Appendix A

Coefficients of the Phase Function

	$g = 0$	$g = 0.8$	$g = 0.85$	$g = 0.86$	$g = 0.9$
$a_1\ \tilde{p}_{1\ 1}$	0.0625	0.65172	0.76432	0.78760	0.87800
$a_2\ \tilde{p}_{1\ 2}$	0.0625	0.13721	0.09725	0.08832	0.05204
$a_3\ \tilde{p}_{1\ 3}$	0.0625	0.02113	0.01201	0.01046	0.00529
$a_4\ \tilde{p}_{1\ 4}$	0.0625	0.00696	0.00382	0.00330	0.00164
$a_5\ \tilde{p}_{1\ 5}$	0.0625	0.00338	0.00183	0.00158	0.00078
$a_6\ \tilde{p}_{1\ 6}$	0.0625	0.00208	0.00112	0.00097	0.00048
$a_7\ \tilde{p}_{1\ 7}$	0.0625	0.00152	0.00081	0.00070	0.00035
$a_8\ \tilde{p}_{1\ 8}$	0.0625	0.00127	0.00068	0.00059	0.00029
$a_9\ \tilde{p}_{1\ 9}$	0.0625	0.00120	0.00064	0.00055	0.00027
$a_{10}\ \tilde{p}_{1\ 10}$	0.0625	0.00127	0.00068	0.00059	0.00029
$a_{11}\ \tilde{p}_{1\ 11}$	0.0625	0.00152	0.00081	0.00070	0.00035
$a_{12}\ \tilde{p}_{1\ 12}$	0.0625	0.00208	0.00112	0.00097	0.00048
$a_{13}\ \tilde{p}_{1\ 13}$	0.0625	0.00338	0.00183	0.00158	0.00078
$a_{14}\ \tilde{p}_{1\ 14}$	0.0625	0.00696	0.00382	0.00330	0.00164
$a_{15}\ \tilde{p}_{1\ 15}$	0.0625	0.02113	0.01201	0.01046	0.00529
$a_{16}\ \tilde{p}_{1\ 16}$	0.0625	0.13721	0.09725	0.08832	0.05204
$\sum_{k'} a_{k'}\ \tilde{p}_{1k'} \approx$	1	1	1	1	1

Table A.1: Coefficients $a_{k'}\tilde{p}_{k=1\ k'}$ for K = 16 ordinates and different anisotropy factors g.

	$g = 0$	$g = 0.8$
$a_1\ \tilde{p}_{1\ 1}$	0.03125	0.39695
$a_2\ \tilde{p}_{1\ 2}$	0.03125	0.19511
$a_3\ \tilde{p}_{1\ 3}$	0.03125	0.05682
$a_4\ \tilde{p}_{1\ 4}$	0.03125	0.02094
$a_5\ \tilde{p}_{1\ 5}$	0.03125	0.00980
$a_6\ \tilde{p}_{1\ 6}$	0.03125	0.00541
$a_7\ \tilde{p}_{1\ 7}$	0.03125	0.00337
$a_8\ \tilde{p}_{1\ 8}$	0.03125	0.00228
$a_9\ \tilde{p}_{1\ 9}$	0.03125	0.00166
$a_{10}\ \tilde{p}_{1\ 10}$	0.03125	0.00128
$a_{11}\ \tilde{p}_{1\ 11}$	0.03125	0.00103
$a_{12}\ \tilde{p}_{1\ 12}$	0.03125	0.00086
$a_{13}\ \tilde{p}_{1\ 13}$	0.03125	0.00075
$a_{14}\ \tilde{p}_{1\ 14}$	0.03125	0.00068
$a_{15}\ \tilde{p}_{1\ 15}$	0.03125	0.00063
$a_{16}\ \tilde{p}_{1\ 16}$	0.03125	0.00060
$a_{17}\ \tilde{p}_{1\ 17}$	0.03125	0.00060
$a_{18}\ \tilde{p}_{1\ 18}$	0.03125	0.00060
$a_{19}\ \tilde{p}_{1\ 19}$	0.03125	0.00063
$a_{20}\ \tilde{p}_{1\ 20}$	0.03125	0.00068
$a_{21}\ \tilde{p}_{1\ 21}$	0.03125	0.00075
$a_{22}\ \tilde{p}_{1\ 22}$	0.03125	0.00086
$a_{23}\ \tilde{p}_{1\ 23}$	0.03125	0.00103
$a_{24}\ \tilde{p}_{1\ 24}$	0.03125	0.00128
$a_{25}\ \tilde{p}_{1\ 25}$	0.03125	0.00166
$a_{26}\ \tilde{p}_{1\ 26}$	0.03125	0.00228
$a_{27}\ \tilde{p}_{1\ 27}$	0.03125	0.00337
$a_{28}\ \tilde{p}_{1\ 28}$	0.03125	0.00541
$a_{29}\ \tilde{p}_{1\ 29}$	0.03125	0.00980
$a_{30}\ \tilde{p}_{1\ 30}$	0.03125	0.02094
$a_{31}\ \tilde{p}_{1\ 31}$	0.03125	0.05682
$a_{32}\ \tilde{p}_{1\ 32}$	0.03125	0.19511
$\sum_{k'} a_{k'}\ \tilde{p}_{1k'} \approx$	1	1

Table A.2: Coefficients $a_{k'}\tilde{p}_{k=1\ k'}$ for $K = 32$ ordinates and different anisotropy factors g.

