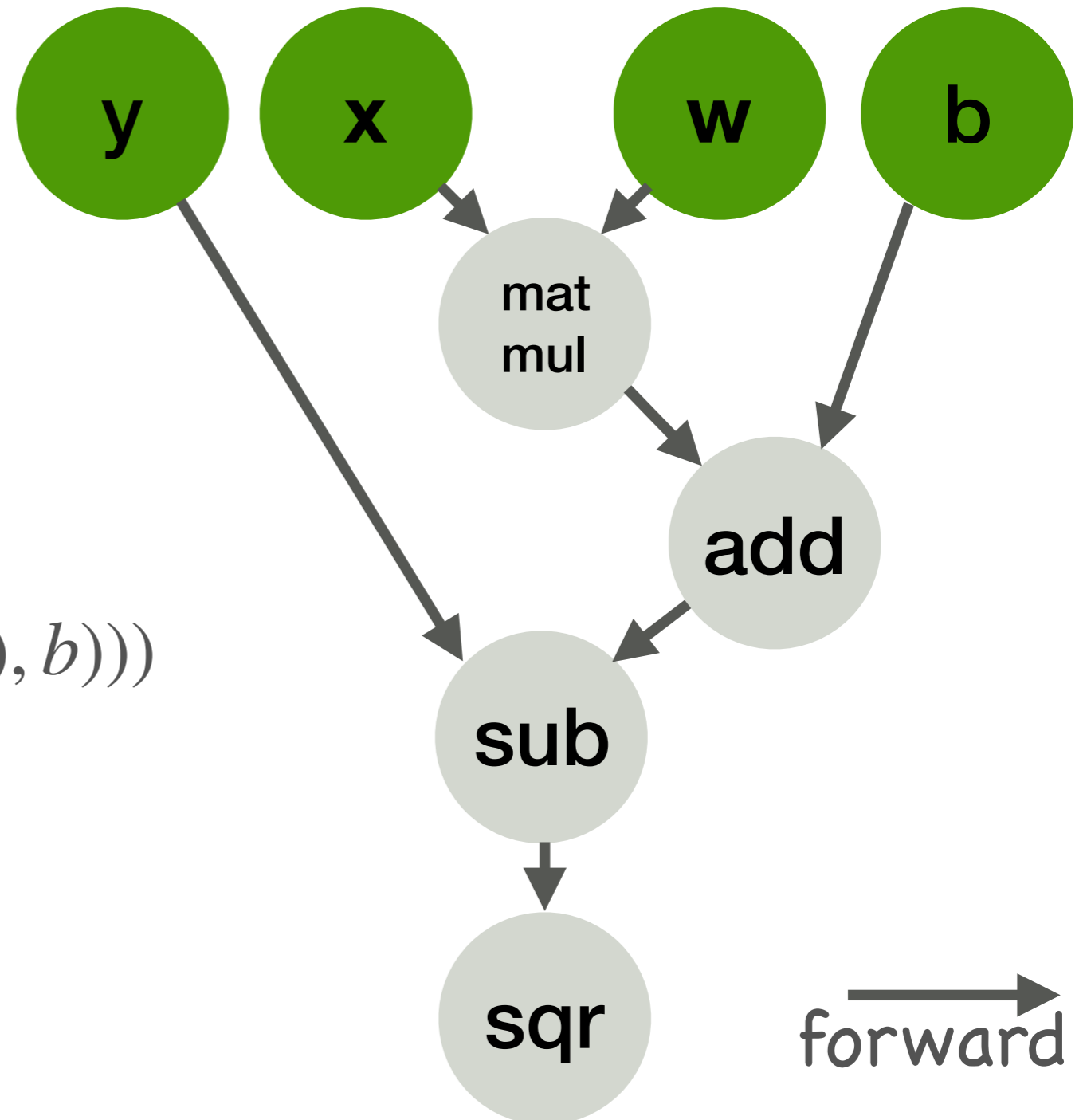


# Gradients on computation graphs

© 2019 Philipp Krähenbühl and Chao-Yuan Wu

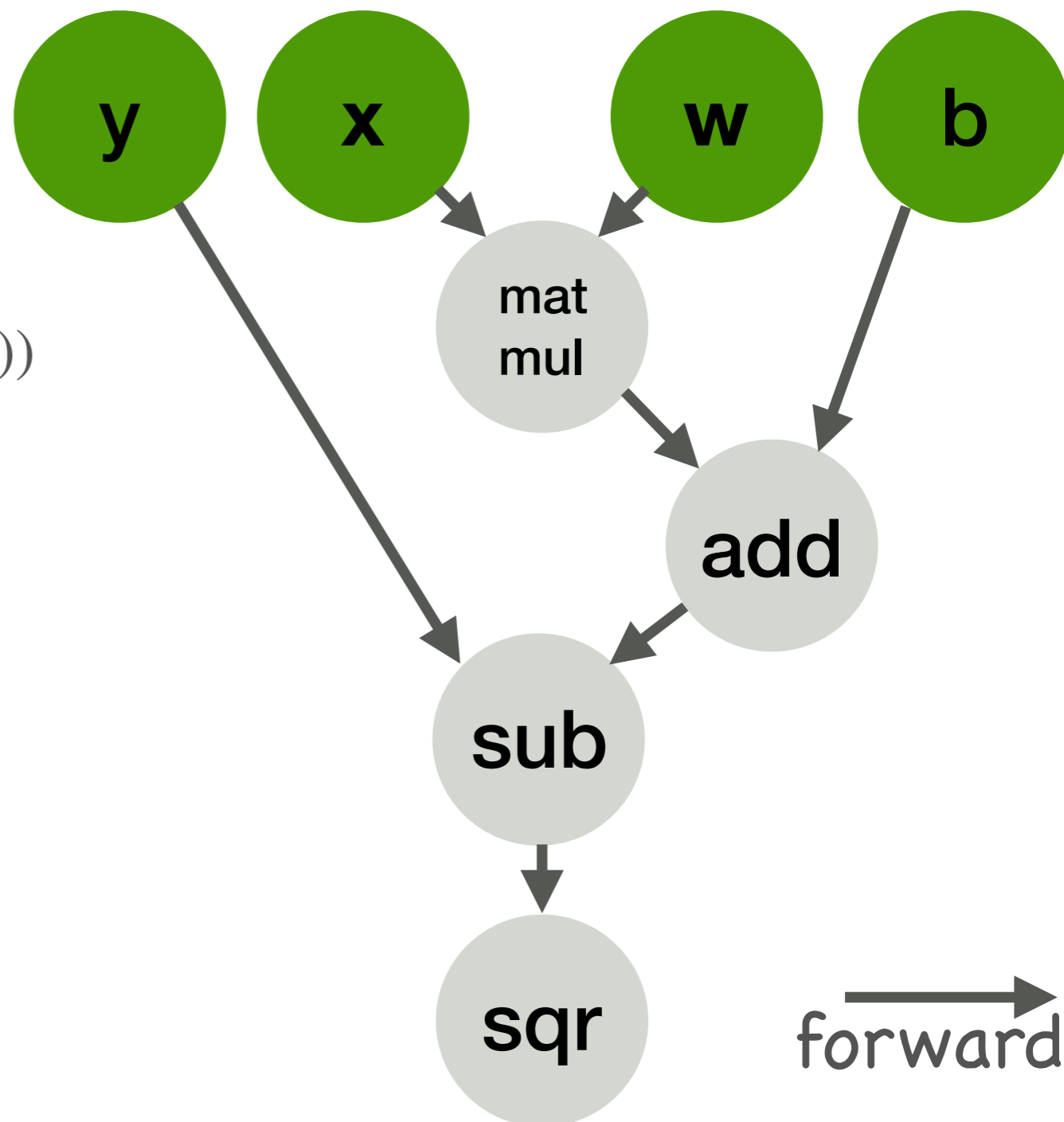
# Linear regression



- $\ell(\mathbf{w}, b) = (y - \mathbf{w}^\top \mathbf{x} + b)^2$
- `sqr(sub(y, add(matmul(w, x), b)))`

