7 APPENDIX—SUPPLEMENTARY MATERIAL

7.1 Proof of Theorem 3

Proof. Without loss of generality we assume that $\{1,2\} \in \mathcal{I}$ and $\mathbf{h}^* = (\alpha, -\beta, h_3, \dots, h_n)^\top$ where $\alpha\beta > 0$.

There must be a bijection ϕ which satisfies $\phi(1) = 2$, $\phi(2) = 1$ and the three requirements of ϕ in *Definition 2*. Consider $\mathbf{h}^\star = (-\beta, \alpha, h_{\phi(3)}, \dots, h_{\phi(n)})^\top$. Obviously $\|\mathbf{h}^\star\|_1 = \|\mathbf{h}^*\|_1$ and $\|\mathbf{h}^\star\|_2 = \|\mathbf{h}^*\|_2$. Moreover, $\forall i$,

$$\begin{aligned}
&[g(\mathbf{h}^\star)]_i \\
&= \gamma \sum\nolimits_{l=1}^n Q_{i,l} h_{\phi(l)} - \eta h_{\phi(i)} - \operatorname{sgn}(h_{\phi(i)}) \\
&= \gamma \sum\nolimits_{k=1}^n Q_{i,\phi^{-1}(k)} h_k - \eta h_{\phi(i)} - \operatorname{sgn}(h_{\phi(i)}) \\
&= \gamma \sum\nolimits_{k=1}^n Q_{i,\phi(k)} h_k - \eta h_{\phi(i)} - \operatorname{sgn}(h_{\phi(i)}) \\
&= \gamma \sum\nolimits_{k=1}^n Q_{\phi(i),k} h_k - \eta h_{\phi(i)} - \operatorname{sgn}(h_{\phi(i)}) \\
&= [g(\mathbf{h}^*)]_{\phi(i)}.
\end{aligned}$$

Hence, $g(\mathbf{h}^\star) = \mathbf{0}$ due to the KKT condition $g(\mathbf{h}^*) = \mathbf{0}$, which means that $\mathbf{h}^\star$ is also a minimum. Similarly we can derive $\mathbf{h}^{\star\top} Q \mathbf{h}^\star = \mathbf{h}^{*\top} Q \mathbf{h}^*$ and thereby we arrive at $G(\mathbf{h}^\star) = G(\mathbf{h}^*)$.

Notice that $d_{\mathcal{H}}(\mathbf{h}^*, \mathbf{h}^\star) \geq 1$, with the only exception $d_{\mathcal{H}}(\mathbf{h}^*, \mathbf{h}^\star) = 0$ when $\operatorname{sgn}(\mathbf{h}^\star) = -\operatorname{sgn}(\mathbf{h}^*)$, i.e., $\forall i, \phi(i) \neq i$ and $\mathcal{I}' = \{\{i, \phi(i)\} | i = 1, \dots, n\}$. This completes the proof. □

7.2 Proof of Theorem 4

Proof. When $n = 2$ it is trivial that X_n is anisotropic.

Suppose that $n > 2$ and $\mathcal{E}(\mathcal{H}_Q)$ has two principal axes $\mathbf{v}_j$ and $\mathbf{v}_k$ with the same length $1/\sqrt{\lambda_j} = 1/\sqrt{\lambda_k}$. Then there is at least one principal axis $\mathbf{v}_l, l \neq j, k$ about which $\mathcal{E}(\mathcal{H}_Q)$ is rotational along the circle spanned by $\mathbf{v}_j$ and $\mathbf{v}_k$.

From $\forall i, Q_{i,i} = \kappa$ we know that $\mathcal{E}(\mathcal{H}_Q)$ intersects the i-th coordinate axis at $\pm\mathbf{e}_i/\sqrt{\kappa}$ with length $1/\sqrt{\kappa}$, where $\mathbf{e}_i$ is the i-th unit vector of $\mathbb{R}^n$. Now $\mathcal{E}(\mathcal{H}_Q)$ has n principal axes with at most $n-1$ different length but another system of n orthogonal axes with length $1/\sqrt{\kappa}$, so $\mathbf{v}_l$ must be in the form of

$$\mathbf{v}_l = \frac{\bar{\mathbf{v}}_l}{\|\bar{\mathbf{v}}_l\|_2},\ \bar{\mathbf{v}}_l = \sum_{i=1}^n \delta_i \mathbf{e}_i \neq \mathbf{0},\ \delta_i \in \{-1, 0, 1\}.$$

In other words, $\mathbf{v}_l$ lies on the central direction of certain quadrant of a subspace of $\mathbb{R}^n$ determined by $\mathbf{v}_j$ and $\mathbf{v}_k$. But this is impossible since X_n is not axisymmetric w.r.t. Q.

