Problem 1  

Hypothesis Testing with Two Observations

(i) The problem of guessing $H$ based on $Y_1$ is of the kind addressed in Section 20.14 (with $\sigma^2$ being 1, $J$ being 2, and the prior being uniform). Using the results of that section we conclude that

$$\langle Y_1, \mu \rangle_E$$

forms a one-dimensional sufficient statistic and that an optimal rule is to guess “$H = 0$” when $\langle Y_1, \mu \rangle_E$ is positive and to guess “$H = 1$” otherwise.

(ii) To find an optimal rule for guessing $H$ based on $Y_2$, we compute the likelihood ratio. We first consider the case $\alpha \geq 0$:

$$\frac{f_{Y_2|H=0}(y_2)}{f_{Y_2|H=1}(y_2)} = \frac{e^{-(y_2-\alpha)} I\{y_2 - \alpha \geq 0\}}{e^{-(y_2+\alpha)} I\{y_2 + \alpha \geq 0\}}$$

$$= \begin{cases} 0 & \text{if } -\infty < y_2 < -\alpha, \\ 0 & \text{if } -\alpha \leq y_2 < \alpha, \\ \exp(2\alpha) & \text{if } \alpha \leq y_2 < \infty, \end{cases} \quad (1)$$

where we have used the convention of (20.39). Similarly for $\alpha \leq 0$:

$$\frac{f_{Y_2|H=0}(y_2)}{f_{Y_2|H=1}(y_2)} = \frac{e^{-(y_2-\alpha)} I\{y_2 - \alpha \geq 0\}}{e^{-(y_2+\alpha)} I\{y_2 + \alpha \geq 0\}}$$

$$= \begin{cases} 1 & \text{if } -\infty < y_2 < -\alpha, \\ \infty & \text{if } \alpha \leq y_2 < -\alpha, \\ \exp(2\alpha) & \text{if } -\alpha \leq y_2 < \infty, \end{cases} \quad (2)$$
An optimal rule for the case $\alpha > 0$ is to guess “$H = 1$” if $-\alpha \leq y_2 < \alpha$, and to guess “$H = 0$” otherwise (because $e^{2\alpha}$ is greater than one when $\alpha$ is positive). An optimal rule for the case $\alpha \leq 0$ is to guess “$H = 0$” if $\alpha \leq y_2 < -\alpha$ and to guess “$H = 1$” otherwise.

(iii) A two dimensional sufficient statistic is the pair

$$\langle Y_1, \mu \rangle_E, Y_2$$

because conditional on $H$ the observations $Y_1$ and $Y_2$ are independent; see Proposition 22.4.6.

(iv) Define

$$T_1 = \langle Y_1, \mu \rangle_E$$

so

$$f_{T_1|H=0}(t_1) = \frac{1}{\sqrt{2\pi ||\mu||^2}} e^{-\frac{(t_1-\mu)^2}{2||\mu||^2}}$$

and

$$f_{T_1|H=1}(t_1) = \frac{1}{\sqrt{2\pi ||\mu||^2}} e^{-\frac{(t_1+||\mu||^2)^2}{2||\mu||^2}}.$$  

Consequently,

$$\frac{f_{T_1|H=0}(t_1)}{f_{T_1|H=1}(t_1)} = e^{2t_1}, \ t_1 \in \mathbb{R}.$$  

By the conditional independence

$$\frac{f_{T_1,Y_2|H=0}(t_1,y_2)}{f_{T_1,Y_2|H=1}(t_1,y_2)} = \frac{f_{T_1|H=0}(t_1) f_{Y_2|H=0}(y_2)}{f_{T_1|H=1}(t_1) f_{Y_2|H=1}(y_2)}$$

$$= e^{2t_1} \frac{f_{Y_2|H=0}(y_2)}{f_{Y_2|H=1}(y_2)},$$

which can be computed using (1) and (2). For the case $\alpha > 0$ an optimal decision rule is to guess “$H = 1$” if $-\alpha \leq y_2 < \alpha$ or $e^{2(\alpha + t_1)} < 1$, i.e., to guess “$H = 1$” if $-\alpha \leq y_2 < \alpha$ or $\alpha + t_1 < 0$. For the case $\alpha \leq 0$ an optimal rule is to guess “$H = 0$” if $\alpha \leq y_2 < -\alpha$ or $e^{2(\alpha + t_1)} > 1$, i.e., to guess “$H = 0$ if $\alpha \leq y_2 < -\alpha$ or $\alpha + t_1 > 0$.

