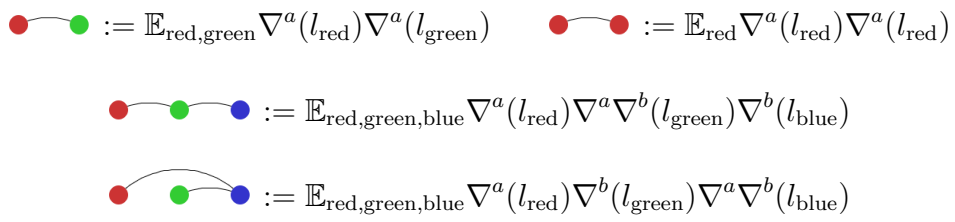
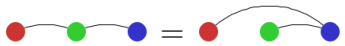
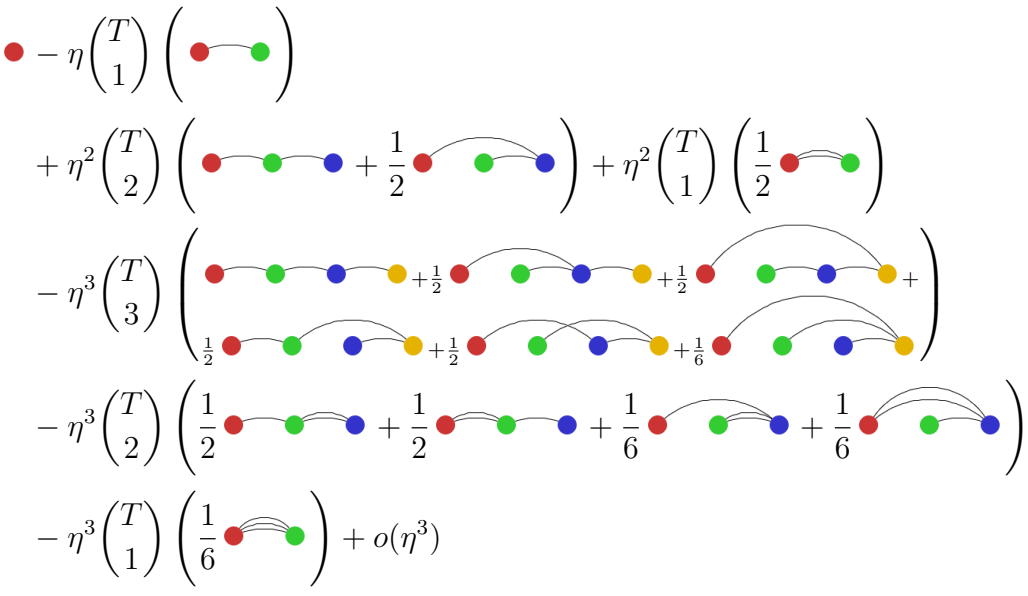
0. Diagrams for SGD Test Loss

We use diagrams for book-keeping of the Taylor terms (of test loss at a fixed initial weight). Each color in a diagram represents a data value and thus selects a particular loss function from the data-parameterized distribution of loss functions. Each node in a diagram represents a tensor obtained from derivatives of the loss function corresponding to that node's color. It is a diagram's edges in a diagram that specify those derivative tensors. We understand the edges as directed from left to right, with the source acting on the target by differentiation. Finally, a diagram's value is the expectation over all iid assigments of data to the colors. Thus,



We see that  gives the trace of the covariance of gradients. Moreover,  , illustrating how diagram notation can streamline computation by helping to group terms. However, we caution that a diagram's value generally depends on a diagram's digraph structure, not just its undirected structure. For example:

Thus prepared, we may expand the test loss after T updates, each with batch-size 1 sampled without replacement. The recipe is to draw all the diagrams whose underlying poset has a unique rightmost element. Each node in the diagram contributes a symmetry factor $1/i!$ where i is the node's in-degree. On top of that, a diagram with a edges and v vertices has an overall combinatorial weight of $(-\eta)^a \binom{T}{v-1}$. We obtain:

$\mathbb{E}\,(\text{SGD Test Loss}) =$ 

And a routine grouping of terms yields:

$$\cdots = \bullet - \eta\binom{T}{1}\left(\bullet\!\!-\!\!\bullet\right) + \eta^2\binom{T}{2}\left(\frac{3}{2}\,\bullet\!\!-\!\!\bullet\!\!-\!\!\bullet\right) + \eta^2\binom{T}{1}\left(\frac{1}{2}\,\bullet\!\!=\!\!\bullet\right) - \eta^3\binom{T}{3}\left(\frac{5}{2}\,\bullet\!\!-\!\!\bullet\!\!-\!\!\bullet\!\!-\!\!\bullet + \frac{1}{2}\,\bullet\!\!-\!\!\bullet\!\!-\!\!\bullet\!\!-\!\!\bullet\;\; \frac{1}{6}\,\bullet\!\!-\!\!\bullet\!\!-\!\!\bullet\!\!-\!\!\bullet\right) - \eta^3\binom{T}{2}\left(\bullet\!\!-\!\!\bullet\!\!-\!\!\bullet + \frac{5}{6}\,\bullet\!\!-\!\!\bullet\!\!-\!\!\bullet\right) - \eta^3\binom{T}{1}\left(\frac{1}{6}\,\bullet\!\!=\!\!\bullet\right) + o(\eta^3)$$