Appendix B

Example: Differentiation of a Composite Function

We demonstrate on an example how a function $f(\boldsymbol{x})$ that is composed of sub-functions is differentiated by using the method of differentiation of algorithms. First, we derive an expression for the total derivative, which is a composition of the derivatives of the sub-functions. Second, we calculate a value of the derivative by using both the forward mode and the reverse mode of differentiation of algorithms.

We define a function f^1 that maps the variable $\boldsymbol{x}$ onto $\boldsymbol{y}^1$:

$$\begin{aligned} f^1 &: \mathbb{R}^{\mathrm{m}} \to \mathbb{R}^{\mathrm{m}_1} \\ \boldsymbol{x} &\mapsto \boldsymbol{y}^1 = f^1(\boldsymbol{x}). \end{aligned} \tag{B.1}$$

The output $\boldsymbol{y}^1$ of function f^1 becomes an input variable of function f^2 that is also explicitly dependent on $\boldsymbol{x}$:

$$\begin{aligned} f^2 &: \mathbb{R}^{\mathrm{m}_1} \times \mathbb{R}^{\mathrm{m}} \to \mathbb{R}^{\mathrm{m}_2} \\ \begin{pmatrix} \boldsymbol{y}^1 \\ \boldsymbol{x} \end{pmatrix} &\mapsto \boldsymbol{y}^2 = f^2(\boldsymbol{y}^1, \boldsymbol{x}). \end{aligned} \tag{B.2}$$

A third function f^3 maps the output $\boldsymbol{y}^2$ and the input variable $\boldsymbol{x}$ onto $\boldsymbol{y}^3$:

$$\begin{aligned} f^3 &: \mathbb{R}^{\mathrm{m}_2} \times \mathbb{R}^{\mathrm{m}} \to \mathbb{R}^{\mathrm{m}_3} \\ \begin{pmatrix} \boldsymbol{y}^2 \\ \boldsymbol{x} \end{pmatrix} &\mapsto \boldsymbol{y}^3 = f^3(\boldsymbol{y}^2, \boldsymbol{x}). \end{aligned} \tag{B.3}$$

Finally, function f^4 maps the output $\boldsymbol{y}^3$ of f^3 and the input variable $\boldsymbol{x}$ onto $\boldsymbol{y}$

$$\begin{aligned} f^4 &: \mathbb{R}^{\mathrm{m}_3} \times \mathbb{R}^{\mathrm{m}} \to \mathbb{R}^{\mathrm{n}} \\ \begin{pmatrix} \boldsymbol{y}^3 \\ \boldsymbol{x} \end{pmatrix} &\mapsto \boldsymbol{y} = f^4(\boldsymbol{y}^3, \boldsymbol{x}). \end{aligned} \tag{B.4}$$

Consequently, the function f can also be written as a composite function:

$$\begin{aligned} f &: \mathbb{R}^{\mathrm{m}} \to \mathbb{R}^{\mathrm{n}} \\ \boldsymbol{x} \mapsto \boldsymbol{y} = f(\boldsymbol{x}) &= (f^4(\boldsymbol{x}) \circ f^3(\boldsymbol{x}) \circ f^2(\boldsymbol{x}) \circ f^1)(\boldsymbol{x}). \end{aligned} \tag{B.5}$$

The computational graph is shown in Figure B.1. A function value $\boldsymbol{y} = f(\boldsymbol{x})$ is evaluated by stepping through the graph from the left to right.

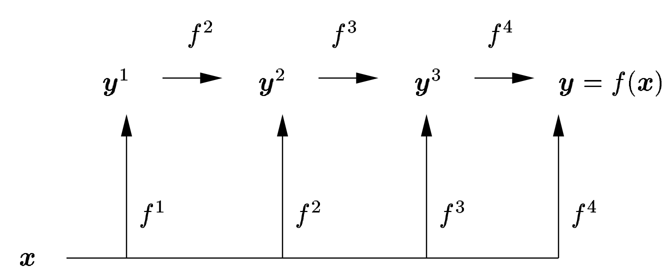

Figure B.1: Computational graph of the composite function f.

Example:

$f^1 = 2x$

$f^2 = x + f^1 \longrightarrow 3x$

$f^3 = 3x + 2f^2 \longrightarrow 5x + 2f^1 \longrightarrow 9x$

$f^4 = x^2 + f^3 \longrightarrow x^2 + 3x + 2f^2 \longrightarrow x^2 + 5x + 2f^1 \longrightarrow x^2 + 9x$

