MCB131: Problem Set 1

Michael Cubeddu

(1) a.) $V_1 = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}$ $V_2 = \begin{bmatrix} -1 \\ 1 \\ 1 \\ -1 \end{bmatrix}$ $V_3 = \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix}$ $V_4 = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}$

$$V \sim Unif(\{V_1, V_2, V_3, V_4\})$$

$$P(V_1) = P(V_2) = P(V_3) = P(V_4) = \frac{1}{4}$$

Covariance matrix C is given by:

$$C_{ij} = Cov(V_i, V_j) = E[(V_i - E[V_i])(V_j - E[V_j])]$$
$$= E[V_i V_j] \cdot \frac{1}{16} V_i^T V_j$$

Thus,

$$C = \frac{1}{4} \begin{array}{c} \begin{array}{cccc} V_1 & V_2 & V_3 & V_4 \end{array} \\ \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ V_1 \\ V_2 \\ V_3 \\ V_4 \end{array}$$

We can see that C is symmetric and its vector columns are linearly independent.

b.) Using the formula $\det(C - \lambda I_4) = 0$, we get the characteristic polynomial:

$$\lambda^2 (\lambda^2 - \lambda + \frac{1}{4}) = \lambda^2 (\lambda - \frac{1}{2})^2$$

Thus, there are two-order eigenvalues:

$$\underline{\lambda_1 = 0, \quad \lambda_2 = \frac{1}{2}}$$

Using $\lambda_1 = 0$ to find eigenvectors:

$$\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

Using $\lambda_2 = \frac{1}{2}$ to get eigenvectors:

$$\begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}$$

c.) $\tilde{V} = V + Z$, $Z \sim N(0, \sigma^2)$

$$\tilde{C}_{ij} = Cov(\tilde{V}_i, \tilde{V}_j) = E[(V_i + Z_i - E[\tilde{V}_i])(V_j + Z_j - E[\tilde{V}_j])]$$

– Note that $E[\tilde{V}] = E[V] + E[Z] = 0$.

$$Cov(\tilde{V}_i, \tilde{V}_j) = E[(V_i + Z_i)(V_j + Z_j)]$$
$$= E[V_i V_j] + \underbrace{E[V_i Z_j]}_{0} + \underbrace{E[V_j Z_i]}_{0} + \underbrace{E[Z_i Z_j]}_{\delta_{ij}\sigma^2}$$

because $\forall i, j, V_i \perp\!\!\!\perp Z_j$. Also, Z_i are i.i.d.

and $E[Z_i Z_j] = \delta_{ij}\sigma^2$

Thus, $\boxed{\tilde{C}_{ij} = \frac{1}{16} V_i^T V_j + \delta_{ij}\sigma^2} \Rightarrow \tilde{C} = C + \sigma^2 I_4$

d.) We can write the probability distribution via the PDF $P(\tilde{V})$.
For a given $\tilde{V}$, there are only 4 ways it could have been constructed:
that is, only 4 possible values of vector V with their
respective additions of $\tilde{V} - V_i = Z$. Thus,

$$P(\tilde{V}) = \frac{1}{4} P(Z = \tilde{V} - V_1) + \frac{1}{4} P(Z = \tilde{V} - V_2) + \frac{1}{4}(P(Z = \tilde{V} - V_3))$$
$$+ \frac{1}{4} P(Z = \tilde{V} - V_4)$$

Where Z is distributed by a Multivariate Gaussian and the value of $P(Z = \tilde{V} - V_i)$ is given by:

$$f(\tilde{V} - V_i) = \frac{1}{(\sqrt{2\pi})^n |\tilde{C}|^{1/2}} e^{-\frac{1}{2}(\tilde{V} - V_i)^T \tilde{C}^{-1} (\tilde{V} - V_i)}$$

$f(\tilde{V} - V_i)$ is the error rate prob. for Z, the noise. For V_i's distant to $\tilde{V}$, $f(\tilde{V} - V_i)$ will be low.

② Want to show: $H = \frac{N}{2}\log(2\pi e) + \frac{1}{2}\log(\det(C))$
$= \frac{N}{2}\ln(2\pi) + \frac{1}{2}\ln(|C|)$

The PDF of the Multivariate Gaussian is given by:

$$f(x) = \frac{1}{(\sqrt{2\pi})^n |C|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)}$$

The entropy H of $f(x)$ is thus given by:

$$H = -\int_{-\infty}^{\infty} f(x)\ln(f(x))\,\delta x$$

Expanding $\ln(f(x))$, we get:

$$H = -\int_{-\infty}^{\infty} f(x)\left[\frac{1}{2}\ln((2\pi)^n|C|) + \frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)\right]\partial x$$

$$= \frac{1}{2}\ln((2\pi)^n|C|) + \frac{1}{2}\ln(e)\cdot\underbrace{E[(x-\mu)^T C^{-1}(x-\mu)]}_{\approx n}$$

