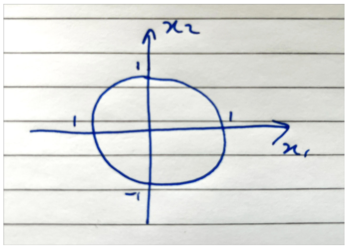
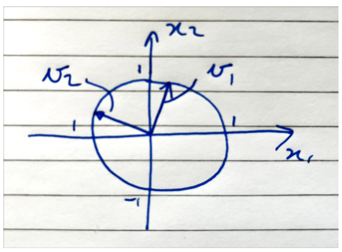
3M1 Mathematical Methods, 2024

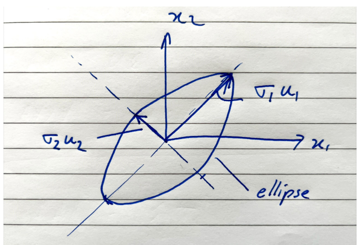
Linear Algebra

1. (a) Unit sphere is a circle in this case.



(b) Unit circle is transformed by a real 2×2 matrix into an ellipse. The major axis is the direction u_1 and has length σ_1. The minor axis is the direction u_2 and has length σ_2. The vectors v_1 and v_2 are the directions on the circle that A rotates to align with u_1 and u_2, i.e. $Av_i = \sigma_i u_i$





(c) – $\kappa_2 = \sigma_1/\sigma_m$ by definition.
$|\det A| = \Pi_{i=1}^m \sigma_i$ since the determinant is invariant under rotation and U and V are unitary matrices.
– Interpretation of the condition number: Ratio of the longest axis of the (hyper)ellipse over the shortest axis of the (hyper)ellipse of the image of S under the mapping of A, i.e. how much A 'stretches' the unit sphere S.
– Interpretation of the determinant: change in the area/volume (factor) of S after mapping by A.

(d) i. The rank is min (k, number of nonzero singuar values), k is an upper bound on the rank.

ii.

$$
\begin{aligned}
A - A_k &= \sum_{i=1}^{n} \sigma_i u_i v_i^H - \sum_{i=1}^{k} \sigma_i u_i v_i^H \\
&= \sum_{i=k+1}^{n} \sigma_i u_i v_i^H
\end{aligned}
$$

The Frobenius norm is given by $\|A\|_F^2 = \text{trace}(AA^H)$. Straightforward to show that $\|A\|_F^2 = \sqrt{\sum \sigma_i^2}$, hence

$$
\|A - A_k\|_F = \left\| \sum_{i=k+1}^{n} \sigma_i u_i v_i^H \right\|_F = \left(\sum_{i=k+1}^{n} \sigma_i^2 \right)^{1/2}.
$$

iii. Need to show that $\|A - A_k\|_F \leq \|A - B_k\|_F$ for all rank k matrices B_k. Setting $X = A - B_k$ and $Y = B_k$, and $j = k + 1$, we have $\sigma_i(A - B_k) \geq \sigma_{k+i}(A)$ since $\sigma_{j>k}(B_k) = 0$, leading to

$$\|A - B_k\|_F^2 \geq \sum_{i=1}^{n-k} \sigma_i^2(A - B_k) \geq \sum_{i=1}^{n-k} \sigma_{i+k}^2(A) = \|A - A_k\|_F^2.$$

Hence

$$\|A - A_k\|_F \leq \|A - B_k\|_F.$$

Optimisation

For $x \in \mathbb{R}$ with probability density function $p(x)$ and function $f(x)$ the integral

$$\mathcal{I} = \int_{-\infty}^{\infty} f(x)p(x)dx$$

can be approximated by a general Monte Carlo estimate

$$\hat{\mathcal{I}} = \sum_{n=1}^{N} w_n f(x_n)$$

where each x_n is a random sample from $p(x)$. This can be written in the vector form

$$\hat{\mathcal{I}} = \mathbf{w}^\mathsf{T}\mathbf{f}$$

by defining the N-dimensional column vectors $\mathbf{f} = [f(x_1), f(x_2), \cdots, f(x_N)]^\mathsf{T} \in \mathbb{R}^N$ and $\mathbf{w} = [w_1, w_2, \cdots, w_N]^\mathsf{T}$ such that $\sum_{n=1}^{N} w_n = \mathbf{1}^\mathsf{T}\mathbf{w} = 1$, where the N-dimensional vector of ones is $\mathbf{1} = [1, 1, \cdots, 1]^\mathsf{T}$.

