1. Appendix

Before we provide the proof for Proposition 1, we list two useful lemmas that are used repeatedly in the following.

Lemma 1 ([2]). *The non-increasingly ordered singular values of a matrix* $\mathbf{M}$ *obey* $0 \le \sigma_i \le \frac{\|M\|_F}{\sqrt{i}}$, *where* $\|\cdot\|_F$ *denotes the matrix Frobenius norm.*

Lemma 2 ([3]). *Let* $\sigma_i(M)$ *and* $\sigma_i(N)$ *be the non-increasingly ordered singular values of matrices* $\mathbf{M}, \mathbf{N} \in \mathbb{R}^{a\times b}$. *Then,* $\mathrm{tr}\{\mathbf{M}\mathbf{N}^T\} \le \sum_i^r \sigma_i(\mathbf{M})\sigma_i(\mathbf{N})$, *where* $r = min(a, b)$.

Proof of Proposition 1

Proof. The Fishers Criterion can be rewritten as $\psi = \mathrm{tr}\{\bar{\mathbf{S}}^{-1}\mathbf{S}_B\}$, where $\bar{\mathbf{S}} = \mathbf{F}\mathbf{F}^T$ ($\mathbf{F}$ is the matrix containing all features of the unlabeled set, arranged in columns) and $\mathbf{S}_B = \sum_{c=1}^N M_c \boldsymbol{\mu}_c \boldsymbol{\mu}_c^T = \sum_{c=1}^N \mathbf{S}_c$ ($\boldsymbol{\mu}_c$ is the mean feature vector of class c). For notation clarity and simplicity, we assume that all data are centered and that data mean does not change after only one sample is removed. This is justifiable when the number of unlabeled data is sufficiently large, which is the case we consider here.

Suppose the removed instance has pseudo-label belonging to class u. After removing the instance $f(\mathbf{x}_u)$, the two scatter matrices becomes: $\bar{\mathbf{S}}' = \bar{\mathbf{S}} - f(\mathbf{x}_u)f(\mathbf{x}_u)^T$ and $\mathbf{S}'_B = \mathbf{S}_B + \mathbf{S}'_u - \mathbf{S}_u = \mathbf{S}_B + \mathbf{E}_B$, where $\mathbf{S}'_u = (M_u - 1)\boldsymbol{\mu}'_u \boldsymbol{\mu}'^T_u$ and $\boldsymbol{\mu}'_u = (\boldsymbol{\mu}_u M_u - f(\boldsymbol{\mu}_u))/(M_u - 1)$. Then, we can rewrite:

$$\mathbf{E}_B = \frac{M_u \boldsymbol{\mu}_u \boldsymbol{\mu}_u^T - M_u \boldsymbol{\mu}_u f(\mathbf{x}_u)^T - M_u f(\mathbf{x}_u)\boldsymbol{\mu}_u^T + f(\mathbf{x}_u)f(\mathbf{x}_u)^T}{M_u - 1} \quad (1)$$

We can then define the IDA as:

$$\begin{aligned} d\psi_u &= \mathrm{tr}\{\bar{\mathbf{S}}^{-1}\mathbf{S}_B - \bar{\mathbf{S}}'^{-1}\mathbf{S}'_B\} \quad (2) \\ &= \mathrm{tr}\{\bar{\mathbf{S}}^{-1}\mathbf{S}_B - (\bar{\mathbf{S}} - f(\mathbf{x}_u)f(\mathbf{x}_u)^T)^{-1}(\mathbf{S}_B + \mathbf{E}_B)\} \end{aligned}$$

The latter term can be reformulated by the Woodbury identity [1]:

$$\begin{aligned} &(\bar{\mathbf{S}} - f(\mathbf{x}_u)f(\mathbf{x}_u)^T)^{-1}(\mathbf{S}_B + \mathbf{E}_B) \quad (3) \\ &= (\bar{\mathbf{S}}^{-1} + \frac{\bar{\mathbf{S}}^{-1}f(\mathbf{x}_u)f(\mathbf{x}_u)^T\bar{\mathbf{S}}^{-1}}{1 - f(\mathbf{x}_u)^T\bar{\mathbf{S}}^{-1}f(\mathbf{x}_u)})(\mathbf{S}_B + \mathbf{E}_B) \end{aligned}$$

