Exam, 2024-2025 Questions and Solutions

Note that many of the answers are far more detailed than required for full credit.

1. Group theory: Normal subgroups, representations

Given a finite group G with a normal subgroup H that is not all of G:

A. *Consider a coset Ha. Show that every element of the form hah' ($h' \in H$) is also an element of the coset Ha.*

Say $g = hah'$. We need to show that $g = hah'$ can be written as $h''a$, for some $h'' \in H$.

But $hah' = hah'\left(a^{-1}a\right) = h\left(ah'a^{-1}\right)a$. $ah'a^{-1} \in H$ since H is normal. So we can take $h'' = hah'a^{-1}$.

B. *Now consider the set C of distinct cosets. Right-multiplication by group elements is a permutation on C, and therefore, yields a representation L_C of G as permutation matrices; the dimension of this representation is $\left|C\right| = \frac{\left|G\right|}{\left|H\right|}$. What is its character?*

For $z \in H$, part A showed that $Haz = Ha$, i.e., right-multiplication by z preserves every coset. So $L_C(z)$ is the identity permutation on C, and $\chi_C(z) = \left|C\right|$.

For $z \notin H$, no coset is preserved by right-multiplication. For if it were the case that $Ha = Haz$, then for some h and h'' in H, we would have $ha = h''az$. Then $h''^{-1}ha = az$, $a^{-1}(h''^{-1}h)a = z$. Since H is normal and $h''^{-1}h \in H$, $z = a^{-1}(h''^{-1}h)a$ is also in H, a contradiction.

So for $z \notin H$, right-multiplication by z maps every coset to a different coset. The corresponding permutation matrix has zeros on the diagonal, so $\chi_C(z) = 0$.

C. *Is L_C irreducible? Why or why not?*

$\dim L_C \geq 2$ since $\left|C\right| = \frac{\left|G\right|}{\left|H\right|}$ and H is not all of G. The number of copies of the identity representation in L_C is given by the trace formula, namely, $\frac{1}{\left|G\right|}\sum_{g \in G}\chi_C(g)$. From part B, there are $\left|H\right|$ terms of the sum whose value is $\left|C\right| = \frac{\left|G\right|}{\left|H\right|}$, and the other terms are all zero. So there is one copy of the identify representation in L_C, and L_C is therefore reducible.

D. *If a finite group G has any subgroup H whose size is exactly half of that of G, find a nontrivial irreducible representation.*

In view of part C, if H is normal, then L_C is 2-dimensional and contains one copy of the identity. So removing this copy will yield a one-dimensional (and therefore irreducible) representation whose character is $+1$ on elements in H (since $2-1=1$), and whose character is -1 on elements not in H.

So it suffices to show that if $|H| = \frac{|G|}{2}$, then H must be normal.

A normal subgroup, by definition, is one whose right cosets are also left cosets. Left and right cosets are both ways of partitioning a group into disjoint subsets, whose size is equal to that of the subgroup H. So the right cosets consist of H and the complement of H in G. The left cosets also consist of H and the complement of H in G. So the partitions provided by the left and right cosets are identical, and H is normal. (Thanks to Math Stack Exchange for this last argument.)

2. Group theory: Conjugate classes, automorphisms, representations

The "quaternion group" Q is an 8-element group that can be defined as follows. Its elements are $\{1,-1,i,-i,j,-j,k,-k\}$. 1 and -1 behave in the standard fashion, i.e., multiplication by 1 is the identity, and multiplication by -1 is a sign change. For the other elements: all of their squares are -1, and they multiply as follows: $ij=k$, $jk=i$, $ki=j$, $ji=-k$, $kj=-i$, $ik=-j$.

A. *What are the conjugate classes of Q?*

Since 1 and -1 commute with all elements, each is its own conjugate class.

For the others: note first the definition of the way that the group elements combine is invariant under cyclic permutation of i, j, and k. Note also that $i^{-1}=-i$, since $i(-i)=-1(i^2)=1$, and similarly $j^{-1}=-j$ and $k^{-1}=-k$. Then $iji^{-1}=-iji=-ki=-j$, and also $kjk^{-1}=-kjk=-ki=-j$, so j and $-j$ are in the same conjugate class, but not in the conjugate class of any of the other letters. Because of the cyclic symmetry of the group definition, this holds for i and k as well.

So the conjugate classes are $\{1\}$, $\{-1\}$, $\{i,-i\}$, $\{j,-j\}$, and $\{k,-k\}$.

B. *What are the inner automorphisms of Q?*

Since 1 and -1 commute with all elements, the adjoint action $g\to z^{-1}gz$ is trivial for these group elements. But as in part A, the adjoint action for i changes the sign of $\pm j$ and $\pm k$, but not the sign of $\pm i$. The adjoint action for $-i$ is identical. So each of i, j, and k yield different adjoint actions (each adjoint action flips the sign of the other two letters). These are the three non-trivial inner automorphisms of Q.

