Errata

The following is the errata for the second edition of **”Machine Learning Refined: Foundations, Algorithms, and Applications”** published by Cambridge University Press in 2020.

Page	Location	Incorrect	Correct
168	Equation (6.83)	model $(x_p, \mathbf{w})$	model $(\mathbf{x}_p, \mathbf{w})$
212	Equation (8.10)	$\mathbf{C}\,\mathbf{C}^T = \mathbf{I}_{N\times N}$	$\mathbf{C}^T\mathbf{C} = \mathbf{I}_{N\times N}$
528	Line 12	” ... we update the partial derivative of each parent by multiplying it by the partial derivative of its children node with respect to that parent. When the backward sweep is completed we will have recursively constructed the gradient of the function with respect to all of its inputs.”	While this is not incorrect per se, the reader should note that in the backward sweep of reverse-mode differentiation, when a parent node has multiple children the accumulated partials should be added, not multiplied, since this is what the chain rule requires. See Example 0.1 below.

Example 0.1 Reverse-mode differentiation of a simple function

Consider the simple function $g(w) = w^2 + \sin(w)$ whose computation graph is drawn in the top panel of Figure 0.1, and consists of nodes w (input node), a, b, and c. In the forward sweep of the automatic differentiation we compute the partial derivatives of each child node with respect to its parent(s) and store them, as illustrated in the bottom panel of Figure 0.1.

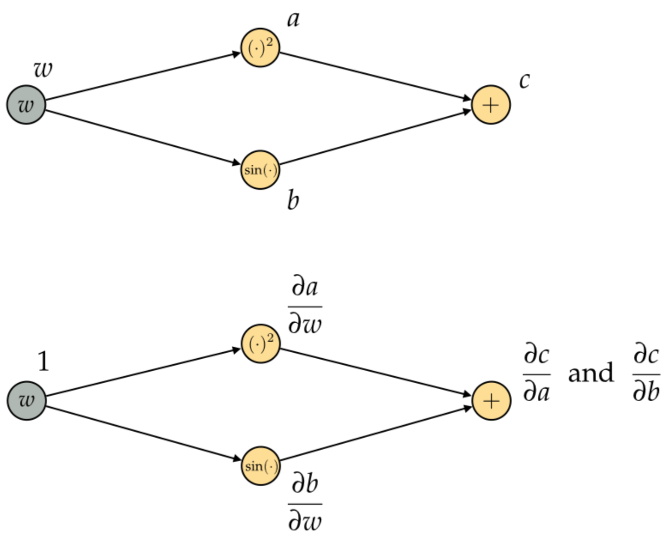

Figure 0.1 (top panel) The computation graph associated with the function $g(w) = w^2 + \sin(w)$. (bottom panel) The result of the full forward sweep of the reverse-mode differentiation of $g(w)$ with respect to w.

The backward sweep then starts at node c, all the way to the right, and at each step, we update the partial derivative of each parent node by multiplying it by the partial derivative of its children node with respect to that parent, as shown from top to bottom in Figure 0.2. Note importantly that at node w – a parent node with more than one child – we must add up the derivative contributions of each of its children to compute the final derivative at this node as

$$\frac{\partial g}{\partial w} = \frac{\partial c}{\partial a}\frac{\partial a}{\partial w} + \frac{\partial c}{\partial b}\frac{\partial b}{\partial w}. \tag{0.1}$$

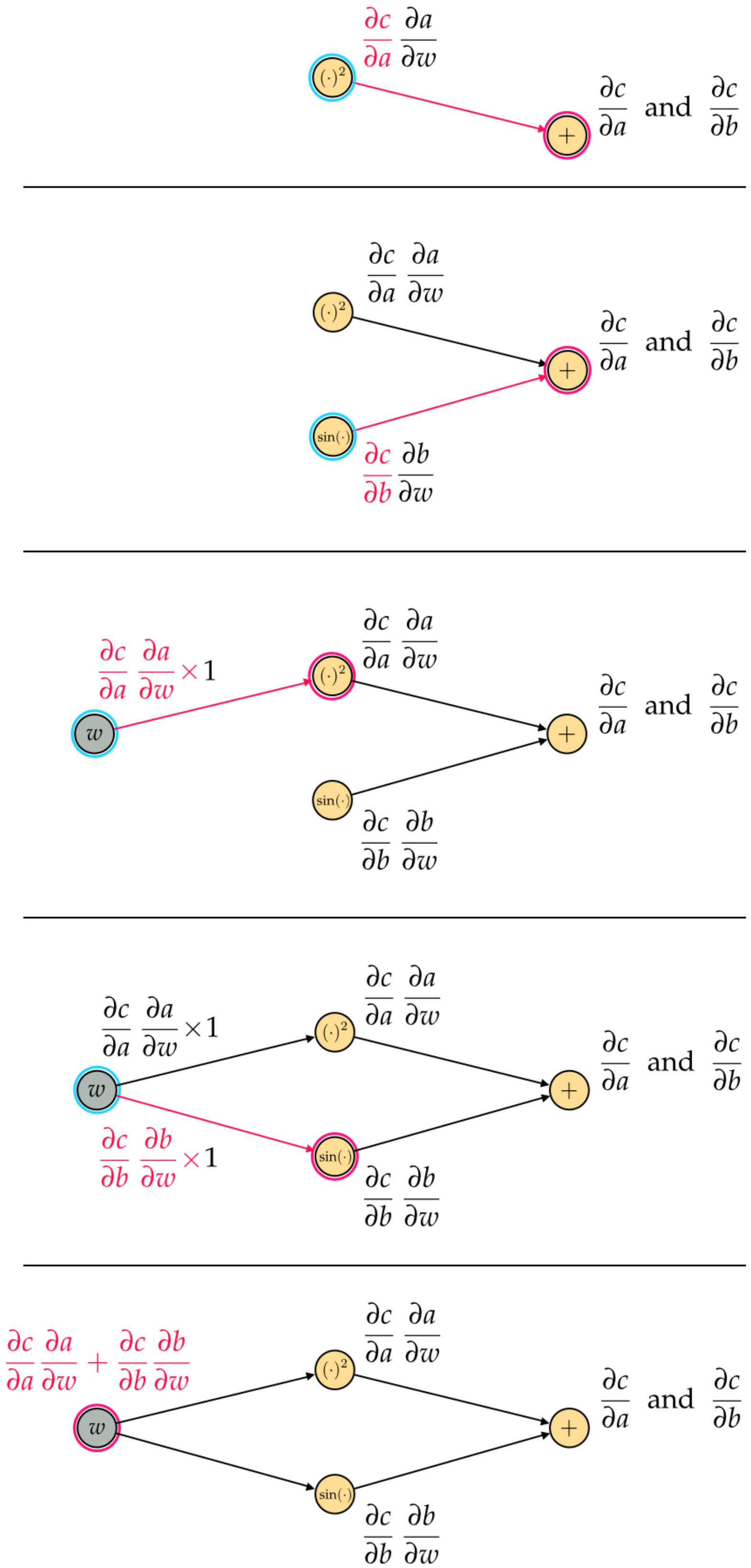

Figure 0.2 From top to bottom, the backward sweep of the reverse-mode differentiation of $g(w)$. Note in the bottom panel that sometimes a parent node might have more than one child (here the node w), in which case the derivative contributions of each of its children must be added – according to the chain rule – to compute the final derivative at that node.