Binary Classification with Estimated Subspaces

June 15, 2011

Unprocessed Binary Classifier with Known Subspaces

We consider the following binary classification problem:

$$y = \begin{cases} U_0 x_0 + z & y \in H_0 \\ U_1 x_1 + z & y \in H_1 \end{cases} \tag{1}$$

where $z \sim \mathcal{N}(0, \sigma^2 I), x_0 \sim \mathcal{N}(0, \Sigma_0), x_1 \sim \mathcal{N}(0, \Sigma_1)$ with known diagonal Σ_0, Σ_1. We assume $U_0 \in \mathbb{C}^{p \times k_0}, U_1 \in \mathbb{C}^{p \times k_1}$ are known and have orthonormal columns. We also assume that x and z are independent. We therefore have that

$$\begin{aligned} y|H_0 &\sim \mathcal{N}(0, \sigma^2 I + U_0 \Sigma_0 U_0^H) \\ y|H_1 &\sim \mathcal{N}(0, \sigma^2 I + U_1 \Sigma_1 U_1^H) \end{aligned} \tag{2}$$

We then consider the likelihood ratio test

$$\begin{aligned} \Lambda(y) &= \frac{f(y|H_1)}{f(y|H_0)} \\ &= \frac{(2\pi)^{-p/2} \det(\sigma^2 I + U_1 \Sigma_1 U_1^H)^{-1/2} \exp\{-\frac{1}{2} y^H (\sigma^2 I + U_1 \Sigma_1 U_1^H)^{-1} y\}}{(2\pi)^{-p/2} \det(\sigma^2 I + U_0 \Sigma_0 U_0^H)^{-1/2} \exp\{-\frac{1}{2} y^H (\sigma^2 I + U_0 \Sigma_0 U_0^H)^{-1} y\}} \\ &= \frac{\det(\sigma^2 I + U_1 \Sigma_1 U_1^H)^{-1/2}}{\det(\sigma^2 I + U_0 \Sigma_0 U_0^H)^{-1/2}} \exp\{-\frac{1}{2} y^H \left[(\sigma^2 I + U_1 \Sigma_1 U_1^H)^{-1} - (\sigma^2 I + U_0 \Sigma_0 U_0^H)^{-1}\right] y\} \end{aligned} \tag{3}$$

This simplifies to

$$\hat{\Lambda}(y) = y^H \left[(\sigma^2 I + U_0 \Sigma_0 U_0^H)^{-1} - (\sigma^2 I + U_1 \Sigma_1 U_1^H)^{-1}\right] y \tag{4}$$

where

$$\begin{aligned} &\text{Declare } H_0 \text{ if } \quad \hat{\Lambda}(y) < 2 \ln \left(\eta \frac{\det(\sigma^2 I + U_0 \Sigma_0 U_0^H)^{1/2}}{\det(\sigma^2 I + U_1 \Sigma_1 U_1^H)^{1/2}} \right) \\ &\text{Declare } H_1 \text{ if } \quad \hat{\Lambda}(y) > 2 \ln \left(\eta \frac{\det(\sigma^2 I + U_0 \Sigma_0 U_0^H)^{1/2}}{\det(\sigma^2 I + U_1 \Sigma_1 U_1^H)^{1/2}} \right) \end{aligned} \tag{5}$$

Using the Sherman-Morrison-Woodberry matrix inversion lemma, we may simplify (4)

$$\hat{\Lambda}(y) = y^H \left[\frac{1}{\sigma^2} I - \frac{1}{\sigma^2} U_0 (\Sigma_0^{-1} + \frac{1}{\sigma^2} U_0^H U_0)^{-1} U_0^H \frac{1}{\sigma^2} - (\frac{1}{\sigma^2} I - \frac{1}{\sigma^2} U_1 (\Sigma_1^{-1} + \frac{1}{\sigma^2} U_1^H U_1)^{-1} U_1^H \frac{1}{\sigma^2}) \right] y \tag{6}$$

Recalling that U_0, U_1 have orthonormal columns, we can simplify this to

$$\hat{\Lambda}(y) = \frac{1}{\sigma^2} y^H \left[U_1(\sigma^2\Sigma_1^{-1} + I)^{-1} U_1^H - U_0(\sigma^2\Sigma_0^{-1} + I)^{-1} U_0^H \right] y \tag{7}$$

Thus, defining

$$\boxed{\tilde{\Lambda}(y) = y^H \left[U_1(\sigma^2\Sigma_1^{-1} + I)^{-1} U_1^H - U_0(\sigma^2\Sigma_0^{-1} + I)^{-1} U_0^H \right] y} \tag{8}$$

we have the decision rule

$$\begin{aligned} &\text{Declare } H_1 \text{ if} \quad \tilde{\Lambda}(y) < \gamma \\ &\text{Declare } H_2 \text{ if} \quad \tilde{\Lambda}(y) > \gamma \end{aligned} \tag{9}$$

where γ is our threshold.

Plug-in Unprocessed Binary Classifier

In most cases, U_0, U_1 are not known and we must make estimates $\hat{U}_0, \hat{U}_1$ from the data. One solution is to simply "plug-in" our estimates into our detector where ever the true U_0, U_1 appear. Doing so yields

$$\boxed{\hat{\Lambda}(y) = y^H \left[\hat{U}_1(\sigma^2\Sigma_1^{-1} + I)^{-1} \hat{U}_1^H - \hat{U}_0(\sigma^2\Sigma_0^{-1} + I)^{-1} \hat{U}_0^H \right] y} \tag{10}$$

where our decision rule is

$$\begin{aligned} &\text{Declare } H_1 \text{ if} \quad \hat{\Lambda}(y) < \gamma \\ &\text{Declare } H_2 \text{ if} \quad \hat{\Lambda}(y) > \gamma \end{aligned} \tag{11}$$

where γ is our threshold.

Processed Binary Classifier with Known Subspaces

We consider the following binary classification problem:

$$y = \begin{cases} U_0 x + z & y \in H_0 \\ U_1 x + z & y \in H_1 \end{cases} \tag{12}$$

where $z \in \mathbb{C}^n, z \sim \mathcal{N}(0, \sigma^2 I)$ under both hypotheses. We assume that $U_0 \in \mathbb{C}^{n \times k_0}$, $U_1 \in \mathbb{C}^{n \times k_1}$ are known and $U_0 \perp U_1$. We assume that $f(x|H_0) \sim \mathcal{N}(0, \Sigma_0)$ and $f(x|H_1) \sim \mathcal{N}(0, \Sigma_1)$ where Σ_0, Σ_1 are diagonal. We also assume that x and z are independent.

By properties of Gaussian random variables, under H_0, $y \sim \mathcal{N}(0, U_0\Sigma_0 U_0^H + \sigma^2 I)$ and Under H_1, $y \sim \mathcal{N}(0, U_1\Sigma_1 U_1^H + \sigma^2 I)$.