(because if we normalize x this is about n)

$$\approx \underline{\frac{n}{2}\ln(2\pi) + \frac{1}{2}\ln(|C|)}$$

③ Entropy Maximization

a.) Our constraints are:

$$\sum_{n=0}^{\infty} p(n) = 1 \quad \text{and} \quad \sum_{n=0}^{\infty} np(n) = \mu$$

Entropy is given by:

$$H(x) = -\sum_{n=0}^{\infty} p(n)\ln(p(n))$$

And thus the Lagrangian for our constraints is:

$$\mathcal{L}(x) = -\sum_{n=0}^{\infty} p(n)\ln(p(n)) + \lambda_1\left(\sum_{n=0}^{\infty} p(n) - 1\right) + \lambda_2\left(\sum_{n=0}^{\infty} np(n) - \mu\right)$$

$$\frac{\partial \mathcal{L}}{\partial p(n)} = -\ln(p(n)) - 1 + \lambda_1 + \lambda_2 n = 0$$

$$p(n) = e^{\lambda_1 + n\lambda_2 - 1} \Rightarrow Ce^{n\lambda_2} \quad \text{where } C \text{ is constant.}$$

To ensure $p(n)$ sums to 1:

$$\sum_{n=0}^{\infty} Ce^{\lambda_2 n} = 1, \quad \text{thus} \quad \sum_{n=0}^{\infty} (e^{\lambda_2})^n = \frac{1}{C}$$

That is, the series only converges if $|e^{\lambda_2}| < 1$. It converges to:

$$\sum_{n=0}^{\infty} (e^{\lambda_2})^n = \frac{1}{1-e^{\lambda_2}}, \quad \text{thus} \quad \boxed{C = 1 - e^{\lambda_2}}$$

Plugging in for the mean constraint we get:

$$\sum_{n=0}^{\infty} nCe^{\lambda_2 n} = \mu, \quad \text{giving us} \quad \boxed{\mu = \frac{Ce^{\lambda_2}}{(1-e^{\lambda_2})^2}}$$

Solving this system of 2 equations for 2 unknowns λ_2 and C we get:

$$\boxed{\lambda_2 = \ln\left(\frac{\mu}{1+\mu}\right)} \quad \text{and} \quad \boxed{C = \frac{1}{1+\mu}}$$

Thus, the entropy-maximizing distribution is $\boxed{p(n) = \frac{\mu n}{(1+\mu)^2}}$ *

b.) I will solve for the general case for a not-necessarily diagonal covariance matrix C with $\mu = 0$ which will satisfy the case for a diagonal C with $C_{ij} = \sigma_i^2 \delta_{ij}$.

For the diagonal case, the distribution can be written as:

$$p(\bar{x}) = \frac{1}{\sqrt{(2\pi)^k |C|}} e^{-\frac{1}{2} \sum_{ij} x_i C_{ij}^{-1} x_j}$$

$$= \prod_i \left[\frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}} \right] \quad \text{(multiplication of k Gaussians)}$$

For the gen'l case, the components of x are correlated. Because C is symmetric positive definite, it can be diagonalized using an orthogonal matrix:

$$C = U \Lambda U^T$$

such that $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_k)$ and λ_i are the eigenvalues of C and U is an orthogonal matrix such that its columns are the eigen vectors of C.

Applying the rotation $\bar{x}' = U^T(\bar{x} - \bar{\mu})$, x' is a Gaussian distribution whose components are now independent. Thus, the distribution of x' is given by:

$$p(x') = \prod_j \frac{1}{\sqrt{2\pi\lambda_j}} e^{\left(-\frac{1}{2}(x_j')^2 \cdot \lambda_j^{-1}\right)}$$

$(x_j' - \mu_j')^2$

where $\mu_j' = 0 \ \forall j$.

Now, we convert back to x from x':

$$U\bar{x}' = \bar{x} - \bar{\mu} \implies x_j' = U_j^T x_j + \mu_j$$

$$\boxed{p(x) = \prod_j \frac{1}{\sqrt{2\pi\lambda_j}} e^{\left(-\frac{1}{2}(U_j^T x_j + \mu_j)^2 \lambda_j^{-1}\right)}}$$

4) Mutual Information/Data Processing Inequality

a.) Let's assume the first definition $I = H(X)+H(Y)-H(X,Y)$ to be true.

- Proof 1: $\cancel{H(X)}+H(Y)-H(X,Y) = \cancel{H(X)} - H(X|Y)$

$$H(X,Y) = H(Y) + H(X|Y)$$

$$-\sum_{x,y} p(x,y)\ln p(x,y) = -\sum_{x,y} p(x,y)\ln p(y) - \sum_{x,y} p(x,y)\ln p(x|y)$$

We know that $p(x,y)=p(y)+p(x|y)$ always. Substituting $p(y)+p(x|y)$ for $p(x,y)$ in the expression above shows that both expressions are equivalent.

