Appendix A. Proofs

Appendix A.1. *Proof of Lemma 3.2*

Proof. Given a quadratic surface $\mathcal{S}$ denoted by (10), for any point $\boldsymbol{x} \in \mathcal{S}$, recall the definitions of $\hat{\boldsymbol{n}}(\boldsymbol{x})$, $\gamma(\boldsymbol{x})$, $\gamma^+(\boldsymbol{x})$, $\gamma^-(\boldsymbol{x})$, $\boldsymbol{x}^+$ and $\boldsymbol{x}^-$ in Section 3.2. By (11), $Q(\boldsymbol{x}^+) = 1$ and $Q(\boldsymbol{x}^-) = -1$, which are equivalent to

$$\begin{aligned}
&\frac{1}{2}(\boldsymbol{x} + \gamma^+(\boldsymbol{x})\hat{\boldsymbol{n}}(\boldsymbol{x}))^T \mathbf{W}(\boldsymbol{x} + \gamma^+(\boldsymbol{x})\hat{\boldsymbol{n}}(\boldsymbol{x})) + \boldsymbol{b}^T(\boldsymbol{x} + \gamma^+(\boldsymbol{x})\hat{\boldsymbol{n}}(\boldsymbol{x})) + c - 1 = 0 \\
&\frac{1}{2}(\boldsymbol{x} - \gamma^-(\boldsymbol{x})\hat{\boldsymbol{n}}(\boldsymbol{x}))^T \mathbf{W}(\boldsymbol{x} - \gamma^-(\boldsymbol{x})\hat{\boldsymbol{n}}(\boldsymbol{x})) + \boldsymbol{b}^T(\boldsymbol{x} - \gamma^-(\boldsymbol{x})\hat{\boldsymbol{n}}(\boldsymbol{x})) + c + 1 = 0
\end{aligned} \tag{A.1}$$

With $Q(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\mathbf{W}\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x} + c = 0$, (A.1) can be simplified as the following:

$$\begin{aligned}
&\frac{1}{2}\gamma^+(\boldsymbol{x})^2 \hat{\boldsymbol{n}}(\boldsymbol{x})^T \mathbf{W}\hat{\boldsymbol{n}}(\boldsymbol{x}) + \gamma^+(\boldsymbol{x})\hat{\boldsymbol{n}}(\boldsymbol{x})^T(\mathbf{W}\boldsymbol{x} + \boldsymbol{b}) - 1 = 0 \\
&\frac{1}{2}\gamma^-(\boldsymbol{x})^2 \hat{\boldsymbol{n}}(\boldsymbol{x})^T \mathbf{W}\hat{\boldsymbol{n}}(\boldsymbol{x}) - \gamma^-(\boldsymbol{x})\hat{\boldsymbol{n}}(\boldsymbol{x})^T(\mathbf{W}\boldsymbol{x} + \boldsymbol{b}) + 1 = 0
\end{aligned} \tag{A.2}$$

Notice that (A.2) are second order equations with respect to $\gamma^+(\boldsymbol{x})$ and $\gamma^-(\boldsymbol{x})$, respectively. Therefore, we will be able to solve out the explicit solutions as the following:

$$\begin{aligned}
\gamma^+(\boldsymbol{x}) &= \frac{-\hat{\boldsymbol{n}}(\boldsymbol{x})^T(\mathbf{W}\boldsymbol{x} + \boldsymbol{b}) \pm \sqrt{[\hat{\boldsymbol{n}}(\boldsymbol{x})^T(\mathbf{W}\boldsymbol{x} + \boldsymbol{b})]^2 + 2\hat{\boldsymbol{n}}(\boldsymbol{x})^T\mathbf{W}\hat{\boldsymbol{n}}(\boldsymbol{x})}}{\hat{\boldsymbol{n}}(\boldsymbol{x})^T\mathbf{W}\hat{\boldsymbol{n}}(\boldsymbol{x})} \\
\gamma^-(\boldsymbol{x}) &= \frac{\hat{\boldsymbol{n}}(\boldsymbol{x})^T(\mathbf{W}\boldsymbol{x} + \boldsymbol{b}) \pm \sqrt{[\hat{\boldsymbol{n}}(\boldsymbol{x})^T(\mathbf{W}\boldsymbol{x} + \boldsymbol{b})]^2 - 2\hat{\boldsymbol{n}}(\boldsymbol{x})^T\mathbf{W}\hat{\boldsymbol{n}}(\boldsymbol{x})}}{\hat{\boldsymbol{n}}(\boldsymbol{x})^T\mathbf{W}\hat{\boldsymbol{n}}(\boldsymbol{x})}
\end{aligned}$$

