A. Proofs

A.1. Proposition 2

Since $B \in \mathcal{B}$, we have

$$\|B(:,r)\|_2 \leq 1, \forall r = 1, \dots, R, \tag{17}$$

so that

$$\|B\|_F \leq \sqrt{R}. \tag{18}$$

Therefore, when

(i) $W_i \in \mathcal{W}_{\ell 1}$: Use (18), we can write

$$\|BW_i(:,k)\|_2 \leq \|B\|_F \|W_i(:,k)\|_2 \leq \sqrt{R}\|W_i(:,k)\|_2.$$

Thus, if $\sqrt{R}\|W_i(:,k)\|_2 \leq 1$, then

$$\|BW_i(:,k)\|_2 \leq \sqrt{R}\|W_i(:,k)\|_2 \leq 1,$$

which means $\|BW_i(:,k)\|_2 \leq 1$.

(ii) $W_i \in \mathcal{W}_{\ell 2}$: First, we have $BW_i(:,k) = \sum_{r=1}^R W_i(r,k)B(:,r)$. Then, by Cauchy-Schwarz inequality, we have

$$\begin{aligned}\|BW_i(:,k)\|_2 &= \left\|\sum_{r=1}^R W_i(r,k)B(:,r)\right\|_2 \\ &\leq \sum_{r=1}^R \|W_i(r,k)B(:,r)\|_2 \\ &\leq \sum_{r=1}^R |W_i(r,k)|\|B(:,r)\|_2 \\ &= \sum_{r=1}^R |W_i(r,k)| = \|W_i(:,k)\|_1, \quad (19)\end{aligned}$$

where (19) is due to (17). Therefore, if $\|W_i(:,k)\|_1 \leq 1$, $\|BW_i(:,k)\|_2 \leq 1$ holds.

A.2. Proposition 3

Let $\tilde{B}(:,r) \equiv \mathcal{F}(B(:,r))$, (12) is equivalent to (11) since the following equations hold:

$$\begin{aligned}&f_i(B, W_i, Z_i) \\ &= \frac{1}{2}\left\|x_i - \sum_{k=1}^K \left(\sum_{r=1}^R W_i(r,k)B(:,r)\right) * Z_i(:,k)\right\|_2^2, \\ &= \frac{1}{2}\left\|x_i - \sum_{r=1}^R B(:,r) * \left(\sum_{k=1}^K W_i(r,k)Z_i(:,k)\right)\right\|_2^2, \quad (20) \\ &= \frac{1}{2P}\left\|\mathcal{F}(x_i) - \sum_{r=1}^R \mathcal{F}(B(:,r)) \odot \mathcal{F}(Z_i W_i^\top(:,r))\right\|_2^2, \quad (21) \\ &= \tilde{f}_i(\tilde{B}, W_i, Z_i),\end{aligned}$$

where (20) is due to

$$\begin{aligned}&\sum_{k=1}^K \left(\sum_{r=1}^R W_t(r,k)B(:,r)\right) * Z_t(:,k) \\ &= \sum_{r=1}^R B(:,r) * \left(\sum_{k=1}^K W_i(r,k)Z_i(:,k)\right).\end{aligned}$$

Then, (21) comes from the convolution theorem (Mallat, 1999), i.e.,

$$\mathcal{F}(B(:,r) * Z_i W_i^\top(:,r)) = \mathcal{F}(B(:,r)) \odot \mathcal{F}(Z_i W_i^\top(:,r)),$$

where $B(:,r)$ and $Z_i W_i^\top(:,r)$ are first zero-padded to P-dimensional, and the Parseval's theorem (Mallat, 1999): $\frac{1}{P}\|\mathcal{F}(x)\|_2^2 = \|x\|_2^2$ where $x \in \mathbb{R}^P$.

As for constraints, when B is transformed to the frequency domain, it is padded from M dimensional to P dimensional. Thus, we use $\mathcal{C}(\mathcal{F}^{-1}(\tilde{B}))$ to crop the extra dimensions to get back the original support.