A Appendix

We now prove Theorem 5.

Proof. First, note that our expected regret, when $\Sigma_t = I$ is just:

$$\mathbb{E}\|\hat{w}_t - \beta_t\|^2$$

We can also choose our coordinate system so that $[x]_{s,\perp}$ is (statistically) uncorrelated with $[x]_{s,t}$. Then our estimate of w_t is just $\Sigma_{s\to t}^{-1}\widehat{\mathbb{E}}[(\Pi_t X)Y]$, where

$$\mathbb{E}[(\Pi_t X)Y] = \frac{1}{n}\sum_{(x,y)\in T}(\Pi_t x)y$$

where (x, y) are the values in our training set. Define η_x for x in our training set by $y_x = \mathbb{E}[Y|x] + \eta_x$, where y_x is the value on training sample x. By assumption, $\mathbb{E}\eta_x^2 \leq 1$. If we rotate to a coordinate system where $\Sigma_{s\to t}$ is diagonal, then:

$$\begin{aligned}
\mathbb{E}\|\hat{w}_t - \beta_t\|^2 &= \mathbb{E}||\widehat{\mathbb{E}}[(\Pi_t X)Y] - \mathbb{E}[(\Pi_t X)Y]||^2_{\Sigma_{s\to t}^{-2}} \\
&= \mathbb{E}\left[\sum_i \frac{(\widehat{\mathbb{E}}[(\Pi_t X)_i Y] - \mathbb{E}[(\Pi_t X)_i Y])^2}{\lambda_i^2}\right] \\
&= \mathbb{E}\left[\sum_i \frac{\left(\frac{1}{n}\sum_{x\in T}\eta_x(\Pi_t x)_i\right)^2}{\lambda_i^2}\right] \\
&= \mathbb{E}\left[\sum_i \frac{\frac{1}{n^2}\sum_{x\in T}\eta_x^2(\Pi_t x)_i^2}{\lambda_i^2}\right] \\
&\leq \sum_i \frac{\frac{1}{n^2}\sum_{x\in T}(\Pi_t x)_i^2}{\lambda_i^2} \\
&= \frac{1}{n}\sum_i \frac{1}{\lambda_i}
\end{aligned}$$

where the third to last step uses independence and the final step uses the definition of λ_i. □