Substitute this term into the above IDA equation, we have:

$$d\psi_u = \mathrm{tr}\{\frac{\bar{\mathbf{S}}^{-1}f(\mathbf{x}_u)f(\mathbf{x}_u)^T\bar{\mathbf{S}}^{-1}\mathbf{S}_B}{f(\mathbf{x}_u)^T\bar{\mathbf{S}}^{-1}f(\mathbf{x}_u) - 1} + \bar{\mathbf{S}}^{-1}\tilde{\mathbf{E}}_B + \frac{\bar{\mathbf{S}}^{-1}f(\mathbf{x}_u)f(\mathbf{x}_u)^T\bar{\mathbf{S}}^{-1}\mathbf{E}_B}{f(\mathbf{x}_u)^T\bar{\mathbf{S}}^{-1}f(\mathbf{x}_u) - 1}\} \quad (4)$$

where $\tilde{\mathbf{E}}_B = -\mathbf{E}_B$. To upper-bound $d\psi_u$, we derive an upper-bound for the three terms respectively, given that trace operation is additive.

Upper-bound for $\mathrm{tr}\{\frac{\bar{\mathbf{S}}^{-1}f(\mathbf{x}_u)f(\mathbf{x}_u)^T\bar{\mathbf{S}}^{-1}\mathbf{S}_B}{f(\mathbf{x}_u)^T\bar{\mathbf{S}}^{-1}f(\mathbf{x}_u) - 1}\}$: From Lemma 2, we have:

$$\begin{aligned} &\mathrm{tr}\{\frac{\bar{\mathbf{S}}^{-1}f(\mathbf{x}_u)f(\mathbf{x}_u)^T\bar{\mathbf{S}}^{-1}\mathbf{S}_B}{f(\mathbf{x}_u)^T\bar{\mathbf{S}}^{-1}f(\mathbf{x}_u) - 1}\} \quad (5) \\ &\le \frac{\sum_i \sigma_i(\bar{\mathbf{S}}^{-1}\mathbf{S}_B\bar{\mathbf{S}}^{-1})\sigma_i(f(\mathbf{x}_u)f(\mathbf{x}_u)^T)}{f(\mathbf{x}_u)^T\bar{\mathbf{S}}^{-1}f(\mathbf{x}_u) - 1} \\ &\le \frac{f(\mathbf{x}_u)^T f(\mathbf{x}_u)\sigma_1(\bar{\mathbf{S}}^{-1}\mathbf{S}_B\bar{\mathbf{S}}^{-1})}{f(\mathbf{x}_u)^T\bar{\mathbf{S}}^{-1}f(\mathbf{x}_u) - 1} \end{aligned}$$

where $\sigma_1(\cdot)$ denotes the largest singular value. Given that the largest singular value is actually the spectral norm, based on the norm submultiplicative, we have:

$$\sigma_1(\bar{\mathbf{S}}^{-1}\mathbf{S}_B\bar{\mathbf{S}}^{-1}) \le \|\bar{\mathbf{S}}^{-1}\|_2^2\|\mathbf{S}_B\|_2 \quad (6)$$

