Appendix A: Proof of the Inequality and the Expression for B^2_{COST} in Thm 1

This appendix proves the following inequality from the proof sketch of Thm 1 and derives the corresponding expression for B^2_{cost}:

$$\mathbb{E}_{\theta'_T\sim\mathcal{D}_{\theta_T}}\left[e^{\lambda(C(K)-\hat{C}(K))}\right]\leq e^{\frac{\lambda^2 B^2_{\text{cost}}}{8m}}, \tag{1}$$

where

$$B^2_{\text{cost}}:=\sup_{K\in\mathcal{K}}\sigma_w^4 d_x\cdot\rho_S(K)^2\cdot\left(128\sum_{i=0}^{T}\rho_M(K)^{4(T-i)}\right.$$
$$\left.+64\sum_{0\leq i<j\leq T}\rho_M(K)^{4T-2(i+j)}\right), \tag{2}$$

for $|\lambda|\leq\inf_{K\in\mathcal{K}}\frac{1}{4\sigma_w^2\cdot\rho_S(K)\cdot\rho_M(K)^{2T}}$.

Proof: For notational simplicity, we write $\mathbb{E}$ in place of $\mathbb{E}_{\theta'_T\sim\mathcal{D}_{\theta_T}}$ throughout the proof. The closed-loop system

$$x_{t+1}=Ax_t+Bu_t+w_t,$$
$$u_t=-Kx_t,$$

can be written in the form

$$x_{t+1}=Mx_t+w_t,$$

where $M:=A-BK$. We assume the system starts from the initial state $x_0=0$. The control performance is measured by the quadratic cost:

$$C_q(K,\theta_T)=\sum_{t=0}^{T}x_t^\top Sx_t,$$

where $S:=Q+K^\top RK$. We first expand x_t in terms of the noise variables:

$$x_t=\sum_{k=0}^{t-1}M^{t-1-k}w_k.$$

Substituting this into the cost expression, we obtain:

$$C_q(K,\theta_T)=\sum_{t=0}^{T}\sum_{i,j=0}^{t-1}w_i^\top M^{t-1-i\top}SM^{t-1-j}w_j.$$

Define $D_{t,ij}:=M^{t-1-i\top}SM^{t-1-j}$. Then we have

$$C_q(K,\theta_T)=\sum_{t=0}^{T}\sum_{i,j=0}^{t-1}w_i^\top D_{t,ij}w_j.$$

By grouping terms over time, we define:

$$H_i=\sum_{t=i+1}^{T}D_{t,ii},\quad H_{ij}=\sum_{t=\max(i,j)+1}^{T}D_{t,ij},\quad i\neq j,$$

and we rewrite the cost as:

$$C_q(K,\theta_T)=\sum_{i=0}^{T-1}w_i^\top H_iw_i+\sum_{0\leq i<j\leq T-1}\left(w_i^\top H_{ij}w_j+w_j^\top H_{ij}^\top w_i\right).$$

Using the spectral decomposition $H_i=U\Lambda U^\top$ and defining $\tilde{w}_i:=U^\top w_i$, we have

$$w_i^\top H_iw_i=\tilde{w}_i^\top\Lambda\tilde{w}_i,$$

where $\Lambda=\text{diag}(\mu_1,\ldots,\mu_{d_x})$ is a diagonal matrix of eigenvalues. Since Λ is diagonal and positive semidefinite, we can upper bound the quadratic form as

$$\tilde{w}_i^\top\Lambda\tilde{w}_i=\sum_{j=1}^{d_x}\mu_j\tilde{w}_{i,j}^2\leq\|\Lambda\|\cdot\sum_{j=1}^{d_x}\tilde{w}_{i,j}^2=\|\Lambda\|\cdot\|\tilde{w}_i\|_2^2,$$

due to $\|\Lambda\|=\max_j\mu_j$. The inequality follows from the fact that each $\mu_j\leq\|\Lambda\|$.

