1.1 Given, $X_1 \rightarrow N(\mu_1, \sigma_1^2)$ and $X_2 \rightarrow N(\mu_2, \sigma_2^2)$

Here Let $f(x)$ and $g(x)$ be Gaussian PDFs with arbitrary means μ_1 and μ_2 and variance σ_1 and σ_2

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \quad \text{and} \quad g(x) = \frac{1}{\sqrt{2\pi}\,\sigma_2} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}$$

Their multiplication is:

$$f(x) \cdot g(x) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\left(\frac{(x-\mu_1)^2}{2\sigma_1^2} + \frac{(x-\mu_2)^2}{2\sigma_2^2}\right)} \quad \text{——— } (1)$$

Examine the term in the exponent:

$$\beta = \frac{(x-\mu_1)^2}{2\sigma_1^2} + \frac{(x-\mu_2)^2}{2\sigma_2^2}$$

Expanding the two quadratics and collecting terms in powers of x gives,

$$\beta = \frac{(\sigma_1^2+\sigma_2^2)x^2 - 2(\mu_1\sigma_2^2 + \mu_2\sigma_1^2)x + \mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2}{2\sigma_1^2\sigma_2^2}$$

Dividing through by the coefficient of x^2 gives,

$$\beta = \frac{x^2 - 2\frac{\mu_1\sigma_2^2 + \mu_2\sigma_1^2}{\sigma_1^2+\sigma_2^2}x + \frac{\mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2}{\sigma_1^2+\sigma_2^2}}{2\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}} \quad \text{——— } (2)$$

This is again a quadratic in x and so equation (1) is a Gaussian function. Compare the terms in equation (2) to a the usual Gaussian form

$$P(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\left(\frac{x^2 - 2\mu x + \mu^2}{2\sigma^2}\right)}$$

The product of two Gaussian PDFs is proportional to a Gaussian PDF with a mean that is half the coefficient of x in equation (2) and a variance that is the square root of half of the denominator:

$$\sigma_{12} = \sqrt{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}} \quad \text{and} \quad \mu_{12} = \frac{\mu_1\sigma_2^2 + \mu_2\sigma_1^2}{\sigma_1^2+\sigma_2^2}$$

The convolution of two Gaussian PDFs

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \quad \text{and} \quad g(x) = \frac{1}{\sqrt{2\pi}\,\sigma_2} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}$$

The convolution of two functions $f(t)$ and $g(t)$ over a finite range is defined as:

$$\int_0^x f(x-\tau)\, g(\tau)\, d\tau = f \otimes g$$

we can use the convolution theorem,

$$F^{-1}[F(f(x))\, F(g(x))] = f(x) \otimes g(x)$$

where F is the fourier Transform,

$$F(f(x)) = \int_{-\alpha}^{\alpha} f(x)\, e^{-2\pi j k x}\, dx$$

and F^{-1} is the inverse Fourier transform

$$F^{-1}(F(k)) = \int_{-\alpha}^{\alpha} F(k)\, e^{2\pi j k x}\, dk$$

using the transformation, $x' = x - \mu_1$

the Fourier transform of $f(x)$ is given by,

$$F(f(x)) = \frac{1}{\sqrt{2\pi}\,\sigma_1} \int_{-\alpha}^{\alpha} e^{-\frac{x'^2}{2\sigma_1^2}}\, e^{-2\pi j k (x'+\mu_1)}\, dx' \qquad (3)$$

$$= \frac{e^{-2\pi j k \mu_1}}{\sqrt{2\pi}\,\sigma_1} \int_{-\alpha}^{\alpha} e^{-\frac{x'^2}{2\sigma_1^2}}\, e^{-2\pi j k x'}\, dx'$$

using Euler's formula, $e^{-j\theta} = \cos\theta - j\sin\theta$

we can split the term in $e^{x'}$ to give

$$F(f(x)) = \frac{e^{-2\pi j k \mu_1}}{\sqrt{2\pi}\,\sigma_1} \int_{-\infty}^{\infty} e^{-\frac{x'^2}{2\sigma_1^2}} \left[\cos(2\pi k x') - j\sin(2\pi k x')\right] dx'$$

The term in $\sin(x')$ is odd and so its integral over all space will be zero, leaving:

$$F(f(x)) = \frac{e^{-2\pi j k \mu_1}}{\sqrt{2\pi}\,\sigma_1} \int_{-\infty}^{\infty} e^{-\frac{x'^2}{2\sigma_1^2}} \cos(2\pi k x')\, dx'$$

The integral is given in standard form

$$\int_0^{\infty} e^{-at^2} \cos(2xt)\, dt = \frac{1}{2}\sqrt{\frac{\pi}{a}}\, e^{-x^2/a}.$$

and so,

$$F(f(x)) = e^{-2\pi j k \mu_1}\, e^{-2\pi^2 \sigma_1^2 k^2} \quad \text{—— (4)}$$