We consider the processed vector

$$w = \begin{bmatrix} U_0^H y \\ U_1^H y \end{bmatrix} \tag{13}$$

Under H_0,

$$w = \begin{bmatrix} \mathcal{N}(0, \Sigma_0 + \sigma^2 I) \\ \mathcal{N}(0, U_1^H U_0 \Sigma_0 U_0^H U_1 + \sigma^2 I) \end{bmatrix} \tag{14}$$

Therefore, under H_0,

$$w \sim \mathcal{N}\left(\left[\begin{array}{c} 0 \\ 0 \end{array}\right], \left[\begin{array}{ll} \Sigma_0 + \sigma^2 I & (\Sigma_0 + \sigma^2) U_0^H U_1 \\ U_1^H U_0 (\Sigma_0 + \sigma^2) & U_1^H U_0 \Sigma_0 U_0^H U_1^H + \sigma^2 I \end{array}\right]\right) \tag{15}$$

And since $U_0 \perp U_1$

$$w \sim \mathcal{N}\left(\left[\begin{array}{c} 0 \\ 0 \end{array}\right], C_0 = \left[\begin{array}{ll} \Sigma_0 + \sigma^2 I & 0 \\ 0 & \sigma^2 I \end{array}\right]\right) \tag{16}$$

Similarly, under H_1,

$$w \sim \mathcal{N}\left(\left[\begin{array}{c} 0 \\ 0 \end{array}\right], C_1 = \left[\begin{array}{ll} \sigma^2 I & 0 \\ 0 & \Sigma_1 + \sigma^2 I \end{array}\right]\right) \tag{17}$$

Our likelihood ratio test (LRT) is

$$\begin{aligned} \Lambda(w) &= \frac{f_1(w)}{f_0(w)} \\ &= \frac{(2\pi^{-(k_0+k_1)} |C_1|^{-1/2} \exp\{-\frac{1}{2} w^H c_1^{-1} w\}}{(2\pi^{-(k_0+k_1)} |C_0|^{-1/2} \exp\{-\frac{1}{2} w^H C_0^{-1} w\}} \\ &= \left(\frac{|C_0|}{|C_1|}\right)^{1/2} \exp\{-\frac{1}{2} w^H (C_1^{-1} - C_0^{-1}) w\} \end{aligned} \tag{18}$$

where our decision is

$$\begin{array}{ll} \text{Declare } H_0 \text{ if} & \Lambda(w) < \eta \\ \text{Declare } H_1 \text{ if} & \Lambda(w) > \eta \end{array}$$

Defining

$$\tilde{\Lambda}(w) = w^H (C_0^{-1} - C_1^{-1}) w \tag{19}$$

we have

$$C_0^{-1} - C_1^{-1} = \left[\begin{array}{ll} \mathbf{diag}\left(\frac{1}{\Sigma_{0_{ii}} + \sigma^2}\right) & 0 \\ 0 & \frac{1}{\sigma^2} I \end{array}\right] - \left[\begin{array}{ll} \frac{1}{\sigma^2} I & 0 \\ 0 & \mathbf{diag}\left(\frac{1}{\Sigma_{1_{jj}} + \sigma^2}\right) \end{array}\right] = \left[\begin{array}{ll} \mathbf{diag}\left(\frac{-\Sigma_{0_{ii}}}{\sigma^4 + \sigma^2 \Sigma_{0_{ii}}}\right) & 0 \\ 0 & \mathbf{diag}\left(\frac{\Sigma_{1_{jj}}}{\sigma^4 + \sigma^2 \Sigma_{1_{jj}}}\right) \end{array}\right]$$

We may factor a $\frac{1}{\sigma^2}$ and our detector becomes

$$\boxed{\tilde{\Lambda}(w) = w^H \left[\begin{array}{ll} \mathbf{diag}\left(\frac{-\Sigma_{0_{ii}}}{\sigma^2 + \Sigma_{0_{ii}}}\right) & 0 \\ 0 & \mathbf{diag}\left(\frac{\Sigma_{1_{jj}}}{\sigma^2 + \Sigma_{1_{jj}}}\right) \end{array}\right] w} \tag{20}$$

Our decision becomes

$$\begin{array}{ll} \text{Declare } H_0 \text{ if} & \tilde{\Lambda}(w) < \gamma \\ \text{Declare } H_1 \text{ if} & \tilde{\Lambda}(w) > \gamma \end{array}$$

where γ is the threshold of our detector.
This is an equivalent detector to (8).

Plug-in Processed Binary Classifier

When U_0, U_1 are not known, a common strategy is to form estimates $\hat{U}_0, \hat{U}_1$ from labeled training data and plug these expressions into the derived detector for known U_0, U_1. Therefore, our new processed vector becomes

$$\hat{w} = \begin{bmatrix} \hat{U}_0^H y \\ \hat{U}_1^H y \end{bmatrix} \tag{21}$$

And our statistic becomes

$$\boxed{\hat{\Lambda}(\hat{w}) = \hat{w}^H \begin{bmatrix} \mathbf{diag}\left(\frac{-\Sigma_{0_{ii}}}{\sigma^2+\Sigma_{0_{ii}}}\right) & 0 \\ 0 & \mathbf{diag}\left(\frac{\Sigma_{1_{jj}}}{\sigma^2+\Sigma_{1_{jj}}}\right) \end{bmatrix} \hat{w}} \tag{22}$$

Our decision becomes

$$\begin{aligned} &\text{Declare } H_0 \text{ if} \quad \hat{\Lambda}(w) < \gamma \\ &\text{Declare } H_v \text{ if} \quad \hat{\Lambda}(w) > \gamma \end{aligned}$$

where γ is our desired threshold.

This is an equivalent detector to (10).

Random Matrix Theory Processed Binary Classifier

We consider the same binary classification problem:

$$y = \begin{cases} U_0 x + z & y \in H_0 \\ U_1 x + z & y \in H_1 \end{cases} \tag{23}$$

where $z \in \mathbb{C}^n, z \sim \mathcal{N}(0, \sigma^2 I)$ under both hypotheses. We assume that $U_0 \in \mathbb{C}^{n\times k_0}$, $U_1 \in \mathbb{C}^{n\times k_1}$ are unknown but we have estimates $\hat{U}_0, \hat{U}_1$ which were estimated from a set of training data. We assume that $f(x|H_0) \sim \mathcal{N}(0, \Sigma_0)$ and $f(x|H_1) \sim \mathcal{N}(0, \Sigma_1)$. We also assume that x and z are independent.

By properties of Gaussian random variables, under H_0, $y \sim \mathcal{N}(0, U_0\Sigma_0 U_0^H + \sigma^2 I)$ and Under H_1, $y \sim \mathcal{N}(0, U_1\Sigma_1 U_1^H + \sigma^2 I)$.