C. *What all the automorphisms of Q?*

Two elements of Q are intrinsically unique: the identity (which is always intrinsically unique), and -1. since it is the only non-identity that commutes with every element. So they cannot be remapped by any automorphism. The conjugate classes are also an intrinsic partitioning, so no automorphism can break up a conjugate class. That is, if it maps i to an element other than $-i$, it must map $-i$ to the negative of that element. So we can analyze the automorphisms in terms of what they do to the conjugate classes as sets, and what they do *within* the conjugate classes.

Action on the conjugate classes as sets: As noted above, the definitions of group multiplication are invariant under cyclic permutation of the group elements, so cyclic permutations of the conjugate classes, with preservation of all signs, are automorphisms. But also, an odd permutation of the conjugate classes (e.g., pairwise swap of the j's with the k's) is consistent with the group properties if all signs are flipped.

Action within the conjugate classes: If the conjugate classes are not permuted, are there any other automorphisms? Of the 2^3 ways of either flipping or not flipping signs within conjugate classes, part B shows that the ones in which two conjugate classes are sign-flipped are automorphisms, as well as the trivial automorphism in which none are flipped. So, of the 2^3 ways of either flipping or not flipping signs within conjugate classes, four are automorphisms.

It remains to check that flipping the signs of only one conjugate class, or of all three, is not an automorphism. This is the case: for example, $-ij=-k$ but $ij=k$.

So there are 24 automorphisms: the six ways of permuting the conjugate classes (and flipping all signs if that permutation is odd), followed by any of the four inner automorphisms, which keeps the conjugate classes invariant.

D. *What is the character table of Q? Use part B, or, part D of Question 1, to find the one-dimensional representations.*

There are five conjugate classes, so there are five irreducible representations. The sum of the squares of the dimensions of these representations is 8, so there must be four irreducible representations of dimension 1, and one of dimension 2.

To find the nontrivial irreducible representations using part B: note that the inner automorphisms form a group. Since each nontrivial inner automorphisms flips the sign of some group elements and not of the others, this sign-flip or lack of flip (multiplication by ± 1) is a representation.

To find these representations using part D of Question 1: Note that $\{1,i,i^2=-1,i^3=-i\}$ is a subgroup. So there's a representation L_i that is $+1$ in this subgroup, and -1 on its complement.

Thus, the character table so far is:

Conj. Class:	$\{1\}$	$\{-1\}$	$\{i,-i\}$	$\{j,-j\}$	$\{k,-k\}$
Representation					
Trivial	$+1$	$+1$	$+1$	$+1$	$+1$
L_i	$+1$	$+1$	$+1$	-1	-1
L_j	$+1$	$+1$	-1	$+1$	-1
L_k	$+1$	$+1$	-1	-1	$+1$
L_2	2	?	?	?	?

The rest of the table can be competed by orthogonality: column orthogonality for the second column, and then row orthogonality for the remaining entries:

Conj. Class:	$\{1\}$	$\{-1\}$	$\{i,-i\}$	$\{j,-j\}$	$\{k,-k\}$
Representation					
Trivial	$+1$	$+1$	$+1$	$+1$	$+1$
L_i	$+1$	$+1$	$+1$	-1	-1
L_j	$+1$	$+1$	-1	$+1$	-1
L_k	$+1$	$+1$	-1	-1	$+1$
L_2	2	-2	0	0	0

3. Fourier analysis as a unitary transformation; generating functions

In its standard form, Fourier transformation is almost a unitary transformation – the inner product of two functions differs from the inner product of their Fourier transforms by a factor of 2π *(Parseval's Theorem). We can make it unitary by a slightly nonstandard formulation, which presents a nearly symmetric relationship between complex-valued functions on the line and their Fourier transforms:*

$$\widehat{f}(x) = (Sf)(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ixu} f(u)du \tag{1}$$

and

$$f(x) = \left(S^{-1}\widehat{f}\right)(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ixu} \widehat{f}(u)du\,. \tag{2}$$

In this formulation, Fourier transformation is truly unitary: $\int_{-\infty}^{\infty} \widehat{f}(x)\overline{\widehat{g}(x)}dx = \int_{-\infty}^{\infty} f(x)\overline{g(x)}dx$ *.We write this "unitarized" Fourier transformation as an operator (i.e.,* $\widehat{f}(x) = (Sf)(x)$*), to emphasize this viewpoint .Here, we find the eigenvalues and eigenvectors of* S *, and use this to define operations that are "fractional" Fourier transforms.*