$f = x^2 + 9x$

B.1 Total Derivative

An expression for the total derivative $\frac{df}{d\boldsymbol{x}}$ is obtained by systematically applying the chain rule of differentiation to the composite function $f(\boldsymbol{x})$. We obtain the final result

$$\begin{aligned} \frac{df(\boldsymbol{x})}{d\boldsymbol{x}} &= \frac{\partial(f^4 \circ f^3 \circ f^2 \circ f^1)(\boldsymbol{x})}{\partial \boldsymbol{x}} \\ &\quad + \frac{\partial(f^4 \circ f^3 \circ f^2)(\boldsymbol{x})}{\partial \boldsymbol{x}} \\ &\quad + \frac{\partial(f^4 \circ f^3)(\boldsymbol{x})}{\partial \boldsymbol{x}} \\ &\quad + \frac{\partial f^4(\boldsymbol{x})}{\partial \boldsymbol{x}}, \end{aligned} \tag{B.6}$$

or more specifically:

$$\begin{aligned} \frac{df(\boldsymbol{x})}{d\boldsymbol{x}} &= \frac{\partial f^4(\boldsymbol{y}^3)}{\partial \boldsymbol{y}^3} \frac{\partial f^3(\boldsymbol{y}^2)}{\partial \boldsymbol{y}^2} \frac{\partial f^2(\boldsymbol{y}^1)}{\partial \boldsymbol{y}^1} \frac{\partial f^1(\boldsymbol{x})}{\partial \boldsymbol{x}} \\ &\quad + \frac{\partial f^4(\boldsymbol{y}^3)}{\partial \boldsymbol{y}^3} \frac{\partial f^3(\boldsymbol{y}^2)}{\partial \boldsymbol{y}^2} \frac{\partial f^2(\boldsymbol{x})}{\partial \boldsymbol{x}} \\ &\quad + \frac{\partial f^4(\boldsymbol{y}^3)}{\partial \boldsymbol{y}^3} \frac{\partial f^3(\boldsymbol{x})}{\partial \boldsymbol{x}} \\ &\quad + \frac{\partial f^4(\boldsymbol{x})}{\partial \boldsymbol{x}}. \end{aligned} \tag{B.7}$$

The transpose $\frac{df}{dx}^T$ of $\frac{df}{dx}$ is obtained by using the equality $(\boldsymbol{A} \cdot \boldsymbol{B} \cdot \boldsymbol{C})^T = \boldsymbol{C}^T \cdot \boldsymbol{B}^T \cdot \boldsymbol{A}^T$. We get from Equation B.6:

$$\begin{aligned} \frac{df(\boldsymbol{x})}{d\boldsymbol{x}}^T &= \frac{\partial(f^4 \circ f^3 \circ f^2 \circ f^1)(\boldsymbol{x})}{\partial \boldsymbol{x}}^T \\ &+ \frac{\partial(f^4 \circ f^3 \circ f^2)(\boldsymbol{x})}{\partial \boldsymbol{x}}^T \\ &+ \frac{\partial(f^4 \circ f^3)(\boldsymbol{x})}{\partial \boldsymbol{x}}^T \\ &+ \frac{\partial f^4(\boldsymbol{x})}{\partial \boldsymbol{x}}^T, \end{aligned} \tag{B.8}$$

or

$$\begin{aligned} \frac{df(\boldsymbol{x})}{d\boldsymbol{x}}^T &= \frac{\partial f^1(\boldsymbol{x})}{\partial \boldsymbol{x}}^T \frac{\partial f^2(\boldsymbol{y}^1)}{\partial \boldsymbol{y}^1}^T \frac{\partial f^3(\boldsymbol{y}^2)}{\partial \boldsymbol{y}^2}^T \frac{\partial f^4(\boldsymbol{y}^3)}{\partial \boldsymbol{y}^3}^T \\ &+ \frac{\partial f^2(\boldsymbol{x})}{\partial \boldsymbol{x}}^T \frac{\partial f^3(\boldsymbol{y}^2)}{\partial \boldsymbol{y}^2}^T \frac{\partial f^4(\boldsymbol{y}^3)}{\partial \boldsymbol{y}^3}^T \\ &+ \frac{\partial f^3(\boldsymbol{x})}{\partial \boldsymbol{x}}^T \frac{\partial f^4(\boldsymbol{y}^3)}{\partial \boldsymbol{y}^3}^T \\ &+ \frac{\partial f^4(\boldsymbol{x})}{\partial \boldsymbol{x}}^T. \end{aligned} \tag{B.9}$$

There are two different ways of computing a value of the total derivative. We call them the forward mode of differentiation for calculating $\frac{df}{dx}$ and the reverse mode of differentiation for calculating $\frac{df}{dx}^T$.

B.2 Forward Mode of Differentiation

The total derivative $\frac{df}{dx}$ (see Equation B.7) is computed in the same order as the sub-functions f^1, f^2, f^3, and f^4 are evaluated (see Equations B.1-B.4). First, we start by calculating the partial derivative $\frac{\partial f^1(\boldsymbol{x})}{\partial \boldsymbol{x}}$ of the first sub-function f^1. The remaining steps of the forward mode involve computing the remaining derivatives by making use of