Since the noise vector w_i has independent sub-Gaussian entries with parameter σ_w^2, and $U\in\mathbb{R}^{d_x\times d_x}$ is an orthogonal matrix, then for any $v\in\mathbb{R}^{d_x}$ with $\|v\|_2=1$, by Def. 1 and Def. 2 in the main body of the paper, which define sub-Gaussian random variables and sub-Gaussian random vectors, we have

$$\mathbb{E}\left[e^{\lambda v^\top\tilde{w}_i}\right]=\mathbb{E}\left[e^{\lambda(Uv)^\top w_i}\right]\leq\exp\left(\frac{\lambda^2\sigma_w^2}{2}\right),$$

where we used that Uv is a unit vector and that linear combinations of independent sub-Gaussian variables remain sub-Gaussian. Hence, $\tilde{w}_i$ is a sub-Gaussian vector with parameter σ_w^2.

We note that the definition of a sub-Gaussian vector guarantees sub-Gaussian concentration along any unit direction $v\in\mathbb{R}^{d_x}$. In particular, taking $v=e_j$, the j-th standard basis vector, selects the j-th coordinate of $\tilde{w}_i$, i.e., $\tilde{w}_{i,j}=e_j^\top\tilde{w}_i$. Since $\|e_j\|_2=1$, by the Def. 1, applying the inequality above with $v=e_j$ gives:

$$\mathbb{E}\left[e^{\lambda\tilde{w}_{i,j}}\right]=\mathbb{E}\left[e^{\lambda e_j^\top\tilde{w}_i}\right]\leq\exp\left(\frac{\lambda^2\sigma_w^2}{2}\right),$$

which shows that each entry $\tilde{w}_{i,j}$ is a sub-Gaussian random variable with parameter σ_w^2.

Since each $\tilde{w}_{i,j}$ is a sub-Gaussian variable with parameter σ_w^2, we apply the result from [1, Appendix B], which gives:

$$\mathbb{E}\left[e^{\lambda(\tilde{w}_{i,j}^2-\mathbb{E}[\tilde{w}_{i,j}^2])}\right]\leq\exp(16\lambda^2\sigma_w^4).$$

Let $\Lambda=\text{diag}(\mu_1,\ldots,\mu_{d_x})$, where μ_i is the i^{th} eigenvalue of H_i. Then we have

$$\tilde{w}_i^\top\Lambda\tilde{w}_i=\sum_{j=1}^{d_x}\mu_j\tilde{w}_{i,j}^2,$$

and we can upper bound the centered sum as

$$\sum_{j=1}^{d_x}\mu_j(\tilde{w}_{i,j}^2-\mathbb{E}[\tilde{w}_{i,j}^2])\leq\|\Lambda\|\cdot\sum_{j=1}^{d_x}(\tilde{w}_{i,j}^2-\mathbb{E}[\tilde{w}_{i,j}^2]),$$

due to $\|\Lambda\|=\max_j|\mu_j|$. Applying the bound for each $\tilde{w}_{i,j}$ and combining, we obtain:

$$\mathbb{E}\left[e^{\lambda(\tilde{w}_i^\top\Lambda\tilde{w}_i-\mathbb{E}[\tilde{w}_i^\top\Lambda\tilde{w}_i])}\right]=\mathbb{E}\left[e^{\lambda(w_i^\top H_iw_i-\mathbb{E}[w_i^\top H_iw_i])}\right]$$
$$\leq\exp\left(16\lambda^2\|\Lambda\|^2\sigma_w^4d_x\right),$$

for all λ such that $|\lambda|\leq\frac{1}{4\sigma_w^2\|\Lambda\|}$, by the result from [1, Appendix B]. That is,

$$\mathbb{E}\left[e^{\lambda(w_i^\top H_iw_i-\mathbb{E}[w_i^\top H_iw_i])}\right]\leq\exp\left(16\lambda^2\|H_i\|^2\sigma_w^4d_x\right),$$

for all λ such that $|\lambda| \leq \frac{1}{4\sigma_w^2\|H_i\|}$, by the result from [1, Appendix B].

