Balanced Setting Nearest Neighbor Proof

We consider the following binary classification problem:

$$y = \begin{cases} U_1 x + z & y \in H_1 \\ U_2 x + z & y \in H_2 \end{cases} \tag{1}$$

where $z \sim \mathcal{N}(0, \sigma^2 I), x \sim \mathcal{N}(0, \Sigma)$ with known diagonal Σ. We assume $U_1, U_2 \in \mathbb{C}^{p \times k}$ are known and have column rank k, which is also known.

As x is unknown, we consider the generalized likelihood ratio test (GLRT)

$$\Lambda(y) = \frac{\max_x f_2(y)}{\max_x f_1(y)} \tag{2}$$

where

$$\begin{aligned} &\text{Declare } H_1 \text{ if} \quad \Lambda(y) < \eta \\ &\text{Declare } H_2 \text{ if} \quad \Lambda(y) > \eta \end{aligned} \tag{3}$$

where $\eta = \frac{\text{Prob}(H_1)}{\text{Prob}(H_2)}$. We assume that H_1 and H_2 are equally likely so that $\eta = 1$.

Now, under $H_1, y_1 \sim \mathcal{N}(U_1 x, \sigma^2 I)$ and under $H_2, y_2 \sim \mathcal{N}(U_2 x, \sigma^2)$.

To solve $\max_x f_i(y)$ we find the MLE estimate of x. We have

$$f_i(y; x) = (2\pi\sigma^2)^{-p/2} \exp\{-\frac{1}{2\sigma^2}(y - U_i x)^H (y - U_i x)\} \tag{4}$$

and

$$\ln(f_i(y; x)) = -(p/2)(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y - U_i x)^H (y - U_i x) \tag{5}$$

The first term of (5) is constant. As we wish to maximize (5), this is equivalent to minimizing $(y - U_i x)^H (y - U_i x)$. Noticing that is this is the error of the the least-squares solution, we recognize that

$$\hat{x}_{MLE} = (U_i^H U_i)^{-1} U_i^H y \tag{6}$$

Plugging our MLE estimates (6) into our GLRT (32) we have

$$\Lambda(y) = \frac{(2\pi\sigma^2)^{-p/2} \exp\{-\frac{1}{2\sigma^2}(y - U_2(U_2^H U_2)^{-1} U_2^H y)^H (y - U_2(U_2^H U_2)^{-1} U_2^H y)\}}{(2\pi\sigma^2)^{-p/2} \exp\{-\frac{1}{2\sigma^2}(y - U_1(U_1^H U_1)^{-1} U_1^H y)^H (y - U_1(U_1^H U_1)^{-1} U_1^H y)\}} \tag{7}$$

which simplifies to

$$\hat{\Lambda}(y) = \|(I - U_1(U_1^H U_1)^{-1} U_1^H) y\|_F^2 - \|(I - U_2(U_2^H U_2)^{-1} U_2^H) y\|_F^2 \tag{8}$$

where

$$\begin{aligned} &\text{Declare } H_1 \text{ if} \quad \hat{\Lambda}(y) < 2\ln(\eta) \\ &\text{Declare } H_2 \text{ if} \quad \hat{\Lambda}(y) > 2\ln(\eta) \end{aligned} \tag{9}$$

Recalling that $\eta = 1$ this reduces to

$$i_{\text{oracle}} = \underset{i \in \{1,2\}}{\text{argmin}} \|((I - U_i(U_i^H U_i)^{-1} U_i^H) y\|_F^2 \tag{10}$$

If we further assume that the columns of U_1, U_2 are orthonormal, this reduces to

$$i_{\text{oracle}} = \underset{i \in \{1,2\}}{\text{argmin}} \|((I - U_i U_i^H) y\|_F^2 \tag{11}$$

Balanced Setting Stochastic Proof

We consider the following binary classification problem:

$$y = \begin{cases} U_1 x_1 + z & y \in H_1 \\ U_2 x_2 + z & y \in H_2 \end{cases} \tag{12}$$

where $z \sim \mathcal{N}(0, \sigma^2 I), x_1 \sim \mathcal{N}(0, \Sigma_1), x_2 \sim \mathcal{N}(0, \Sigma_2)$ with known diagonal Σ. We assume $U_1, U_2 \in \mathbb{C}^{p\times k}$ are known and have orthonormal columns and thus have rank k, which is also known. We also assume that x and z are independent. We therefore have that

$$\begin{aligned} y|H_1 &\sim \mathcal{N}(0, \sigma^2 I + U_1 \Sigma_1 U_1^H) \\ y|H_2 &\sim \mathcal{N}(0, \sigma^2 I + U_2 \Sigma_2 U_2^H) \end{aligned} \tag{13}$$