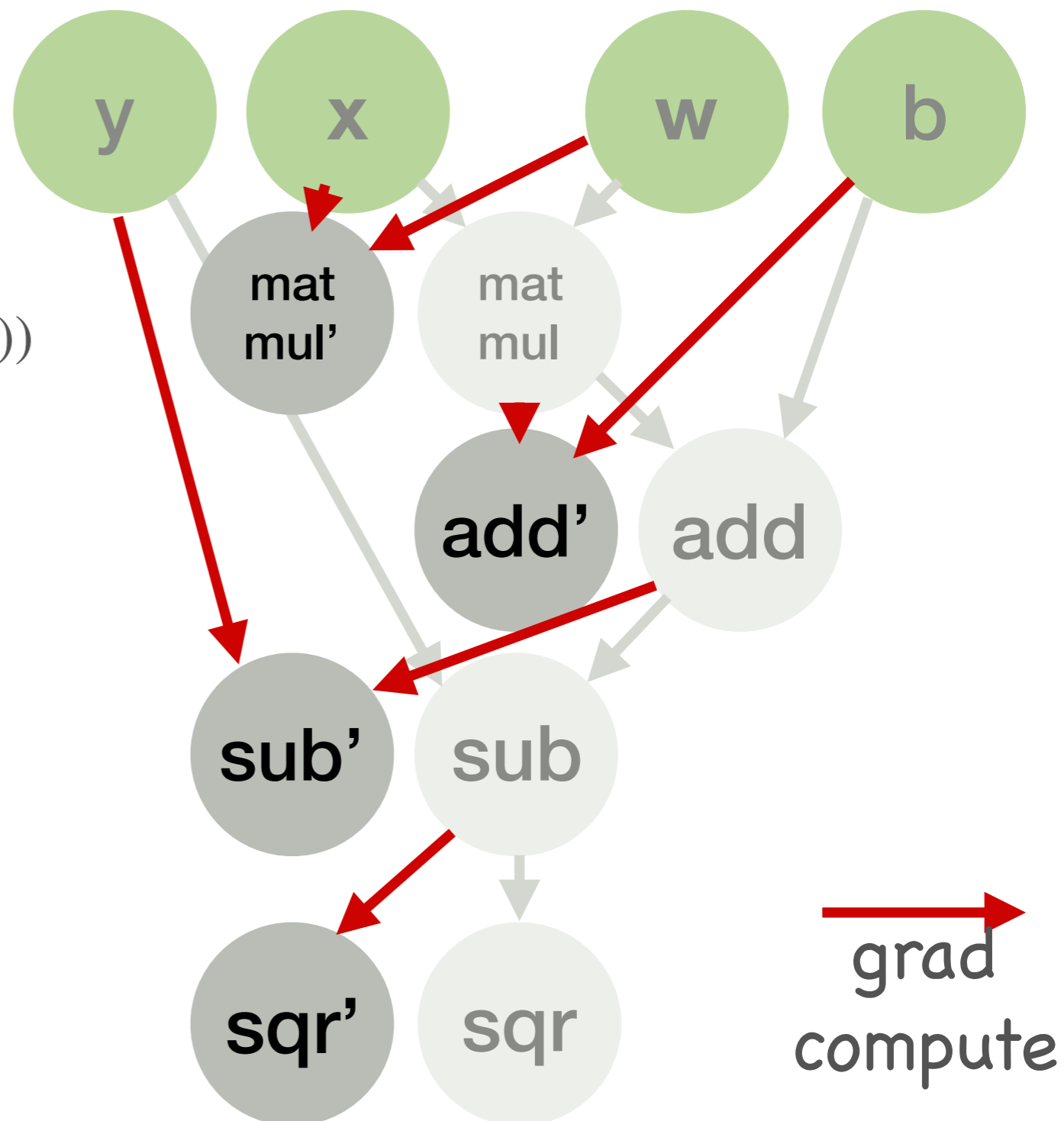
# Gradient of linear regression

- $$\frac{d\mathcal{L}(\mathbf{w}, b)}{d\mathbf{w}} = \frac{d(y - \mathbf{w}^\top \mathbf{x} + b)^2}{d\mathbf{w}}$$
- $$\begin{aligned} & \frac{d}{d\mathbf{w}} \text{sqr}(\text{sub}(y, \text{add}(\text{matmul}(\mathbf{w}, \mathbf{x}), b))) \\ &= \text{sqr}'(\text{sub}(y, \text{add}(\text{matmul}(\mathbf{w}, \mathbf{x}), b))) \\ & \quad \frac{d}{d\mathbf{w}} \text{sub}(y, \text{add}(\text{matmul}(\mathbf{w}, \mathbf{x}), b)) \\ &= \dots \\ &= \text{sqr}'(\text{sub}(y, \text{add}(\text{matmul}(\mathbf{w}, \mathbf{x}), b))) \\ & \quad \text{sub}'(y, \text{add}(\text{matmul}(\mathbf{w}, \mathbf{x}), b)) \\ & \quad \text{add}'(\text{matmul}(\mathbf{w}, \mathbf{x}), b) \\ & \quad \frac{d}{d\mathbf{w}} \text{matmul}(\mathbf{w}, \mathbf{x}) \end{aligned}$$