Hence all principal axes of $\mathcal{E}(\mathcal{H}_Q)$ have different length, which is exactly what we were to prove. □

7.3 Proof of Theorem 5

Proof. The KKT condition $g(\mathbf{h}) = \mathbf{0}$ implies

$$\mathbf{h} = \hat{Q}\mathbf{y}, \tag{12}$$

where $\mathbf{y} = \operatorname{sgn}(\mathbf{h}), \hat{Q} = (\gamma Q - \eta I)^{-1}$, and the constant $\eta < \gamma\lambda_1$, λ_1 is the smallest eigenvalue of Q. Substitute (12) into $\|\mathbf{h}\|_2 = 1$, and note that $\hat{Q}^\top = \hat{Q}$,

$$(\hat{Q}\mathbf{y})^\top (\hat{Q}\mathbf{y}) = 1 \Longrightarrow \mathbf{y}^\top \hat{Q}^2 \mathbf{y} = 1.$$

All eigenvalues of Q are different and positive, so are all eigenvalues of $\hat{Q}$. Consequently, $\hat{Q}^2$ has a unique spectral decomposition. Let $\hat{Q}^2 = \sum_{i=1}^n \mu_i \mathbf{u}_i \mathbf{u}_i^\top$, then $\mathbf{y}^\top \hat{Q}^2 \mathbf{y} = \sum_{i=1}^n \mu_i \|\mathbf{u}_i^\top \mathbf{y}\|_2^2$.

We assert that $\forall \mathbf{y}_1, \mathbf{y}_2 \in \{-1, +1\}^n$, the only possibility of $\mathbf{y}_1^\top \hat{Q}^2 \mathbf{y}_1 = \mathbf{y}_2^\top \hat{Q}^2 \mathbf{y}_2$ is either $\mathbf{y}_1 = \mathbf{y}_2$ or $\mathbf{y}_1 = -\mathbf{y}_2$. Otherwise, there exist two nonempty disjoint indices $\mathcal{J}$ and $\mathcal{K}$, such that $\forall j \in \mathcal{J}, k \in \mathcal{K}, [\mathbf{y}_1]_j = -[\mathbf{y}_1]_k = -[\mathbf{y}_2]_j = [\mathbf{y}_2]_k$. Moreover, $\forall i$, $\sum_{j\in\mathcal{J}} [\mathbf{u}_i]_j [\mathbf{y}_1]_j + \sum_{k\in\mathcal{K}} [\mathbf{u}_i]_k [\mathbf{y}_1]_k = \sum_{j\in\mathcal{J}} [\mathbf{u}_i]_j [\mathbf{y}_2]_j + \sum_{k\in\mathcal{K}} [\mathbf{u}_i]_k [\mathbf{y}_2]_k$ since the spectral decomposition of $\hat{Q}^2$ is unique. Hence, $\sum_{j\in\mathcal{J}} [\mathbf{u}_i]_j = \sum_{k\in\mathcal{K}} [\mathbf{u}_i]_k$ for all $i = 1, \dots, n$. This means that the row rank of the matrix $U = (\mathbf{u}_1 | \dots | \mathbf{u}_n)$ is $n-1$, which contradicts the linear independence of $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$.

Therefore, all minima of (4) are equivalent w.r.t. $d_{\mathcal{H}}$. □

7.4 Proof of Lemma 7

Proof. For any $\mathbf{h} \in \widetilde{\mathcal{H}}_Q$, $\exists \boldsymbol{\alpha} \in \mathbb{R}^n$ such that $\mathbf{h} = U\boldsymbol{\alpha}$, where U consists of n orthonormal eigenvectors of Q, and $\|\boldsymbol{\alpha}\|_2 = 1$ since $\|\mathbf{h}\|_2 = 1$ and $U^\top U = I$. This $\mathbf{h} = U\boldsymbol{\alpha}$ is a UL decomposition (El-Yaniv & Pechyony, 2007) since U has only information about unlabeled samples. Each column of U has unit length, and thus $\|U\|_{\text{Fro}}^2 = n$. The first part of the bound comes from *inequality (20)* in El-Yaniv and Pechyony (2007).

Another UL decomposition is shown in (12). The equation (12) holds for η^* since it holds for any constant η smaller than γ times the smallest eigenvalue of Q. It is also a kernel UL decomposition since the matrix $\hat{Q}$ is symmetric positive definite. Then, the other part of the bound is derived from *inequality (20)* and *inequality (23)* in El-Yaniv and Pechyony (2007) with $\mu_1 = \sqrt{n}$ and $\mu_2 = \sqrt{\mu}$, respectively. □