- Proof 2: $H(X) + \cancel{H(Y)} - H(X,Y) = \cancel{H(Y)} - H(Y|X)$

$$H(X,Y) = H(X) + H(Y|X)$$

$$-\sum_{x,y} p(x,y)\ln p(x,y) = -\sum_{x,y} p(x,y)\ln p(x) - \sum_{x,y} p(x,y)\ln p(y|x)$$

We know that $p(x,y)=p(x)+p(y|x)$ always. Substituting in for the joint $p(x,y)$ above we show the expressions are equivalent.

Finally, because $p(x,y)=p(x)+p(y|x)=p(y)+p(x|y)$, we can show

$$\begin{aligned} I &= H(X)+H(Y)-H(X,Y) \\ &= H(X)-H(X|Y) \\ &= H(Y)-H(Y|X) \end{aligned}$$

QED.

b.) (Bonus)

Want to show: $I(X;Z) \le I(X;Y)$

$$I(X;Z) = H(X) - H(X|Z)$$

We know that $H(X|Z) \le H(X|Y,Z)$ because conditioning over another variable can not possibly increase entropy.

Thus,

$$I(X;Z) \le H(X) - H(X|Y,Z)$$

Because $p(x|y,z) = p(x|y)$, we get:

$$I(X;Z) \le H(X) - H(X|Y) = I(X;Y).$$ QED

(5) a.) First, we note that

$$f(x) = \begin{cases} e^{-\lambda x} & ; x \ge 0 \\ e^{\lambda x} & ; x < 0 \end{cases}$$

Thus,

$$\hat{f}(\omega) = F(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\lambda|x|} e^{-i\omega x} \partial x$$

$$= \frac{1}{\sqrt{2\pi}} \left(\int_{-\infty}^{0} e^{x(\lambda - i\omega)} \partial x + \int_{0}^{\infty} e^{-\lambda x - i\omega x} \partial x \right)$$

$$= \frac{1}{\sqrt{2\pi}} \left(\frac{1}{\lambda - i\omega} + \frac{1}{\lambda + i\omega} \right)$$

$$\boxed{F(\omega) = \frac{2\lambda}{\sqrt{2\pi}(\lambda^2 + \omega^2)}}$$

b.) The max of f is at $x = 0$ where $f(0) = 1$.
The max of F is at $\omega = 0$ where $F(0) = \frac{2}{\sqrt{2\pi}\,\lambda}$

$$\text{Width}(f) = x \text{ s.t. } f(x) = .5 \max\{f\} = .5(1) = 1/2$$
$$= 0$$

$$\text{Width}(F) = \omega \text{ s.t. } F(\omega) = .5 \max\{F\} = \frac{1}{\lambda\sqrt{2\pi}}$$
$$= 0$$

c.) $F(g(x)) = G(\omega)$ implies $F^{-1}(G(\omega)) = g(x)$

$$g(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} G(\omega) e^{i\omega x}\,\partial\omega$$

$$\frac{\partial}{\partial x} g(x) = \frac{1}{2\pi}\frac{\partial}{\partial x}\left(\int_{-\infty}^{\infty} G(\omega)\, e^{i\omega x}\,\partial\omega\right) = \frac{1}{2\pi}\int_{-\infty}^{\infty} G(\omega)\frac{\partial}{\partial x}\left(e^{i\omega x}\right)\partial\omega$$

$$= \frac{i\omega}{2\pi}\int_{-\infty}^{\infty} G(\omega) e^{i\omega x}\,\partial\omega = \underbrace{F^{-1}\left(i\omega G(\omega)\right)}_{\Downarrow}$$

$$\boxed{F\left(\frac{\partial}{\partial x} g(x)\right) = i\omega G(\omega)}$$

d.)

$$f(x) = \begin{cases} e^{-\lambda x} & ; x \geq 0 \\ e^{\lambda x} & ; x < 0 \end{cases}$$

$$\frac{\partial f}{\partial x} = \begin{cases} -\lambda e^{-\lambda x} & ; x > 0 \\ \text{undefined} & ; x = 0 \\ \lambda e^{\lambda x} & ; x < 0 \end{cases} \equiv -\frac{\lambda x e^{-\lambda|x|}}{|x|}$$

Let the Fourier transform of $\partial f/\partial x$ be $F'(\omega)$:

$$F'(\omega) = \frac{1}{\sqrt{2\pi}}\left(\int_{-\infty}^{0} \lambda e^{\lambda x} e^{-i\omega x}\,\partial x + \int_{0}^{\infty} -\lambda e^{-\lambda x} e^{i\omega x}\,\partial x\right)$$

$$= \frac{1}{\sqrt{2\pi}}\left(\frac{\lambda}{\lambda - i\omega} - \frac{\lambda}{\lambda + i\omega}\right) = \frac{1}{\sqrt{2\pi}}\left(\frac{\lambda^2 + \lambda i\omega - \lambda^2 + \lambda i\omega}{\lambda^2 + \omega^2}\right)$$

$$= \frac{2\lambda i\omega}{\lambda^2 + \omega^2} = \boxed{i\omega F[f(x)]}$$

Our is equivalent to that of c.) for $F'(\omega)$.