# Gradient of linear regression

- $$\frac{d\ell(\mathbf{w}, b)}{d\mathbf{w}} = \frac{d(y - \mathbf{w}^\top \mathbf{x} + b)^2}{d\mathbf{w}}$$
- $$\begin{aligned} & \frac{d}{d\mathbf{w}} \text{sqr}(\text{sub}(y, \text{add}(\text{matmul}(\mathbf{w}, \mathbf{x}), b))) \\ &= \text{sqr}'(\text{sub}(y, \text{add}(\text{matmul}(\mathbf{w}, \mathbf{x}), b))) \\ & \quad \frac{d}{d\mathbf{w}} \text{sub}(y, \text{add}(\text{matmul}(\mathbf{w}, \mathbf{x}), b)) \\ &= \dots \\ &= \text{sqr}'(\text{sub}(y, \text{add}(\text{matmul}(\mathbf{w}, \mathbf{x}), b))) \\ & \quad \text{sub}'(y, \text{add}(\text{matmul}(\mathbf{w}, \mathbf{x}), b)) \\ & \quad \text{add}'(\text{matmul}(\mathbf{w}, \mathbf{x}), b) \\ & \quad \frac{d}{d\mathbf{w}} \text{matmul}(\mathbf{w}, \mathbf{x}) \end{aligned}$$



