

# Continuous Optimization: Assignment 7

Due on June 11, 2024

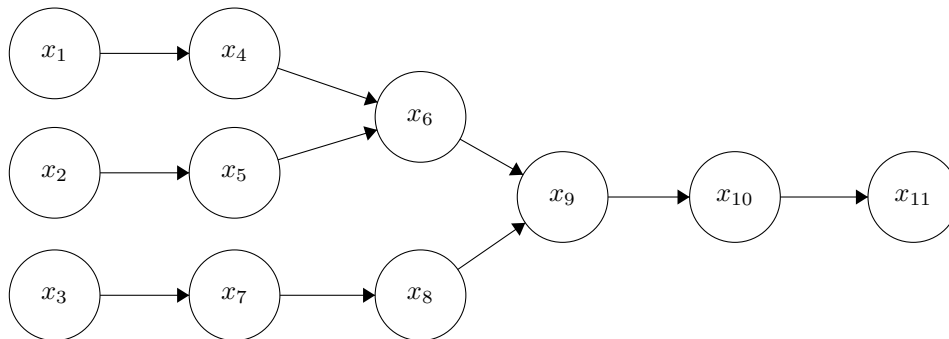
Honglu Ma

Hiroyasu Akada

Mathivathana Ayyappan

## Exercise 1

The computational graph for function  $f$  is expressed as such



where

$$\begin{aligned}
 x_4 &= x_1^2 \\
 x_5 &= x_2^3 \\
 x_7 &= x_3^4 \\
 x_6 &= x_4 \cdot x_5 \\
 x_8 &= \sin(x_7) \\
 x_9 &= x_6 + x_8 \\
 x_{10} &= \exp(x_9) \\
 x_{11} &= x_{10}^2
 \end{aligned}$$

and we have the following derivatives

$$\begin{aligned}
 \frac{\partial x_4}{\partial x_1} &= 2x_1 \\
 \frac{\partial x_5}{\partial x_2} &= 3x_2^2 \\
 \frac{\partial x_7}{\partial x_3} &= 4x_3^3 \\
 \frac{\partial x_6}{\partial x_4} &= x_5 \\
 \frac{\partial x_6}{\partial x_5} &= x_4 \\
 \frac{\partial x_8}{\partial x_7} &= \cos(x_7) \\
 \frac{\partial x_9}{\partial x_6} &= 1 \\
 \frac{\partial x_9}{\partial x_8} &= 1 \\
 \frac{\partial x_{10}}{\partial x_9} &= \exp(x_9) \\
 \frac{\partial x_{11}}{\partial x_{10}} &= 2x_{10}
 \end{aligned}$$

## Forward Mode

We propagate the tangents through the computational graph to compute the derivative of  $f$  with respect to  $x_1$ ,  $x_2$  and  $x_3$ . Normally, we would have to propagate all the bases i.e.  $\hat{i}$ ,  $\hat{j}$  and  $\hat{k}$  one after another in order to all the partial derivatives but here each operation of the computational graph is a scalar operation and the starting node  $x_1$ ,  $x_2$  and  $x_3$  are also scalas so we can simply set  $\dot{x}_1 = 1$  when calculating  $\frac{\partial f}{\partial x_1}$  and so on. To calculate the partial derivatives at point  $(x_1, x_2, x_3) = (\tilde{x}_1, \tilde{x}_2, \tilde{x}_3)$ , we have

(i)  $\frac{\partial f}{\partial x_1}$  Set  $\dot{x}_1 = 1, \dot{x}_2 = 0, \dot{x}_3 = 0$

$x_1 = \tilde{x}_1$	$\dot{x}_1 = 1$
$x_2 = \tilde{x}_2$	$\dot{x}_2 = 0$
$x_3 = \tilde{x}_3$	$\dot{x}_3 = 0$
$x_4 = x_1^2 = \tilde{x}_1^2$	$\dot{x}_4 = 2x_1\dot{x}_1 = 2\tilde{x}_1\dot{x}_1 = 2\tilde{x}_1$
$x_5 = x_2^3 = \tilde{x}_2^3$	$\dot{x}_5 = 3x_2^2\dot{x}_2 = 0$
$x_6 = x_4 \cdot x_5 = \tilde{x}_1^2 \cdot \tilde{x}_2^3$	$\dot{x}_6 = x_5\dot{x}_4 + x_4\dot{x}_5 = 2\tilde{x}_1\tilde{x}_2^3$
$x_7 = x_3^4 = \tilde{x}_3^4$	$\dot{x}_7 = 4x_3^3\dot{x}_3 = 0$
$x_8 = \sin(x_7) = \sin(\tilde{x}_3^4)$	$\dot{x}_8 = \cos(x_7)\dot{x}_7 = 0$
$x_9 = x_6 + x_8 = \tilde{x}_1^2 \cdot \tilde{x}_2^3 + \sin(\tilde{x}_3^4)$	$\dot{x}_9 = \dot{x}_6 + \dot{x}_8 = 2\tilde{x}_1\tilde{x}_2^3$
$x_{10} = \exp(x_9) = \exp(\tilde{x}_1^2 \cdot \tilde{x}_2^3 + \sin(\tilde{x}_3^4))$	$\dot{x}_{10} = \exp(x_9)\dot{x}_9 = 2\exp(\tilde{x}_1^2 \cdot \tilde{x}_2^3 + \sin(\tilde{x}_3^4))\tilde{x}_1\tilde{x}_2^3$
$x_{11} = x_{10}^2 = \exp(2(\tilde{x}_1^2 \cdot \tilde{x}_2^3 + \sin(\tilde{x}_3^4)))$	$\dot{x}_{11} = 2x_{10}\dot{x}_{10} = 4\exp(2(\tilde{x}_1^2 \cdot \tilde{x}_2^3 + \sin(\tilde{x}_3^4)))\tilde{x}_1\tilde{x}_2^3$