We then consider the likelihood ratio test

$$\begin{aligned} \Lambda(y) &= \frac{f(y|H_2)}{f(y|H_1)} \\ &= \frac{(2\pi)^{-p/2} \det(\sigma^2 I + U_2\Sigma_2 U_2^H)^{-1/2} \exp\{-\frac{1}{2} y^H (\sigma^2 I + U_2 \Sigma_2 U_2^H)^{-1} y\}}{(2\pi)^{-p/2} \det(\sigma^2 I + U_1\Sigma_1 U_1^H)^{-1/2} \exp\{-\frac{1}{2} y^H (\sigma^2 I + U_1 \Sigma_1 U_1^H)^{-1} y\}} \\ &= \frac{C}{\det(\sigma^2 I + U_1 \Sigma_1 U_1^H)^{-1/2}} \exp\{-\frac{1}{2} y^H \left[(\sigma^2 I + U_2\Sigma_2 U_2^H)^{-1} - (\sigma^2 I + U_1 \Sigma_1 U_1^H)^{-1}\right] y\} \end{aligned} \tag{14}$$

This simplifies to

$$\hat{\Lambda}(y) = y^H \left[(\sigma^2 I + U_1 \Sigma_1 U_1^H)^{-1} - (\sigma^2 I + U_2 \Sigma_2 U_2^H)^{-1}\right] y \tag{15}$$

where

$$\begin{aligned} &\text{Declare } H_1 \text{ if} \quad \hat{\Lambda}(y) < 2\ln\left(\eta \frac{\det(\sigma^2 I + U_1\Sigma_1 U_1^H)^{1/2}}{\det(\sigma^2 I + U_2 \Sigma_2 U_2^H)^{1/2}}\right) \\ &\text{Declare } H_2 \text{ if} \quad \hat{\Lambda}(y) > 2\ln\left(\eta \frac{\det(\sigma^2 I + U_1\Sigma_1 U_1^H)^{1/2}}{\det(\sigma^2 I + U_2 \Sigma_2 U_2^H)^{1/2}}\right) \end{aligned} \tag{16}$$

Using the Sherman-Morrison-Woodberry matrix inversion lemma, we may simplify (15)

$$\hat{\Lambda}(y) = y^H \left[\frac{1}{\sigma^2} I - \frac{1}{\sigma^2} U_1 (\Sigma_1^{-1} + \frac{1}{\sigma^2} U_1^H U_1)^{-1} U_1^H \frac{1}{\sigma^2} - (\frac{1}{\sigma^2} I - \frac{1}{\sigma^2} U_2 (\Sigma_2^{-1} + \frac{1}{\sigma^2} U_2^H U_2)^{-1} U_2^H \frac{1}{\sigma^2})\right] y \tag{17}$$

Recalling that U_1, U_2 have orthonormal columns, we can simplify this to

$$\hat{\Lambda}(y) = \frac{1}{\sigma^2} y^H \left[U_2 (\sigma^2 \Sigma_2^{-1} + I)^{-1} U_2^H - U_1 (\sigma^2 \Sigma_1^{-1} + I)^{-1} U_1^H\right] y \tag{18}$$

We may also simplify our expressions in (16) by noting that since U_1 and U_2 have orthonormal columns by noting that

$$\det(\sigma^2 I + U_i \Sigma_i U_i^H)^{1/2} = (\sigma^2)^{p-k} \det(\sigma^2 I_k + \Sigma_i) \quad i = 1, 2 \tag{19}$$

Thus, defining

$$\tilde{\Lambda}(y) = y^H \left[U_2 (\sigma^2 \Sigma_2^{-1} + I)^{-1} U_2^H - U_1 (\sigma^2 \Sigma_1^{-1} + I)^{-1} U_1^H\right] y \tag{20}$$

we have the decision rule

$$
\begin{array}{ll}
\text{Declare } H_1 \text{ if} & \tilde{\Lambda}(y) < \sigma^2 \ln\left(\eta^2 \dfrac{\det(\sigma^2 I_k + \Sigma_1)}{\det(\sigma^2 I_k + \Sigma_2)}\right) \\
\text{Declare } H_2 \text{ if} & \tilde{\Lambda}(y) > \sigma^2 \ln\left(\eta^2 \dfrac{\det(\sigma^2 I_k + \Sigma_1)}{\det(\sigma^2 I_k + \Sigma_2)}\right)
\end{array}
\tag{21}
$$

We now consider 4 different cases

1. Equal Covariances of I_k
2. Equal Covariances of Σ
3. $k_1 \neq k_2$, but both covariances are identity ($\Sigma_1 = I_{k_1}, \Sigma_2 = I_{k_2}$)
4. $k_1 \neq k_2$, covariances are not identity (Σ_1, Σ_2)

