (25 %) Common case

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{x}^T \mathbf{G} \mathbf{x} + \boldsymbol{\lambda}^T (\mathbf{C}\mathbf{x} - \mathbf{d}) + \boldsymbol{\mu}^T (\mathbf{E}\mathbf{x} - \mathbf{h})$$

a

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \underbrace{2\mathbf{G}\mathbf{x}}_{\text{See 3}\mathbf{d})} + \underbrace{\mathbf{C}^T \boldsymbol{\lambda}}_{\text{See 3}\mathbf{c})} + \underbrace{\mathbf{E}^T \boldsymbol{\mu}}_{\text{See 3}\mathbf{c})}$$

b

$$\nabla_{\boldsymbol{\mu}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \underbrace{\mathbf{E}\mathbf{x} - \mathbf{h}}_{\text{See 3}\mathbf{b})}$$

c

$$\nabla_{\boldsymbol{\lambda}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \underbrace{\mathbf{C}\mathbf{x} - \mathbf{d}}_{\text{See 3}\mathbf{b})}$$