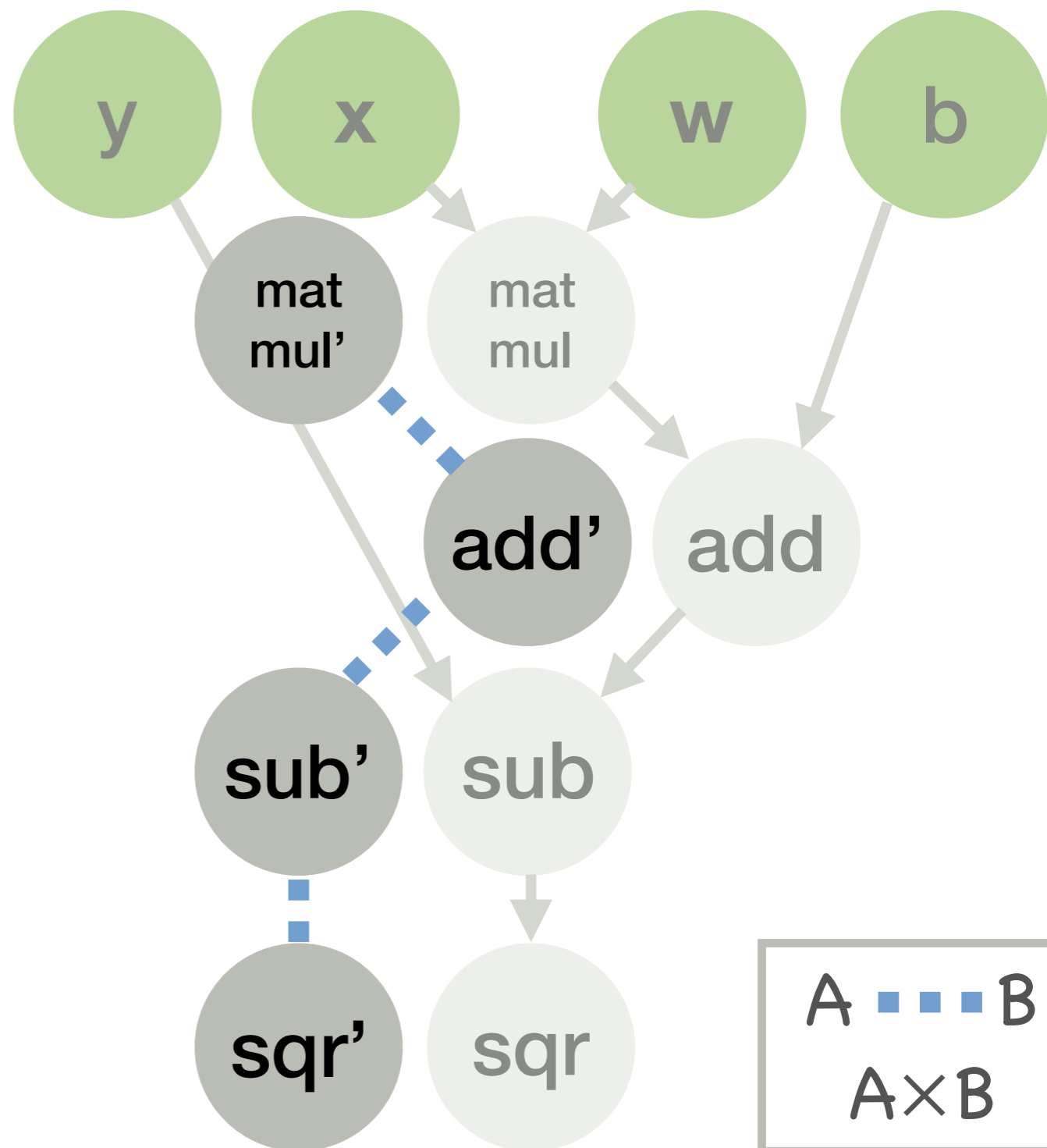
# Gradient of linear regression

$\text{sqr}'(\text{sub}(y, \text{add}(\text{matmul}(\mathbf{w}, \mathbf{x}), b)))$