We consider each case individually,

Case 1

When both covariances are equal and identity, (20) becomes

$$
\begin{aligned}
\tilde{\Lambda}(y) &= y^H \left[U_2(\sigma^2 I + I)^{-1} U_2^H - U_1(\sigma^2 I + I)^{-1} U_1^H\right] y \\
&= \frac{1}{1+\sigma^2} y^H \left[U_2 U_2^H - U_1 U_1^H\right] y
\end{aligned}
\tag{22}
$$

Thus, defining

$$
\tilde{\Lambda}_1(y) = y^H \left[U_2 U_2^H - U_1 U_1^H\right] y \tag{23}
$$

we have the decision rule

$$
\begin{array}{ll}
\text{Declare } H_1 \text{ if} & \tilde{\Lambda}_1(y) < 2(\sigma^2 + 1)\sigma^2 \ln(\eta) \\
\text{Declare } H_2 \text{ if} & \tilde{\Lambda}_1(y) > 2(\sigma^2 + 1)\sigma^2 \ln(\eta)
\end{array}
\tag{24}
$$

Case 2

When both covariances are equal but not identity ($\Sigma_1 = \Sigma_2 = \Sigma$), (20) becomes

$$
\begin{aligned}
\tilde{\Lambda}(y) &= y^H \left[U_2(\sigma^2 \Sigma + I)^{-1} U_2^H - U_1(\sigma^2 \Sigma + I)^{-1} U_1^H\right] y \\
&= y^H \left([U_2 - U_1] (\sigma^2 \Sigma + I)^{-1} \begin{bmatrix} U_2^H \\ U_1^H \end{bmatrix} \right) y
\end{aligned}
\tag{25}
$$

Thus, defining

$$
\tilde{\Lambda}_2(y) = y^H \left([U_2 - U_1] (\sigma^2 \Sigma + I)^{-1} \begin{bmatrix} U_2^H \\ U_1^H \end{bmatrix} \right) y \tag{26}
$$

we have the decision rule

$$
\begin{array}{ll}
\text{Declare } H_1 \text{ if} & \tilde{\Lambda}_2(y) < 2\sigma^2 \ln(\eta) \\
\text{Declare } H_2 \text{ if} & \tilde{\Lambda}_2(y) > 2\sigma^2 \ln(\eta)
\end{array}
\tag{27}
$$

Case 3

We now consider the case when $k_1 \neq k_2$ but $\Sigma_1 = I_{k_1}$ and $\Sigma_2 = I_{k_2}$. (20) becomes

$$
\begin{aligned}
\tilde{\Lambda}(y) &= y^H \left[U_2(\sigma^2 I_{k_2} + I_{k_2})^{-1} U_2^H - U_1(\sigma^2 I_{k_1} + I_{k_1})^{-1} U_1^H\right] y \\
&= \frac{1}{1+\sigma^2} y^H \left[U_2 U_2^H - U_1 U_1^H\right] y
\end{aligned}
\tag{28}
$$

Thus, defining

$$\tilde{\Lambda}_3(y) = y^H \left[U_2 U_2^H - U_1 U_1^H \right] y \tag{29}$$

we have the decision rule

$$\begin{aligned} &\text{Declare } H_1 \text{ if } \quad \tilde{\Lambda}_3(y) < 2(\sigma^2+1)\sigma^2 \ln(\eta) \\ &\text{Declare } H_2 \text{ if } \quad \tilde{\Lambda}_3(y) > 2(\sigma^2+1)\sigma^2 \ln(\eta) \end{aligned} \tag{30}$$

which is clearly the same solution as case 1.

Case 4

Finally, we consider when $k_1 \neq k_2$ and $\Sigma_1 \neq \Sigma_2$.

Equations (20) and (21) specify our detector as they are as general as possible.

Finally we consider the case when $\eta = 1$. In all cases, this forces the right side of all our decision rules to 0. For cases 1 and 3, this simply becomes the nearest neighbor detector. For cases 2 and 4, this becomes a weighted nearest neighbor problem, where we scale our U matrices by a matrix which incorporates our noise variance, σ^2 and our signal covariance.

Isotropically Random Noise Detector Proof

We consider the following binary classification problem:

$$Y = \begin{cases} U_1 X + W & Y \in H_1 \\ U_2 X + W & Y \in H_2 \end{cases} \tag{31}$$

where $W \in \mathbb{C}^{p \times n}$ is isotropically random noise such that $f(W) = f(Z^H W V)$ where $Z \in \mathbb{C}^{p \times p}, V \in \mathbb{C}^{n \times n}$ are orthogonal. X=$[x_1, \ldots, x_n]$ such that $x_i \sim \mathcal{N}(0, \Sigma)$ for $i = 1, \ldots n$ with known diagonal $\Sigma \in \mathbb{C}^{p \times p}$. We assume $U_1, U_2 \in \mathbb{C}^{p \times k}$ are known and have column rank k, which is also known.