(ii)  $\frac{\partial f}{\partial x_2}$  Set  $\dot{x}_1 = 0, \dot{x}_2 = 1, \dot{x}_3 = 0$ . Repeat the same process as above we get

$$\frac{\partial f}{\partial x_2} = 6\tilde{x}_1^2\tilde{x}_2^2 \exp(2(\tilde{x}_1^2 \cdot \tilde{x}_2^3 + \sin(\tilde{x}_3^4)))$$

(iii)  $\frac{\partial f}{\partial x_3}$  Set  $\dot{x}_1 = 0, \dot{x}_2 = 0, \dot{x}_3 = 1$ . Repeat the same process as above we get

$$\frac{\partial f}{\partial x_3} = 8\tilde{x}_3^3 \cos(\tilde{x}_3^4) \exp(2(\tilde{x}_1^2 \cdot \tilde{x}_2^3 + \sin(\tilde{x}_3^4)))$$

## Backward Mode

We propagate normal vector of the hyperplane defined by  $\langle \bar{y}, \nabla F(x) \rangle = c$  backward through the computational graph where  $\bar{y}$  is the desired variation of the function  $F$ . For each partial derivatives, we set the normal vector to be the corresponding basis vector, i.e.  $\bar{y} = \hat{i}$  for  $\frac{\partial f}{\partial x_1}$  and so on.

(i)  $\frac{\partial f}{\partial x_1}$ 

$$\begin{aligned}
x_1 &= \tilde{x}_1 \\
x_2 &= \tilde{x}_2 \\
x_3 &= \tilde{x}_3 \\
x_4 &= \tilde{x}_1^2 \\
x_5 &= \tilde{x}_2^3 \\
x_6 &= \tilde{x}_1^2 \cdot \tilde{x}_2^3 \\
x_7 &= \tilde{x}_3^4 \\
x_8 &= \sin(\tilde{x}_3^4) \\
x_9 &= \tilde{x}_1^2 \cdot \tilde{x}_2^3 + \sin(\tilde{x}_3^4) \\
x_{10} &= \exp(\tilde{x}_1^2 \cdot \tilde{x}_2^3 + \sin(\tilde{x}_3^4)) \\
x_{11} &= \exp(2(\tilde{x}_1^2 \cdot \tilde{x}_2^3 + \sin(\tilde{x}_3^4)))
\end{aligned}$$

$$\begin{aligned}
\bar{x}_{11} &= 1 \\
\bar{x}_{10} &= 2 \exp(\tilde{x}_1^2 \cdot \tilde{x}_2^3 + \sin(\tilde{x}_3^4)) \\
\bar{x}_9 &= 2 \exp(2\tilde{x}_1^2 \cdot \tilde{x}_2^3 + \sin(\tilde{x}_3^4)) \\
\bar{x}_6 &= 2 \exp(2\tilde{x}_1^2 \cdot \tilde{x}_2^3 + \sin(\tilde{x}_3^4)) \\
\bar{x}_4 &= 2\tilde{x}_2^3 \exp(2\tilde{x}_1^2 \cdot \tilde{x}_2^3 + \sin(\tilde{x}_3^4)) \\
\bar{x}_1 &= 4\tilde{x}_1\tilde{x}_2^3 \exp(2\tilde{x}_1^2 \cdot \tilde{x}_2^3 + \sin(\tilde{x}_3^4))
\end{aligned}$$

Similarly we can calculate the other partial derivatives. The result is the same as the forward mode.

## Exercise 2

(a)

We construct a matrix  $\tilde{P}_j$  for each column  $j$  of  $P$  such that  $\tilde{P}_j \in \mathbb{R}^{2 \times 6}$

$$\tilde{P}_j := \begin{pmatrix} P_{1j} & 0 & P_{2j} & 0 & 1 & 0 \\ 0 & P_{1j} & 0 & P_{2j} & 0 & 1 \end{pmatrix}$$

and we have  $\tilde{P} \in \mathbb{R}^{2n \times 6}$  as following

$$\tilde{P} := \begin{pmatrix} \tilde{P}_1 \\ \vdots \\ \tilde{P}_n \end{pmatrix}$$

At last, we construct  $\tilde{Q} \in \mathbb{R}^{2n \times 1}$ :

$$\tilde{Q} := \begin{pmatrix} Q_1 \\ \vdots \\ Q_n \end{pmatrix}$$

where  $Q_j$  denotes the  $j$ -th column of  $Q$ .

(b)

The objective function has to be linear as an assumption of Gauss-Newton Method. Thus we have to approximate  $A$  with its first order Taylor expansion. The linear least squares problem can be formulated as

$$\min_{x \in \mathbb{R}^3} \frac{1}{2} \|F(P, Q)\|^2$$

where

$$x := \begin{pmatrix} t_1 \\ t_2 \\ \theta \end{pmatrix} \quad F(P, Q) := \tilde{P}A(x) - \tilde{Q}$$

with  $\tilde{P}$  and  $\tilde{Q}$  defined in (a).

The linear system of equations that need to be solved is

$$\nabla F(x^{(k)}) \nabla F(x^{(k)})^\top d^{(k)} = -\nabla F(x^{(k)}) F(x^{(k)})$$

in order to calculate the descent direction  $d^{(k)}$ .  $\nabla F(x^{(k)})$  can be calculated analytically. It can also be approximated by finite difference.