We consider the processed vector

$$w = \begin{bmatrix} \hat{U}_0^H y \\ \hat{U}_1^H y \end{bmatrix} \tag{24}$$

Under H_0,

$$w = \begin{bmatrix} \mathcal{N}(0, \hat{U}_0^H U_0 \Sigma_0 U_0^H \hat{U}_0 + \sigma^2 I) \\ \mathcal{N}(0, \hat{U}_1^H U_0 \Sigma_0 U_0^H \hat{U}_1 + \sigma^2 I) \end{bmatrix} \tag{25}$$

Therefore, under H_0,

$$w \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, C_0 = \begin{bmatrix} \hat{U}_0^H U_0 \Sigma_0 U_0^H \hat{U}_0 + \sigma^2 I & \hat{U}_0^H U_0 \Sigma_0 U_0^H \hat{U}_1 + \sigma^2 \hat{U}_0^H \hat{U}_1 \\ \hat{U}_1^H U_0 \Sigma_0 U_0^H \hat{U}_0 + \sigma^2 \hat{U}_1^H \hat{U}_0 & \hat{U}_1^H U_0 \Sigma_0 U_0^H \hat{U}_1^+ \sigma^2 I \end{bmatrix}\right) \tag{26}$$

Similarly, under H_1,

$$w \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, C_1 = \begin{bmatrix} \hat{U}_0^H U_1 \Sigma_1 U_1^H \hat{U}_0 + \sigma^2 I & \hat{U}_0^H U_1 \Sigma_1 U_1^H \hat{U}_1 + \sigma^2 \hat{U}_0^H \hat{U}_1 \\ \hat{U}_1^H U_1 \Sigma_1 U_1^H \hat{U}_0 + \sigma^2 \hat{U}_1^H \hat{U}_0 & \hat{U}_1^H U_1 \Sigma_1 U_1^H \hat{U}_1 + \sigma^2 I \end{bmatrix}\right) \tag{27}$$

Our likelihood ratio test (LRT) is

$$\begin{aligned}\Lambda(w) &= \frac{f_1(w)}{f_0(w)} \\ &= \frac{(2\pi^{-(k_0+k_1)}|C_1|^{-1/2}\exp\{-\frac{1}{2}w^H c_1^{-1}w\}}{(2\pi^{-(k_0+k_1)}|C_0|^{-1/2}\exp\{-\frac{1}{2}w^H C_0^{-1}w\}} \\ &= \left(\frac{|C_0|}{|C_1|}\right)^{1/2}\exp\{-\frac{1}{2}w^H(C_1^{-1}-C_0^{-1})w\}\end{aligned} \tag{28}$$

where our decision is

$$\begin{aligned}&\text{Declare } H_0 \text{ if} \quad \Lambda(w) < \eta \\ &\text{Declare } H_1 \text{ if} \quad \Lambda(w) > \eta\end{aligned}$$

Defining

$$\tilde{\Lambda}(w) = w^H(C_0^{-1} - C_1^{-1})w \tag{29}$$

Our decision becomes

$$\begin{aligned}&\text{Declare } H_0 \text{ if} \quad \tilde{\Lambda}(w) < 2\ln(\eta) + \ln\left(\frac{|C_1|}{|C_0|}\right) \\ &\text{Declare } H_v \text{ if} \quad \tilde{\Lambda}(w) > 2\ln(\eta) + \ln\left(\frac{|C_1|}{|C_0|}\right)\end{aligned}$$

Let

$$C_0 = \begin{bmatrix} \hat{U}_0^H U_0 \Sigma_0 U_0^H \hat{U}_0 + \sigma^2 I & \hat{U}_0^H U_0 \Sigma_0 U_0^H \hat{U}_1 + \sigma^2 \hat{U}_0^H \hat{U}_1 \\ \hat{U}_1^H U_0 \Sigma_0 U_0^H \hat{U}_0 + \sigma^2 \hat{U}_1^H \hat{U}_0 & \hat{U}_1^H U_0 \Sigma_0 U_0^H \hat{U}_1^+ \sigma^2 I \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \tag{30}$$

Then by the general matrix inversion in block form

$$C_0^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix} \tag{31}$$

We make the following approximation for A. We may ignore off-diagonal terms as these are of the order $O(1/n^2)$. Similarly, we may ignore terms on the diagonal which do not take the form $| < \hat{u}_{0_i}, u_{0_i} > |^2$ as these are also on the order of $O(1/n^2)$. Our approximation is a diagonal matrix:

$$A = \hat{U}_0^H U_0 \Sigma_0 U_0^H \hat{U}_0 + \sigma^2 I \approx \mathbf{diag}\left(\Sigma_{0_{ii}} | < \hat{u}_{0_i}, u_{0_i} > |^2 + \sigma^2\right) \tag{32}$$

We may also approximate D as a diagonal matrix. We may similarly ignore off-diagonal terms as these are of the order $O(1/n^2)$. Our approximation is:

$$D = \hat{U}_1^H U_0 \Sigma_0 U_0^H \hat{U}_1^H + \sigma^2 I \approx \mathbf{diag}\left(\sigma^2 + \sum_{i=1}^{k_0} \Sigma_{0_{ii}} | < \hat{u}_{1_j}, u_{0_i} > |^2\right) \tag{33}$$

We now form an approximate of $(A - BD^{-1}C)^{-1}$. We first consider $BD^{-1}C$:

$$BD^{-1}C = \hat{U}_0^H U_0 \Sigma_0 U_0^H \hat{U}_1 D^{-1} \hat{U}_1^H U_0 \Sigma_0 U_0^H \hat{U}_0 + \sigma^2 \hat{U}_0^H U_0 \Sigma_0 U_0^H \hat{U}_1 D^{-1} \hat{U}_1^H \hat{U}_0 + \sigma^2 \hat{U}_0^H \hat{U}_1 D^{-1} \hat{U}_1^H U_0 \Sigma_0 U_0^H \hat{U}_0 + \sigma^4 \hat{U}_0^H \hat{U}_1 \hat{U}_1^H \hat{U}_0$$

We ignore the two middle terms of this expression as the entries $\hat{U}_1^H \hat{U}_0$ are $\sim \mathcal{N}(0, 1/n)$. We then have the approximation:

$$BD^{-1}C = \hat{U}_0^H U_0 \Sigma_0 U_0^H \hat{U}_1 D^{-1} \hat{U}_1^H U_0 \Sigma_0 U_0^H \hat{U}_0 + \sigma^4 \hat{U}_0^H \hat{U}_1 \hat{U}_1^H \hat{U}_0$$

Now we have the approximation by ignoring the off-diagonal entries in the second term which are of the order $O(1/n^2)$

$$\sigma^4 \hat{U}_0^H \hat{U}_1 D^{-1} \hat{U}_1^H \hat{U}_0 \approx \sigma^4 \,\mathbf{diag}\left(\sum_{j=1}^{k_1} \frac{|<\hat{u}_{0_i}, \hat{u}_{1_j}>|^2}{\sigma^2 + \sum_{i=1}^{k_0} \Sigma_{0_{ii}} |<\hat{u}_{1_j}, u_{0_i}>|^2} \right) \tag{34}$$

We have the following approximations by similarly ignoring off-diagonal terms:

$$\Sigma_0 U_0^H \hat{U}_1 D^{-1} \hat{U}_1^H U_0 \Sigma_0 \approx \mathbf{diag}\left(\Sigma_{0_{ii}}^2 \sum_{j=1}^{k_1} \frac{|<\hat{u}_{1_j}, u_{0_i}>|^2}{\sigma^2 + \sum_{i=1}^{k_0} \Sigma_{0_{ii}} |<\hat{u}_{1_j}, u_{0_i}>|^2} \right) \tag{35}$$

$$\hat{U}_0^H U_0 \Sigma_0 U_0^H \hat{U}_1 D^{-1} \hat{U}_1^H U_0 \Sigma_0 U_0^H \hat{U}_0 \approx \mathbf{diag}\left(|<\hat{u}_{0_i}, u_{0_i}>|^2 \Sigma_{0_{ii}}^2 \sum_{j=1}^{k_1} \frac{|<\hat{u}_{1_j}, u_{0_i}>|^2}{\sigma^2 + \sum_{i=1}^{k_0} \Sigma_{0_{ii}} |<\hat{u}_{1_j}, u_{0_i}>|^2} \right) \tag{36}$$

