1 Brief Overview of Problem

Suppose we have an unconstrained freqency update $\boldsymbol{B}$.

$$\hat{\boldsymbol{B}} = \begin{bmatrix} \boldsymbol{B}_{1,1} & \dots & \boldsymbol{B}_{1,K} \\ \vdots & \ddots & \vdots \\ \boldsymbol{B}_{C,1} & \dots & \boldsymbol{B}_{C,K} \end{bmatrix} \tag{1}$$

where all $\boldsymbol{B}_{c,k}$ are diagonal.

We want a similar but constrained frequency update $\Delta\hat{\boldsymbol{D}}$:

$$\Delta\hat{\boldsymbol{D}} = \begin{bmatrix} \Delta\hat{\boldsymbol{D}}_{1,1} & \dots & \Delta\hat{\boldsymbol{D}}_{1,K} \\ \vdots & \ddots & \vdots \\ \Delta\hat{\boldsymbol{D}}_{C,1} & \dots & \Delta\hat{\boldsymbol{D}}_{C,K} \end{bmatrix} \tag{2}$$

where all $\Delta\hat{\boldsymbol{D}}_{c,k}$ are diagonal and some additional constraints are met.

Let $\Delta\hat{\boldsymbol{D}}_{c,k}[\ell]$ be the ℓth diagonal element of $\Delta\hat{\boldsymbol{D}}_{c,k}$ (in other words, the ℓth frequency component of the cth channel of the kth dictionary).

Let

$$\Delta\hat{\boldsymbol{D}}[\ell] = \begin{bmatrix} \Delta\hat{\boldsymbol{D}}_{1,1}[\ell] & \dots & \Delta\hat{\boldsymbol{D}}_{1,K}[\ell] \\ \vdots & \ddots & \vdots \\ \Delta\hat{\boldsymbol{D}}_{C,1}[\ell] & \dots & \Delta\hat{\boldsymbol{D}}_{C,K}[\ell] \end{bmatrix} \tag{3}$$

We want $\Delta\hat{\boldsymbol{D}}[\ell]$ to be low rank for all ℓ.

We also want $\Delta\hat{\boldsymbol{D}}_{c,k}$ to have local and small spatial support (This is an affine constraint.).

2 Proposed Solution

Currently, we have CK low-rank constraints. Let's work with a single, rank-one constraint.

We can choose from the following:

1. $$\operatorname{rank}(\begin{bmatrix} \Delta\hat{\boldsymbol{D}}[0] & \dots & \Delta\hat{\boldsymbol{D}}[L-1] \end{bmatrix}) = 1 \tag{4}$$

   Or

2. $$\operatorname{rank}\left(\begin{bmatrix} \Delta\hat{\boldsymbol{D}}[0] \\ \vdots \\ \Delta\hat{\boldsymbol{D}}[L-1] \end{bmatrix}\right) = 1 \tag{5}$$

Either of these options is sufficient to ensure that the rank-one constraints from the original problem are met.

Now, using either of these options, we can apply the affine projection first, and then apply low-rank constraint, and the result should meet both constraints.

If we want to work with a higher rank, we can solve for a rank-one approximation, subtract it off, and then solve for a second rank-one approximation. I suspect any rank higher than two wouldn't be worth the effort, but rank 2 might be worthwhile, since we could use both low-rank options.

To summarize:

$$\begin{aligned} \min_{\Delta\hat{\boldsymbol{D}}} \| \begin{bmatrix}\boldsymbol{B}[0] & \dots & \boldsymbol{B}[L-1]\end{bmatrix} - \begin{bmatrix}\Delta\hat{\boldsymbol{D}}[0] & \dots \Delta\hat{\boldsymbol{D}}[L-1]\end{bmatrix} \|_F^2 \\ \text{subject to } \operatorname{rank}(\begin{bmatrix}\Delta\hat{\boldsymbol{D}}[0] & \dots & \Delta\hat{\boldsymbol{D}}[L-1]\end{bmatrix}) = 1 \\ (\mathbf{I}-\boldsymbol{P})\mathcal{F}^{-1}(\Delta\hat{\boldsymbol{D}}_{c,k}) = \mathbf{0} \end{aligned} \tag{6}$$

3 Additional Discussion

So, there are a couple pieces missing here. One is normalization. Normalizing the update would be easy. Unfortunately, that's not what we want. What we want is a normalized dictionary. Perhaps we can track normalization separately from the dictionary itself? That is, if we keep track of the norm of the atoms, we can just divide the corresponding coefficients by that number. Or we could somehow rescale the update and the dictionary to get a normalized result?

Regardless, I think if we can solve everything else, we should be able to come up with some way to handle the normalization problem.

I want to address the more glaring issue: replacing the many rank-one constraints with a single, larger, rank-one constraint may be pretty far from optimal. However, I suspect the approximation may actually be very close, if not optimal.

Consider a simplified problem with similar properties. Let's keep things rank 1, and keep all of our channels and filters. Suppose we replace the affine constraint with a simpler restriction: we require the diagonals of the submatrices to be in the span of a set of two vectors:

$$f_1 = \begin{bmatrix}1\\0\\1\end{bmatrix} \tag{7}$$

$$f_2 = \begin{bmatrix}0\\1\\-1\end{bmatrix} \tag{8}$$

Then, our problem simplifies to the following:

$$
\begin{aligned}
&\min_{\boldsymbol{D}} \|\boldsymbol{B}[0] - \Delta\hat{\boldsymbol{D}}[0]\|_F^2 + \|\boldsymbol{B}[1] - \Delta\hat{\boldsymbol{D}}[1]\|_F^2 + \|\boldsymbol{B}[2] - (\Delta\hat{\boldsymbol{D}}[0] - \Delta\hat{\boldsymbol{D}}[1])\|_F^2 \\
\text{subject to } \; &\operatorname{rank}(\Delta\hat{\boldsymbol{D}}[0]) = 1 \\
&\operatorname{rank}(\Delta\hat{\boldsymbol{D}}[1]) = 1 \\
&\operatorname{rank}(\Delta\hat{\boldsymbol{D}}[0] - \Delta\hat{\boldsymbol{D}}[1]) = 1
\end{aligned}
\tag{9}
$$

In this case, my approach will yield the optimal solution. (I convinced myself looking at a simple 4×4 case with some illustrations and contemplating similar triangles.)

How good is this approximation under more complicated conditions? I'm not sure. I wouldn't be surprised if it's optimal there too, but it's easier to conjecture than to prove.