**Problem 2**

**Optimality Does Not Imply Sufficiency**

(i) We shall establish the sufficiency by showing that for every nondegenerate prior ($\pi_0, \pi_1$) the a posteriori distribution of $H$ given $Y_1, Y_2$, and $\Theta$ is computable from ($Y_1 + Y_2, \Theta, \pi_0, \pi_1$) (Proposition 20.12.4).

$$\Pr[H = 0|Y_1 = y_1, Y_2 = y_2, \Theta = \theta]$$

$$= \frac{\pi_0 f_{Y_1,Y_2|H=0,\Theta=\theta}(y_1, y_2)}{\pi_0 f_{Y_1,Y_2|H=0,\Theta=\theta}(y_1, y_2) + \pi_1 f_{Y_1,Y_2|H=1,\Theta=\theta}(y_1, y_2)}$$

$$= \pi_0 \frac{1}{2\pi \theta^2} \exp \left( -\frac{(y_1 - 1)^2}{2\theta^2} - \frac{(y_2 - 1)^2}{2\theta^2} \right)$$

$$\pi_0 \frac{1}{2\pi \theta^2} \exp \left( -\frac{(y_1 - 1)^2}{2\theta^2} - \frac{(y_2 - 1)^2}{2\theta^2} \right) + \pi_1 \frac{1}{2\pi \theta^2} \exp \left( -\frac{(y_1 + 1)^2}{2\theta^2} - \frac{(y_2 + 1)^2}{2\theta^2} \right)$$

$$= \pi_0 \exp \left( \frac{2y_1 + y_2}{\theta^2} \right) + \pi_1 \exp \left( -\frac{2y_1 + y_2}{\theta^2} \right).$$  

(3)
Similarly,
\[
\Pr[H = 1|Y_1 = y_1, Y_2 = y_2, \Theta = \theta] = \frac{\pi_1 \exp\left(-y_1 + y_2 \frac{y_2}{\theta^2}\right)}{\pi_1 \exp\left(-y_1 + y_2 \frac{y_2}{\theta^2}\right) + \pi_0 \exp\left(y_1 + y_2 \frac{y_2}{\theta^2}\right)},
\]
thus demonstrating that the a posteriori distribution of $H$ is computable from $(y_1 + y_2, \theta)$ and the prior $(\pi_0, \pi_1)$.

(ii) Let us write
\[
\begin{align*}
\frac{f_{Y|H=0,\Theta=\sigma_i}(y_1, y_2)}{f_{Y|H=1,\Theta=\sigma_i}(y_1, y_2)} &= \frac{h^{(i)}(||y||^2) g_0^{(i)}(t)}{h^{(i)}(||y||^2) g_1^{(i)}(t)}, \quad i = 0, 1 \\
\end{align*}
\]
where $t = y_1 + y_2$:
\[
\begin{align*}
h^{(i)}(\xi) &= \frac{1}{2\pi\sigma_i^2} \exp\left(-\frac{\xi + 2}{2\sigma_i^2}\right); \\
g_0^{(i)}(t) &= \exp\left(-\frac{t}{\sigma_i^2}\right); \\
g_1^{(i)}(t) &= \exp\left(-\frac{t}{\sigma_i^2}\right).
\end{align*}
\]
When $\Theta$ is not observed,
\[
\frac{f_{Y|H=0}(y_1, y_2)}{f_{Y|H=1}(y_1, y_2)} = \frac{\rho_0 h^{(0)}(||y||^2) g_0^{(0)}(t)}{\rho_0 h^{(0)}(||y||^2) g_0^{(0)}(t) + \rho_1 h^{(1)}(||y||^2) g_0^{(1)}(t) + \rho_1 h^{(1)}(||y||^2) g_1^{(1)}(t)}
\]
and this — when $\sigma_0^2 \neq \sigma_1^2$ — depends both on $t(= y_1 + y_2)$ and $||y||^2(= y_1^2 + y_2^2)$.

(iii) For a uniform prior we have by (3) that when $\Theta$ is observed
\[
\Pr[H = 0|Y_1 = y_1, Y_2 = y_2, \Theta = \theta] > \frac{1}{2} \Leftrightarrow \exp\left(y_1 + y_2 \frac{y_2}{\theta^2}\right) > \exp\left(-y_1 + y_2 \frac{y_2}{\theta^2}\right) \Leftrightarrow \frac{y_1 + y_2}{\theta^2} > \frac{-y_1 + y_2}{\theta^2} \Leftrightarrow (y_1 + y_2 > 0).
\]
Consequently, for a uniform prior it is optimal to guess “$H = 0$” whenever $y_1 + y_2 > 0$. Since this rule is implementable also when $\Theta$ is not observed, and since the optimal rule when $\Theta$ is not observed cannot outperform the optimal rule when it is, this rule is also optimal when $\Theta$ is not observed.

Problem 3

Covariance Matrices

Denote $A = \begin{pmatrix} 5 & 0 \\ 0 & -1 \end{pmatrix}$, $B = \begin{pmatrix} 5 & 1 \\ 2 & 2 \end{pmatrix}$, $C = \begin{pmatrix} 2 & 10 \\ 10 & 1 \end{pmatrix}$, and $D = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$.