By combining (34) and (36) we have

$$\begin{aligned} BD^{-1}C \quad &\approx \mathbf{diag}\left(|<\hat{u}_{0_i}, u_{0_i}>|^2 \Sigma_{0_{ii}}^2 \sum_{j=1}^{k_1} \frac{|<\hat{u}_{1_j}, u_{0_i}>|^2}{\sigma^2 + \sum_{i=r}^{k_0} \Sigma_{0_{rr}} |<\hat{u}_{1_j}, u_{0_r}>|^2} + \sum_{j=1}^{k_1} \frac{\sigma^4 |<\hat{u}_{0_i}, \hat{u}_{1_j}>|^2}{\sigma^2 + \sum_{i=r}^{k_0} \Sigma_{0_{rr}} |<\hat{u}_{1_j}, u_{0_r}>|^2} \right) \\ &= \mathbf{diag}\left(\sum_{j=1}^{k_1} \frac{\Sigma_{0_{ii}}^2 |<\hat{u}_{0_i}, u_{0_i}>|^2 |<\hat{u}_{1_j}, u_{0_i}>|^2 + \sigma^4 |<\hat{u}_{0_i}, \hat{u}_{1_j}>|^2}{\sigma^2 + \sum_{r=1}^{k_0} \Sigma_{0_{rr}} |<\hat{u}_{1_j}, u_{0_r}>|^2} \right) \end{aligned} \tag{37}$$

We then have

$$A - BD^{-1}C \approx \mathbf{diag}\left(\sigma^2 + \Sigma_{0_{ii}} |<\hat{u}_{0_i}, u_{0_i}>|^2 - \sum_{j=1}^{k_1} \frac{\Sigma_{0_{ii}}^2 |<\hat{u}_{0_i}, u_{0_i}>|^2 |<\hat{u}_{1_j}, u_{0_i}>|^2 + \sigma^4 |<\hat{u}_{0_i}, \hat{u}_{1_j}>|^2}{\sigma^2 + \sum_{r=1}^{k_0} \Sigma_{0_{rr}} |<\hat{u}_{1_j}, u_{0_r}>|^2} \right) \tag{38}$$

We may form $(A - BD^{-1}C)^{-1}$ by simply inverting the diagonal elements of (38)
We now form an approximation of $(D - CA^{-1}B)^{-1}$

$$CA^{-1}B = \hat{U}_1^H U_0 \Sigma_0 U_0^H \hat{U}_0 A^{-1} \hat{U}_0^H U_0 \Sigma_0 U_0^H \hat{U}_1 + \sigma^2 \hat{U}_1^H U_0 \Sigma_0 U_0^H \hat{U}_0 A^{-1} \hat{U}_0^H \hat{U}_1 + \sigma^2 \hat{U}_1^H \hat{U}_0 A^{-1} \hat{U}_0^H U_0 \Sigma_0 U_0^H \hat{U}_1 + \sigma^4 \hat{U}_1^H \hat{U}_0 A^{-1} \hat{U}_0^H \hat{U}_1$$

Again, we may ignore the middle two terms for the same reason above to yield the approximation

$$CA^{-1}B = \hat{U}_1^H U_0 \Sigma_0 U_0^H \hat{U}_0 A^{-1} \hat{U}_0^H U_0 \Sigma_0 U_0^H \hat{U}_1 + \sigma^4 \hat{U}_1^H \hat{U}_0 A^{-1} \hat{U}_0^H \hat{U}_1 \tag{39}$$

Now we have the approximation by ignoring the off-diagonal entries in the second term which are of the order $O(1/n^2)$

$$\sigma^4 \hat{U}_1^H \hat{U}_0 A^{-1} \hat{U}_0^H \hat{U}_1 \approx \mathbf{diag}\left(\sigma^4 \sum_{i=1}^{k_0} \frac{|<\hat{u}_{1_j}, \hat{u}_{0_i}>|^2}{\Sigma_{0_{ii}} |<\hat{u}_{0_i}, u_{0_i}>|^2 + \sigma^2} \right) \tag{40}$$

By similarly ignoring off-diagonal entries in the first term, we have the approximations:

$$\Sigma_0 U_0^H \hat{U}_0 A^{-1} \hat{U}_0^H U_0 \Sigma_0 \approx \mathbf{diag}\left(\frac{\Sigma_{0_{ii}} |<\hat{u}_{0_i}, u_{0_i}>|^2}{\Sigma_{0_{ii}} |<\hat{u}_{0_i}, u_{0_i}>|^2 + \sigma^2} \right) \tag{41}$$

$$\hat{U}_1^H U_0 \Sigma_0 U_0^H \hat{U}_0 A^{-1} \hat{U}_0^H U_0 \Sigma_0 U_0^H \hat{U}_1 \approx \mathbf{diag}\left(\sum_{i=1}^{k_0} \frac{\Sigma_{0_{ii}} |<\hat{u}_{0_i}, u_{0_i}>|^2 |<\hat{u}_{1_j}, u_{0_i}>|^2}{\Sigma_{0_{ii}} |<\hat{u}_{0_i}, u_{0_i}>|^2 + \sigma^2} \right) \tag{42}$$

Combining (40) and (42) we have

$$\begin{aligned} CA^{-1}B \quad &\approx \mathbf{diag}\left(\sum_{i=1}^{k_0} \frac{\Sigma_{0_{ii}} |<\hat{u}_{0_i}, u_{0_i}>|^2 |<\hat{u}_{1_j}, u_{0_i}>|^2}{\Sigma_{0_{ii}} |<\hat{u}_{0_i}, u_{0_i}>|^2 + \sigma^2} + \sigma^4 \sum_{i=1}^{k_0} \frac{|<\hat{u}_{1_j}, \hat{u}_{0_i}>|^2}{\Sigma_{0_{ii}} |<\hat{u}_{0_i}, u_{0_i}>|^2 + \sigma^2} \right) \\ &= \mathbf{diag}\left(\sum_{i=1}^{k_0} \frac{\Sigma_{0_{ii}} |<\hat{u}_{0_i}, u_{0_i}>|^2 |<\hat{u}_{1_j}, u_{0_i}>|^2 + \sigma^4 |<\hat{u}_{1_j}, \hat{u}_{0_i}>|^2}{\Sigma_{0_{ii}} |<\hat{u}_{0_i}, u_{0_i}>|^2 + \sigma^2} \right) \end{aligned} \tag{43}$$

We then have that

$$D - CA^{-1}B \approx \mathbf{diag}\left(\sigma^2 + \sum_{i=1}^{k_0} \Sigma_{0_{ii}} |<\hat{u}_{1_j}, u_{0_i}>|^2 - \sum_{i=1}^{k_0} \frac{\Sigma_{0_{ii}} |<\hat{u}_{0_i}, u_{0_i}>|^2 |<\hat{u}_{1_j}, u_{0_i}>|^2 + \sigma^4 |<\hat{u}_{1_j}, \hat{u}_{0_i}>|^2}{\Sigma_{0_{ii}} |<\hat{u}_{0_i}, u_{0_i}>|^2 + \sigma^2} \right) \tag{44}$$