Equation 6.21 until we arrive at $\frac{\partial(f^4 \circ f^3 \circ f^2 \circ f^1)(\boldsymbol{x})}{\partial \boldsymbol{x}}$:

$$\frac{\partial(f^2 \circ f^1)(\boldsymbol{x})}{\partial \boldsymbol{x}} = \frac{\partial f^2(\boldsymbol{y}^1)}{\partial \boldsymbol{y}^1} \frac{\partial f^1(\boldsymbol{x})}{\partial \boldsymbol{x}} \tag{B.10}$$

$$\frac{\partial(f^3 \circ f^2 \circ f^1)(\boldsymbol{x})}{\partial \boldsymbol{x}} = \frac{\partial f^3(\boldsymbol{y}^2)}{\partial \boldsymbol{y}^2} \frac{\partial(f^2 \circ f^1)(\boldsymbol{x})}{\partial \boldsymbol{x}} \tag{B.11}$$

$$\frac{\partial(f^4 \circ f^3 \circ f^2 \circ f^1)(\boldsymbol{x})}{\partial \boldsymbol{x}} = \frac{\partial f(\boldsymbol{y}^3)}{\partial \boldsymbol{y}^3} \frac{\partial(f^3 \circ f^2 \circ f^1)(\boldsymbol{x})}{\partial \boldsymbol{x}}. \tag{B.12}$$

Equations B.10 to B.12 yield

$$\frac{\partial(f^4 \circ f^3 \circ f^2 \circ f^1)(\boldsymbol{x})}{\partial \boldsymbol{x}}, \tag{B.13}$$

which is the first term on the right-hand side of Equation B.6. All other remaining terms in Equation B.6 are calculated in the same manner.

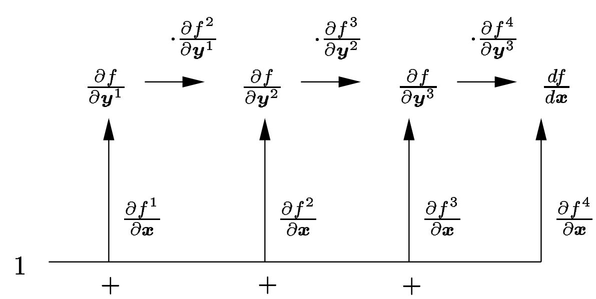

Figure B.2: Computational graph of the forward mode of differentiation.

Example:

$\frac{\partial f^1}{\partial x} = 2,\ \frac{\partial f^2}{\partial x} = 1,\ \frac{\partial f^3}{\partial x} = 3,\ \frac{\partial f^4}{\partial x} = 2x$

$\frac{\partial f^2}{\partial y^1} = 1,\ \frac{\partial f^3}{\partial y^2} = 2,\ \frac{\partial f^4}{\partial y^3} = 1$

$\frac{\partial (f^4 \circ f^3 \circ f^2 \circ f^1)}{\partial x} = \frac{\partial f^4}{\partial y^3}\frac{\partial f^3}{\partial y^2}\frac{\partial f^2}{\partial y^1}\frac{\partial f^1}{\partial x} = 1 \cdot 2 \cdot 1 \cdot 2 = 4$

$\frac{\partial (f^4 \circ f^3 \circ f^2)}{\partial x} = \frac{\partial f^4}{\partial y^3}\frac{\partial f^3}{\partial y^2}\frac{\partial f^2}{\partial x} = 1 \cdot 2 \cdot 1 = 2$

$\frac{\partial (f^4 \circ f^3)}{\partial x} = \frac{\partial f^4}{\partial y^3}\frac{\partial f^3}{\partial x} = 1 \cdot 3 = 3$

$\frac{\partial f^4}{\partial x} = 2x$

$\frac{df}{dx} = \frac{\partial (f^4 \circ f^3 \circ f^2 \circ f^1)}{\partial x} + \frac{\partial (f^4 \circ f^3 \circ f^2)}{\partial x} + \frac{\partial (f^4 \circ f^3)}{\partial x} + \frac{\partial f^4}{\partial x} = 2x + 9$

B.3 Reverse Mode of Differentiation

The reverse mode of differentiation calculates the derivative $\frac{df}{d\boldsymbol{x}}^T$ (see Equation B.9) in the reverse order than the sub-functions f^1, f^2, f^3, and f^4 are evaluated (see Equations B.1-B.4). The derivatives of Equation B.9 are calculated by applying successively Equation 6.23. We start with the last partial derivative $\frac{\partial f^4(\boldsymbol{y}^3)}{\partial \boldsymbol{y}^3}^T$ by proceeding in the reverse direction and arrive finally at $\frac{\partial (f^4 \circ f^3 \circ f^2 \circ f^1)(\boldsymbol{x})}{\partial \boldsymbol{x}}^T$:

$$\frac{\partial (f^4 \circ f^3)(\boldsymbol{y}^2)}{\partial \boldsymbol{y}^2}^T = \frac{\partial f^3(\boldsymbol{y}^2)}{\partial \boldsymbol{y}^2}^T \frac{\partial f^4(\boldsymbol{y}^3)}{\partial \boldsymbol{y}^3}^T \tag{B.14}$$

$$\frac{\partial (f^4 \circ f^3 \circ f^2)(\boldsymbol{y}^1)}{\partial \boldsymbol{y}^1}^T = \frac{\partial f^2(\boldsymbol{y}^1)}{\partial \boldsymbol{y}^1}^T \frac{\partial (f^4 \circ f^3)(\boldsymbol{y}^2)}{\partial \boldsymbol{y}^2}^T \tag{B.15}$$

$$\frac{\partial (f^4 \circ f^3 \circ f^2 \circ f^1)(\boldsymbol{x})}{\partial \boldsymbol{x}}^T = \frac{\partial f^1(\boldsymbol{x})}{\partial \boldsymbol{x}}^T \frac{\partial (f^4 \circ f^3 \circ f^2)(\boldsymbol{y}^1)}{\partial \boldsymbol{y}^1}^T . \tag{B.16}$$