⑥ a.) $f(x) = e^{-x^2/2\sigma^2} \cdot \cos(\omega_0 x)$

$$F[e^{-x^2/\sigma^2}] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{\frac{-x^2}{2\sigma^2}} e^{-i\omega x}\,\partial x = \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x^2 - 2\sigma^2 i\omega x)}\,\partial x \left(\frac{1}{\sqrt{2\pi}}\right)$$

Completing the square in the exponent we get:

$$F[e^{-x^2/\sigma^2}] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x + 2\sigma^2 i\omega x - \sigma^4\omega^2) - \frac{\sigma^4\omega^2}{2\sigma^2}}\,\partial x$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x + i\omega\sigma^2)^2 - \frac{\sigma^4\omega^2}{2\sigma^2}}\,\partial x$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x + i\omega\sigma^2) - \frac{\sigma^2\omega^2}{2}}\,\partial x$$

$$= \sqrt{2\pi\sigma^2}\, e^{-\frac{\sigma^2\omega^2}{2}}\left(\frac{1}{\sqrt{2\pi}}\right) = \boxed{\sigma e^{-\frac{\sigma^2\omega^2}{2}}}$$

Thus,

$$\boxed{F(\omega) = \frac{\sigma}{2}\left(e^{\frac{-\sigma^2(\omega-\omega_0)^2}{2}} + e^{\frac{-\sigma^2(\omega+\omega_0)^2}{2}}\right)} \ *$$

b.) F is symmetric and is centered around $\omega = 0$.
The width of F can be described the sum of the widths of the two Gaussians, call them g_1 and g_2.

$$g_1 = e^{\frac{-\sigma^2(\omega-\omega_0)^2}{2}}, \quad g_2 = e^{\frac{-\sigma^2(\omega+\omega_0)^2}{2}}$$

The Gaussians are maximized at $\omega = \omega_0$ and $\omega = -\omega_0$, respectively. Plugging in for either of these options to maximize F, we get that

$$\max\{F\} = \frac{\sigma}{2}$$

Thus, we want to find the ω_1 and ω_2 values that make the Fourier Transform F equal $\frac{\sigma}{4}$. There are 2 ω values in question because we are considering two independent Gaussians in our definition of F.

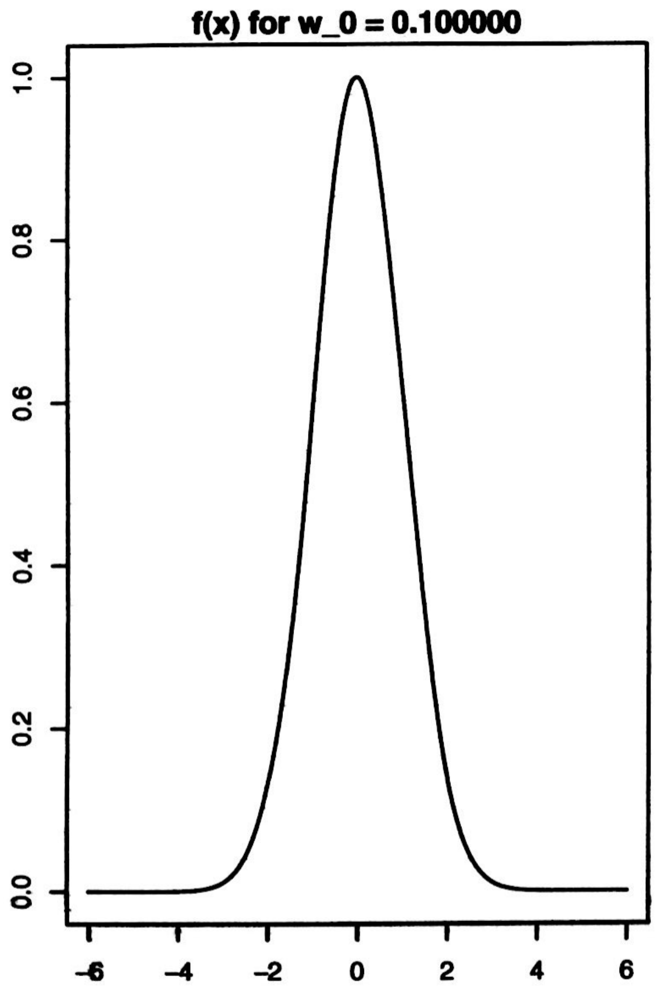
For g_1:
$$\frac{\sigma}{4} = \frac{\sigma}{2}\left(e^{\frac{-\sigma^2(\omega_1-\omega_0)^2}{2}}\right)$$
$$\frac{1}{2} = e^{\frac{-\sigma^2(\omega_1-\omega_0)^2}{2}}$$
$$\omega_1 = \omega_0 + \frac{1}{\sigma}\sqrt{2\ln(2)}$$