A. *Hermite polynomials* $h_n(x)$ *may be defined via a generating function* $\sum_{n=0}^{\infty} \frac{h_n(x)}{n!} t^n \triangleq h(x,t) = \exp\left(xt - \frac{t^2}{2}\right)$. *Instead we consider the Hermite functions, defined (here) by* $H_n(x) = \frac{1}{2^{n/2}} h_n(x\sqrt{2}) \exp\left(-\frac{x^2}{2}\right)$. *Determine the generating function* $H(x,t) \triangleq \sum_{n=0}^{\infty} \frac{H_n(x)}{n!} t^n$.

$$H(x,t) = \sum_{n=0}^{\infty} \frac{H_n(x)}{n!} t^n = \sum_{n=0}^{\infty} \frac{h_n(x\sqrt{2})}{2^{n/2} n!} \exp\left(-\frac{x^2}{2}\right) t^n = \exp\left(-\frac{x^2}{2}\right) \sum_{n=0}^{\infty} \frac{h_n(x\sqrt{2})}{2^{n/2} n!} t^n = \exp\left(-\frac{x^2}{2}\right) \sum_{n=0}^{\infty} \frac{h_n(x\sqrt{2})}{n!} \left(\frac{t}{\sqrt{2}}\right)^n$$
$$= \exp\left(-\frac{x^2}{2}\right) h(x\sqrt{2}, \frac{t}{\sqrt{2}}) = \exp\left(-\frac{x^2}{2}\right) \exp\left(x\sqrt{2}\frac{t}{\sqrt{2}} - \frac{1}{2}\left(\frac{t}{\sqrt{2}}\right)^2\right) = \exp\left(-\frac{x^2}{2} + xt - \frac{t^2}{4}\right)$$

B. *What is* $SH(x,t)$*, the unitarized Fourier transform (eq. 1) of the generating function* $H(x,t)$ (with respect to x)?

$$(SH)(x,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ixu} H(u,t)du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-ixu) \exp\left(-\frac{u^2}{2} + ut - \frac{t^2}{4}\right) du\,.$$
$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{4}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{u^2}{2} + ut - ixu\right) du$$

Completing the square,

$$
\begin{aligned}
(SH)(x,t) &= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{4}\right)\int_{-\infty}^{\infty}\exp\left(-\frac{u^2}{2}+ut-ixu\right)du \\
&= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{4}\right)\exp\left(\frac{1}{2}(t-ix)^2\right)\int_{-\infty}^{\infty}\exp\left(-\frac{u^2}{2}+ut-ixu-\frac{1}{2}(t-ix)^2\right)du \\
&= \exp\left(-\frac{t^2}{4}\right)\exp\left(\frac{1}{2}(t-ix)^2\right)\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}(u+t-ix)^2\right)du \\
&= \exp\left(-\frac{t^2}{4}\right)\exp\left(\frac{1}{2}(t-ix)^2\right)
\end{aligned},
$$

where the last step is the standard Gaussian integral.

So, $(SH)(x,t)=\exp\left(-\frac{x^2}{2}-ixt+\frac{t^2}{4}\right)$.

Comparing this with part A, $(SH)(x,t)=H(x,-it)$.

C. *By comparing the generating functions $H(x,t)$ and $SH(x,t)$, determine eigenvalues and eigenvectors of S.*

By linearity and part B, $\sum_{n=0}^{\infty}\frac{1}{n!}SH_n(x)t^n=(SH)(x,t)=H(x,-it)$. And from part A, $H(x,it)=\sum_{n=0}^{\infty}\frac{H_n(x)}{n!}(-it)^n$

So $\sum_{n=0}^{\infty}\frac{1}{n!}SH_n(x)t^n=\sum_{n=0}^{\infty}\frac{1}{n!}H_n(x)(-it)^n$.

Equating like powers of t yields $SH_n(x)=(-i)^nH_n(x)$. So H_n is an eigenvector, and $(-i)^n$ is its eigenvalue.

D. *Provide a recipe for an operator R that is the operator-kth-root of a Fourier transform, namely, an operator R for which $R^k=S$ (other than the trivial $R=S$). (Technical: the recipe should make sense for all functions f that can be written as an absolutely convergent sum of Hermite functions.)*

