3M1 Mathematical Methods, 2022

Linear Algebra

(a) A standard method to solve this question consists in finding the eigenvalues and show that they are all strictly positive.

The eigenvalues are $3 \pm \sqrt{4 + v^2}$. One of them is always strictly positive. The smallest is strictly positive if $4 + v^2 < 9$.

$$\implies |v| < \sqrt{5}$$

(b) (i) The relationship $\mathbf{Ax} = \mathbf{b}$ can be rewritten as $\mathbf{x} = \mathbf{N}^{-1}\mathbf{Px} + \mathbf{N}^{-1}\mathbf{b}$. This assumes in particular that $\mathbf{N}$ has an inverse.

For a suitable choice of $\mathbf{N}$ and $\mathbf{P}$, it is possible to calculate the solution iteratively using:

$$\mathbf{x}_{k+1} = \mathbf{N}^{-1}\mathbf{P}\mathbf{x}_k + \mathbf{N}^{-1}\mathbf{b}$$

The process can be initiated with an arbitrary value for $\mathbf{x}_0$. If the method converges, $\mathbf{x}_n$ will tend towards the solution $\mathbf{x}$.

(ii) We can consider here the evolution of the error from one iteration to the next, i.e. the difference between the current estimate of the solution and the true solution $\mathbf{x}$: $\mathbf{e}_k = \mathbf{x}_k - \mathbf{x}$. We know that:

$$\mathbf{x} = \mathbf{N}^{-1}\mathbf{Px} + \mathbf{N}^{-1}\mathbf{b}$$

$$\mathbf{x}_{k+1} = \mathbf{N}^{-1}\mathbf{P}\mathbf{x}_k + \mathbf{N}^{-1}\mathbf{b}$$

Hence:

$$\mathbf{x}_{k+1} - \mathbf{x} = \mathbf{N}^{-1}\mathbf{P}\mathbf{x}_k - \mathbf{N}^{-1}\mathbf{Px}$$

$$\implies \mathbf{e}_{k+1} = \mathbf{N}^{-1}\mathbf{P}\mathbf{e}_k$$

In order to converge to the solution, we must have $\mathbf{e}_k$ tend towards 0. This is achieved if all eigenvalues of $\mathbf{N}^{-1}\mathbf{P}$ are strictly less than 1 in absolute value, i.e. that the spectral radius of $\mathbf{N}^{-1}\mathbf{P}$ is less than one.

(c) (i) In this case, $\mathbf{N} = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$ and $\mathbf{P} = \begin{bmatrix} 0 & -v \\ -v & 0 \end{bmatrix}$.

This leads to $\mathbf{N}^{-1}\mathbf{P} = \begin{bmatrix} 0 & -v \\ -v/5 & 0 \end{bmatrix}$

The eigenvalues are $\pm v/\sqrt{(}5)$ We therefore need $|v| < \sqrt{5}$

(ii) Here, $\mathbf{N} = \mathbf{I}$ and $\mathbf{P} = \begin{bmatrix} 0 & -v \\ -v & -4 \end{bmatrix}$.

The eigenvalues of $\mathbf{N}^{-1}\mathbf{P}$ are $-2 \pm \sqrt{4 + v^2}$.

One of them is always lower than -4, so much larger than 1 in absolute value. This iteration scheme would therefore be unstable for all values of v.

Optimisation

For D-dimensional column vectors $\mathbf{x} \in \mathbb{R}^D$ and $\mathbf{b} \in \mathbb{R}^D$ consider the function of $\mathbf{x}$

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{Q}\mathbf{x} - \mathbf{x}^\mathsf{T}\mathbf{b} + \mathbf{1}^\mathsf{T}\mathbf{g}(\mathbf{x})$$

where the $D \times D$ matrix $\mathbf{Q}$ has elements $q_{i,j}$, each $i = 1, \cdots, D$ and $j = 1, \cdots, D$, the D-dimensional vector of ones is denoted as $\mathbf{1} = [1, 1, \cdots, 1]^\mathsf{T}$ and $\mathbf{g}(\mathbf{x})$ is an element-wise application of the function $g(\cdot)$ which acts on each component of the vector $\mathbf{x}$ such that $\mathbf{g}(\mathbf{x}) = [g(x_1), g(x_2), \cdots, g(x_D)]^\mathsf{T}$.