We may then form $(D - CA^{-1}B)^{-1}$ by inverting the diagonal elements of (44).
We now turn to the last two submatrices $D^{-1}C(A - BD^{-1}C)^{-1}$ and $A^{-1}B(D - CA^{-1}B)^{-1}$. We have expressions for $D^{-1}, A^{-1}, (A - BD^{-1}C)^{-1}, (D - CA^{-1}B)^{-1}$ which can all be approximated with diagonal matrices. We use these and multiply them with C, B. Now $D^{-1}C(A - BD^{-1}C)^{-1}$ is simply the matrix C whose entries are scaled by the appropriate entries in the flanking diagonal matrices. However, $C = \hat{U}_1^H U_0 \Sigma_0 U_0^H \hat{U}_0 + \sigma^2 \hat{U}_1^H \hat{U}_0$ where every entry is on the order of $O(1/n^2)$. Therefore, we can approximate this as a zero matrix. Similarly, we can approximate $A^{-1}B(D - CA^{-1}B)^{-1}$ as the zero matrix.
Therefore, we have our approximation for C_0^{-1}.

$$C_0^{-1} \approx \begin{bmatrix} (A - BD^{-1}C)^{-1} & 0 \\ 0 & (D - CA^{-1}B)^{-1} \end{bmatrix} \approx \begin{bmatrix} (38)^{-1} & 0 \\ 0 & (44)^{-1} \end{bmatrix} \tag{45}$$

We can make an even further approximation if we want to approximate C_0 as diagonal to start with. We have

$$C_0^{-1} \approx \begin{bmatrix} A^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix} \approx \begin{bmatrix} \mathbf{diag}\left(\frac{1}{\Sigma_{0_{ii}} |<\hat{u}_{0_i}, u_{0_i}>|^2 + \sigma^2} \right) & 0 \\ 0 & \mathbf{diag}\left(\frac{1}{\sigma^2 + \sum_{i=1}^{k_0} \Sigma_{0_{ii}} |<\hat{u}_{1_j}, u_{0_i}>|^2} \right) \end{bmatrix} \tag{46}$$

We can use similar derivation to approximate C_1^{-1}.
Define

$$A - BD^{-1}C \approx \mathbf{diag}\left(\sigma^2 + \sum_{j=1}^{k_1} \Sigma_{1_{jj}} |<\hat{u}_{0_i}, u_{1_j}>|^2 - \sum_{j=1}^{k_1} \frac{\Sigma_{1_{jj}} |<\hat{u}_{1_j}, u_{1_j}>|^2 |<\hat{u}_{0_i}, u_{1_j}>|^2 + \sigma^4 |<\hat{u}_{1_j}, \hat{u}_{0_i}>|^2}{\Sigma_{1_{jj}} |<\hat{u}_{1_j}, u_{1_j}>|^2 + \sigma^2} \right) \tag{47}$$

and

$$D - CA^{-1}B \approx \mathbf{diag}\left(\sigma^2 + \Sigma_{1_{jj}} |<\hat{u}_{1_j}, u_{1_j}>|^2 - \sum_{i=1}^{k_0} \frac{\Sigma_{1_{jj}}^2 |<\hat{u}_{1_j}, u_{1_j}>|^2 |<\hat{u}_{0_i}, u_{1_j}>|^2 + \sigma^4 |<\hat{u}_{0_i}, \hat{u}_{1_j}>|^2}{\sigma^2 + \sum_{r=1}^{k_1} \Sigma_{1_{rr}} |<\hat{u}_{0_i}, u_{1_r}>|^2} \right) \tag{48}$$

Then

$$C_1^{-1} \approx \begin{bmatrix} (47)^{-1} & 0 \\ 0 & (48)^{-1} \end{bmatrix} \tag{49}$$

Approximating C_1 as a diagonal matrix, we arrive at

$$C_1^{-1} \approx \begin{bmatrix} \mathbf{diag}\left(\frac{1}{\sigma^2+\sum_{j=1}^{k_1}\Sigma_{1_{jj}}|<\hat{u}_{0_i},u_{1_j}>|^2}\right) & 0 \\ 0 & \mathbf{diag}\left(\frac{1}{\Sigma_{1_{jj}}|<\hat{u}_{1_i},u_{1_i}>|^2+\sigma^2}\right) \end{bmatrix} \tag{50}$$

We then form our detector by using expressions (46) and (50) in (29). Our detector is:

$$\boxed{\tilde{\Lambda}(w) = w^H((46)-(50))w} \tag{51}$$

Our decision becomes

$$\begin{aligned} &\text{Declare } H_0 \text{ if} \quad \tilde{\Lambda}(w) < \gamma \\ &\text{Declare } H_v \text{ if} \quad \tilde{\Lambda}(w) > \gamma \end{aligned}$$

where γ is our threshold.

Simulations

We now use simulations to verify the accuracy of (45), (46), (49) and (50).

We increase the dimension of our problem to ensure that our inverse covariance approximation error does tend to 0 as $n \to \infty$. We test our approximation for $n = 50, 100, 200, 500, 1000, 5000$. For each n, 20 trials are performed and the approximation error is averaged over these trials.

We set $m = n$ for the purpose of this simulation, where m is the number of training samples used to form our subspace estimates. Also for the purpose of our simulations, the vectors in our true subspaces are all orthogonal and for simplicity's sake are the elementary basis vectors. Vectors were generated via (23) where $\Sigma_0 = \mathbf{diag}(1 : k0), \Sigma_1 = \mathbf{diag}(k_0 + 1 : k_0 + k_1)$.

Our inverse covariance matrix approximations were then constructed via (45), (46), (49) and (50). The Frobenius norm was then calculated between these approximations and the true inverse covariance matrix. This was repeated for a few settings to compare the effect of changing k_0, k_1 and σ^2. The results follow.

Note that for all of the above figures, the red and green curves are essentially the same and the blue and black curves are essentially the same. This corresponds to both approximations for C_0^{-1} being essentially equal and both approximations for C_1^{-1} being approximately equal.

We also note that the "noisier" hypothesis has the lowest approximation error. This makes sense as we expect our approximations to work best when our hypotheses behave more randomly.

We also note that as $n \to \infty$ our approximation error $\to 0$ which leads us to conclude that our approximations are indeed valid approximations.

We also note that the imbalance of k_1, k_0 does not seem to have any visible effect on the the approximation error. Instead, the relative "noisy-ness" of each hypothesis is what has the largest effect of approximation error. In fact, as the noise increases, the approximation error decreases.