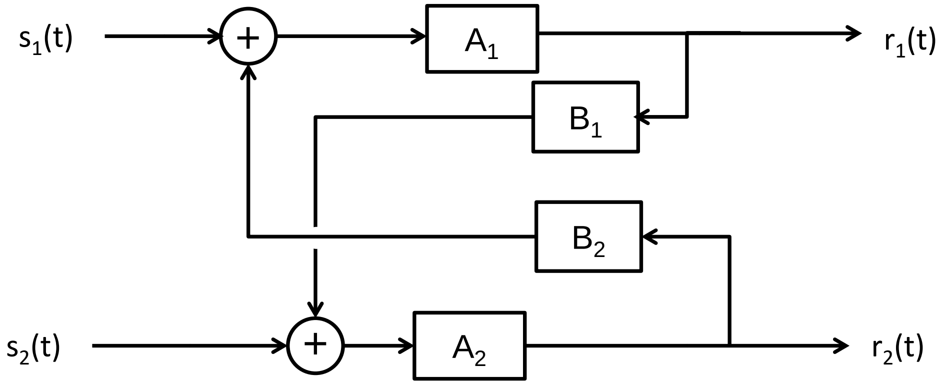
The Hermite function basis diagonalizes S, and the diagonal elements are the eigenvectors $(-i)^n=e^{-\frac{2\pi i}{4}n}$, so we set R to be the kth root of this matrix. To compute Rf, express f in the basis of Hermite functions, i.e., $f(x)=\sum_{n=0}^{\infty}a_nH_n(x)$. (The a_n can be computed in the standard dot-product, as $a_n=\frac{\int f(x)H_n(x)dx}{\int (H_n(x))^2\,dx}$; both integrals will be convergent for any reasonably-behaved f because of the Gaussian built into H_n.

Then, set $a_n{}'=a_ne^{-\frac{2\pi i}{4k}n}$, and $Rf(x)=\sum_{n=0}^{\infty}a_n{}'H_n(x)$. (Technical: If $f(x)=\sum_{n=0}^{\infty}a_nH_n(x)$ is absolutely convergent, then so is the sum for Rf, since $|a_n'|=|a_n|$.)

Note that the Hermite functions, which are oscillatory polynomials multiplied by Gaussians, can be thought of as wavelets. Thus, this construction gives an interpretation of a wavelet representation as intermediate between time and frequency.

4. Input-output systems, noise, covariance

Consider the two-input, two-output system below, where the transformations A_i and B_i are linear.

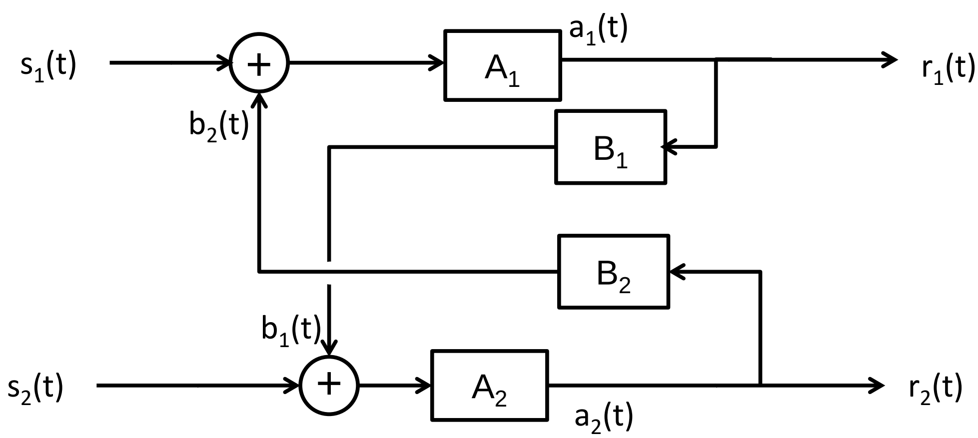


A. *Determine the transfer functions $\hat{L}_{ij}$ between input s_i and output r_j*

B. *If s_1 and s_2 are uncorrelated Gaussian white noises with unit spectral density, what is the power spectrum of r_1 and of r_2?*

C. *What is the coherence of r_1 and r_2?*

A. With the internal signals labeled as follows:



$\hat{r}_1(\omega) = \hat{A}_1(\omega)\left(\hat{s}_1(\omega) + \hat{B}_2(\omega)\hat{r}_2(\omega)\right)$ and $\hat{r}_2(\omega) = \hat{A}_2(\omega)\left(\hat{s}_2(\omega) + \hat{B}_1(\omega)\hat{r}_1(\omega)\right)$.