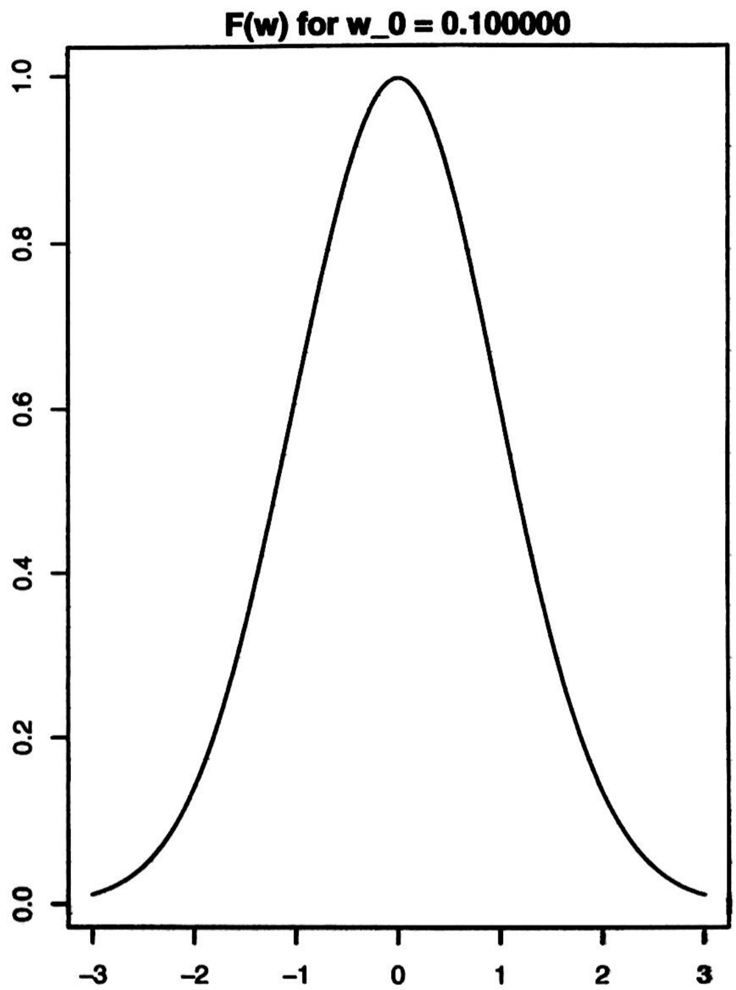
But because g_1 is centered at ω_0 we have to subtract ω_0 to get
$$\omega_1 = \boxed{\frac{1}{\sigma}\sqrt{2\ln(2)}}$$

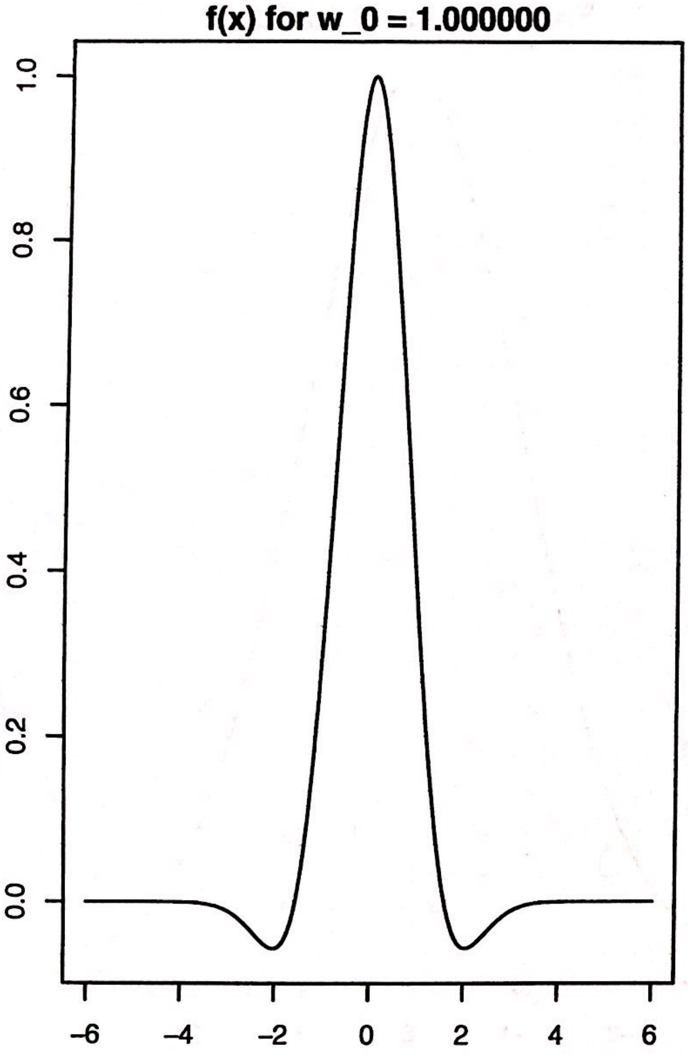
Repeating for g_2:
$$\frac{\sigma}{4} = \frac{\sigma}{2}\left(e^{\frac{-\sigma^2(\omega_2+\omega_0)^2}{2}}\right)$$
$$\omega_2 = \frac{1}{\sigma}\sqrt{2\ln(2)} - \omega_0$$