1. Noting that the variance of $\hat{\mathcal{I}}$ can be written as $\mathbf{w}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{w}$ where $\boldsymbol{\Sigma}$ is the covariance matrix of the vector $\mathbf{f}$, derive the Lagrangian, $\mathcal{L}(\mathbf{w}, \lambda)$, corresponding to the constrained optimisation problem to find the $\mathbf{w}$ that minimises the variance of $\hat{\mathcal{I}}$. [10%]

   The Lagrangian is $\mathcal{L}(\mathbf{w}, \lambda) = \mathbf{w}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{w} + \lambda(1 - \mathbf{1}^\mathsf{T}\mathbf{w})$ 10 marks

2. Denoting the N-dimensional column vector of zeros as $\mathbf{0}$, rewrite the two systems of equations

$$\begin{aligned} \frac{\partial}{\partial \mathbf{w}}\mathcal{L}(\mathbf{w}, \lambda) &= \mathbf{0} \\ \mathbf{1}^\mathsf{T}\mathbf{w} &= 1 \end{aligned}$$

   in matrix form, $\mathbf{A}\mathbf{c} = \mathbf{d}$, explicitly stating the form of matrix $\mathbf{A}$ when the two $N+1$ column vectors are given as $\mathbf{c} = [\mathbf{w}^\mathsf{T}, \lambda]^\mathsf{T}$ and $\mathbf{d} = [\mathbf{0}^\mathsf{T}, 1]^\mathsf{T}$. [20%]

   The system is represented in partitioned matrix form as given in the question with the required matrix

$$\mathbf{A} = \begin{pmatrix} 2\boldsymbol{\Sigma} & \mathbf{1} \\ \mathbf{1}^\mathsf{T} & \mathbf{0} \end{pmatrix}$$

   15 marks for the matrix A and 5 marks for the whole system

$$\begin{pmatrix} 2\boldsymbol{\Sigma} & \mathbf{1} \\ \mathbf{1}^\mathsf{T} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{w} \\ \lambda \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix}$$

3. Noting that the matrix $\mathbf{A}$ and its inverse can be written in partitioned form such that

$$\mathbf{A}^{-1} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{C}_1^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{C}_2^{-1} \\ -\mathbf{C}_2^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{C}_2^{-1} \end{pmatrix}$$

   where $\mathbf{C}_1 = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}$ and $\mathbf{C}_2 = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$, solve the system for $\mathbf{c} = [\mathbf{w}^\mathsf{T}, \lambda]^\mathsf{T}$ and state the conditions for which this solution is unique. [40%]

Require to obtain explicit form of inverse for matrix $\mathbf{A}$

$$\mathbf{A}^{-1} = \begin{pmatrix} \frac{1}{2}\mathbf{\Sigma}^{-1} & \frac{\mathbf{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\mathsf{T}\mathbf{\Sigma}^{-1}\mathbf{1}} \\ \frac{\mathbf{1}^\mathsf{T}\mathbf{\Sigma}^{-1}}{\mathbf{1}^\mathsf{T}\mathbf{\Sigma}^{-1}\mathbf{1}} & -\frac{2}{\mathbf{1}^\mathsf{T}\mathbf{\Sigma}^{-1}\mathbf{1}} \end{pmatrix} 10 marks$$

and then solving for both $\mathbf{w} = \frac{\mathbf{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\mathsf{T}\mathbf{\Sigma}^{-1}\mathbf{1}}$ and $\lambda = -\frac{2}{\mathbf{1}^\mathsf{T}\mathbf{\Sigma}^{-1}\mathbf{1}}$ 10 marks for the solution to be unique then the matrix $\mathbf{A}$ must be invertible and as such $\mathbf{\Sigma}$ must be invertible - which by definition it will be 20 marks

4. Show that the variance minimising Monte Carlo estimator $\hat{\mathcal{I}}^*$ for $\mathcal{I}$ takes the form

$$\hat{\mathcal{I}}^* = \frac{\mathbf{1}^\mathsf{T}\mathbf{\Sigma}^{-1}\mathbf{f}}{\mathbf{1}^\mathsf{T}\mathbf{\Sigma}^{-1}\mathbf{1}}$$

and the corresponding minimal variance is equal to

$$\frac{1}{\mathbf{1}^\mathsf{T}\mathbf{\Sigma}^{-1}\mathbf{1}}$$

[20%]