By the definition, we have $\hat{\boldsymbol{n}}(\boldsymbol{x})^T(W\boldsymbol{x} + \boldsymbol{b}) = \|\mathbf{W}\boldsymbol{x} + \boldsymbol{b}\|$. Then the above equations can be simplified as the following:

$$\gamma^+(\boldsymbol{x}) = \frac{-\|\mathbf{W}\boldsymbol{x}+\boldsymbol{b}\| \pm \sqrt{\|\mathbf{W}\boldsymbol{x}+\boldsymbol{b}\|^2 + 2\hat{\boldsymbol{n}}(\boldsymbol{x})^T\mathbf{W}\hat{\boldsymbol{n}}(\boldsymbol{x})}}{\hat{\boldsymbol{n}}(\boldsymbol{x})^T\mathbf{W}\hat{\boldsymbol{n}}(\boldsymbol{x})}, \quad \gamma^-(\boldsymbol{x}) = \frac{\|\mathbf{W}\boldsymbol{x}+\boldsymbol{b}\| \pm \sqrt{\|\mathbf{W}\boldsymbol{x}+\boldsymbol{b}\|^2 - 2\hat{\boldsymbol{n}}(\boldsymbol{x})^T\mathbf{W}\hat{\boldsymbol{n}}(\boldsymbol{x})}}{\hat{\boldsymbol{n}}(\boldsymbol{x})^T\mathbf{W}\hat{\boldsymbol{n}}(\boldsymbol{x})}$$

By eliminating two useless roots, we have the explicit solutions of $\gamma^+(\boldsymbol{x})$ and $\gamma^-(\boldsymbol{x})$.

$$\begin{aligned}
\gamma^+(\boldsymbol{x}) &= \frac{-\|\mathbf{W}\boldsymbol{x} + \boldsymbol{b}\| + \sqrt{\|\mathbf{W}\boldsymbol{x} + \boldsymbol{b}\|^2 + 2\hat{\boldsymbol{n}}(\boldsymbol{x})^T\mathbf{W}\hat{\boldsymbol{n}}(\boldsymbol{x})}}{\hat{\boldsymbol{n}}(\boldsymbol{x})^T\mathbf{W}\hat{\boldsymbol{n}}(\boldsymbol{x})} \\
\gamma^-(\boldsymbol{x}) &= \frac{\|\mathbf{W}\boldsymbol{x} + \boldsymbol{b}\| - \sqrt{\|\mathbf{W}\boldsymbol{x} + \boldsymbol{b}\|^2 - 2\hat{\boldsymbol{n}}(\boldsymbol{x})^T\mathbf{W}\hat{\boldsymbol{n}}(\boldsymbol{x})}}{\hat{\boldsymbol{n}}(\boldsymbol{x})^T\mathbf{W}\hat{\boldsymbol{n}}(\boldsymbol{x})}
\end{aligned} \tag{A.3}$$

By (12), there exists an orthonormal matrix $\mathbf{U} \in \mathbb{R}^{n\times n}$ such that $\mathbf{W} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^T$, where $\mathbf{\Sigma}$ is a diagonal matrix of the singular values of $\mathbf{W}$. Recall that the singular values of $\mathbf{W}$ are in a decreasing order: $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r \geqslant \sigma_{r+1} = \cdots = \sigma_n = 0$. Denote $\mathbf{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n]$. Then $\{\boldsymbol{u}_i \in \mathbb{R}^n\}_{1\leqslant i\leqslant n}$ forms an orthonormal basis on $\mathbb{R}^n$. Hence, there exists $\boldsymbol{\alpha}(\boldsymbol{x}) = [\alpha_1(\boldsymbol{x}), \alpha_2(\boldsymbol{x}), \ldots, \alpha_n(\boldsymbol{x})]^T \in \mathbb{R}^n$ such that

$$\mathbf{W}\boldsymbol{x} + \boldsymbol{b} = \sum_{i=1}^{n} \alpha_i(\boldsymbol{x})\boldsymbol{u}_i = \mathbf{U}\boldsymbol{\alpha}(\boldsymbol{x}). \tag{A.4}$$