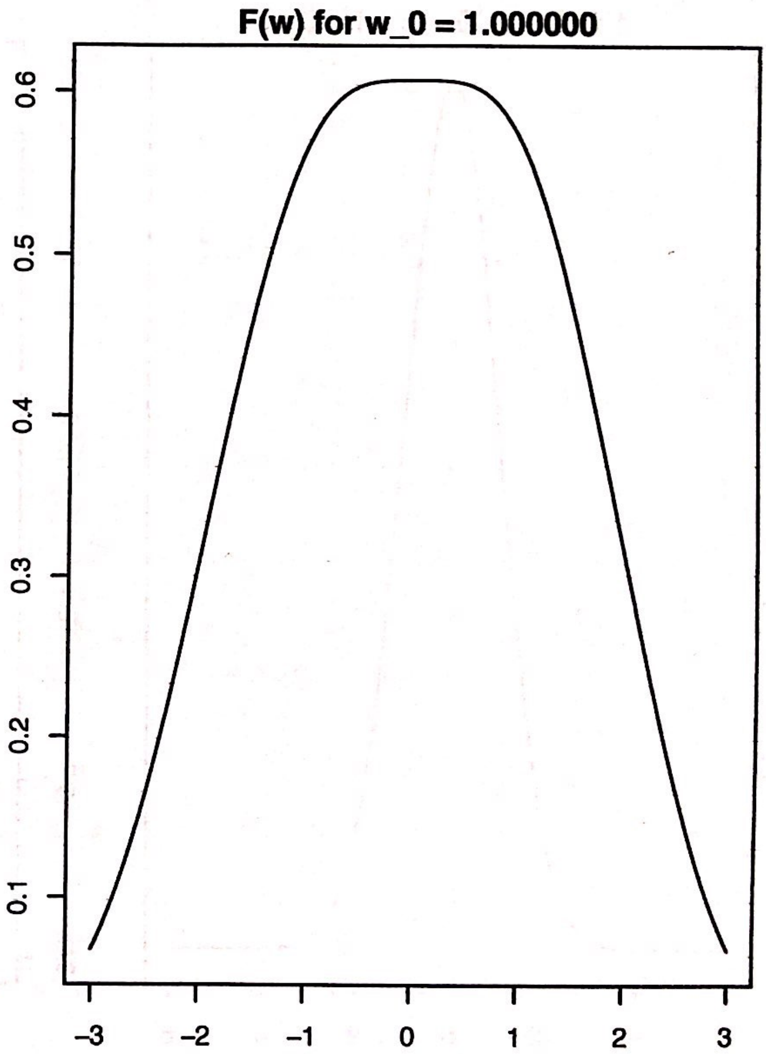
Because g_2 is centered at $-\omega_0$ we have to add ω_0 to get
$$\omega_2 = \boxed{\frac{1}{\sigma}\sqrt{2\ln(2)}}$$
← width of F. Depending on parameters σ and ω_0, the width of F compared to that of f will change.