Finally, we conclude that we should use (46) and(50) rather than (45) (49) for our approximations as the first two expressions are much more simple to construct. $\Sigma_1, \Sigma_0, \sigma^2$ are either assumed to be known or may be estimated via MLE. $| < \hat{u}_{0_i}, u_{0_i} > |^2, | < \hat{u}_{1_j}, u_{1_j} > |^2$ may be calculated using random matrix theory as we did in ASILOMAR. That leaves the last task of estimating $| < \hat{u}_{0_i}, u_{1_j} > |^2, | < \hat{u}_{1_j}, u_{0_i} > |^2$

ROC Simulations

We simulate each of the five detectors above given in equations (8) (10) (20) (22) (51). During the simulation, the dimension of the testing and training points was $n = 100$. 100 training points were used to create our

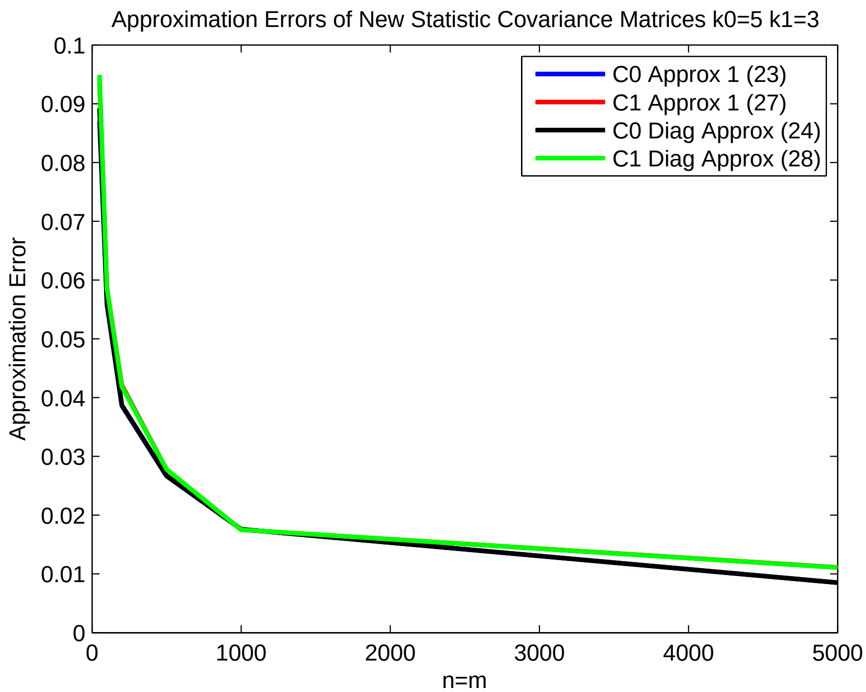

Figure 1: Here $k_0 = 5, k_1 = 3, \sigma = 3$. We see that the approximation of the inverse covariance matrix has roughly the same approximation error for both hypotheses (0 vs. 1)

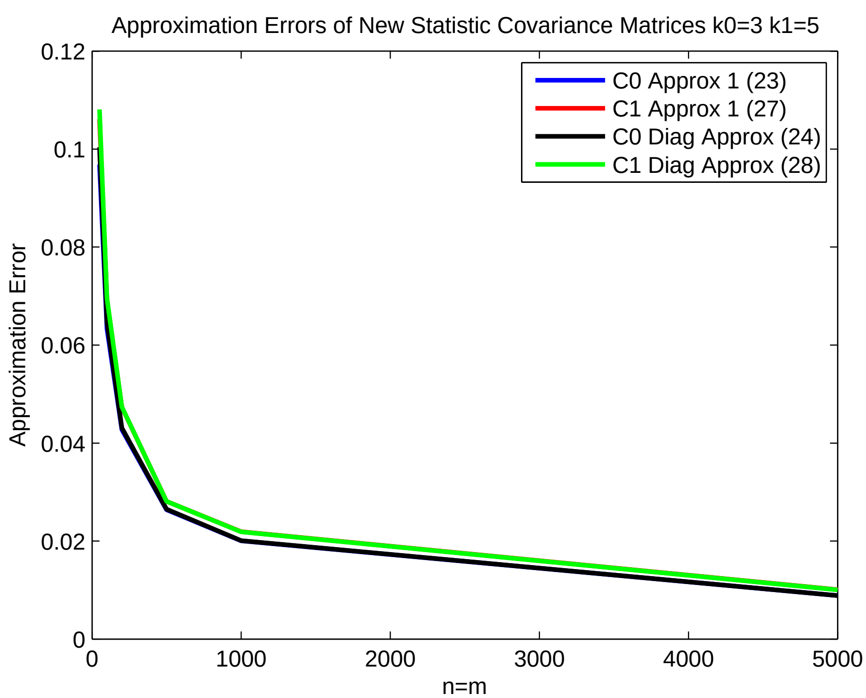

Figure 2: Here $k_0 = 3, k_1 = 5, \sigma = 3$. We see that the approximation of the inverse covariance matrix has a lower error for $C_0^{-}1$ as it is more noisy.

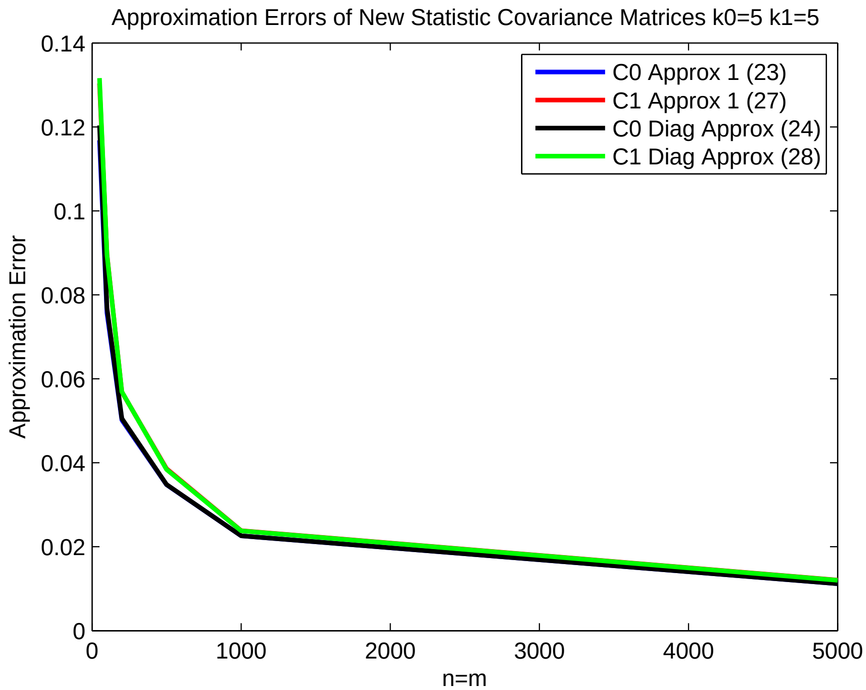

Figure 3: Here $k_0 = 5, k_1 = 5, \sigma = 3$. We see that the approximation of the inverse covariance matrix has a lower error for C_0^-1 as it is more noisy.

