Let $\Xi^*_{i,n} = \Xi(X^*_{i,n})$. By the triangle inequality,

$$\begin{aligned}
& \mathbb{P}\left(\|\textstyle\sum_{i=1}^n \Xi_{i,n}\| \geq 6t\right) \\
\leq\ & \mathbb{P}(\|\textstyle\sum_{i=1}^{[n/q]q} \Xi^*_{i,n}\| + \|\sum_{i\in I_r} \Xi_{i,n}\| + \|\sum_{i=1}^{[n/q]q}(\Xi^*_{i,n} - \Xi_{i,n})\| \geq 6t) \qquad (118)\\
\leq\ & \tfrac{n}{q}\beta(q) + \mathbb{P}\left(\|\textstyle\sum_{i\in I_r} \Xi_{i,n}\| \geq t\right) + \mathbb{P}\left(\|\sum_{i\in I_e} \Xi^*_{i,n}\| \geq t\right) + \mathbb{P}\left(\|\sum_{i\in I_o} \Xi^*_{i,n}\| \geq t\right)
\end{aligned}$$

To control the last two terms we apply Theorem 4.1, recognizing that $\sum_{i\in I_e} \Xi^*_{i,n}$ and $\sum_{i\in I_o} \Xi^*_{i,n}$ are each the sum of fewer than $[n/q]$ independent $d_1 \times d_2$ matrices, namely $W^*_k = \sum_{i=(k-1)q+1}^{kq} \Xi^*_{i,n}$. Moreover each W^*_k satisfies $\|W^*_k\| \leq qR_n$ and $\max\{\|E[W^*_k W^{*\prime}_k]\|, \|E[W^{*\prime}_k W^*_k]\|\} \leq q^2 s_n$. Theorem 4.1 then yields

$$\mathbb{P}\left(\left\|\sum_{i\in I_e} \Xi^*_{i,n}\right\| \geq t\right) \leq (d_1 + d_2)\text{exp}\left(\frac{-t^2/2}{nqs_n^2 + qR_n t/3}\right) \qquad (119)$$

and similarly for I_o. ■

Proof of Corollary 4.2. Follows from Theorem 4.2 with $t = Cs_n\sqrt{nq\log(d_1+d_2)}$ for sufficiently large C, and the conditions $\frac{n}{q}\beta(q) = o(1)$ and $R_n\sqrt{q\log(d_1+d_2)} = o(s_n\sqrt{n})$. ■

Proof of Lemma 4.1. Let $G = E[b^K_w(X_i)b^K_w(X_i)']$. Since $B_{K,w} = clsp\{b_{K1}w_n, \ldots, b_{KK}w_n\}$, we have:

$$\begin{aligned}
& \sup\{|\tfrac{1}{n}\textstyle\sum_{i=1}^n b(X_i)^2 - 1| : b \in B_{K,w}, E[b(X)^2] = 1\} \\
=\ & \sup\{|c'(B_w'B_w/n - G)c| : c \in \mathbb{R}^K, \|G^{1/2}c\| = 1\} \qquad (120)\\
=\ & \sup\{|c'G^{1/2}(G^{-1/2}(B_w'B_w/n)G^{-1/2} - I_K)G^{1/2}c| : c \in \mathbb{R}^K, \|G^{1/2}c\| = 1\} \qquad (121)\\
=\ & \sup\{|c'(\widetilde{B}_w'\widetilde{B}_w/n - I_K)c| : c \in \mathbb{R}^K, \|c\| = 1\} \qquad (122)\\
=\ & \|\widetilde{B}_w'\widetilde{B}_w/n - I_K\|_2^2 \qquad (123)
\end{aligned}$$

as required. ■

7.4 Proofs for Section 5

We first present a general result that allows us to bound the L^∞ operator norm of the $L^2(X)$ projection P_K onto a linear sieve space $B_K \equiv clsp\{b_{K1}, \ldots, b_{KK}\}$ by the ℓ^∞ norm of the inverse of its corresponding Gram matrix.

Lemma 7.1 *If there exists a sequence of positive constants $\{c_K\}$ such that (i) $\sup_{x\in\mathcal{X}} \|b^K(x)\|_{\ell^1} \lesssim c_K$*