Combining these yields

$\hat{r}_1(\omega) = \hat{A}_1(\omega)\left(\hat{s}_1(\omega) + \hat{B}_2(\omega)\left[\hat{A}_2(\omega)\left(\hat{s}_2(\omega) + \hat{B}_1(\omega)\hat{r}_1(\omega)\right)\right]\right)$, or

$$\hat{r}_1(\omega) = \hat{A}_1(\omega)\hat{s}_1(\omega) + \hat{A}_1(\omega)\hat{A}_2(\omega)\hat{B}_2(\omega)\hat{s}_2(\omega) + \hat{A}_1(\omega)\hat{A}_2(\omega)\hat{B}_1(\omega)\hat{B}_2(\omega)\hat{r}_1(\omega)\text{, or}$$

$$\hat{r}_1(\omega) = \hat{A}_1(\omega)\frac{\hat{s}_1(\omega) + \hat{A}_2(\omega)\hat{B}_2(\omega)\hat{s}_2(\omega)}{1 - \hat{A}_1(\omega)\hat{A}_2(\omega)\hat{B}_1(\omega)\hat{B}_2(\omega)}\text{; similarly,}$$

$$\hat{r}_2(\omega) = \hat{A}_2(\omega)\frac{\hat{s}_2(\omega) + \hat{A}_1(\omega)\hat{B}_1(\omega)\hat{s}_1(\omega)}{1 - \hat{A}_1(\omega)\hat{A}_2(\omega)\hat{B}_1(\omega)\hat{B}_2(\omega)}.$$

We can find the desired single-input, single-output transfer functions by setting one of the s_i to zero. With $\hat{D}(\omega) = 1 - \hat{A}_1(\omega)\hat{A}_2(\omega)\hat{B}_1(\omega)\hat{B}_2(\omega)$

$$\hat{L}_{11}(\omega) = \frac{\hat{A}_1(\omega)}{\hat{D}(\omega)},\ \hat{L}_{12}(\omega) = \frac{\hat{A}_1(\omega)\hat{A}_2(\omega)\hat{B}_1(\omega)}{\hat{D}(\omega)},\ \hat{L}_{21}(\omega) = \frac{\hat{A}_1(\omega)\hat{A}_2(\omega)\hat{B}_2(\omega)}{\hat{D}(\omega)},\ \hat{L}_{22}(\omega) = \frac{\hat{A}_2(\omega)}{\hat{D}(\omega)}.$$

B. Using the Fourier-estimates formalism, in the uncorrelated case, the power spectra $\hat{P}_i(\omega)$ are the magnitude-squared of $\left|\hat{r}_i(\omega)\right|^2$ when $\left\langle\left|\hat{s}_i(\omega)\right|^2\right\rangle = 1$ and $\left\langle \hat{s}_1(\omega)\overline{\hat{s}_2(\omega)}\right\rangle = 0$:

$$\hat{P}_1(\omega) = \left\langle\left|\hat{r}_1(\omega)\right|^2\right\rangle = \left|\frac{\hat{A}_1(\omega)}{\hat{D}(\omega)}\right|^2 \left\langle\left(\hat{s}_1(\omega) + \hat{A}_2(\omega)\hat{B}_2(\omega)\hat{s}_2(\omega)\right)\overline{\left(\hat{s}_1(\omega) + \hat{A}_2(\omega)\hat{B}_2(\omega)\hat{s}_2(\omega)\right)}\right\rangle$$

$$= \left|\frac{\hat{A}_1(\omega)}{\hat{D}(\omega)}\right|^2 \left\langle \hat{s}_1(\omega)\overline{\hat{s}_1(\omega)} + \left(\hat{A}_2(\omega)\hat{B}_2(\omega)\hat{s}_2(\omega)\right)\overline{\left(\hat{A}_2(\omega)\hat{B}_2(\omega)\hat{s}_2(\omega)\right)}\right\rangle .$$

$$= \left|\frac{\hat{A}_1(\omega)}{\hat{D}(\omega)}\right|^2 \left(1 + \left|\hat{A}_2(\omega)\hat{B}_2(\omega)\right|^2\right)$$

(Cross-terms involving $\left\langle \hat{s}_1(\omega)\overline{\hat{s}_2(\omega)}\right\rangle$ vanish because of independence).

Similarly $\hat{P}_2(\omega) = \left|\frac{\hat{A}_2(\omega)}{\hat{D}(\omega)}\right|^2 \left(1 + \left|\hat{A}_1(\omega)\hat{B}_1(\omega)\right|^2\right)$.

C. We compute the coherency (coherence and its phase) as $\hat{C}(\omega) = \frac{\hat{P}_{12}(\omega)}{\sqrt{\hat{P}_1(\omega)\hat{P}_2(\omega)}}$, where

$$\hat{P}_{12}(\omega) = \left\langle \hat{r}_1(\omega)\overline{\hat{r}_2(\omega)}\right\rangle.$$

$$\hat{P}_{12}(\omega) = \left\langle \hat{r}_1(\omega)\overline{\hat{r}_2(\omega)}\right\rangle = \frac{\hat{A}_1(\omega)}{\hat{D}(\omega)}\overline{\frac{\hat{A}_2(\omega)}{\hat{D}(\omega)}}\left\langle\left(\hat{s}_1(\omega) + \hat{A}_2(\omega)\hat{B}_2(\omega)\hat{s}_2(\omega)\right)\overline{\left(\hat{s}_2(\omega) + \hat{A}_1(\omega)\hat{B}_1(\omega)\hat{s}_1(\omega)\right)}\right\rangle$$