The second term in this expression is a Gaussian PDF in k; the fourier transform of a Gaussian PDF is another Gaussian PDF. The first term is a phase term accounting for the mean of $f(x)$ i.e. its offset from zero. The fourier transform of $g(x)$ will give a similar expression, and so,

$$F(f(x))F(g(x)) = e^{-2\pi j k \mu_1}\, e^{-2\pi^2 \sigma_1^2 k^2}\, e^{-2\pi j k \mu_2}\, e^{-2\pi^2 \sigma_2^2 k^2}$$

$$= e^{-2\pi j k(\mu_1+\mu_2)}\, e^{-2\pi^2(\sigma_1^2+\sigma_2^2)k^2} \quad \text{—— (5)}$$

Comparing equation (5) to equation (4), we can see that it is the fourier transform of a Gaussian PDF with mean and variance,

$$\mu_{1\otimes 2} = \mu_1 + \mu_2 \quad \text{and} \quad \sigma_{1\otimes 2} = \sqrt{\sigma_1^2 + \sigma_2^2}$$

since, the fourier transform is invertible,

$$P_{f\otimes g}(x) = F^{-1}\left[F(f(x))\,F(g(x))\right] = \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}\, e^{-\frac{(x-(\mu_1+\mu_2))^2}{2(\sigma_1^2+\sigma_2^2)}}$$

So, the convolution of two Gaussian PDFs is also a Gaussian PDF.

1.2 given $\vec{y} = A\vec{x} + \vec{e}$

to prove minimizing trace of error covariance $Tr\left(E\left(\vec{e}\vec{e}^T\right)\right)$ is maximizing $P(y|x)$.

Given the equation overall goal is to estimate $\vec{x}$. Let $\hat{x}$ be the predicted bearing signal.

the error, $f(e_k) = f(x_k - \hat{x}_k)$.

for MSE, $f(e_k) = (x_k - \hat{x}_k)^2$

$= E\left(\vec{e}_k \vec{e}_k^T\right)$.

Let, P_k ~~which~~ be the error covariance matrix at time k

which is equivalent $P_k = E\left(\vec{e}_k \vec{e}_k^T\right)$.

$$= E\left((x_k - \hat{x}_k)(x_k - \hat{x}_k)^T\right) \quad \text{— (1)}$$

let, the prior estimate of $\hat{x}_k$ be $\hat{x}'_k$. using kalman gain K_k we can write, $\hat{x}_k = \hat{x}'_k + K_k(z_k - H\hat{x}'_k)$.

where $(z_k - H\hat{x}'_k)$ is the measurment residual

$$\therefore v_k = z_k - H\hat{x}_k$$

$$z_k = H\hat{x}_k + v_k.$$

$$\therefore \hat{x}_k = \hat{x}'_k + K_k\left(H\hat{x}_k + v_k - H\hat{x}'_k\right).$$

Substituting in (1),

$$P_k = E\left(\left(x_k - \left(\hat{x}'_k + K_k\left(H\hat{x}_k + v_k - H\hat{x}'_k\right)\right)\right)\left(x_k - \left(\hat{x}'_k + K_k\left(H\hat{x}_k + v_k - H\hat{x}'_k\right)\right)\right)^T\right)$$

$$= E\left[\begin{array}{l}\left[(I - K_k H)(x_k - \hat{x}'_k) - K_k v_k\right] \\ \left[(I - K_k H)(x_k - \hat{x}'_k) - K_k v_k\right]^T\end{array}\right]$$

here we can see, $(x_k - \hat{x}'_k)$ is the error of the prior estimate.

$$\therefore P_k = (I - K_k H)\,\overbrace{E\left((x_k - \hat{x}'_k)(x_k - \hat{x}'_k)^T\right)}^{P'_k \,\text{— prior estimate of } P_k}\,(I - K_k H)$$

$$+ K_k E(v_k v_k^T) K_k^T$$

$$= (I - K_k H) P'_k (I - K_k H) + K_k E(v_k v_k^T) K_k^T$$

And P_{kk} is the error covariance matrix,

$$P_{kk} = \begin{bmatrix} E[e_{k-1} e_{k-1}^T] & E[e_k e_{k-1}^T] & E[e_{k+1} e_{k-1}^T] \\ E[e_{k-1} e_k^T] & E[e_k e_k^T] & E[e_{k+1} e_k^T] \\ E[e_{k-1} e_{k+1}^T] & E[e_k e_{k+1}^T] & E[e_{k+1} e_{k+1}^T] \end{bmatrix}$$

the sum of the diagonal elements of the matrix is the Trace of a matrix. Here we can see, the Trace of error covariance matrix is the sum of mean squared error. which can be minimised by minimising the trace; which essentially maximising $p(y|x)$, ~~ea~~

[proved].

Reference: http://web.mit.edu/kirtley/kirtley/binlustuff/literature/control/Kalman%20filter.pdf