For the first norm, $\|\bar{\mathbf{S}}^{-1}\|_2 = 1/\sigma_{min}(\bar{\mathbf{S}})$. Typically, $\bar{\mathbf{S}}$ is regularized by a ridge parameter $\rho > 0$, i.e. $\bar{\mathbf{S}} + \rho\mathbf{I}$, it can be said that $\sigma_{min}(\bar{\mathbf{S}}) > \rho$, so that $\|\bar{\mathbf{S}}^{-1}\|_2 < 1/\rho$. For the second norm, $\|\mathbf{S}_B\|_2 = \|\sum_{c=1}^N M_c\boldsymbol{\mu}_c\boldsymbol{\mu}_c^T\|_2 \le \sum_{c=1}^N M_c\|\boldsymbol{\mu}_c\boldsymbol{\mu}_c^T\|_2 = \sum_{c=1}^N M_c\boldsymbol{\mu}_c^T\boldsymbol{\mu}_c = \delta$. It follows that $\sigma_1(\bar{\mathbf{S}}^{-1}\mathbf{S}_B\bar{\mathbf{S}}^{-1}) \le \delta/\rho^2$. Finally, based on the von Neumann [3] property, $f(\mathbf{x}_u)^T\bar{\mathbf{S}}^{-1}f(\mathbf{x}_u) - 1 = \mathrm{tr}\{f(\mathbf{x}_u)^T\bar{\mathbf{S}}^{-1}f(\mathbf{x}_u)\} - 1 = C\sigma_1(\bar{\mathbf{S}}^{-1})f(\mathbf{x}_u)^T f(\mathbf{x}_u) - 1$, where $C \in [-1, 1]$. Hence, for simplicity, we use the following approximation: $f(\mathbf{x}_u)^T\bar{\mathbf{S}}^{-1}f(\mathbf{x}_u) - 1 \approx f(\mathbf{x}_u)^T f(\mathbf{x}_u)/\rho - 1$. Then, we can derive the upper-bound for $\mathrm{tr}\{\frac{\bar{\mathbf{S}}^{-1}f(\mathbf{x}_u)f(\mathbf{x}_u)^T\bar{\mathbf{S}}^{-1}\mathbf{S}_B}{f(\mathbf{x}_u)^T\bar{\mathbf{S}}^{-1}f(\mathbf{x}_u) - 1}\}$ as:

$$\mathrm{tr}\{\frac{\bar{\mathbf{S}}^{-1}f(\mathbf{x}_u)f(\mathbf{x}_u)^T\bar{\mathbf{S}}^{-1}\mathbf{S}_B}{f(\mathbf{x}_u)^T\bar{\mathbf{S}}^{-1}f(\mathbf{x}_u) - 1}\} \le \frac{\delta f(\mathbf{x}_u)^T f(\mathbf{x}_u)}{\rho(f(\mathbf{x}_u)^T f(\mathbf{x}_u) - \rho)} \quad (7)$$

Upper-bound for $\mathrm{tr}\{\bar{\mathbf{S}}^{-1}\tilde{\mathbf{E}}_B\}$: From Lemma 2, we have:

$$\mathrm{tr}\{\bar{\mathbf{S}}^{-1}\tilde{\mathbf{E}}_B\} \le \sum_{i=1}^4 \sigma_i(\bar{\mathbf{S}}^{-1})\sigma_i(\tilde{\mathbf{E}}_B) \quad (8)$$

since rank$(\tilde{\mathbf{E}}_B) \le 4$ [1]. Then, with Lemma 1, we have $\sigma_i(\tilde{\mathbf{E}}_B) \le \frac{\|\tilde{\mathbf{E}}_B\|_F}{\sqrt{i}} = \frac{\|\mathbf{E}_B\|_F}{\sqrt{i}}$. By substituting the definition of $\mathbf{E}_B$ and using the triangular inequality, we have:

$$\sigma_i(\tilde{\mathbf{E}}_B) \le \frac{\|M_u\boldsymbol{\mu}_u\boldsymbol{\mu}_u^T - M_u\boldsymbol{\mu}_u f(\mathbf{x}_u)^T - M_u f(\mathbf{x}_u)\boldsymbol{\mu}_u^T\|_F + \|f(\mathbf{x}_u)f(\mathbf{x}_u)^T\|_F}{(M_u - 1)\sqrt{i}} \quad (9)$$