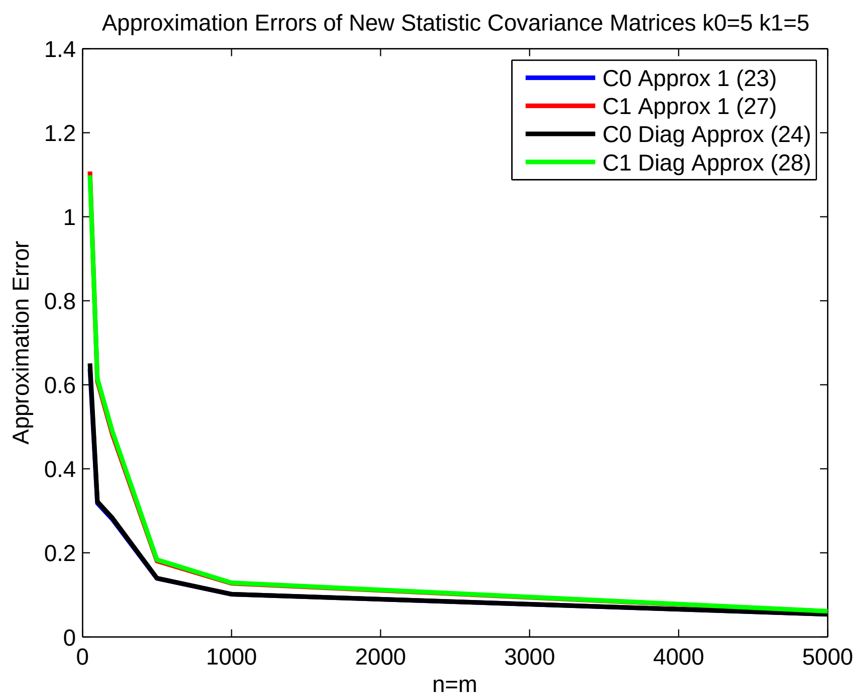

Figure 4: Here $k_0 = 3, k_1 = 5, \sigma = 1$. We see that the approximation of the inverse covariance matrix has a lower error for C_0^-1 because the noise is so low that it only affects C_0.

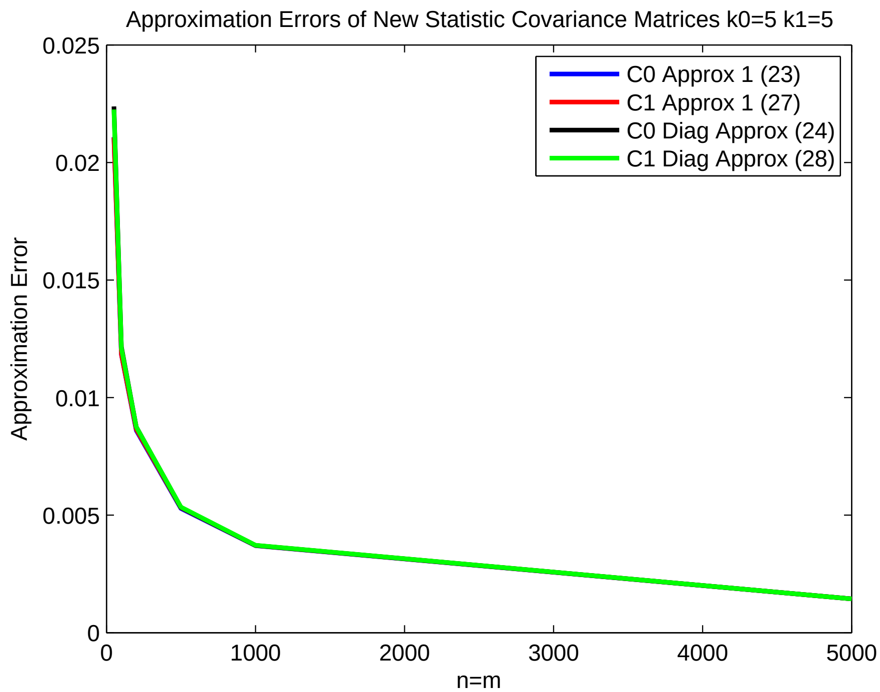

Figure 5: Here $k_0 = 3, k_1 = 5, \sigma = 8$. We see that the approximation of the inverse covariance matrix has roughly the same approximation error for both hypotheses (0 vs. 1) as almost all dimensions are affected by noise in both hypotheses.

subspace estimates $\hat{U}_0, \hat{U}_1$. 10000 testing points were used to generate AUC curves. The results are shown below for different Σ_0, Σ_1.

As of now, we see little improvement. Also of note is that the optimal unprocessed and processed ROC curves are exactly the same and the two plug-in detectors both produce the same ROC curves.

Angle Approximations

We have the following expressions for our angles used in (45), (46), (49) and (50). Defining $c = \frac{n}{m}$,

$$|<u_{0_i}, \hat{u}_{0_i}>|^2 \to \begin{cases} \dfrac{\Sigma_{0_{ii}}^2 - \sigma^4 c}{\Sigma_{0_{ii}}^2 + \sigma^2 \Sigma_{0_{ii}} c} < 1 & \text{if } \Sigma_{0_{ii}} \geq \sigma^2 \sqrt{c} \\ 0 & \text{o.w.} \end{cases} \tag{52}$$

$$|<u_{1_j}, \hat{u}_{1_j}>|^2 \to \begin{cases} \dfrac{\Sigma_{1_{jj}}^2 - \sigma^4 c}{\Sigma_{1_{jj}}^2 + \sigma^2 \Sigma_{1_{jj}} c} < 1 & \text{if } \Sigma_{1_{jj}} \geq \sigma^2 \sqrt{c} \\ 0 & \text{o.w.} \end{cases} \tag{53}$$

$$\begin{aligned} |<\hat{u}_{0_i}, u_{1_j}>|^2 = \quad & |<\hat{u}_{0_i}, \hat{u}_{1_j}>|^2 |<u_{1_j}, \hat{u}_{1_j}>|^2 + \left(1 - |<\hat{u}_{0_i}, \hat{u}_{1_j}>|^2\right)\left(1 - |<u_{1_j}, \hat{u}_{1_j}>|^2\right) \pm \\ & 2\sqrt{|<\hat{u}_{0_i}, \hat{u}_{1_j}>|^2 |<u_{1_j}, \hat{u}_{1_j}>|^2 \left(1 - |<\hat{u}_{0_i}, \hat{u}_{1_j}>|^2\right)\left(1 - |<u_{1_j}, \hat{u}_{1_j}>|^2\right)} \end{aligned} \tag{54}$$

$$\begin{aligned} |<u_{0_i}, \hat{u}_{1_j}>|^2 = \quad & |<\hat{u}_{0_i}, \hat{u}_{1_j}>|^2 |<u_{0_i}, \hat{u}_{0_i}>|^2 + \left(1 - |<\hat{u}_{0_i}, \hat{u}_{1_j}>|^2\right)\left(1 - |<u_{0_i}, \hat{u}_{0_i}>|^2\right) \pm \\ & 2\sqrt{|<\hat{u}_{0_i}, \hat{u}_{1_j}>|^2 |<u_{0_i}, \hat{u}_{0_i}>|^2 \left(1 - |<\hat{u}_{0_i}, \hat{u}_{1_j}>|^2\right)\left(1 - |<u_{0_i}, \hat{u}_{0_i}>|^2\right)} \end{aligned} \tag{55}$$