$$= \frac{\hat{A}_1(\omega)\overline{\hat{A}_2(\omega)}}{\left|\hat{D}(\omega)\right|^2}\left\langle \hat{s}_1(\omega)\left(\overline{\hat{A}_1(\omega)\hat{B}_1(\omega)\hat{s}_1(\omega)}\right) + \hat{A}_2(\omega)\hat{B}_2(\omega)\hat{s}_2(\omega)\overline{\hat{s}_2(\omega)}\right\rangle$$

(terms involving $\left\langle \hat{s}_1(\omega)\overline{\hat{s}_2(\omega)}\right\rangle$ vanish because of independence). So

$$\hat{P}_{12}(\omega) = \frac{\hat{A}_1(\omega)\overline{\hat{A}_2(\omega)}}{\left|\hat{D}(\omega)\right|^2}\left(\overline{\hat{A}_1(\omega)\hat{B}_1(\omega)} + \hat{A}_2(\omega)\hat{B}_2(\omega)\right)$$

and

$$\hat{C}_{12}(\omega)=\frac{\hat{P}_{12}(\omega)}{\sqrt{\hat{P}_1(\omega)\hat{P}_2(\omega)}}=\frac{\frac{\hat{A}_1(\omega)\overline{\hat{A}_2(\omega)}}{\left|\hat{D}(\omega)\right|^2}\left(\overline{\hat{A}_1(\omega)\hat{B}_1(\omega)}+\hat{A}_2(\omega)\hat{B}_2(\omega)\right)}{\sqrt{\left|\frac{\hat{A}_1(\omega)}{\hat{D}(\omega)}\right|^2\left(1+\left|\hat{A}_2(\omega)\hat{B}_2(\omega)\right|^2\right)\left|\frac{\hat{A}_2(\omega)}{\hat{D}(\omega)}\right|^2\left(1+\left|\hat{A}_1(\omega)\hat{B}_1(\omega)\right|^2\right)}}$$

$$=\frac{\hat{A}_1(\omega)\overline{\hat{A}_2(\omega)}}{\left|\hat{A}_1(\omega)\right|\left|\hat{A}_2(\omega)\right|}\frac{\overline{\hat{A}_1(\omega)\hat{B}_1(\omega)}+\hat{A}_2(\omega)\hat{B}_2(\omega)}{\sqrt{\left(1+\left|\hat{A}_2(\omega)\hat{B}_2(\omega)\right|^2\right)\left(1+\left|\hat{A}_1(\omega)\hat{B}_1(\omega)\right|^2\right)}}$$

The coherence is the magnitude of the above.

5. Principal components analysis

Consider a data matrix X with elements $x_{i,j}$ (R rows, C columns, $C \gg R$), in which all rows and columns are linearly independent (so there are R principal components). Indicate to what extent these manipulations can change the number of principal components.

A. The rows of X are permuted.

This cannot change the number of principal components, as it merely permutes the row and column labels of XX^*, which cannot change the eigenvalues.

B. The mean of each column is subtracted from each element of X, i.e., $x'_{i,j}=x_{i,j}-\frac{1}{R}\sum_k x_{k,j}$.

This introduces a linear dependence among the rows $\vec{x}_i = x_{i,j}$: with $\vec{x}'_i = x'_{i,j}$, $\frac{1}{R}\sum_k \vec{x}'_k = 0$, so the dimension spanned by the rows, and hence the number of principal components, is now $R-1$.

C. The mean of each row is subtracted from each element of X, i.e., $x'_{i,j}=x_{i,j}-\frac{1}{C}\sum_k x_{i,k}$, [Note, the 1/C term was mistakenly left out in the original questions.]

The rows typically remain linearly independent if $C \gg R$, but a linear dependence could be introduced and the rank could be reduced to $R-1$. This will happen if a row of ones (here, $\vec{u}$) is in the linear span of the rows, i.e., $\vec{u}=\sum_i c_i\vec{x}_i$ with not all $c_i=0$. We can then choose some row j with $c_j \neq 0$, and replace it by a linear combination of $\vec{u}$ and the other c_i's, i.e., $\vec{x}_j=\vec{u}-\frac{1}{c_j}\sum_{i\neq j}c_i\vec{x}_i$. Then, subtracting the row means would yield $\vec{x}'_j=-\frac{1}{c_j}\sum_{i\neq j}c_i\vec{x}'_i$, demonstrating that in the new data matrix, the rows are no longer independent.

