

let's assume we have a subspace  $S$  of rank  $r \ll n$  out of  $R^n$ , and are randomly generating a weight matrix to get a mask.

the mask will NOT be horizontally dependent: to get the  $(i,j)$  th member of the mask, we take  $S_{i,:} * w_j$ , return 1 if it is positive, and return 0 if it is negative. The  $(i,k)$ th member of the mask is similarly based on the sign of  $S_{i,:} * w_k$ . This is equivalent to determining whether the angle between  $S_{i,:}$  and each of two different independent random vectors is  $<$  or  $> \frac{\pi}{2}$ . It's pretty clear from here that since the two random vectors are independent from one another, their alignments with some fixed vector will also be independent from one another, so the horizontal elements are independent. (kinda like how if  $a$  and  $b$  are independent RVs,  $a+1$  and  $b+1$  are also independent)

that was just pairwise independence but the same logic holds for any given row of mask members

However, the mask is vertically dependent. the  $(i,j)$  th member of the mask is based on the sign of  $S_{i,:} * w_j$  and the  $(k,j)$ th member of the mask is based on the sign of  $S_{k,:} * w_j$

The issue here is that  $S_{i,:}$  and  $S_{k,:}$  are not always independent. Let's try some examples to build some intuition on the subject. Suppose that  $S$  is some collection  $[e_{x_1}, \dots, e_{x_r}]$ , where  $e_i$  is the  $i$ th elementary vector s.t.  $I_n = [e_1, \dots, e_n]$ , and  $x \in [n]$ . So  $S$  is a column subset of  $I_n$ . This is important because we are taking the ReLU across  $I_n$  essentially, as our dividing planes are oriented perpendicular to each  $e_i$ .

Well,  $S_{i,:} * w_j = w_{x_i,j}$  and  $S_{k,:} * w_j = w_{x_k,j}$ . These are pretty clearly independent, and by this logic, the entire column is independent given this  $S$ .

So this is primarily dependent on the orthogonality of the rows of  $S$ . If we only have orthogonal rows, then we see different aspects of the information of the random point, so there is no information leakage on those rows. However, there's heavy overlap on the remaining rows (which are all 1)

Of course, presumably this matrix is tall, so there is necessarily some overlap. I think that this overlap, and hence the distribution of the covariance of the rows, can become more or less pronounced depending on the deformation of the space. For example, the above example has a cluster of extremely dependent rows (the 0s) and a cluster of extremely independent rows (the rows which contain a 1).

On the other hand, we can compose the matrix of

$$\begin{bmatrix} I_r \\ I_r \end{bmatrix}$$

...

$I_r]$

This

It feels like we can say that the dependence of the entire row can be defined by the distribution of pairwise dependencies among individual entries (so k-wise independence implies total independence). I have no idea how to do that, but it seems right. TODO actually look into that lol

The covariance between two of these random variables, through the ReLU mask, is dependent on the angle between the corresponding rows of the projection matrix.

Because we want the vertical covariance of the columns to be distributed evenly to prevent outliers in the mask, this implies that subspaces with a small variance of "row angle from the mean" should perform better in PRL-RMC. I actually don't know whether it's better to have weak dependence across the mask or strong dependence in fewer places in the mask.

to get a more exact idea: we can get each column by sampling a random element in  $R^r$ . We are essentially

Basically, the problem we will have once we prove the mask-based bounds is that we don't know how much the mask varies for a fixed subspace which we are trying to recover. If we can get an idea of the mean & variance of the (mask variance within a given subspace) across all subspaces, then we can get a statistical bound that way?