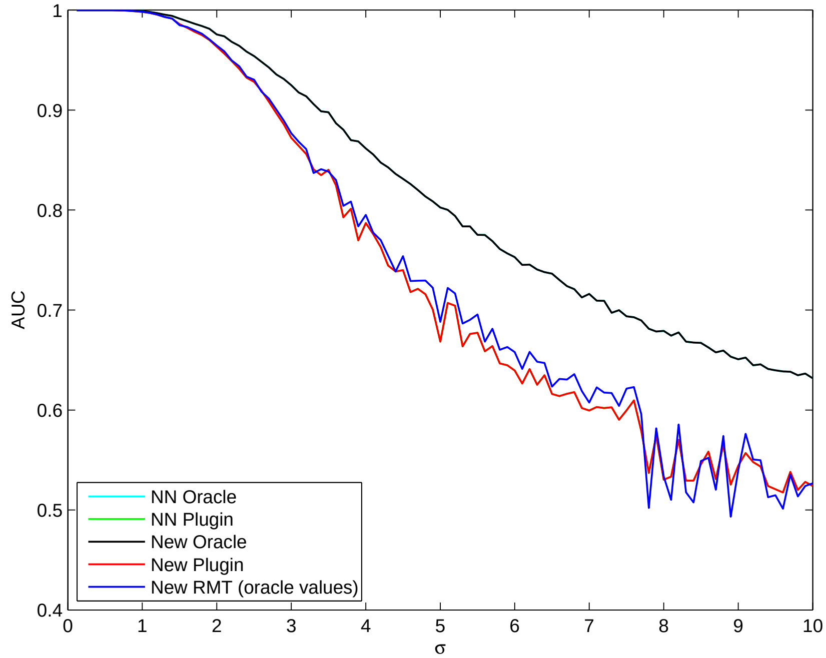

Figure 6: $\Sigma_0 = \mathbf{diag}(10, 5, 4, 2, 1), \Sigma_1 = \mathbf{diag}(5, 2, 1, .2, .1)$

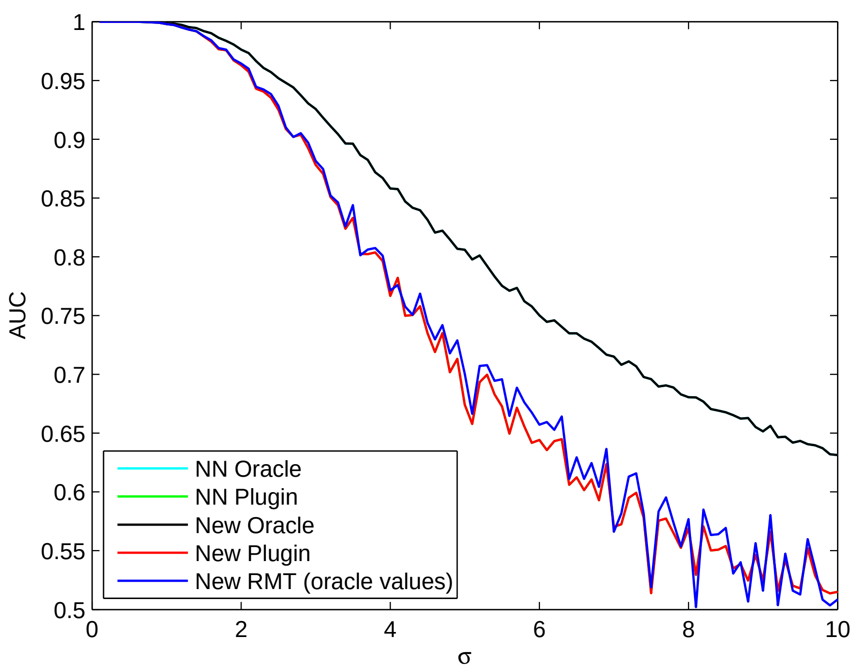

Figure 7: $\Sigma_0 = \mathbf{diag}(10, 5, 4, 2, 1), \Sigma_1 = \mathbf{diag}(5, 2, 1, .2, .1)$

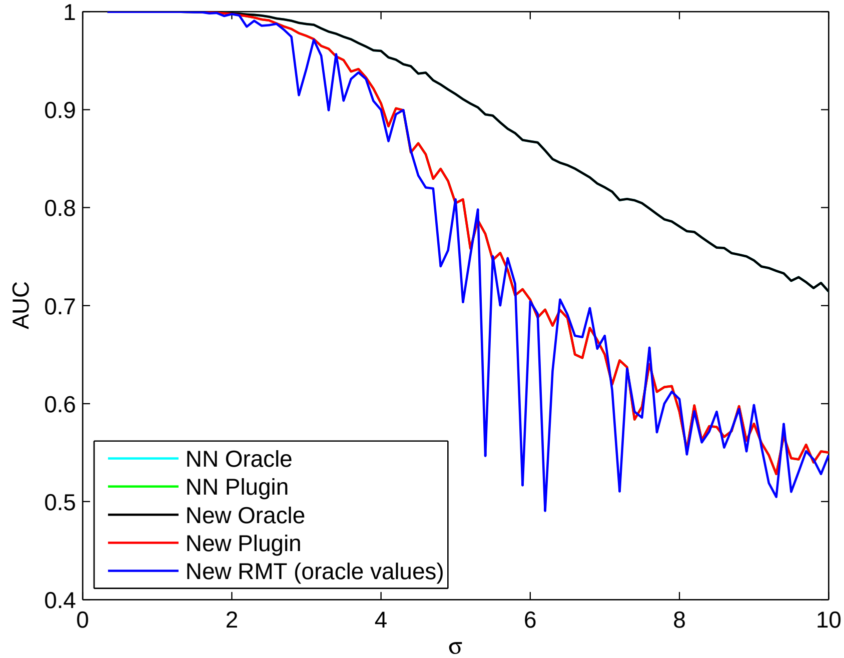

Figure 8: $\Sigma_0 = \mathbf{diag}(10, 9, 8, 7, 6), \Sigma_1 = \mathbf{diag}(5, 4, 3, 2, 1)$

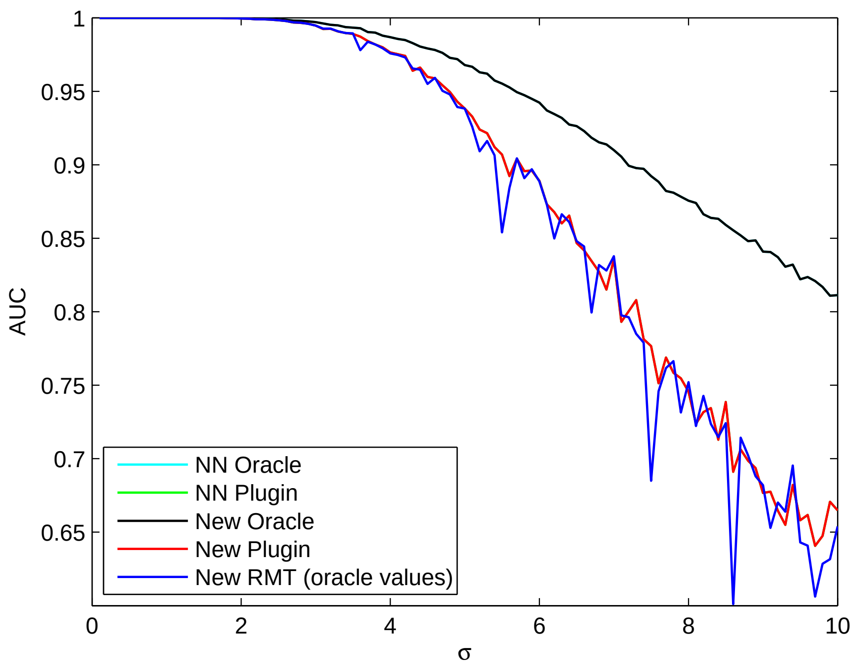

Figure 9: $\Sigma_0 = \mathbf{diag}(10, 5, 4, 2, 1), \Sigma_1 = \mathbf{diag}(10, 8, 5, 2, 1)$