A cannot be a covariance matrix because the entry “$-1$” cannot be the variance of a random variable.
B cannot be a covariance matrix because it is not symmetric.
C cannot be a covariance matrix because for \( \alpha = (1, -1)^T \) we have that \( \alpha^T C \alpha < 0 \) and thus C is not positive semidefinite.  
D is the only possible covariance matrix, since it is symmetric and its two eigenvalues 0 and 2 are both nonnegative.

**Problem 4**  
*Multivariate Gaussians*

If \( Z \) is a Gaussian random variable, then it is also a Gaussian 1-vector; see Section 23.6.1 and in particular (i). And since the vector \((Z, Z)^T\) can be written as a linear transformation of the random 1-vector \( Z \)  
\[
(Z, Z)^T = \begin{pmatrix} 1 \\ 1 \end{pmatrix} (Z),
\]
it follows that \((Z, Z)^T\) is the result of applying a deterministic linear transformation to a Gaussian vector, and is thus a Gaussian vector (Proposition 23.6.3).  

We next turn to the canonical representation of this vector. Denote the vector by \( Z \), so \( Z = (Z, Z)^T \). Its covariance matrix is  
\[
K_{ZZ} = \sigma^2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},
\]
whose eigenvalues are 0 and \( 2\sigma^2 \) with corresponding eigenvectors \((1, -1)^T\) and \((1, 1)^T\). Normalizing the eigenvectors to unit norm we obtain that \( K_{ZZ} U = \Sigma \), where  
\[
U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 1 \\ -1 \\ 1 \end{pmatrix}, \quad \Sigma = \sigma^2 \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}.
\]

The canonical representation is thus  
\[
\begin{pmatrix} Z' \\ Z \end{pmatrix} - \begin{pmatrix} E[Z] \\ E[Z] \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{2} \sigma \end{pmatrix} \begin{pmatrix} W^{(1)} \\ W^{(2)} \end{pmatrix}.
\]

**Problem 5**  
*Manipulating Gaussians*

In matrix notation, the vector \((Y, Z)^T\) is \( AW \), where \( W \) is a standard Gaussian 5-vector and the matrix \( A \) is  
\[
A = \begin{pmatrix} 3 & 4 & -2 & 1 & -1 \\ 1 & -4 & -2 & 3 & -1 \end{pmatrix}.
\]
Consequently, this vector is a centered Gaussian of covariance matrix \( AA^T \), i.e.  
\[
\begin{pmatrix} 31 & -5 \\ -5 & 31 \end{pmatrix}.
\]
Problem 6

Independence, Uncorrelatedness and Gaussianity

(i) By symmetry it is not difficult to see that the distribution of $Y$ is identical to that of $X$, so $Y \sim \mathcal{N}(0,1)$. Formally this can be proved by considering the cumulative distribution function:

$$\Pr[Y \leq \xi] = \Pr[H = 1] \Pr[Y \leq \xi | H = 1] + \Pr[H = -1] \Pr[Y \leq \xi | H = -1]$$

$$= \Pr[H = 1] \Pr[X \leq \xi | H = 1] + \Pr[H = -1] \Pr[X \geq -\xi | H = -1]$$

$$= \frac{1}{2} \Pr[X \leq \xi] + \frac{1}{2} \Pr[X \geq -\xi]$$

$$= \frac{1}{2} F_X(\xi) + \frac{1}{2} - \frac{1}{2} F_X(-\xi), \quad \xi \in \mathbb{R},$$

where the second equality follows from the relation $Y = HX$ and the third because $X$ and $H$ are independent. Taking the derivative of the cumulative distribution function yields that the density of $Y$ is equal to the density of $X$.

(ii) No, $X$ and $Y$ are not correlated. Indeed, since they are both of zero mean, their correlation is $\mathbb{E}[XY]$, which is zero because

$$\mathbb{E}[XY] = \mathbb{E}[HX^2]$$

$$= \mathbb{E}[H] \mathbb{E}[X^2]$$

$$= 0,$$

where the second equality follows because $H$ and $X$ are independent, and the third because $H$ is zero mean.

(iii) Since $X$ and $Y$ are of equal law and both are $\mathcal{N}(0,1)$,

$$\Pr[|X| \geq 1] = \Pr[|Y| \geq 1] = 2 \Phi(1).$$

(iv) Since $|X| = |Y|$ (deterministically),

$$\Pr[|X| \geq 1, \ |Y| \geq 1] = \Pr[|X| \geq 1]$$

$$= 2 \Phi(1).$$

(v) The random variables $X$ and $Y$ are not independent because, as seen above,

$$\Pr[|X| \geq 1, \ |Y| \geq 1] \neq \Pr[|X| \geq 1] \Pr[|Y| \geq 1].$$

(Had $X$ and $Y$ been independent, then so would have been $|X|$ and $|Y|$, and the above inequality shows that this is not the case.)

(vi) The vector $(X, Y)^T$ is not a Gaussian vector because its components are uncorrelated but not independent. This vector is thus an example of a vector whose components are Gaussian and that is yet not a Gaussian vector.