Optimalisering og regulering TTK4135 - Assignment 0

Ola Kristoffer Hoff

13. January 2022

Problem 1: Definitions

a What is the definition of the gradient of a function?

The definition of a gradient is the vector of the partial derivatives of the function f.
given $f(x,y) = x^3 - xy^2 + 2$

$$\nabla f(x,y) = \begin{bmatrix} 3x^2 - y^2 \\ -2xy \end{bmatrix} \tag{1}$$

or in more general terms:

$$\nabla f(x_0, x_1, ...) = \begin{bmatrix} \frac{\partial f}{\partial x_0} \\ \frac{\partial f}{\partial x_1} \\ \vdots \end{bmatrix} \tag{2}$$

b What is the definition of the Jacobian of a function?

The Jacobian matrix, J, of a function f is the gradient of the functions of f.

What this means is that we have a column vector of all the functions of f that we take the gradient of, making row vectors, this results in a matrix. Given m functions in f and n variables in f.

$$J = \begin{bmatrix} \nabla f_0 \\ \nabla f_1 \\ \vdots \\ \nabla f_m \end{bmatrix} = \begin{bmatrix} \frac{\partial f_0}{\partial x_0} & \cdots & \frac{\partial f_0}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_0} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \tag{3}$$

An example:

$$f(x,y) = \begin{bmatrix} f_0 \\ f_1 \end{bmatrix} = \begin{bmatrix} x^3 - 2xy + y^2 + 1 \\ x + y^2 - 1 \end{bmatrix} \tag{4}$$

$$J(x,y) = \begin{bmatrix} 3x^2 - 2y & 2y - 2x \\ 1 & 2y \end{bmatrix} \tag{5}$$

c Let $f(x)$ be a scalar and $x \in \mathbb{R}^n$. What will the size of the gradient be?

The size of the gradient is determined by the number of variables in f, hence the dimension of the x vector. Here $x \in \mathbb{R}^n$ gives us the answer: n.

d Let $f(x)$ be a column vector of length m and $x \in \mathbb{R}^n$. What will the size of the Jacobian be?

Using the same logic as above and the statements about the Jacobian matrix we see that we get a $m \times n$ matrix.

Problem 2: Linear

Let $f(x) = Ax$, where

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{6}$$

a Use the definition and calculate $\frac{\partial f(x)}{\partial x}$. After calculating it, simplify it into a matrix form. Is this the Jacobian or the gradient of $f(x)$?

The common matrix notation $Ax = b$ is in use here, this commonly describes a linear system of equations. So $f(x)$ is a vector of equations, but written in Ax matrix format. This means that when we take the partial derivatives of it we get the Jacobian matrix, which in this case is equal to A. Given that x is a vector of first order variables.

$$f(x) = Ax = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix} \tag{7}$$

$$\frac{\partial f(x)}{\partial x} = \begin{bmatrix} \frac{a_{11}x_1+a_{12}x_2}{\partial x} \\ \frac{a_{21}x_1+a_{22}x_2}{\partial x} \end{bmatrix} = \begin{bmatrix} \frac{a_{11}x_1+a_{12}x_2}{\partial x_1} & \frac{a_{11}x_1+a_{12}x_2}{\partial x_2} \\ \frac{a_{21}x_1+a_{22}x_2}{\partial x_1} & \frac{a_{21}x_1+a_{22}x_2}{\partial x_2} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = A \tag{8}$$

b Can you, without doing any calculations, find $\frac{\partial Ax}{\partial x}$ when x is a column vector of length n, and A is a matrix of dimension $m \times n$ (i.e., m rows and n columns)?

This is the same thing as we did above. We have substituted the $f(x)$ with is equality Ax, hence the answer will be the same. And we can also see the general case of the answer being A as shown above.

Problem 3: Nonlinear

Let $f(x, y) = x^T G y$, where

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, G = \begin{bmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}. \tag{9}$$

a What is the dimension of $f(x, y)$? Is $\nabla_x f(x, y)$ equal to $\frac{\partial f(x,y)}{\partial x}$ (no calculations are needed)?

Since a $m \times n$ matrix multiplied with a $i \times j$ matrix creates a $m \times j$. This gives us that $x^T G$ becomes a 1×3 and lastly with y it becomes a 1×1, hence a singular value.

Yes, $\nabla_x f(x, y) = \frac{\partial f(x,y)}{\partial x}$ is correct, both are derivations of f with respect to x and give the answer Gy.

b Use the definition and calculate $\nabla_x f(x, y)$. Then write the answer in matrix form.

Given the general form:

$$\nabla f(x, y, z, \cdots) = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} & \cdots \end{bmatrix}^T \tag{10}$$

We have that:

$$\nabla_x f(x, y, z, \cdots) = \begin{bmatrix} \frac{\partial f}{\partial x} \end{bmatrix}^T \tag{11}$$

Given the definition of a partial derivative with respect to vector x.

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f_0}{\partial x_0} & \cdots & \frac{\partial f_0}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_0} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \tag{12}$$

We get that:

$$\nabla_x f(x, y, z, \cdots) = \left[Gy\right]^T = \left[Gy\right] = Gy = \begin{bmatrix} y_1 g_{11} + y_2 g_{12} + y_3 g_{13} \\ y_1 g_{21} + y_2 g_{22} + y_3 g_{23} \end{bmatrix} \tag{13}$$

c Use the definition and calculate $\nabla_y f(x, y)$. Then write the answer in matrix form.

Following the same logic as above we get:

$$\nabla_y f(x, y) = \left[x^T G\right] = x^T G = (x^T G)^T = G^T x \tag{14}$$

d Let $f(x) = x^T Hx$, where $x \in \mathbb{R}^n$ and $H \in \mathbb{R}^{n \times n}$. Find $\nabla f(x)$ using the results from the previous exercises. What will the answer be if H is symmetric?

To find the gradient we simply find the partial derivatives of x and y as we have done before and combine the answers.

$$\nabla f(x, y) = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = Hx + H^T x \tag{15}$$

If H is symmetric then we have: $H = H^T$:

$$Hx + H^T x => 2Hx \tag{16}$$

Problem 4: Common case

$\mathcal{L}(x, \lambda, \mu) = x^T Gx + \lambda^T (Cx - d) + \mu^T (Ex - h)$

a Find $\nabla_x \mathcal{L}(x, \lambda, \mu)$.

$$\nabla_x \mathcal{L}(x, \lambda, \mu) = Gx + G^T x + \lambda^T C + \mu^T E \tag{17}$$

b Find $\nabla_\lambda \mathcal{L}(x, \lambda, \mu)$.

$$\nabla_\lambda \mathcal{L}(x, \lambda, \mu) = Cx - d \tag{18}$$

c Find $\nabla_\mu \mathcal{L}(x, \lambda, \mu)$.

$$\nabla_\mu \mathcal{L}(x, \lambda, \mu) = Ex - h \tag{19}$$