All terms yield together the first term on the right-hand side of Equation B.8:

$$\frac{\partial (f^4 \circ f^3 \circ f^2 \circ f^1)(\boldsymbol{x})}{\partial \boldsymbol{x}}^T = \frac{\partial f^1(\boldsymbol{x})}{\partial \boldsymbol{x}}^T \frac{\partial f^2(\boldsymbol{y}^1)}{\partial \boldsymbol{y}^1}^T \frac{\partial f^3(\boldsymbol{y}^2)}{\partial \boldsymbol{y}^2}^T \frac{\partial f^4(\boldsymbol{y}^3)}{\partial \boldsymbol{y}^3}^T . \tag{B.17}$$

The remaining derivatives of Equation B.8 are calculated in the same way.

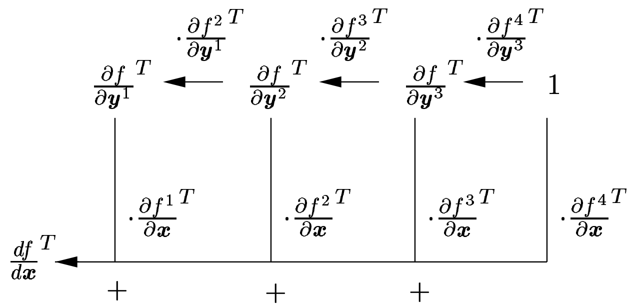

Figure B.3: Computational graph of the reverse mode of differentiation.

Example:

$\frac{\partial f^1}{\partial x}^T = 2,\ \frac{\partial f^2}{\partial x}^T = 1,\ \frac{\partial f^3}{\partial x}^T = 3,\ \frac{\partial f^4}{\partial x}^T = 2x$

$\frac{\partial f^2}{\partial y^1}^T = 1,\ \frac{\partial f^3}{\partial y^2}^T = 2,\ \frac{\partial f^4}{\partial y^3}^T = 1$

$\frac{\partial (f^4 \circ f^3 \circ f^2 \circ f^1)}{\partial x}^T = \frac{\partial f^1}{\partial x}^T \frac{\partial f^2}{\partial y^1}^T \frac{\partial f^3}{\partial y^2}^T \frac{\partial f^4}{\partial y^3}^T = 2 \cdot 1 \cdot 2 \cdot 1 = 4$

$\frac{\partial (f^4 \circ f^3 \circ f^2)}{\partial x}^T = \frac{\partial f^2}{\partial x}^T \frac{\partial f^3}{\partial y^2}^T \frac{\partial f^4}{\partial y^3}^T = 1 \cdot 2 \cdot 1 = 2$

$\frac{\partial (f^4 \circ f^3)}{\partial x}^T = \frac{\partial f^3}{\partial x}^T \frac{\partial f^4}{\partial y^3}^T = 3 \cdot 1 = 3$

$\frac{\partial f^4}{\partial x}^T = 2x$

$\frac{df}{x}^T = \frac{\partial (f^4 \circ f^3 \circ f^2 \circ f^1)}{\partial x}^T + \frac{\partial (f^4 \circ f^3 \circ f^2)}{\partial x}^T + \frac{\partial (f^4 \circ f^3)}{\partial x}^T + \frac{\partial f^4}{\partial x}^T = 2x + 9$

Appendix C

Total Derivative of Objective Function

The total derivative $\frac{d\Phi}{d\boldsymbol{\mu}}$ was obtained by applying systematically the chain rule of differentiation to the composed objective function (see Equation 6.24 on page 96):

$$\Phi(\boldsymbol{\mu}) = \left(\tilde{\Phi} \circ \boldsymbol{\psi}^{\mathrm{Z}}(\boldsymbol{\mu}) \circ \boldsymbol{\psi}^{\mathrm{Z}-1}(\boldsymbol{\mu}) \circ \boldsymbol{\psi}^{\mathrm{Z}-2}(\boldsymbol{\mu}) \circ \ldots \circ \boldsymbol{\psi}^{2}(\boldsymbol{\mu}) \circ \boldsymbol{\psi}^{1}\right)(\boldsymbol{\mu}) .$$

The chain rule of differentiation yields (see also example in Appendix B.1):

