Solutions to 3M1, 2018

1. *Linear Algebra*

   See separate handwritten pages.

2. *Discrete Markov Processes*

   (a)(i) Each row of $\mathbf{P}$ is the probability of transition from a state to all other states. As the process must be in one of the states at the next time instance, each row must add to one. [10%]

   (a)(ii) Writing out

$$\mathbf{P}-\mathbf{I}=\begin{bmatrix} p_{11}-1 & p_{12} & \dots & p_{1N} \\ p_{21} & p_{22}-1 & \dots & p_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ p_{N1} & p_{N2} & \dots & p_{NN}-1 \end{bmatrix}$$

   Summing the first $N-1$ columns yields the (negative) the last column. Thus the columns of $\mathbf{P}-\mathbf{I}$ are not independent, so $\det(\mathbf{P}-\mathbf{I})=0$. This means that an eigenvalue of $\mathbf{P}$ must be 1. [20%]

   (a)(iii) For a stationary distribution need to satisfy

$$\boldsymbol{\pi}\mathbf{P}=\boldsymbol{\pi}$$

   This will be the normalised left eigenvector of the largest eignvalue. Thus

$$\boldsymbol{\pi}^{(\infty)}\mathbf{P}=\lambda\boldsymbol{\pi}^{(\infty)}$$

   where λ is the eigenvalue. If the magnitude is greater than 1, then the resulting distribution will not be a valid distribution. [15%]

   (b)(i) Transition matrices are

$$\mathbf{P}=\begin{bmatrix} 0 & 0.5 & 0.5 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix};\quad \mathbf{A}=\frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

   [15%]

   (b)(ii) To find the stationary need to solve

$$\boldsymbol{\pi}\tilde{\mathbf{P}}=\boldsymbol{\pi}$$

   For $d=0$ it is simple to show this yields the following equations

$$\begin{aligned} \pi_3 &= \pi_1 \\ 0.5\pi_1 &= \pi_2 \\ 0.5\pi_1+\pi_2 &= \pi_3 \end{aligned}$$

   Solving these yields

$$\boldsymbol{\pi}=\begin{bmatrix} 0.4 & 0.2 & 0.4 \end{bmatrix}$$

   For $d=1$ it is simple to see that all elements of the stationary distribution must be the same

$$\boldsymbol{\pi}=\frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$$

   [25%]

   (b)(iii) As N gets large it is not practical to find the eigenvalues/vectors for the matrix $\tilde{\mathbf{P}}$. However the standard approach for finding largest eigenvalues can be applied. The process is:

(a) initialise distribution

$$\boldsymbol{\pi}^{(0)} = \frac{1}{N}\left[\begin{array}{cccc} 1 & 1 & \dots & 1 \end{array}\right]$$

(b) update distribution

$$\boldsymbol{\pi}^{(i+1)} = (1-d)\boldsymbol{\pi}^{(i)}\mathbf{P} + d\boldsymbol{\pi}^{(0)}$$

(c) repeat until converged

As the matrix $\mathbf{P}$ should be sparse this can be more efficient than simply cost N^2 from a matrix vector multiplication. [15%]

3. *Optimisation: gradient descent* The gradient of the function is given by

$$\nabla V = \begin{pmatrix} 8x - 6y + 1 \\ -6x + 8y - 2 \end{pmatrix},$$

At the origin, the gradient direction is $(1, -2)$. So the function value $f(\alpha)$ along the gradient with step α starting at the origin is given by

$$f(\alpha) = \alpha^2 + 4\alpha^2 + 3(3\alpha)^2 + \alpha + \alpha = 32\alpha^2 + 5\alpha$$

which takes on its minimum value at $\alpha = -5/64$, so the updated location is $(-5, 10)/64$, with a function value ≈ -0.19.

Subsequent steps of Steepest Descent would get closer and closer to the minimum. Since the function is a quadratic form, one more Conjugate Gradient step would get to the exact minimum (assuming no rounding and truncation errors).

The Taylor expansion of a function in two dimensions, $f(z)$ with $z \equiv (x, y)$, is

$$f(z) = f(0) + z^T\nabla f(0) + \frac{1}{2}z^T\nabla^2 f(0)z + \dots$$

Truncating at second order and setting the gradient of this to zero gives

$$\begin{aligned} 0 &= \nabla f(0) + \nabla^2 f(0)z \\ z &= -[\nabla^2 f(0)]^{-1}\nabla f(0) \end{aligned}$$

So the Newton update rule is

$$z_{k+1} = z_k - [\nabla^2 f(z_k)]^{-1}\nabla f(z_k)$$

The second derivative (Hessian) matrix is given by

$$H = \begin{pmatrix} 8 & -6 \\ -6 & 8 \end{pmatrix},$$

its determinant is 28, and its inverse is

$$H^{-1} = \frac{1}{14}\begin{pmatrix} 4 & 3 \\ 3 & 4 \end{pmatrix},$$

Therefore the result of a Newton step starting from origin is

$$-\frac{1}{14}\begin{pmatrix} 4 & 3 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 1 \\ -2 \end{pmatrix} = \frac{1}{14}\begin{pmatrix} 2 \\ 5 \end{pmatrix}$$

This gives a stationary point of the function, because it gives zero gradient when substituted into the above formula for ∇V.