We now provide a detailed derivation for bounding the moment generating function (MGF) of the cross term $w_i^\top H_{ij} w_j$, where w_i and w_j are independent random vectors with sub-Gaussian entries and $i \neq j$. Our goal is to upper bound the centered MGF:

$$\mathbb{E}\left[e^{\lambda(w_i^\top H_{ij} w_j - \mathbb{E}[w_i^\top H_{ij} w_j])}\right].$$

We show how to express the bilinear form $w_i^\top H_{ij} w_j$ as a quadratic form, which allows us to apply the result from [1, Appendix B].

Let $\zeta_{ij} \in \mathbb{R}^{2d_x}$ be the concatenated vector

$$\zeta_{ij} := \begin{bmatrix} w_i \\ w_j \end{bmatrix},$$

and define the symmetric block matrix $\Psi_{ij} \in \mathbb{R}^{2d_x \times 2d_x}$ as

$$\Psi_{ij} := \begin{bmatrix} 0 & H_{ij} \\ H_{ij}^\top & 0 \end{bmatrix}.$$

We compute the quadratic form $\zeta_{ij}^\top \Psi_{ij} \zeta_{ij}$:

$$\begin{aligned} \zeta_{ij}^\top \Psi_{ij} \zeta_{ij} &= \begin{bmatrix} w_i^\top & w_j^\top \end{bmatrix} \begin{bmatrix} 0 & H_{ij} \\ H_{ij}^\top & 0 \end{bmatrix} \begin{bmatrix} w_i \\ w_j \end{bmatrix} \\ &= w_i^\top H_{ij} w_j + w_j^\top H_{ij}^\top w_i \\ &= 2 w_i^\top H_{ij} w_j, \end{aligned}$$

where the last step uses that both terms are scalars and equal.

Thus, we have:

$$w_i^\top H_{ij} w_j = \frac{1}{2} \zeta_{ij}^\top \Psi_{ij} \zeta_{ij}.$$

This shows that the centered bilinear form can be written as a centered quadratic form over the augmented vector ζ_{ij}.

$$w_i^\top H_{ij} w_j - \mathbb{E}[w_i^\top H_{ij} w_j] = \frac{1}{2}\left(\zeta_{ij}^\top \Psi_{ij} \zeta_{ij} - \mathbb{E}[\zeta_{ij}^\top \Psi_{ij} \zeta_{ij}]\right).$$

Then we have

$$\mathbb{E}\left[e^{\lambda(w_i^\top H_{ij} w_j - \mathbb{E}[w_i^\top H_{ij} w_j])}\right] = \mathbb{E}\left[e^{\frac{\lambda}{2}(\zeta_{ij}^\top \Psi_{ij} \zeta_{ij} - \mathbb{E}[\zeta_{ij}^\top \Psi_{ij} \zeta_{ij}])}\right].$$

We now perform a spectral decomposition of the symmetric matrix Ψ_{ij}: let

$$\Psi_{ij} = \mathcal{U}_{ij} \Sigma_{ij} \mathcal{U}_{ij}^\top,$$

where $\Sigma_{ij} = \mathrm{diag}(\nu_1, \ldots, \nu_{2d_x})$ is a diagonal matrix of eigenvalues and $\mathcal{U}_{ij} \in \mathbb{R}^{2d_x \times 2d_x}$ is orthogonal. Define the rotated noise vector

$$\tilde{\zeta}_{ij} := \mathcal{U}_{ij}^\top \zeta_{ij} \in \mathbb{R}^{2d_x}.$$

Then we can rewrite the quadratic form as

$$\zeta_{ij}^\top \Psi_{ij} \zeta_{ij} = \tilde{\zeta}_{ij}^\top \Sigma_{ij} \tilde{\zeta}_{ij} = \sum_{k=1}^{2d_x} \nu_k \tilde{\zeta}_{ij,k}^2.$$

Since w_i and w_j are independent and each has i.i.d. sub-Gaussian entries with parameter σ_w^2, the concatenated vector ζ_{ij} is also sub-Gaussian with parameter σ_w^2. Moreover, similar to the analysis of $\tilde{w}_i$, since $\mathcal{U}_{ij}$ is orthogonal, $\tilde{\zeta}_{ij} = \mathcal{U}_{ij}^\top \zeta_{ij}$ is sub-Gaussian vector with parameter σ_w^2.

