Working through least-squares math:

Let's say $S=\begin{bmatrix} s_1 & 1 \\ s_2 & 1 \end{bmatrix}$, $\theta=\begin{bmatrix} k \\ b \end{bmatrix}$, and $y=\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$. To be explicit, the equations we are considering are:

$$\begin{aligned} y_1 &= s_1k + b + \epsilon_1 \\ y_2 &= s_2k + b + \epsilon_2 \end{aligned} \longrightarrow \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} s_1 & 1 \\ s_2 & 1 \end{bmatrix}\begin{bmatrix} k \\ b \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}$$

① Start w/ $\|S\theta - y\|^2$. Show this equals $(S\theta - y)^T(S\theta - y)$.

$$\rightarrow \left\| \begin{bmatrix} s_1 & 1 \\ s_2 & 1 \end{bmatrix}\begin{bmatrix} k \\ 0 \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \right\|^2 = \left\| \begin{matrix} s_1k + b - y_1 \\ s_2k + b - y_2 \end{matrix} \right\|^2 = (s_1k + b - y_1)^2 + (s_2k + b - y_2)^2$$

same thing!

$$\rightarrow (S\theta - y)^T(S\theta - y) = \begin{bmatrix} s_1k + b - y_1 \\ s_2k + b - y_2 \end{bmatrix}^T \begin{bmatrix} s_1k + b - y_1 \\ s_2k + b - y_2 \end{bmatrix}$$

$$= \begin{bmatrix} s_1k + b - y_1 & s_2k + b - y_2 \end{bmatrix} \begin{bmatrix} s_1k + b - y_1 \\ s_2k + b - y_2 \end{bmatrix} = (s_1k + b - y_1)^2 + (s_2k + b - y_2)^2$$

② Expand $(S\theta - y)^T(S\theta - y)$

$$\Rightarrow (\theta^T S^T - y^T)(S\theta - y) = \theta^T S^T S\theta - \underbrace{\theta^T S^T y}_{= (S\theta)^T y \,=\, y^T(S\theta) \,=\, y^T S\theta} - y^T S\theta + y^T y$$

$$= \theta^T S^T S\theta - 2y^T S\theta + y^T y$$

③ Take gradient of $\theta^T S^T S\theta$ (w/ respect to $\theta$)

$$\Rightarrow \begin{bmatrix} k & b \end{bmatrix} \begin{bmatrix} s_1 & s_2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} s_1 & 1 \\ s_2 & 1 \end{bmatrix} \begin{bmatrix} k \\ b \end{bmatrix} = \begin{bmatrix} ks_1 + b & ks_2 + b \end{bmatrix} \begin{bmatrix} ks_1 + b \\ ks_2 + b \end{bmatrix} = (ks_1 + b)^2 + (ks_2 + b)^2$$

$$\frac{d}{dk} \rightarrow \frac{d}{dk}\left[(ks_1 + b)^2 + (ks_2 + b)^2\right] = 2(ks_1 + b)s_1 + 2(ks_2 + b)s_2$$

$$\frac{d}{db} \rightarrow \frac{d}{db}\left[(ks_1 + b)^2 + (ks_2 + b)^2\right] = 2(ks_1 + b) + 2(ks_2 + b)$$

$$\begin{bmatrix} \frac{d}{dk} \\ \frac{d}{db} \end{bmatrix} = \begin{bmatrix} 2\left[(ks_1 + b)s_1 + (ks_2 + b)s_2\right] \\ 2\left[(ks_1 + b) + (ks_2 + b)\right] \end{bmatrix}$$

$$= 2\begin{bmatrix} ks_1^2 + bs_1 + ks_2^2 + bs_2 \\ ks_1 + b + ks_2 + b \end{bmatrix}$$

$$= 2\underbrace{\begin{bmatrix} s_1^2 + s_2^2 & s_1 + s_2 \\ s_1 + s_2 & 1 + 1 \end{bmatrix}}\underbrace{\begin{bmatrix} k \\ b \end{bmatrix}}_{\theta}$$

$$\underbrace{\begin{bmatrix} s_1 & s_2 \\ 1 & 1 \end{bmatrix}}_{S^T}\underbrace{\begin{bmatrix} s_1 & 1 \\ s_2 & 1 \end{bmatrix}}_{S} = \begin{bmatrix} s_1^2 + s_2^2 & s_1 + s_2 \\ s_1 + s_2 & 1 + 1 \end{bmatrix}$$

$$= S^T S \theta \quad ✓$$

(4) Take gradient of $y^T S \theta$

$$\rightarrow \begin{bmatrix} y_1 & y_2 \end{bmatrix}\begin{bmatrix} s_1 & 1 \\ s_2 & 1 \end{bmatrix}\begin{bmatrix} k \\ b \end{bmatrix}$$

$$= \begin{bmatrix} y_1 & y_2 \end{bmatrix}\begin{bmatrix} s_1 k + b \\ s_2 k + b \end{bmatrix} = y_1 s_1 k + y_1 b + y_2 s_2 k + y_2 b$$

$\frac{d}{dk} \rightarrow y_1 s_1 + y_2 s_2$

$\frac{d}{db} \rightarrow y_1 + y_2$

$$\begin{bmatrix} \frac{d}{dk} \\ \frac{d}{db} \end{bmatrix} = \begin{bmatrix} y_1 s_1 + y_2 s_2 \\ y_1 + y_2 \end{bmatrix} = \underbrace{\begin{bmatrix} s_1 & s_2 \\ 1 & 1 \end{bmatrix}}_{S^T}\underbrace{\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}}_{y} = S^T y \quad ✓$$