The second order optimality conditions are that the the eigenvalues of the Hessian need to be positive for a stationary point to be a minimum. The eigenvalues of the Hessian are 2 and 14, so the stationary point we found is indeed a minimum. The corresponding eigenvectors are $(1, 1)$ and $(1, -1)$.

4. The objective is to maximise $2x_1 + x_2$, subject to the following constraints,

$$
\begin{aligned}
&2x_1 \geq x_2 \\
&2x_1 + 3x_2 \leq 16 \\
&x_1 \leq x_2 + 3 \\
&x_1, x_2 \geq 0
\end{aligned}
$$

We introduce slack variables s_1, s_2, s_3, one for each of the constraints, to obtain the following equality constraints, in addition to all variables being positive,

$$
\begin{aligned}
-2x_1 + x_2 + s_1 &= 0 \\
2x_1 + 3x_2 + s_2 &= 16 \\
x_1 - x_2 + s_3 &= 3 \\
x_1, x_2, s_1, s_2, s_3 &\geq 0
\end{aligned}
$$

The corresponding tableau is canonical,

$$
\begin{pmatrix}
1 & -2 & -1 & 0 & 0 & 0 & 0 \\
0 & -2 & 1 & 1 & 0 & 0 & 0 \\
0 & 2 & 3 & 0 & 1 & 0 & 16 \\
0 & 1 & -1 & 0 & 0 & 1 & 3
\end{pmatrix}
$$

Pivoting on column 2 (x_1) and row 4 (s_3), we get

$$
\begin{pmatrix}
1 & 0 & -3 & 0 & 0 & 2 & 6 \\
0 & 0 & -1 & 1 & 0 & 2 & 6 \\
0 & 0 & 5 & 0 & 1 & -2 & 10 \\
0 & 1 & -1 & 0 & 0 & 1 & 3
\end{pmatrix}
$$

Now pivoting on column 3 (x_2) and row 3 (s_2), we get

$$
\begin{pmatrix}
5 & 0 & 0 & 0 & 3 & 4 & 60 \\
0 & 0 & 0 & 5 & 0 & 8 & 40 \\
0 & 0 & 5 & 0 & 1 & -2 & 10 \\
0 & 5 & 0 & 0 & 1 & 3 & 25
\end{pmatrix}
$$

This is now the optimal solution, giving 60/5=12 as the optimal solution for $x_1 = 5$ and $x_2 = 2$.

1. (a) (i) If $A = A^T$, then $U = V$.

Since eigvecs of A^TA and A^TA are the same, and the same as A, problem reduces to a regular eigen decomposition / diagonalisation

(ii) Errors can become large when $K(A)$ is large.

In l_2, $K_2 = \frac{\sqrt{\lambda_{max}(AA^T)}}{\sqrt{\lambda_{min}(AA^T)}} = \frac{\sigma_{max}}{\sigma_{min}} = \frac{\sigma_1}{\sigma_k}$

$k = \min(m, n)$

(b) l_2 norm will lead to linear problem in λ^*:

$$\lambda^* = \min \| Ax - \lambda^* x \|_2$$

$$\begin{aligned} e &= \| Ax - \lambda^* x \|_2^2 \\ &= (Ax - \lambda^* x)^T (Ax - \lambda^* x) \\ &= x^T A^T A x - \lambda^* x^T A x - \lambda^* x^T A^T x + \lambda^{*2} x^T x \end{aligned}$$

Differentiate e w.r.t. λ^* and set equal to zero,

$$\frac{de}{d\lambda^*} = -2x^T A x + 2\lambda^* x^T x = 0$$

$$\Rightarrow \lambda^* = \frac{x^T A x}{x^T x}$$

Using l_2 norm is convenient because error expression is quadratic in λ^*, hence optimality condition will be linear.

(c) $x_{k+1} = N^{-1} P x_k + N^{-1} b \qquad (Ax = b)$

$$e_k = x - x_k$$

$$e_{k+1} = x - x_{k+1} = x - N^{-1} P x_k + N^{-1} b$$

$$e_{k+1} = N^{-1}(Nx - Px_k \mp b)$$

$$b = Ax = (N-P)x$$

$$= N^{-1}(Nx - Px_k + -A(N-P)x)$$

$$= N^{-1}Pe_k$$

$$e_{k+1} = N^{-1}Pe_k = (N^{-1}P)^{k+1}\underline{e}_0$$

Require that $\|e_{k+1}\| < \|e_k\|$

Expand $\underline{e}_0$ using eigenvectors of $N^{-1}P$

$$\underline{e}_0 = \sum_{i=1}^{n} c_i \underline{u}_i$$

$$\underline{e}_k \cong \sum c_i \lambda_i^k \underline{u}_i$$

$\Rightarrow$ require that $|\lambda_i|_{max} < 1$

In terms of norms, $\|M\|_2 = \|N^{-1}P\|_2 < 1$