$$\begin{aligned}
\frac{d\Phi(\boldsymbol{\mu})}{d\boldsymbol{\mu}} = \; & \frac{\partial\tilde{\Phi}(\boldsymbol{\psi}^{\mathrm{Z}})}{\partial\boldsymbol{\psi}^{\mathrm{Z}}} \frac{\partial\boldsymbol{\psi}^{\mathrm{Z}}(\boldsymbol{\psi}^{\mathrm{Z}-1})}{\partial\boldsymbol{\psi}^{\mathrm{Z}-1}} \frac{\partial\boldsymbol{\psi}^{\mathrm{Z}-1}(\boldsymbol{\psi}^{\mathrm{Z}-2})}{\partial\boldsymbol{\psi}^{\mathrm{Z}-2}} \cdots \frac{\partial\boldsymbol{\psi}^{3}(\boldsymbol{\psi}^{2})}{\partial\boldsymbol{\psi}^{2}} \frac{\partial\boldsymbol{\psi}^{2}(\boldsymbol{\psi}^{1})}{\partial\boldsymbol{\psi}^{1}} \frac{\partial\boldsymbol{\psi}^{1}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}} \\
& + \frac{\partial\tilde{\Phi}(\boldsymbol{\psi}^{\mathrm{Z}})}{\partial\boldsymbol{\psi}^{\mathrm{Z}}} \frac{\partial\boldsymbol{\psi}^{\mathrm{Z}}(\boldsymbol{\psi}^{\mathrm{Z}-1})}{\partial\boldsymbol{\psi}^{\mathrm{Z}-1}} \frac{\partial\boldsymbol{\psi}^{\mathrm{Z}-1}(\boldsymbol{\psi}^{\mathrm{Z}-2})}{\partial\boldsymbol{\psi}^{\mathrm{Z}-2}} \cdots \frac{\partial\boldsymbol{\psi}^{3}(\boldsymbol{\psi}^{2})}{\partial\boldsymbol{\psi}^{2}} \frac{\partial\boldsymbol{\psi}^{2}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}} \\
& + \frac{\partial\tilde{\Phi}(\boldsymbol{\psi}^{\mathrm{Z}})}{\partial\boldsymbol{\psi}^{\mathrm{Z}}} \frac{\partial\boldsymbol{\psi}^{\mathrm{Z}}(\boldsymbol{\psi}^{\mathrm{Z}-1})}{\partial\boldsymbol{\psi}^{\mathrm{Z}-1}} \frac{\partial\boldsymbol{\psi}^{\mathrm{Z}-1}(\boldsymbol{\psi}^{\mathrm{Z}-2})}{\partial\boldsymbol{\psi}^{\mathrm{Z}-2}} \cdots \frac{\partial\boldsymbol{\psi}^{3}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}} \\
& + \ldots + \\
& + \frac{\partial\tilde{\Phi}(\boldsymbol{\psi}^{\mathrm{Z}})}{\partial\boldsymbol{\psi}^{\mathrm{Z}}} \frac{\partial\boldsymbol{\psi}^{\mathrm{Z}}(\boldsymbol{\psi}^{\mathrm{Z}-1})}{\partial\boldsymbol{\psi}^{\mathrm{Z}-1}} \frac{\partial\boldsymbol{\psi}^{\mathrm{Z}-1}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}} \\
& + \frac{\partial\tilde{\Phi}(\boldsymbol{\psi}^{\mathrm{Z}})}{\partial\boldsymbol{\psi}^{\mathrm{Z}}} \frac{\partial\boldsymbol{\psi}^{\mathrm{Z}}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}} .
\end{aligned} \tag{C.1}$$

We obtain with the short-hand notation:

$$\begin{aligned}\frac{d\Phi(\boldsymbol{\mu})}{d\boldsymbol{\mu}} &= \frac{\partial(\tilde{\Phi}\circ\boldsymbol{\psi}^{\mathrm{Z}}\circ\boldsymbol{\psi}^{\mathrm{Z}-1}\circ\ldots\circ\boldsymbol{\psi}^{3}\circ\boldsymbol{\psi}^{2}\circ\boldsymbol{\psi}^{1})(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}}\\ &+\frac{\partial(\tilde{\Phi}\circ\boldsymbol{\psi}^{\mathrm{Z}}\circ\boldsymbol{\psi}^{\mathrm{Z}-1}\circ\ldots\circ\boldsymbol{\psi}^{3}\circ\boldsymbol{\psi}^{2})(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}}\\ &+\frac{\partial(\tilde{\Phi}\circ\boldsymbol{\psi}^{\mathrm{Z}}\circ\boldsymbol{\psi}^{\mathrm{Z}-1}\circ\ldots\circ\boldsymbol{\psi}^{3})(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}}\\ &+\ldots+\\ &+\frac{\partial(\tilde{\Phi}\circ\boldsymbol{\psi}^{\mathrm{Z}}\circ\boldsymbol{\psi}^{\mathrm{Z}-1})(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}}\\ &+\frac{\partial(\tilde{\Phi}\circ\boldsymbol{\psi}^{\mathrm{Z}})(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}}.\end{aligned}\tag{C.2}$$