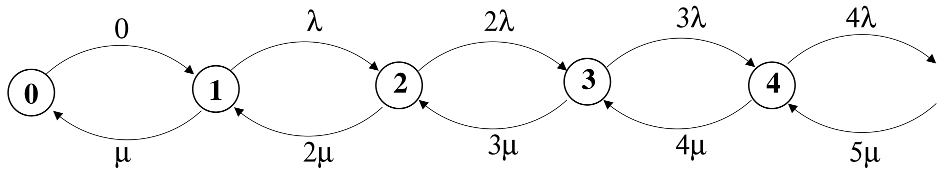
This first part requires the solution for $\mathbf{w}$ from the previous question and taking the inner product with $\mathbf{f}$ - 10 marks The second part requires using the solution for $\mathbf{w}$ and inserting it into the expression for the variance of the estimator - 10 marks

5. If the x_n are sampled independently and identically from $p(x)$ the covariance matrix for $\mathbf{f}$ takes the form $\sigma_f^2\mathbf{I}$, where $\mathbf{I}$ is an $N \times N$ dimensional identity matrix, show that the minimum variance estimator is of the form $\hat{\mathcal{I}}^* = \frac{1}{N}\sum_{n=1}^{N} f(x_n)$. [10%]

This requires evaluation the optimal $\mathbf{w}$ with the matrix $\sigma_f^2\mathbf{I}$ which leads to each w_n being equal to $\frac{1}{N}$ - 10 marks

1. *Birth Death Processes*

(a) The state-space and transition rates are shown below.



[10%]

(b)(i) In time Δt for the probability of number of bacteria is 2 the following operations can occur in terms of the changes in the probability mass

- births and deaths from the current state yields: $-2(\lambda+\mu)\Delta t\pi_2(t)$
- a "death" from three customers yields: $3\mu\Delta t\pi_3(t)$
- a "birth" from one customer yields: $\lambda\Delta t\pi_1(t)$

The assumptions behind this are that the probability of multiple events occurring is very small (ignored) as the size of Δt is very small.

Combining all of these together the probability mass associated with state 2

$$\pi_2(t+\Delta t) = (1-2\lambda\Delta t - 2\mu\Delta t)\pi_2(t) + \lambda\Delta t\pi_1(t) + 3\mu\Delta t\pi_3(t)$$

[15%]

(b)(ii) Taking the limit as $\Delta t \to 0$ yields the following form for $\mathbf{Q}$

$$\mathbf{Q} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0\dots \\ \mu & -\lambda-\mu & \lambda & 0 & 0 & 0\dots \\ 0 & 2\mu & -2\lambda-2\mu & 2\lambda & 0 & 0\dots \\ 0 & 0 & 3\mu & -3\lambda-3\mu & 3\lambda & 0\dots \end{bmatrix}$$

[15%]

(c)(i) Setting $t=0$ in the expression yields $\alpha(0)=0, \beta(0)=0$. This yields

$$\pi_0(0) = 0, \pi_i(0) = 0^{i-1}$$

which is only non-zero when $i=1$, consistent with the starting conditions. [10%]

(c)(ii) For extinction require $\alpha(\infty)=1$. Thus

$$\pi_0(\infty) = \frac{\mu\left(\exp((\lambda-\mu)\infty)-1\right)}{\lambda\exp((\lambda-\mu)\infty)-\mu}$$

Thius is split into three conditions

- $\mu > \lambda$: $\alpha(\infty) = -\mu/-\mu = 1$ - satisfied
- $\mu = \lambda$: Need to get a new expression for α

$$\alpha(t) = \frac{\mu t}{1+\lambda t} = \frac{\lambda t}{1+\lambda t}$$

as $t\to\infty$, $\alpha(\infty)=1$ - satisfied

- $\mu < \lambda$: $\alpha(\infty) = \mu/\lambda$ - not satisfied

[25%]

(c)(iii) This is simple to obtain from (c)(ii) (if complete). Probability given by $\min(1, \mu/\lambda)$ [10%]

(c)(iv) Each of the initial N bacteria can be treated seperately, the overall number of bacteria will be the sum of the N independent processes. Thus for extinction ll indpendent processes must become extinct so probability given by $(\min(1, \mu/\lambda))^N = \min(1, (\mu/\lambda)^N)$

[15%]