By the definition of a sub-Gaussian vector Def. 2, for any $v \in \mathbb{R}^{2d_x}$ with $\|v\|_2 = 1$, we have

$$\mathbb{E}\left[e^{\lambda v^\top \tilde{\zeta}_{ij}}\right] = \mathbb{E}\left[e^{\lambda(\mathcal{U}_{ij} v)^\top \zeta_{ij}}\right] \leq \exp\left(\frac{\lambda^2 \sigma_w^2}{2}\right).$$

Hence, $\tilde{\zeta}_{ij}$ is a sub-Gaussian vector with parameter σ_w^2.

In particular, for each k, taking $v = e_k$ (the k-th standard basis vector in $\mathbb{R}^{2d}$), we obtain

$$\mathbb{E}\left[e^{\lambda \tilde{\zeta}_{ij,k}}\right] = \mathbb{E}\left[e^{\lambda e_k^\top \tilde{\zeta}_{ij}}\right] \leq \exp\left(\frac{\lambda^2 \sigma_w^2}{2}\right),$$

which shows that each coordinate $\tilde{\zeta}_{ij,k}$ is a sub-Gaussian random variable with parameter σ_w^2.

Since each $\tilde{\zeta}_{ij,k}$ is sub-Gaussian with parameter σ_w^2, we apply the scalar MGF bound from [1, Appendix B], which gives:

$$\mathbb{E}\left[e^{\lambda(\tilde{\zeta}_{ij,k}^2 - \mathbb{E}[\tilde{\zeta}_{ij,k}^2])}\right] \leq \exp(16\lambda^2 \sigma_w^4).$$

Then, the centered quadratic form becomes:

$$\tilde{\zeta}_{ij}^\top \Sigma_{ij} \tilde{\zeta}_{ij} - \mathbb{E}[\tilde{\zeta}_{ij}^\top \Sigma_{ij} \tilde{\zeta}_{ij}] = \sum_{k=1}^{2d_x} \nu_k (\tilde{\zeta}_{ij,k}^2 - \mathbb{E}[\tilde{\zeta}_{ij,k}^2]),$$

which can be bounded as

$$\sum_{k=1}^{2d_x} \nu_k (\tilde{\zeta}_{ij,k}^2 - \mathbb{E}[\tilde{\zeta}_{ij,k}^2]) \leq \|\Sigma_{ij}\| \cdot \sum_{k=1}^{2d_x} (\tilde{\zeta}_{ij,k}^2 - \mathbb{E}[\tilde{\zeta}_{ij,k}^2]),$$

where $\|\Sigma_{ij}\| := \max_k |\nu_k|$.

To bound the moment generating function of the centered bilinear form, we recall that

$$\mathbb{E}\left[e^{\lambda(w_i^\top H_{ij} w_j - \mathbb{E}[w_i^\top H_{ij} w_j])}\right] = \mathbb{E}\left[e^{\frac{\lambda}{2}(\zeta_{ij}^\top \Psi_{ij} \zeta_{ij} - \mathbb{E}[\zeta_{ij}^\top \Psi_{ij} \zeta_{ij}])}\right].$$

By the result from [1, Appendix B], applying the composition of independent sub-Gaussian quadratic bounds and combining the $2d_x$ terms, we obtain:

$$\mathbb{E}\left[\exp\left(\lambda \sum_{k=1}^{2d_x} (\tilde{\zeta}_{ij,k}^2 - \mathbb{E}[\tilde{\zeta}_{ij,k}^2])\right)\right] \leq \exp\left(16 \cdot 2d_x \lambda^2 \sigma_w^4\right).$$