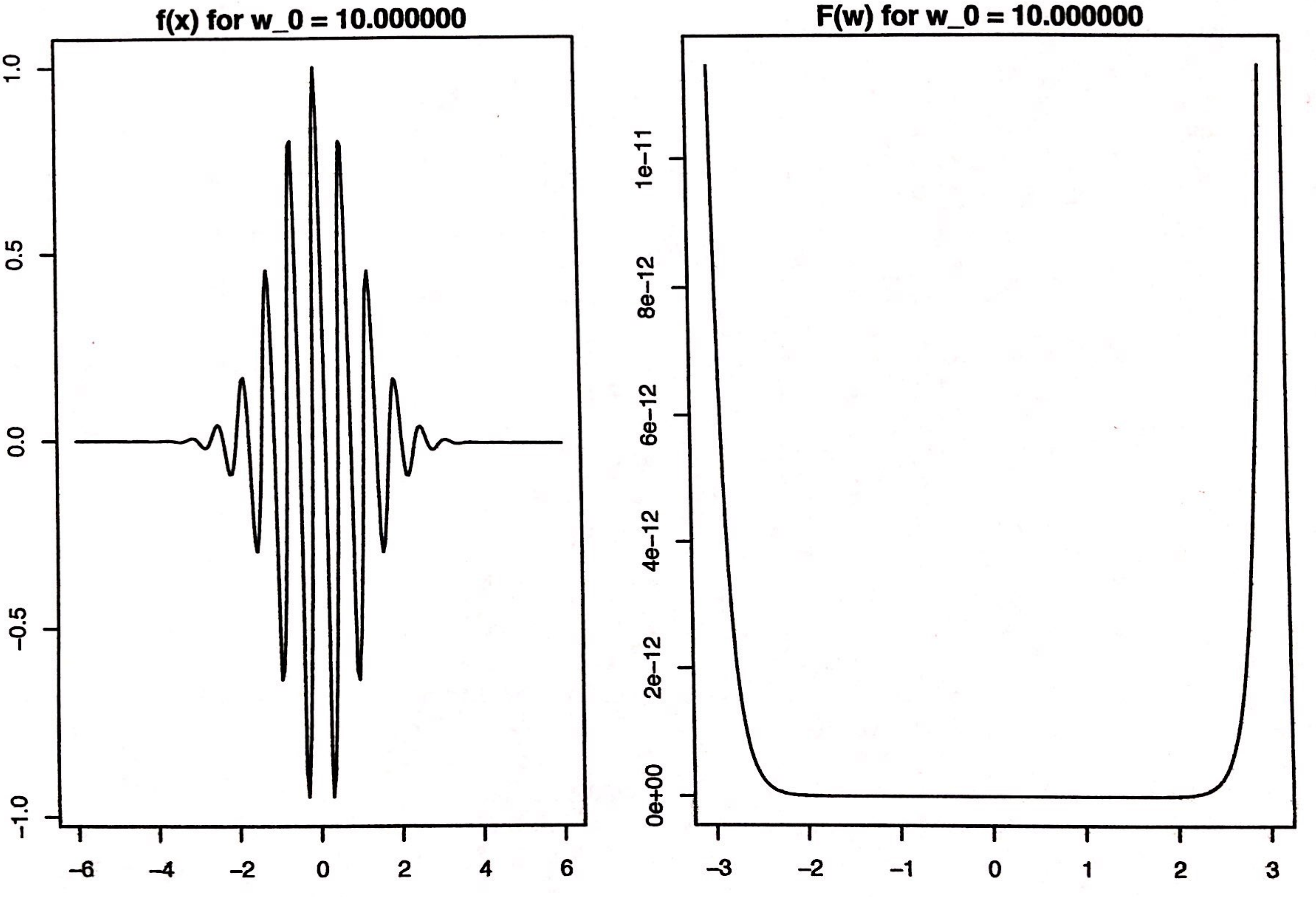
c.) See next page


f(x) for w_0 = 0.100000
0.0
0.2
0.4
0.6
0.8
1.0
−6
−4
−2
0
2
4
6


F(w) for w_0 = 0.100000
0.0
0.2
0.4
0.6
0.8
1.0
−3
−2
−1
0
1
2
3


f(x) for w_0 = 1.000000
1.0
0.8
0.6
0.4
0.2
0.0
−6
−4
−2
0
2
4
6


F(w) for w_0 = 1.000000
0.6
0.5
0.4
0.3
0.2
0.1
−3
−2
−1
0
1
2
3


f(x) for w_0 = 10.000000
F(w) for w_0 = 10.000000
1.0
0.5
0.0
−0.5
−1.0
−6
−4
−2
0
2
4
6
1e−11
8e−12
6e−12
4e−12
2e−12
0e+00
−3
−2
−1
0
1
2
3

d.) $G(\omega) = F[g(x)] \Rightarrow g(x) = F^{-1}[G(\omega)]$

$h(x) = \int_{-\infty}^{\infty} f(y) g(x-y)\, dy$ by definition of convolution

$$g(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} G(\omega) e^{i\omega x}\, d\omega$$

Thus,

$$h(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(y) \int_{-\infty}^{\infty} G(\omega) e^{i\omega(x-y)}\, \partial\omega\, dy$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} G(\omega) \underbrace{\left[\int_{-\infty}^{\infty} f(y) e^{-i\omega y}\, dy\right]}_{\sqrt{2\pi}\, F[f(y)] = F(\omega)\cdot\sqrt{2\pi}} e^{i\omega x}\, \partial\omega$$

$$h(x) = \int_{-\infty}^{\infty} G(\omega) F(\omega) e^{i\omega x}\, \partial\omega = F^{-1}[G(\omega)F(\omega)] \cdot \sqrt{2\pi}$$

$$\Rightarrow F[h(x)] = \boxed{G(\omega)F(\omega)\sqrt{2\pi}}$$

The frequencies present in $g(x)$ will be the same as those present after the filtering.

7 a.) DFT $F[f_x]$ where $L=100$ is the length of the array we get:

$$F[f_x] = \sum_{x=0}^{L-1} f_x e^{-\frac{i2\pi k x}{Lu}} \quad \text{for } k = 0, 1, \ldots, L-1$$

where u is the pixel length 0.01 mm

Let I_j be the intensity of pixel j we can rewrite:

$$F[I] = \sum_{j=1}^{100} I_j e^{-\frac{i2\pi k j}{100(.01)}} = \sum_{j=1}^{100} I_j e^{-i2\pi k j} \quad \text{for } k = 0, \ldots, L-1$$

Thus, the frequency $\omega = 2\pi k$ for $k = 0, \ldots, L-1$.