$\text{sub}'(y, \text{add}(\text{matmul}(\mathbf{w}, \mathbf{x}), b))$

$\text{add}'(\text{matmul}(\mathbf{w}, \mathbf{x}), b)$

$\frac{d}{d\mathbf{w}} \text{matmul}(\mathbf{w}, \mathbf{x})$



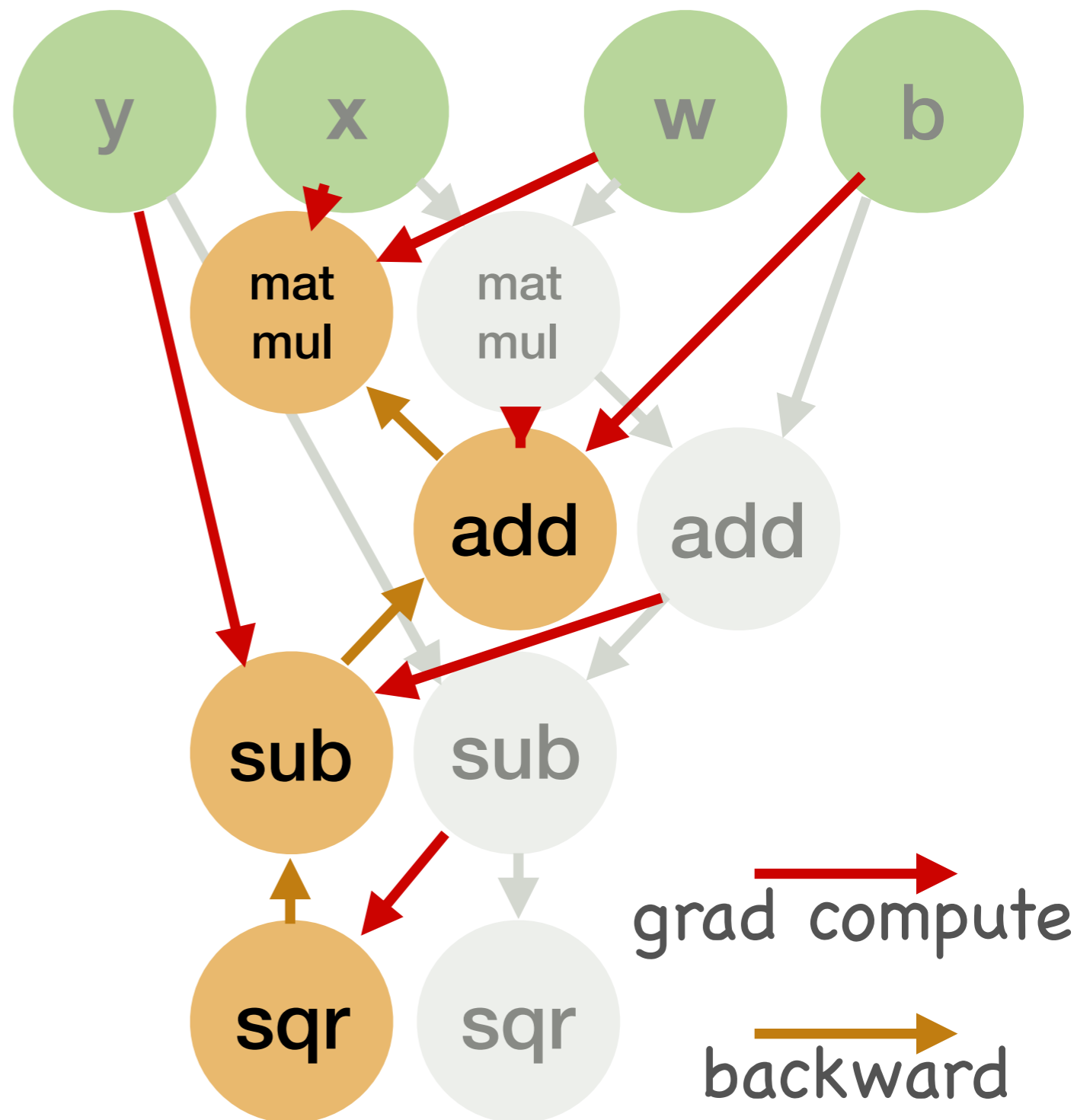
# Gradient of linear regression

$\text{sqr}'(\text{sub}(y, \text{add}(\text{matmul}(\mathbf{w}, \mathbf{x}), b)))$