As x is unknown, we consider the generalized likelihood ratio test (GLRT)

$$\Lambda(y) = \frac{\max_x f_2(y)}{\max_x f_1(y)} \tag{32}$$

where

$$\begin{aligned} &\text{Declare } H_1: \quad \Lambda(y) < \eta \\ &\text{Declare } H_2: \quad \Lambda(y) > \eta \end{aligned} \tag{33}$$

where $\eta = \frac{\text{Prob}(H_1)}{\text{Prob}(H_2)}$. We assume that H_1 and H_2 are equally likely so that $\eta = 1$.

Gaussian Case

We assume that $W = [w_1, \ldots, w_n]$ is Gaussian such that $w_i \sim \mathcal{N}(0, \sigma^2 I)$

We also have $Y = [y_1, \ldots, y_n]$. Under H_1, we have that $y_i \sim \mathcal{N}(U_1 x_i, \sigma^2 I)$ and under H_2, we have that $y_i \sim \mathcal{N}(U_2 x_i, \sigma^2 I)$ where the y_i are independent regardless of the hypothesis. Therefore, we have

$$
\begin{aligned}
cllf_j(Y) &= \prod_{i=1}^{n} f_j(y_i) \\
&= \prod_{i=1}^{n}(2\pi\sigma^2)^{-p/2}\exp\{-\frac{1}{2\sigma^2}(y_i - U_j x_i)^H(y_i - U_j x_i)\} \\
&= (2\pi\sigma^2)^{-pn/2}\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - U_j x_i)^H(y_i - U_j x_i)\} \\
&= (2\pi\sigma^2)^{-pn/2}\exp\{-\frac{1}{2\sigma^2}/Tr((Y - U_j X)^H(Y - U_j X))\} \\
&= (2\pi\sigma^2)^{-pn/2}\exp\{-\frac{1}{2\sigma^2}\|Y - U_j X\|_F^2\}
\end{aligned} \tag{34}
$$

Our GLRT becomes

$$
\Lambda(Y) = \frac{f_2(Y)}{f_1(Y)} = \frac{(2\pi\sigma^2)^{-pn/2}\exp\{-\frac{1}{2\sigma^2}\|Y - U_2 X\|_F^2\}}{(2\pi\sigma^2)^{-pn/2}\exp\{-\frac{1}{2\sigma^2}\|Y - U_1 X\|_F^2\}} \tag{35}
$$

which simplifies to

$$
\hat{\Lambda}(Y) = \|Y - U_1 X\|_F^2 - \|Y - U_2 X\|_F^2 \tag{36}
$$

where

$$
\begin{aligned}
&\text{Declare } H_1 \text{ if} \quad \hat{\Lambda}(y) < 2\ln(\eta) \\
&\text{Declare } H_2 \text{ if} \quad \hat{\Lambda}(y) > 2\ln(\eta)
\end{aligned} \tag{37}
$$

Recognizing (36) as a difference of matrix least squares, we may use the orthogonal invariance of the Frobenius norm to rewrite (36) as

$$
\hat{\Lambda}(Y) = \|U_1^H Y - X\|_F^2 - \|U_2^H Y - X\|_F^2 \tag{38}
$$

Let $U_1^H Y = A = U_A \Sigma_A V_B^T$ and $U_2^H Y = B = U_B \Sigma_B V_B^T$, where $\Sigma_A = \mathbf{diag}(\sigma_{A1}, \ldots, \sigma_{An}), \Sigma_A = \mathbf{diag}(\sigma_{B1}, \ldots, \sigma_{Bn})$. X is simply the best rank-k matrix which minimizes the Frobenius norm difference between A and B. Then the Eckart Young Mirsky Theorem states that these errors are the sum the last $p - k$ singular values. Therefore, (38) becomes

$$
\hat{\Lambda}(Y) = \sum_{i=k+1}^{p} \sigma_{Ai}^2 - \sum_{i=k+1}^{p} \sigma_{Bi}^2 \tag{39}
$$

Since $\eta = 1$, we have the decision

$$
\begin{aligned}
&\text{Declare } H_1 \text{ if} \quad \hat{\Lambda}(y) < 0 \\
&\text{Declare } H_2 \text{ if} \quad \hat{\Lambda}(y) > 0
\end{aligned} \tag{40}
$$

which simply chooses the hypothesis with the lowest reconstruction error.