D. Each element of X is replaced by its square.

This is a nonlinear transformation of each row. While it typically does not introduce linear dependencies (if the elements of X have no special relationships), the number of principal components to as few as 1 in special cases – the extreme being that each element of X is ± 1.

E. Each element of X is replaced by its first N Fourier components (not including DC).

This is a projection of X onto a subspace of dimension $2N$ (each Fourier component requires two dimensions, spanned by the cosine component and the sine component) so the number of principal components is reduced to at most $\max(2N,R)$. It could as small as zero, if the rows of the original X were entirely contained in higher Fourier components.

F. New rows of X are adjoined to X, with each new row equal to the sum of the first N Fourier components (not including DC) of one row in the original X.

This cannot reduce the number of principal components, as span of the rows of the new matrix includes the span of the original. It may add as many as $2N$ new principal components, if the higher Fourier components are linearly independent of the first N components.

G. Each element of X is replaced by $\max(0,x_{i,j})$, and new rows are adjoined, given by $\min(0,x_{i,j})$. That is, each row is split into two rows, one containing only the positive values and one containing only the negative values.

This cannot reduce the number of principal components, as the span of the rows of the new matrix includes the span of the original (since $x_{i,j}=\max(0,x_{i,j})+\min(0,x_{i,j})$). Typically, it will increase the number to the as much as twice the number of rows in the original X, if there is a no linear relationship between the negative-going and positive-going portions of the original matrix. Anything in between is possible – for example, if some of the original rows are only positive-going, and for the others, there is no linear relationship between the positive-going and negative-going components.

H. New rows are adjoined to X, consisting of averages of random-with-replacement selections from X.

The number of principal components cannot change. It cannot increase since each new row is a linear combination of existing rows. It also cannot decrease the number of principal components, since the original rows of X are retained.

6. Maximum-entropy distributions

A. What is the functional form for a maximum-entropy distribution for a continuous non-negative scalar variable, with constrained (arithmetic) mean and geometric mean?

We make use of the Lagrange-multiplier formalism for maximum-entropy distributions on a continuous domain. The constraints are linear in the probability distribution: the mean constrains $\langle x\rangle = \int_0^\infty xp(x)dx$; the geometric mean constrains $\langle \log(x)\rangle = \int_0^\infty \log(x)p(x)dx$. So, $p(x)$ is of the form $p(x) = \exp\left(\lambda_0 + \lambda_{mean}x + \lambda_{geomean}\log(x)\right)$. This can be put into a more familiar form – the gamma distribution: with $K = \exp(\lambda_0)$, $m = -1/\lambda_{mean}$, $\alpha = \exp(\lambda_{geomean})$, $p(x) = Kx^\alpha e^{-x/m}$

B. Show how the unique parameter values for the distribution can be determined from the given arithmetic and geometric means.

The parameter values can be found from the constraint equations, but with some effort:

$$\int_0^\infty Kx^\alpha e^{-x/m}dx = K\int_0^\infty (um)^\alpha e^{-u}mdu = Km^{\alpha+1}\int_0^\infty u^\alpha e^{-u}du = Km^{\alpha+1}\Gamma(\alpha+1), \text{ so } K = \frac{1}{m^{\alpha+1}\Gamma(\alpha+1)}$$

For the mean, $\mu = \int_0^\infty Kx^{\alpha+1}e^{-x/m}dx$. This has the same form as the integral above, so

$$\int_0^\infty Kx^{\alpha+1}e^{-x/m}dx = Km^{\alpha+2}\Gamma(\alpha+2) = \frac{m^{\alpha+2}\Gamma(\alpha+2)}{m^{\alpha+1}\Gamma(\alpha+1)} = m\left(\alpha+1\right), \text{ so } m\left(\alpha+1\right) = \mu .$$

The constraint on the geometric mean allows for determination of α from the constrained arithmetic mean μ and the geometric mean γ. The constraint on the geometric mean is $\log\gamma = \int_0^\infty Kx^\alpha \log(x)e^{-x/m}dx$. This integral can be done via the same trick that was used to write an alternative definition of entropy. Since $\frac{d}{d\alpha}x^\alpha = \frac{d}{d\alpha}e^{\alpha\log x} = x^\alpha\log(x)$

$$\log\gamma = \int_0^\infty Kx^\alpha \log(x)e^{-x/m}dx = \frac{d}{d\alpha}\left(\int_0^\infty Kx^\alpha e^{-x/m}dx\right) = K\frac{d}{d\alpha}\left(m^{\alpha+1}\Gamma(\alpha+1)\right).$$