$\omega_{max} = 2\pi \cdot 99$

$\omega_{min} = 2\pi \cdot (0)$

Scanned with CamScanner

b.) The DFT of $f(i)$ as a function of v is:

$$F(f(i)) = \frac{1}{N}\sum_{x=0}^{N-1}\left(\sum_{j=0}^{N-i} v_{j+i}\cdot v_j\right)e^{\frac{-i2\pi k x}{N}} \quad \text{for } k=0,\dots,N-1$$

The inner sum $\sum_{j=0}^{N-i} v_{j+i}\cdot v_j$ given implies that there are $\underline{N+1}$ elements in v because the last term is $\underbrace{v_{N-i+i}}_{v_N}\cdot v_{N-i}$, so I was unable to rewrite the sum so it is from 0 to $N-1$.

8 a.) $y = W^T(x+z_{in}) + z_{out}$

Let C_y be the covariance matrix of y. From our solution to Problem 2, and because we know y is also Gaussian, of dim. $M\times 1$, we can write the entropy and conditional entropy of y given x, for which $C_{y|x}$ is the covariance matrix of $y|x$.

$$H(Y) = \frac{M}{2}\ln(2\pi) + \frac{1}{2}\ln(|C_y|)$$

$$H(Y|X) = \frac{M}{2}\ln(2\pi) + \frac{1}{2}\ln(|C_{y|x}|)$$

We can define C_y as follows:

$$C_y = W^T C W + \sigma_{in} W^T W + \sigma_{out} I_M$$

And $C_{y|x}$ can be decomposed as:

$$C_{y|x} = \sigma_{in} W^T W + \sigma_{out} I_M$$

Thus, the information $I(y,x)$ is given by:

$$I(y,x) = H(y) - H(y|x)$$

$$= \frac{M}{2}\cancel{\ln(2\pi)} + \frac{1}{2}\ln(|C_y|) - \left(\frac{M}{2}\cancel{\ln(2\pi)} + \frac{1}{2}\ln(|C_{y|x}|)\right)$$

$$= \frac{1}{2}\ln(|C_y|) - \frac{1}{2}\ln(|C_{y|x}|)$$

$$I(y,x) = \frac{1}{2}\ln(|W^T C W + \sigma_{in} W^T W + \sigma_{out} I_M|) - \frac{1}{2}\ln(|\sigma_{in} W^T W + \sigma_{out} I_M|)$$

Scanned with CamScanner

b.) Constraint: $Tr(C_y) = M\sigma^2$

The Lagrangian to maximize the mutual information subject to the constraint is thus given by:

$$\mathcal{L} = \frac{1}{2}\ln(|C_y|) - \frac{1}{2}\ln(|C_{y|x}|) + \lambda\left[Tr(C_y) - M\sigma^2\right]$$

$$\begin{aligned}\mathcal{L} = &\frac{1}{2}\ln(|W^TCW + \sigma_{in}W^TW + \sigma_{out}I_M|) \\ &- \frac{1}{2}\ln(|\sigma_{in}W^TW + \sigma_{out}I_M|) \\ &+ \lambda\left[Tr(W^TCW) + \sigma_{in}Tr(W^TW) - M(\sigma^2 - \sigma_{out})\right]\end{aligned}$$

To minimize $\mathcal{L}$ we take the derivative w.r.t W and set $=0$.

c.) We substitute WU for W in $\mathcal{L}$ above and U^TW^T for W^T.

Because U is orthogonal M dim. matrix, we know $U^TU = I_M$.

$$\begin{aligned}\mathcal{L} = &\frac{1}{2}\ln(|U^TW^TCWU + \sigma_{in}U^TW^TWU + \sigma_{out}I_M|) \\ &- \frac{1}{2}\ln(|\sigma_{in}U^TW^TWU + \sigma_{out}I_M|) \\ &+ \lambda\left[Tr(U^TW^TCWU) + \sigma_{in}Tr(U^TW^TWU) - M(\sigma^2 - \sigma_{out})\right]\end{aligned}$$

We can move U to the beginning of the expression for:

$$\underset{(M\times M)}{U^T}\,\underset{(M\times N)}{W^T}\,\underset{(N\times N)}{C}\,\underset{(N\times M)}{W}\,\underset{(M\times M)}{U} = U^TUW^TCW = W^TCW$$

Also, $U^TW^TWU = U^TUW^TW = W^TW$

Thus,

$$\begin{aligned}\mathcal{L} = &\frac{1}{2}\ln(|W^TCW + \sigma_{in}W^TW + \sigma_{out}I_M|) \\ &- \frac{1}{2}\ln(|\sigma_{in}W^TW + \sigma_{out}I_M|) \\ &+ \lambda\left[Tr(W^TCW) + \sigma_{in}Tr(W^TW) - M(\sigma^2 - \sigma_{out})\right]\end{aligned}$$

Which is equivalent to the Lagrangian $\mathcal{L}$ from b.)

This is significant because any transformation $W' = WU$ minimizes $\mathcal{L}$ if W minimizes $\mathcal{L}$. That is, there is an entire hyperplane of W's that minimize $\mathcal{L}$.