Taking (A.4) into $(A.3)$ and we have

$$\gamma^+(\boldsymbol{x}) = \frac{-\|\mathbf{U}\boldsymbol{\alpha}(\boldsymbol{x})\| + \sqrt{\|\mathbf{U}\boldsymbol{\alpha}(\boldsymbol{x})\|^2 + 2\frac{\boldsymbol{\alpha}(\boldsymbol{x})^T\mathbf{\Sigma}\boldsymbol{\alpha}(\boldsymbol{x})}{\boldsymbol{\alpha}(\boldsymbol{x})^T\boldsymbol{\alpha}(\boldsymbol{x})}}}{\frac{\boldsymbol{\alpha}(\boldsymbol{x})^T\mathbf{\Sigma}\boldsymbol{\alpha}(\boldsymbol{x})}{\boldsymbol{\alpha}(\boldsymbol{x})^T\boldsymbol{\alpha}(\boldsymbol{x})}}, \quad \gamma^-(\boldsymbol{x}) = \frac{\|\mathbf{U}\boldsymbol{\alpha}(\boldsymbol{x})\| - \sqrt{\|\mathbf{U}\boldsymbol{\alpha}(\boldsymbol{x})\|^2 - 2\frac{\boldsymbol{\alpha}(\boldsymbol{x})^T\mathbf{\Sigma}\boldsymbol{\alpha}(\boldsymbol{x})}{\boldsymbol{\alpha}(\boldsymbol{x})^T\boldsymbol{\alpha}(\boldsymbol{x})}}}{\frac{\boldsymbol{\alpha}(\boldsymbol{x})^T\mathbf{\Sigma}\boldsymbol{\alpha}(\boldsymbol{x})}{\boldsymbol{\alpha}(\boldsymbol{x})^T\boldsymbol{\alpha}(\boldsymbol{x})}}$$

Notice that $\frac{\boldsymbol{\alpha}(\boldsymbol{x})^T \boldsymbol{\Sigma} \boldsymbol{\alpha}(\boldsymbol{x})}{\boldsymbol{\alpha}(\boldsymbol{x})^T \boldsymbol{\alpha}(\boldsymbol{x})}$ is the Rayleigh quotient of $\boldsymbol{\Sigma}$ at $\boldsymbol{\alpha}(\boldsymbol{x})$. Denote $\frac{\boldsymbol{\alpha}(\boldsymbol{x})^T \Sigma \boldsymbol{\alpha}(\boldsymbol{x})}{\boldsymbol{\alpha}(\boldsymbol{x})^T \boldsymbol{\alpha}(\boldsymbol{x})} = R(\Sigma, \boldsymbol{\alpha}(\boldsymbol{x}))$. By Parseval's identity, $\|\mathbf{U}\boldsymbol{\alpha}(\boldsymbol{x})\|_2^2 = \|\boldsymbol{\alpha}(\boldsymbol{x})\|_2^2$. Hence, the G-margin at $\boldsymbol{x}$ can be written as

$$\gamma(\boldsymbol{x}) = \gamma^+(\boldsymbol{x}) + \gamma^-(\boldsymbol{x}) = \frac{\sqrt{\|\boldsymbol{\alpha}(\boldsymbol{x})\|^2 + 2R(\boldsymbol{\Sigma}, \boldsymbol{\alpha}(\boldsymbol{x}))} - \sqrt{\|\boldsymbol{\alpha}(\boldsymbol{x})\|^2 - 2R(\boldsymbol{\Sigma}, \boldsymbol{\alpha}(\boldsymbol{x}))}}{R(\boldsymbol{\Sigma}, \boldsymbol{\alpha}(\boldsymbol{x}))} \tag{A.5}$$

Notice that inequality $\sqrt{x} - \sqrt{y} \leqslant \sqrt{x-y}$ holds for any $x \geqslant y \geqslant 0$. Therefore, we have the following inequality.

$$\frac{1}{\gamma(\boldsymbol{x})} = \frac{1}{\gamma^+(\boldsymbol{x}) + \gamma^-(\boldsymbol{x})} \geqslant \frac{\sqrt{R(\boldsymbol{\Sigma}, \boldsymbol{\alpha}(\boldsymbol{x}))}}{2}. \tag{A.6}$$

Since $\boldsymbol{\Sigma}$ is known when $\mathcal{S}$ is given, we proved Lemma 3.2.