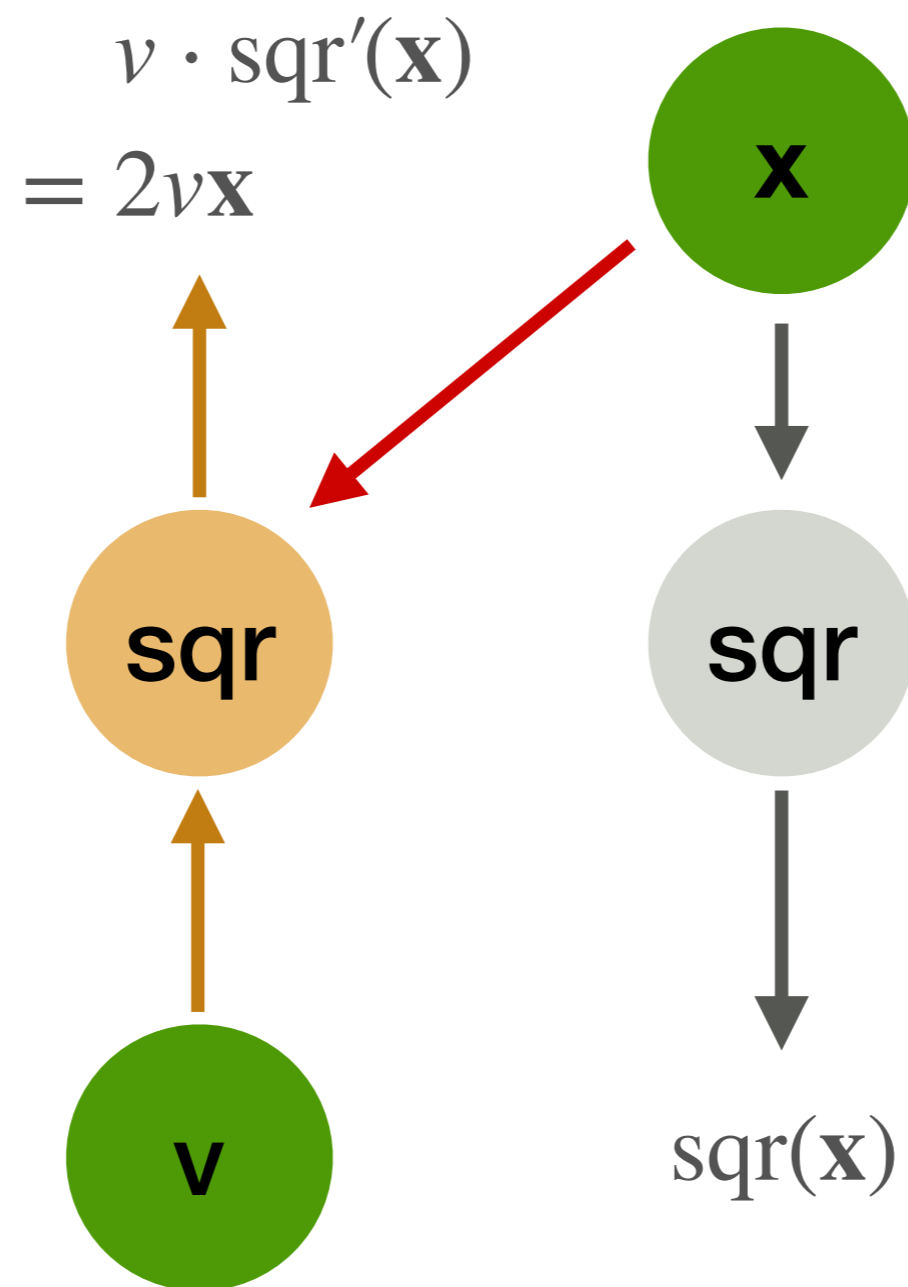
$\text{sub}'(y, \text{add}(\text{matmul}(\mathbf{w}, \mathbf{x}), b))$