This can be simplified:

$$\log\gamma = K\frac{d}{d\alpha}\left(m^{\alpha+1}\Gamma(\alpha+1)\right) = \frac{1}{m^{\alpha+1}\Gamma(\alpha+1)}\frac{d}{d\alpha}\left(m^{\alpha+1}\Gamma(\alpha+1)\right)$$

$$= \frac{d}{d\alpha}\log\left(m^{\alpha+1}\Gamma(\alpha+1)\right) = \frac{d}{d\alpha}\left(\log m^{\alpha+1}\right) + \frac{d}{d\alpha}\left(\log\left(\Gamma(\alpha+1)\right)\right) .$$

$$= \log m + \frac{\Gamma'(\alpha+1)}{\Gamma(\alpha+1)} = \log\mu - \log\left(\alpha+1\right) + \frac{\Gamma'(\alpha+1)}{\Gamma(\alpha+1)}$$

So finally, $\log(\alpha+1)-\frac{\Gamma'(\alpha+1)}{\Gamma(\alpha+1)}=\log\mu-\log\gamma$. Crucially, the arithmetic mean is always larger than the geometric mean (but this ratio is not bounded), and, for $\alpha>-1$ the left-hand side of the above is a monotonically-decreasing function whose range is all positive numbers. So, given any self-consistent values of μ and γ, $\log(\alpha+1)-\frac{\Gamma'(\alpha+1)}{\Gamma(\alpha+1)}=\log\mu-\log\gamma$ yields a unique value for the shape parameter α. Then $m(\alpha+1)=\mu$ yields the scale parameter m, and the normalization constant is $K=\frac{1}{m^{\alpha+1}\Gamma(\alpha+1)}$.

C. As in A, but now, constrained arithmetic mean and root-mean-square.

Again make use of the Lagrange-multiplier formalism for maximum-entropy distributions on a continuous domain. The constraints are linear in the probability distribution: the mean constrains $\langle x\rangle=\int_0^\infty xp(x)dx$; the root-mean-square constrains $\sqrt{\langle x^2\rangle}$, but this is equivalent to constraining $\langle x^2\rangle=\int_0^\infty x^2p(x)dx$. So, $p(x)$ is of the form $p(x)=\exp\left(\lambda_0+\lambda_{mean}x+\lambda_{rms}x^2\right)$, i.e., $p(x)$ is a Gaussian restricted to the positive reals.

D. Determine the maximum-entropy distribution for a bivariate process $\vec{x}=\begin{pmatrix}x_1\\x_2\end{pmatrix}$, in which both variables are constrained to be non-negative, and the mean value of their sum is constrained to be s. Are the variables independent?

Other than normalization, the only constraint is $\langle x_1+x_2\rangle=\int_0^\infty\int_0^\infty(x_1+x_2)p(x_1,x_2)dx_1dx_2$, so $p(x_1,x_2)$ is of the form $p(x_1,x_2)=\exp\left(\lambda_0+\lambda_{sum}(x_1+x_2)\right)$ where both arguments are non-negative, and zero elsewhere. Take $K=\exp(\lambda_0)$ and $\lambda_{sum}=-1/m$ so $p(x_1,x_2)=Ke^{-(x_1+x_2)/m}$

To determine the parameters:

$1=K\int_0^\infty\int_0^\infty e^{-(x_1+x_2)/m}dx_1dx_2=K\int_0^\infty e^{-x_1/m}dx_1\int_0^\infty e^{-x_2/m}dx_2=Km^2$, since $\int_0^\infty e^{-x/m}dx=m\int_0^\infty e^{-u}du=m$. So $K=\frac{1}{m^2}$.

$$s=\langle x_1+x_2\rangle=K\int_0^\infty\int_0^\infty(x_1+x_2)e^{-(x_1+x_2)/m}dx_1dx_2=K\left(\int_0^\infty x_1e^{-x_1/m}dx_1\int_0^\infty e^{-x_2/m}dx_2+\int_0^\infty e^{-x_1/m}dx_1\int_0^\infty x_2e^{-x_2/m}dx_2\right)$$
.

Since $\int_0^\infty xe^{-x/m}dx=m^2\int_0^\infty ue^{-u}du=m^2\left(-ue^{-u}-e^{-u}\right)\Big|_0^\infty=m^2$,

$s = \langle x_1 + x_2 \rangle = K\left(m^2 \bullet m + m \bullet m^2\right) = 2Km^3 = 2m$, so $m = s/2$ and $p(x_1, x_2) = \frac{4}{s^2} e^{-2(x_1+x_2)/s}$. x_1 and x_2 are independent, since $p(x_1, x_2)$ factors into terms that depend only on each variable individually: $p(x_1, x_2) = \frac{4}{s^2} e^{-2(x_1+x_2)/s} = \left(\frac{2}{s} e^{-2x_1/s}\right)\left(\frac{2}{s} e^{-2x_2/s}\right)$