By replacing λ as $\frac{\lambda\|\Sigma_{ij}\|}{2}$ to account for the factor in front of the quadratic form, we conclude:

$$\mathbb{E}\left[e^{\frac{\lambda}{2}(\zeta_{ij}^\top \Psi_{ij} \zeta_{ij} - \mathbb{E}[\zeta_{ij}^\top \Psi_{ij} \zeta_{ij}])}\right] \leq \exp\left(8\lambda^2 \|\Sigma_{ij}\|^2 \sigma_w^4 d_x\right),$$

which implies

$$\mathbb{E}\left[e^{\lambda(w_i^\top H_{ij} w_j - \mathbb{E}[w_i^\top H_{ij} w_j])}\right] \leq \exp\left(8\lambda^2 \|\Sigma_{ij}\|^2 \sigma_w^4 d_x\right),$$

for all λ such that $|\lambda| \leq \frac{1}{4\sigma_w^2\|\Sigma_{ij}\|}$. That is,

$$\mathbb{E}\left[e^{\lambda\left(w_i^\top H_{ij} w_j - \mathbb{E}[w_i^\top H_{ij} w_j]\right)}\right] \leq \exp\left(8\lambda^2 \|H_{ij}\|^2 \sigma_w^4 d_x\right),$$

for all λ such that $|\lambda| \leq \frac{1}{4\sigma_w^2\|H_{ij}\|}$.

We first consider the deviation of the quadratic cost for a single trajectory $\theta_T' \sim \mathcal{D}_{\theta_T}$. From the previous derivation using the structure of the cost in terms of $\{H_i, H_{ij}\}$ and the sub-Gaussianity of w_t, we obtain:

$$
\begin{aligned}
&\mathbb{E}\left[e^{\lambda\left(\mathbb{E}[C_q(K,\theta_T')]-C_q(K,\theta_T')\right)}\right] \\
&= \mathbb{E}\left[\exp\left(\lambda\sum_{i=0}^{T-1}\left(\mathbb{E}[w_i^\top H_i w_i]-w_i^\top H_i w_i\right)\right.\right. \\
&\quad \left.\left. + \sum_{0\le i<j\le T-1}\left(\mathbb{E}[w_i^\top H_{ij} w_j]-w_i^\top H_{ij} w_j\right)\right)\right], \\
&\le \mathbb{E}\left[\exp\left(\lambda\sum_{i=0}^{T}\left(\mathbb{E}[w_i^\top H_i w_i]-w_i^\top H_i w_i\right)\right.\right. \\
&\quad \left.\left. + \sum_{0\le i<j\le T}\left(\mathbb{E}[w_i^\top H_{ij} w_j]-w_i^\top H_{ij} w_j\right)\right)\right],
\end{aligned}
$$

for all $|\lambda| \le \frac{1}{4\sigma_w^2 \max\left\{\max_i \|H_i\|, \max_{i<j}\|H_{ij}\|\right\}}$. We apply the previous bounds for the diagonal and off-diagonal terms. By Jensen's inequality and the fact that the moment generating function of a sum is upper bounded by the product of the individual MGFs (for sub-exponential variables), we obtain:

$$
\begin{aligned}
&\mathbb{E}\left[e^{\lambda\left(\mathbb{E}[C_q(K,\theta_T')]-C_q(K,\theta_T')\right)}\right] \\
&\le \exp\left(\lambda^2\sigma_w^4 d_x\left(\sum_{i=0}^{T}16\|H_i\|^2+\sum_{0\le i<j\le T}8\|H_{ij}\|^2\right)\right),
\end{aligned}
$$

Let $\hat{C}(K) := \frac{1}{m}\sum_{i=1}^m C_q(K,\theta_{i,T})$ denote the empirical cost computed from m i.i.d. trajectories $\{\theta_{i,T}\}_{i=1}^m$.

