Answer: Since $s_j \sim N(0,1)$ we have that $s_j = w_j^T x_i$ and $g'(s_j) = (2\pi)^{-1/2} e^{-\frac{1}{2}s_j^2}$. So then we get

$$l(W) = \sum_{i=1}^{n} \left(\log |W| - \frac{1}{2} \sum_{j=1}^{d} \log 2\pi + (w_j^T x_i)^2) \right)$$

So then taking ∇_W of the log likelihood to maximise it, noting that $\nabla_{W_{ab}} \sum_{j=1}^{d} (w_j^T x_i)^2 = 2(w_a^T x^{(i)}) x_b^{(i)}$ we get

$$\begin{aligned}
\nabla_W l(W) &= \sum_{i=1}^{n} \left(\frac{1}{|W|} \mathsf{adj}(W)^T - (W x_i) x_i^T \right) \\
&= nW^{-T} - W \sum_{i=1}^{n} x_i x_i^T \\
&= nW^{-T} - W X^T X
\end{aligned}$$

Setting this to zero we have an optimal matrix W satisfies the equation $nI = W^T W X^T X$, which we see is invariant under transformations that don't change $W^T W$, such as $\tilde{W} = PW$ for some orthogonal matrix P, then $\tilde{W}^T \tilde{W} = W^T P^T P W = W^T W$.