Based on the property that $\|M\|_F^2 = \mathrm{tr}(M^T M)$:

$$\sigma_i(\tilde{\mathbf{E}}_B) \le \frac{\nu_u + f(\mathbf{x}_u)^T f(\mathbf{x}_u)}{(M_u - 1)\sqrt{i}} \quad (10)$$

where the definition of ν_u is listed in Theorem 3 of our paper. With the bound on $\sigma_1(\bar{\mathbf{S}}^{-1}) < 1/\rho$, we can derive the upper-bound for $\mathrm{tr}\{\bar{\mathbf{S}}^{-1}\tilde{\mathbf{E}}_B\}$ as:

$$\mathrm{tr}\{\bar{\mathbf{S}}^{-1}\tilde{\mathbf{E}}_B\} \leq \sum_{i=1}^{4} \frac{\nu_u + f(\mathbf{x}_u)^T f(\mathbf{x}_u)}{\rho(M_u-1)\sqrt{i}} \leq \frac{H_{4,1/2}(\nu_u + f(\mathbf{x}_u)^T f(\mathbf{x}_u))}{\rho(M_u-1)} \tag{11}$$

Upper-bound for $\mathrm{tr}\{\frac{\bar{\mathbf{S}}^{-1} f(\mathbf{x}_u) f(\mathbf{x}_u)^T \bar{\mathbf{S}}^{-1}\mathbf{E}_B}{f(\mathbf{x}_u)^T \bar{\mathbf{S}}^{-1} f(\mathbf{x}_u) - 1}\}$: With similar derivation as in the first term, we have:

$$\begin{aligned} &\mathrm{tr}\{\frac{\bar{\mathbf{S}}^{-1} f(\mathbf{x}_u) f(\mathbf{x}_u)^T \bar{\mathbf{S}}^{-1}\mathbf{E}_B}{f(\mathbf{x}_u)^T \bar{\mathbf{S}}^{-1} f(\mathbf{x}_u) - 1}\} \\ &\leq \frac{f(\mathbf{x}_u)^T f(\mathbf{x}_u)\sigma_1(\bar{\mathbf{S}}^{-1}\mathbf{E}_B\bar{\mathbf{S}}^{-1})}{f(\mathbf{x}_u)^T f(\mathbf{x}_u)/\rho - 1} \end{aligned} \tag{12}$$

Again, based on the norm submultiplicative, $\sigma_1(\bar{\mathbf{S}}^{-1}\mathbf{E}_B\bar{\mathbf{S}}^{-1}) \leq \|\bar{\mathbf{S}}^{-1}\|_2^2\|\mathbf{E}_B\|_2$. From the derivation in the second term, we readily get $\|\mathbf{E}_B\|_2 = \sigma_1(\|\mathbf{E}_B\|_2) \leq \|\mathbf{E}_B\|_F \leq \frac{\nu_u + f(\mathbf{x}_u)^T f(\mathbf{x}_u)}{(M_u-1)}$. Using the upper-bound for $\|\bar{\mathbf{S}}^{-1}\|_2$, we can obtain the bound $\sigma_1(\bar{\mathbf{S}}^{-1}\mathbf{E}_B\bar{\mathbf{S}}^{-1}) \leq \|\mathbf{E}_B\|_F \leq \frac{\nu_u + f(\mathbf{x}_u)^T f(\mathbf{x}_u)}{(M_u-1)\rho^2}$. Finally, we can derive the upper-bound for the third term:

$$\mathrm{tr}\{\frac{\bar{\mathbf{S}}^{-1} f(\mathbf{x}_u) f(\mathbf{x}_u)^T \bar{\mathbf{S}}^{-1}\mathbf{E}_B}{f(\mathbf{x}_u)^T \bar{\mathbf{S}}^{-1} f(\mathbf{x}_u) - 1}\} \leq \frac{f(\mathbf{x}_u)^T f(\mathbf{x}_u)(\nu_u + f(\mathbf{x}_u)^T f(\mathbf{x}_u))}{\rho(f(\mathbf{x}_u)^T f(\mathbf{x}_u) - \rho)(M_u-1)} \tag{13}$$

Finally, we can conclude the upper-bound for $d\psi_u$ by combining the upper-bounds for three additive terms together. □

References

[1] Roger A Horn and Charles R Johnson. Matrix analysis. 2012. 1

[2] Jorma Kaarlo Merikoski, Humberto Sarria, and Pablo Tarazaga. Bounds for singular values using traces. *Linear Algebra and its Applications*, 210:227–254, 1994. 1

[3] John Von Neumann. Some matrix-inequalities and metrization of matric space. 1937. 1