$\text{add}'(\text{matmul}(\mathbf{w}, \mathbf{x}), b)$

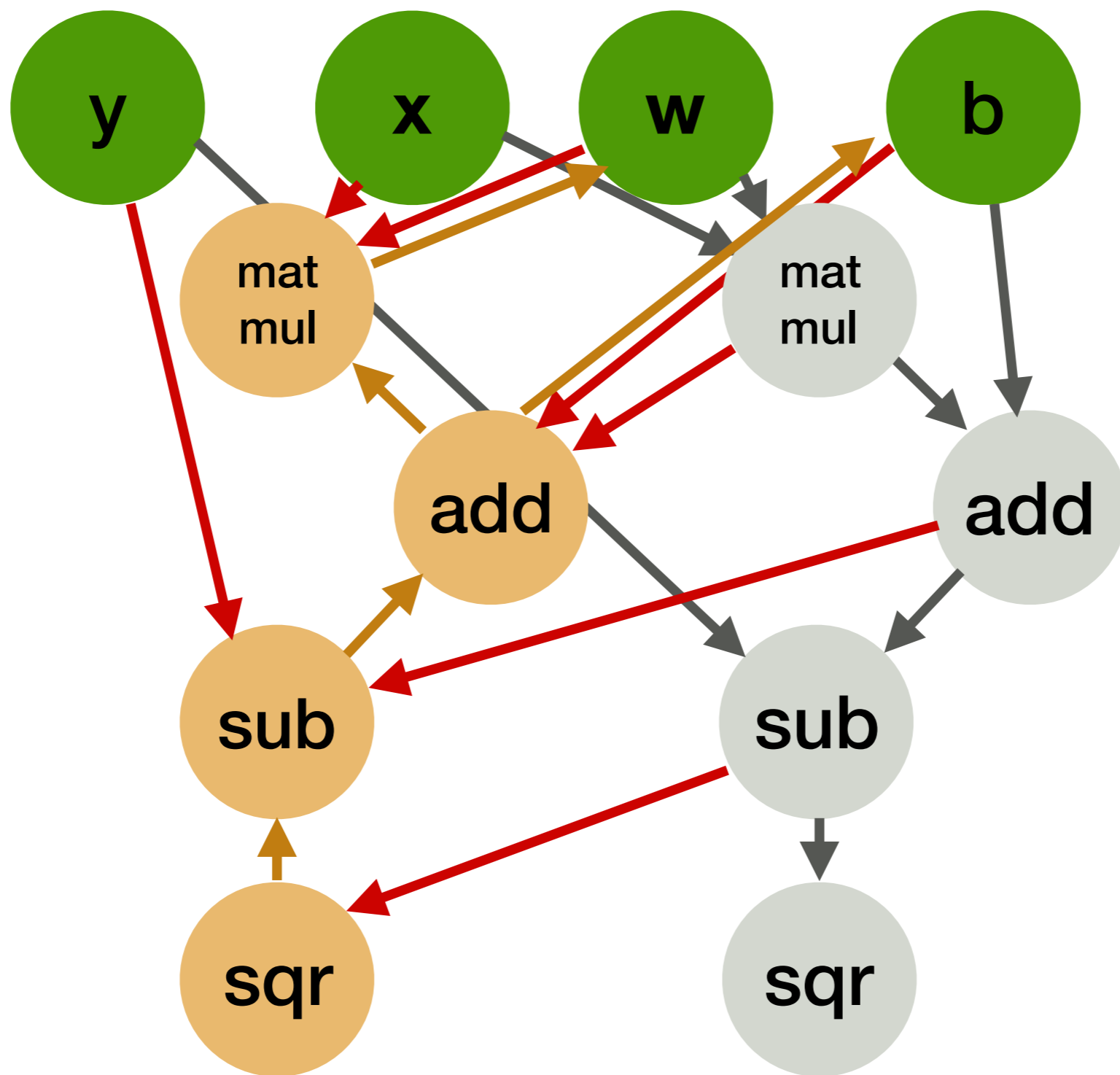
$\frac{d}{d\mathbf{w}} \text{matmul}(\mathbf{w}, \mathbf{x})$



# Gradient of linear regression



# Gradient of linear regression



forward

grad compute

backward