2. We know that Group Normalisation normalizes over a group of channels as follows:

$$\mu_i = \frac{1}{m}\sum_{k \in C_i} x_k \quad , \quad \sigma_i = \sqrt{\frac{1}{m}\sum_{k \in C_i}(x_k - \mu_i)^2}$$

Where C_i is the channel group from the parameter vector C, $m = |C_i|$ (size)

Which gives us for $i \in S_k$

$$\hat{x}_i = \frac{1}{\sigma_i}(x_i - \mu_i)$$ as the normalized feature for C_k group of channels

It is easy to see that if all channels are put into a single group, we normalize the whole feature This means we have achieved Layer Normalization for $N = 1$.

Alternatively, if all channels are in separate groups, we are normalizing for each channel in the feature vector. This means we have achieved Instance Normalization for $|C_i| = 1 \ \forall \ 1 \leq i \leq N$.
In this case N = ~~total~~ number of channels, i.e. $N = \sum |C_i|$