The transpose of Equation C.1 is

$$\begin{aligned}\frac{d\Phi(\boldsymbol{\mu})}{d\boldsymbol{\mu}}^{T} &= \frac{\partial\boldsymbol{\psi}^{1}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}}^{T}\frac{\partial\boldsymbol{\psi}^{2}(\boldsymbol{\psi}^{1})}{\partial\boldsymbol{\psi}^{1}}^{T}\frac{\partial\boldsymbol{\psi}^{3}(\boldsymbol{\psi}^{2})}{\partial\boldsymbol{\psi}^{2}}^{T}\ldots\frac{\partial\boldsymbol{\psi}^{\mathrm{Z}-1}(\boldsymbol{\psi}^{\mathrm{Z}-2})}{\partial\boldsymbol{\psi}^{\mathrm{Z}-2}}^{T}\frac{\partial\boldsymbol{\psi}^{\mathrm{Z}}(\boldsymbol{\psi}^{\mathrm{Z}-1})}{\partial\boldsymbol{\psi}^{\mathrm{Z}-1}}^{T}\frac{\partial\tilde{\Phi}(\boldsymbol{\psi}^{\mathrm{Z}})}{\partial\boldsymbol{\psi}^{\mathrm{Z}}}^{T}\\ &+\frac{\partial\boldsymbol{\psi}^{2}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}}^{T}\frac{\partial\boldsymbol{\psi}^{3}(\boldsymbol{\psi}^{2})}{\partial\boldsymbol{\psi}^{2}}^{T}\ldots\frac{\partial\boldsymbol{\psi}^{\mathrm{Z}-1}(\boldsymbol{\psi}^{\mathrm{Z}-2})}{\partial\boldsymbol{\psi}^{\mathrm{Z}-2}}^{T}\frac{\partial\boldsymbol{\psi}^{\mathrm{Z}}(\boldsymbol{\psi}^{\mathrm{Z}-1})}{\partial\boldsymbol{\psi}^{\mathrm{Z}-1}}^{T}\frac{\partial\tilde{\Phi}(\boldsymbol{\psi}^{\mathrm{Z}})}{\partial\boldsymbol{\psi}^{\mathrm{Z}}}^{T}\\ &+\frac{\partial\boldsymbol{\psi}^{3}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}}^{T}\ldots\frac{\partial\boldsymbol{\psi}^{\mathrm{Z}-1}(\boldsymbol{\psi}^{\mathrm{Z}-2})}{\partial\boldsymbol{\psi}^{\mathrm{Z}-2}}^{T}\frac{\partial\boldsymbol{\psi}^{\mathrm{Z}}(\boldsymbol{\psi}^{\mathrm{Z}-1})}{\partial\boldsymbol{\psi}^{\mathrm{Z}-1}}^{T}\frac{\partial\tilde{\Phi}(\boldsymbol{\psi}^{\mathrm{Z}})}{\partial\boldsymbol{\psi}^{\mathrm{Z}}}^{T}\\ &+\ldots+\\ &+\frac{\partial\boldsymbol{\psi}^{\mathrm{Z}-1}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}}^{T}\frac{\partial\boldsymbol{\psi}^{\mathrm{Z}}(\boldsymbol{\psi}^{\mathrm{Z}-1})}{\partial\boldsymbol{\psi}^{\mathrm{Z}-1}}^{T}\frac{\partial\tilde{\Phi}(\boldsymbol{\psi}^{\mathrm{Z}})}{\partial\boldsymbol{\psi}^{\mathrm{Z}}}^{T}\\ &+\frac{\partial\boldsymbol{\psi}^{\mathrm{Z}}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}}^{T}\frac{\partial\tilde{\Phi}(\boldsymbol{\psi}^{\mathrm{Z}})}{\partial\boldsymbol{\psi}^{\mathrm{Z}}}^{T}.\end{aligned}\tag{C.3}$$

The short-hand notation yields:

$$\begin{aligned}
\frac{d\Phi(\boldsymbol{\mu})}{d\boldsymbol{\mu}}^T = & \frac{\partial(\tilde{\Phi} \circ \boldsymbol{\psi}^{\mathrm{Z}} \circ \boldsymbol{\psi}^{\mathrm{Z}-1} \circ ... \circ \boldsymbol{\psi}^3 \circ \boldsymbol{\psi}^2 \circ \boldsymbol{\psi}^1)(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}^T \\
& + \frac{\partial(\tilde{\Phi} \circ \boldsymbol{\psi}^{\mathrm{Z}} \circ \boldsymbol{\psi}^{\mathrm{Z}-1} \circ ... \circ \boldsymbol{\psi}^3 \circ \boldsymbol{\psi}^2)(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}^T \\
& + \frac{\partial(\tilde{\Phi} \circ \boldsymbol{\psi}^{\mathrm{Z}} \circ \boldsymbol{\psi}^{\mathrm{Z}-1} \circ ... \circ \boldsymbol{\psi}^3)(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}^T \\
& + ... + \\
& + \frac{\partial(\tilde{\Phi} \circ \boldsymbol{\psi}^{\mathrm{Z}} \circ \boldsymbol{\psi}^{\mathrm{Z}-1})(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}^T \\
& + \frac{\partial(\tilde{\Phi} \circ \boldsymbol{\psi}^{\mathrm{Z}})(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}^T .
\end{aligned} \tag{C.4}$$