□

Appendix A.2. ***Proof of theorem 4.1***

Proof. Assume $(\boldsymbol{v}^*, \boldsymbol{b}^*, q^*, \boldsymbol{\zeta}^*)$ and $(\hat{\boldsymbol{v}}, \hat{\boldsymbol{b}}, \hat{q}, \hat{\boldsymbol{\zeta}})$ are both optimal solutions to problem (DWPSVM$''$). For a convex program, its optimal solution set is convex. In other words, $(\forall \alpha \in (0,1))\ \alpha(\boldsymbol{v}^*, \boldsymbol{b}^*, q^*, \boldsymbol{\zeta}^*) + (1-\alpha)(\hat{\boldsymbol{v}}, \hat{\boldsymbol{b}}, \hat{q}, \hat{\boldsymbol{\zeta}})$ is optimal as well. Denote the optimal value as $\bar{z}$.

Define function $p : \mathbb{R}^{\frac{l(l+1)+n(n+1)}{2}+N} \to \mathbb{R}$, such that

$$p(\boldsymbol{v}, \boldsymbol{b}, \boldsymbol{\zeta}) \triangleq \frac{1}{2}\boldsymbol{v}^T \mathbf{H} \boldsymbol{v} + \frac{1}{2}\|\boldsymbol{b}\|_2^2 + C\sum_{i=1}^{N} \zeta_i. \tag{A.7}$$

which leads to

$$\begin{aligned}
\bar{z} =& \frac{1}{2}(\alpha \boldsymbol{v}^* + (1-\alpha)\hat{\boldsymbol{v}})^T \mathbf{H} (\alpha \boldsymbol{v}^* + (1-\alpha)\hat{\boldsymbol{v}}) + \frac{1}{2}\left(\alpha \boldsymbol{b}^* + (1-\alpha)\hat{\boldsymbol{b}}\right)^T \left(\alpha \boldsymbol{b}^* + (1-\alpha)\hat{\boldsymbol{b}}\right) + C\sum_{i=1}^{N}\left(\alpha\zeta_i^* + (1-\alpha)\hat{\zeta}_i\right) \\
=& \frac{\alpha^2}{2}\left(\boldsymbol{v}^{*T}\mathbf{H}\boldsymbol{v}^* + \boldsymbol{b}^{*T}\boldsymbol{b}^*\right) + \frac{(1-\alpha)^2}{2}\left(\hat{\boldsymbol{v}}^T \mathbf{H}\hat{\boldsymbol{v}} + \hat{\boldsymbol{b}}^T\hat{\boldsymbol{b}}\right) + \alpha(1-\alpha)\left(\boldsymbol{v}^{*T}\mathbf{H}\hat{\boldsymbol{v}} + \boldsymbol{b}^{*T}\hat{\boldsymbol{b}}\right) + C\sum_{i=1}^{N}\left(\alpha\zeta_i^* + (1-\alpha)\hat{\zeta}_i\right) \\
=& \alpha p(\boldsymbol{v}^*, \boldsymbol{b}^*, \boldsymbol{\zeta}^*) + (1-\alpha)p(\hat{\boldsymbol{v}}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{\zeta}}) + \alpha(\alpha-1)\left((\boldsymbol{v}^* - \hat{\boldsymbol{v}})^T \mathbf{H} (\boldsymbol{v}^* - \hat{\boldsymbol{v}}) + (\boldsymbol{b}^* - \hat{\boldsymbol{b}})^T(\boldsymbol{b}^* - \hat{\boldsymbol{b}})\right).
\end{aligned}$$

Since $\bar{z} = p(\boldsymbol{v}^*, \boldsymbol{b}^*, \boldsymbol{\zeta}^*) = p(\hat{\boldsymbol{v}}$, it forces $\hat{\boldsymbol{b}}, \hat{\boldsymbol{\zeta}})$ forces $(\boldsymbol{v}^* - \hat{\boldsymbol{v}})^T \mathbf{H} (\boldsymbol{v}^* - \hat{\boldsymbol{v}}) + (\boldsymbol{b}^* - \hat{\boldsymbol{b}})^T(\boldsymbol{b}^* - \hat{\boldsymbol{b}}) = 0$, which implies $(\boldsymbol{v}^* - \hat{\boldsymbol{v}})^T \mathbf{H} (\boldsymbol{v}^* - \hat{\boldsymbol{v}}) = 0$ due to the positive definiteness of $\mathbf{H}$ and $(\boldsymbol{b}^* - \hat{\boldsymbol{b}})^T(\boldsymbol{b}^* - \hat{\boldsymbol{b}}) = 0$.

In conclusion, $\boldsymbol{v}^* = \hat{\boldsymbol{v}}$ and $\boldsymbol{b}^* = \hat{\boldsymbol{b}}$.

□