Then the deviation between the expected and empirical cost can be written as

$$
\mathbb{E}\left[e^{\lambda(C(K)-\hat{C}(K))}\right] = \mathbb{E}\left[e^{\lambda\left(\mathbb{E}[C_q(K,\theta_T)]-\frac{1}{m}\sum_{i=1}^m C_q(K,\theta_{i,T})\right)}\right].
$$

We now use linearity of expectation to rewrite:

$$
\begin{aligned}
&\mathbb{E}[C_q(K,\theta_T)] - \frac{1}{m}\sum_{i=1}^m C_q(K,\theta_{i,T}) \\
&= \frac{1}{m}\sum_{i=1}^m \left(\mathbb{E}[C_q(K,\theta_T)]-C_q(K,\theta_{i,T})\right).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\mathbb{E}\left[e^{\lambda(C(K)-\hat{C}(K))}\right] \\
&= \mathbb{E}\left[e^{\lambda\cdot\frac{1}{m}\sum_{i=1}^m\left(\mathbb{E}[C_q(K,\theta_T)]-C_q(K,\theta_{i,T})\right)}\right] \\
&= \mathbb{E}\left[\prod_{i=1}^m e^{\frac{\lambda}{m}\left(\mathbb{E}[C_q(K,\theta_T)]-C_q(K,\theta_{i,T})\right)}\right].
\end{aligned}
$$

Since the $C_q(K,\theta_{i,T})$ are independent (due to the independent system parameters and noise sequence), the random variable $\mathbb{E}[C_q(K,\theta_T)]-C_q(K,\theta_{i,T})$ is also independent. Hence, the product of exponentials factorizes:

$$
\begin{aligned}
&\mathbb{E}\left[\prod_{i=1}^m e^{\frac{\lambda}{m}\left(\mathbb{E}[C_q(K,\theta_T)]-C_q(K,\theta_{i,T})\right)}\right] \\
&= \prod_{i=1}^m \mathbb{E}\left[e^{\frac{\lambda}{m}\left(\mathbb{E}[C_q(K,\theta_T)]-C_q(K,\theta_{i,T})\right)}\right].
\end{aligned}
$$

Thus, we conclude:

$$
\mathbb{E}\left[e^{\lambda(C(K)-\hat{C}(K))}\right] = \prod_{i=1}^m \mathbb{E}\left[e^{\frac{\lambda}{m}\left(\mathbb{E}[C_q(K,\theta_T)]-C_q(K,\theta_{i,T})\right)}\right].
$$

From the previous derivation for a single trajectory, we know that for all $\lambda' \in \left[0, \frac{1}{4\sigma_w^2\max\left\{\max_i\|H_i\|,\max_{i<j}\|H_{ij}\|\right\}}\right]$, we have:

$$
\begin{aligned}
&\mathbb{E}\left[e^{\lambda'\left(\mathbb{E}[C_q(K,\theta_T)]-C_q(K,\theta_{i,T})\right)}\right] \\
&\le \exp\left(\lambda'^2\sigma_w^4 d_x\cdot\left(\sum_{i=0}^{T}16\|H_i\|^2+\sum_{0\le i<j\le T}8\|H_{ij}\|^2\right)\right).
\end{aligned}
$$

We apply this bound with $\lambda' = \lambda/m$, and use that all m terms are identical. This gives:

$$
\begin{aligned}
&\mathbb{E}\left[e^{\lambda(C(K)-\hat{C}(K))}\right] = \prod_{i=1}^m \mathbb{E}\left[e^{\frac{\lambda}{m}\left(\mathbb{E}[C_q(K,\theta_T)]-C_q(K,\theta_{i,T})\right)}\right] \\
&\le \left(\exp\left(\frac{\lambda^2}{m^2}\cdot\sigma_w^4 d_x\cdot\left(\sum_{i=0}^{T}16\|H_i\|^2+\sum_{0\le i<j\le T}8\|H_{ij}\|^2\right)\right)\right)^m \\
&= \exp\left(\frac{\lambda^2}{m}\cdot\sigma_w^4 d_x\cdot\left(\sum_{i=0}^{T}16\|H_i\|^2+\sum_{0\le i<j\le T}8\|H_{ij}\|^2\right)\right).
\end{aligned}
$$