We rewrite Equation C.3 and make use of the equality $\boldsymbol{A}^T \cdot \boldsymbol{B}^T \cdot \boldsymbol{C}^T = (\boldsymbol{C} \cdot \boldsymbol{B} \cdot \boldsymbol{A})^T$

$$\begin{aligned}
\frac{d\Phi(\boldsymbol{\mu})}{d\boldsymbol{\mu}}^T = & \frac{\partial \boldsymbol{\psi}^1(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}^T \left(\frac{\partial \tilde{\Phi}}{\partial \boldsymbol{\psi}^{\mathrm{Z}}} \frac{\partial \boldsymbol{\psi}^{\mathrm{Z}}}{\partial \boldsymbol{\psi}^{\mathrm{Z}-1}} \frac{\partial \boldsymbol{\psi}^{\mathrm{Z}-1}}{\partial \boldsymbol{\psi}^{\mathrm{Z}-2}} \cdots \frac{\partial \boldsymbol{\psi}^4}{\partial \boldsymbol{\psi}^3} \frac{\partial \boldsymbol{\psi}^3}{\partial \boldsymbol{\psi}^2} \frac{\partial \boldsymbol{\psi}^2}{\partial \boldsymbol{\psi}^1} \right)^T \\
& + \frac{\partial \boldsymbol{\psi}^2(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}^T \left(\frac{\partial \tilde{\Phi}}{\partial \boldsymbol{\psi}^{\mathrm{Z}}} \frac{\partial \boldsymbol{\psi}^{\mathrm{Z}}}{\partial \boldsymbol{\psi}^{\mathrm{Z}-1}} \frac{\partial \boldsymbol{\psi}^{\mathrm{Z}-1}}{\partial \boldsymbol{\psi}^{\mathrm{Z}-2}} \cdots \frac{\partial \boldsymbol{\psi}^4}{\partial \boldsymbol{\psi}^3} \frac{\partial \boldsymbol{\psi}^3}{\partial \boldsymbol{\psi}^2} \right)^T \\
& + \frac{\partial \boldsymbol{\psi}^3(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}^T \left(\frac{\partial \tilde{\Phi}}{\partial \boldsymbol{\psi}^{\mathrm{Z}}} \frac{\partial \boldsymbol{\psi}^{\mathrm{Z}}}{\partial \boldsymbol{\psi}^{\mathrm{Z}-1}} \frac{\partial \boldsymbol{\psi}^{\mathrm{Z}-1}}{\partial \boldsymbol{\psi}^{\mathrm{Z}-2}} \cdots \frac{\partial \boldsymbol{\psi}^4}{\partial \boldsymbol{\psi}^3} \right)^T \\
& + ... + \\
& + \frac{\partial \boldsymbol{\psi}^{\mathrm{Z}-1}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}^T \left(\frac{\partial \tilde{\Phi}}{\partial \boldsymbol{\psi}^{\mathrm{Z}}} \frac{\partial \boldsymbol{\psi}^{\mathrm{Z}}}{\partial \boldsymbol{\psi}^{\mathrm{Z}-1}} \right)^T \\
& + \frac{\partial \boldsymbol{\psi}^{\mathrm{Z}}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}^T \left(\frac{\partial \tilde{\Phi}}{\partial \boldsymbol{\psi}^{\mathrm{Z}}} \right)^T .
\end{aligned} \tag{C.5}$$

By using the short-hand notation for the terms within parentheses

$$\left(\frac{\partial \tilde{\Phi}}{\partial \boldsymbol{\psi}^{\mathrm{Z}}} \frac{\partial \boldsymbol{\psi}^{\mathrm{Z}}}{\partial \boldsymbol{\psi}^{\mathrm{Z}-1}} \frac{\partial \boldsymbol{\psi}^{\mathrm{Z}-1}}{\partial \boldsymbol{\psi}^{\mathrm{Z}-2}} \cdots \frac{\partial \boldsymbol{\psi}^{\mathrm{z}+1}}{\partial \boldsymbol{\psi}^{\mathrm{z}}} \right)^T = \left(\frac{\partial(\tilde{\Phi} \circ F^{\mathrm{Z}} \circ ... \circ F^{\mathrm{z}+1})(\boldsymbol{\psi}^{\mathrm{z}})}{\partial \boldsymbol{\psi}^{\mathrm{z}}} \right)^T = \left(\frac{\partial \Phi}{\partial \boldsymbol{\psi}^{\mathrm{z}}} \right)^T , \tag{C.6}$$

we rewrite Equation C.5 and obtain

$$\begin{aligned}\frac{d\Phi(\boldsymbol{\mu})}{d\boldsymbol{\mu}}^T = & \frac{\partial \boldsymbol{\psi}^1}{\partial \boldsymbol{\mu}}^T \left(\frac{\partial \Phi}{\partial \boldsymbol{\psi}^1}\right)^T \\ & + \frac{\partial \boldsymbol{\psi}^2}{\partial \boldsymbol{\mu}}^T \left(\frac{\partial \Phi}{\partial \boldsymbol{\psi}^2}\right)^T \\ & + \frac{\partial \boldsymbol{\psi}^3}{\partial \boldsymbol{\mu}}^T \left(\frac{\partial \Phi}{\partial \boldsymbol{\psi}^3}\right)^T \\ & + \ldots + \\ & + \frac{\partial \boldsymbol{\psi}^{\mathrm{Z}-1}}{\partial \boldsymbol{\mu}}^T \left(\frac{\partial \Phi}{\partial \boldsymbol{\psi}^{\mathrm{Z}-1}}\right)^T \\ & + \frac{\partial \boldsymbol{\psi}^{\mathrm{Z}}}{\partial \boldsymbol{\mu}}^T \left(\frac{\partial \Phi}{\partial \boldsymbol{\psi}^{\mathrm{Z}}}\right)^T .\end{aligned} \qquad (C.7)$$

Equation C.7 is equivalent to Equation 6.29 on page 98.