This implies the quadratic cost is sub-exponential with variance proxy:

$$
B^2 := \sigma_w^4 d_x\left(128\sum_{i=0}^{T}\|H_i\|^2 + 64\sum_{0\le i<j\le T}\|H_{ij}\|^2\right).
$$

We now derive an explicit upper bound on B^2 in terms of system-dependent quantities $\rho_S(K)$ and $\rho_M(K)$. Recall the definitions:

$$
\rho_S(K) := \|Q\|_F + \|K\|^2\|R\|_F, \quad \rho_M(K) := \rho_A + \rho_B\|K\|,
$$

where $\rho_A := \max\{|a_1|,|a_2|\}$ and $\rho_B := \max\{|b_1|,|b_2|\}$. These provide bounds on $\|S\|$ and $\|M\|$:

$$
\|S\| \le \rho_S(K), \quad \|M\| \le \rho_M(K).
$$

Each H_i corresponds to the contribution of w_i to the quadratic cost:

$$
H_i = M^{T-i\top}\cdots M^\top S M\cdots M^{T-i}.
$$

Hence, we can bound:

$$
\|H_i\| \le \|S\|\cdot\|M\|^{2(T-i)} \le \rho_S(K)\cdot\rho_M(K)^{2(T-i)}.
$$

The cross-term matrix H_{ij} arises from the interaction between w_i and w_j:

$$
H_{ij} = M^{T-i\top}\cdots M^\top S M\cdots M^{T-j},
$$

so that

$$
\begin{aligned}
\|H_{ij}\| &\le \|S\|\cdot\|M\|^{T-i+T-j} = \|S\|\cdot\|M\|^{2T-i-j} \\
&\le \rho_S(K)\cdot\rho_M(K)^{2T-i-j}.
\end{aligned}
$$

Substituting into the expression:

$$B^2 \le \sigma_w^4 d_x \cdot \rho_S(K)^2 \cdot \Bigg(128 \sum_{i=0}^{T} \rho_M(K)^{4(T-i)}$$
$$+64 \sum_{0\le i<j\le T} \rho_M(K)^{4T-2(i+j)}\Bigg).$$

Then we define

$$B_{\text{cost}}^2 := \sup_{K\in\mathcal{K}} \sigma_w^4 d_x \cdot \rho_S(K)^2 \cdot \Bigg(128 \sum_{i=0}^{T} \rho_M(K)^{4(T-i)}$$
$$+64 \sum_{0\le i<j\le T} \rho_M(K)^{4T-2(i+j)}\Bigg).$$

This gives a compact upper bound on B_{cost}^2 in terms of $\rho_S(K)$, $\rho_M(K)$, σ_w^2, d_x, and T.

For the admissible range of λ, we need the maximum spectral norm across all H_i and H_{ij}. Since

$$\max_i \|H_i\| = \|H_0\| \le \rho_S(K) \cdot \rho_M(K)^{2T},$$

and

$$\max_{i<j} \|H_{ij}\| = \|H_{01}\| \le \rho_S(K) \cdot \rho_M(K)^{2T-1},$$

we get

$$|\lambda| \le \inf_{K\in\mathcal{K}} \frac{1}{4\sigma_w^2 \cdot \rho_S(K) \cdot \rho_M(K)^{2T}}.$$

Hence, we have obtained Eq. (1) and the expression for B_{cost} Eq. (2) used in Thm 1, which completes the proof.

□

References

[1] J. Honorio and T. Jaakkola, "Tight bounds for the expected risk of linear classifiers and PAC-Bayes finite-sample guarantees," in *Proc. of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 384–392, 2014.