1. Derive and justify the necessary condition for a point $\mathbf{x}^* \in \mathbb{R}^D$ to be a strong local minimum of the function $f(\mathbf{x})$. [15%]

   For a point $\mathbf{x}^*$ to be a local minimum and as there are no constraints so that all directions in $\mathbb{R}^D$ are feasible then the necessary condition is that $\nabla f(\mathbf{x}^*) = 0$ which for the function should be $\mathbf{Q}\mathbf{x}^* - \mathbf{b} + \mathbf{g}'(\mathbf{x}^*) = 0$ where $\mathbf{g}'(\mathbf{x}^*)$ is a component wise defined vector with elements $g'(.)$.

2. Derive and justify the sufficient condition for a point $\mathbf{x}^* \in \mathbb{R}^D$ to be a strong local minimum of the function $f(\mathbf{x})$. [15%]

   For an arbitrary vector $\mathbf{d}$ then $\mathbf{d}^\mathsf{T}(\mathbf{Q} + g''(\mathbf{x}^*))\mathbf{d} > 0$ must be satisfied for $\mathbf{x}^*$ to be a strong local minimum. Where the non-zero elements of the diagonal matrix $g''(\mathbf{x}^*)$ are the second derivatives $g''(x_i^*)$.

3. Derive a steepest descent method for the D-dimensional function $f(\mathbf{x})$ and provide an expression for the step-size for each iteration based on a second-order Taylor expansion of $f(\mathbf{x})$. [40%]

   The steepest descent method requires an iteration such that $\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \nabla f(\mathbf{x}_k)$ where $\nabla f(\mathbf{x}_k) = \mathbf{Q}\mathbf{x}_k - \mathbf{b} + \mathbf{g}'(\mathbf{x}_k)$. Based on the Taylor-expansion of the function and noting that $\mathbf{x} - \mathbf{x}_k = -\alpha_k \nabla f(\mathbf{x}_k)$ then it follows that

   $$\alpha_k = -\frac{\nabla f(\mathbf{x}_k)^\mathsf{T} \nabla f(\mathbf{x}_k)}{\nabla f(\mathbf{x}_k)^\mathsf{T} \mathbf{H}_k \nabla f(\mathbf{x}_k)}$$

   with $\mathbf{H}_k = \mathbf{Q} + g^{''}(\mathbf{x}_k)$ where $g^{''}(\mathbf{x}_k)$ is a diagonal matrix.

4. For the specific case where $D = 2$, $\mathbf{Q}$ is given as an identity matrix, and the nonlinear term is the exponential function, i.e. $g(\cdot) = \exp(\cdot)$, assess whether $f(\mathbf{x})$ is convex in $\mathbb{R}^2$ and state the implication on the nature of the point $\mathbf{x}^*$. [30%]

   The Hessian is $\mathbf{I} + g^{''}(\mathbf{x}_k)$ which yields a determinant $(1 + g^{''}(x_1))(1 + g^{''}(x_2))$ that has to be strictly greater than zero for all $\mathbf{x}$ for it to be convex. It is clear that when the exponential function is used for $g(\cdot)$ then this will be the case, ie. the strict positivity will hold due to the positive value of the exponential function for all values of argument. This therefore implies that the minimum is in fact a unique global minimum.

3M1 Mathematical Methods, 2022

1. *Importance Sampling*

(a)(i) This is a standard Monte-Carlo approximation so

$$\hat{\mu}_N = \frac{1}{N}\sum_{i=1}^{N} h(\boldsymbol{x}^{(i)})$$

[10%]

(a)(ii) Initially compute the expected value of the integral estimate from a single sample drawn from $p(\boldsymbol{x})$. The expected value of the estimate is unbiased so

$$\mu = \mathbb{E}\left\{\hat{\mu}_N\right\}$$

for all values of N. Now computing the variance

$$\text{var}_p(\hat{\mu}_1) = \mathbb{E}\left\{(h(\boldsymbol{x}) - \mu)^2\right\} = \sigma^2$$

from the question. The variance from N independent estimators is then given by

$$\text{var}_p(\hat{\mu}_N) = \frac{1}{N^2}\sum_{i=1}^{N} \text{var}_p(\hat{\mu}_1) = \frac{\sigma^2}{N}$$

The variance does not depend on the dimensionality of $\boldsymbol{x}$, just the number of samples N and the variance σ^2. [30%]

(b)(i) Re-expressing the original expectation

$$\mu = \int h(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x} = \int h(\boldsymbol{x})\left(\frac{p(\boldsymbol{x})}{q(\boldsymbol{x})}\right)q(\boldsymbol{x})d\boldsymbol{x}$$

As samples are now being drawn from $q(\boldsymbol{x})$ the Monte-Carlo approximation becomes

$$\tilde{\mu}_N = \frac{1}{N}\sum_{i=1}^{N} h(\tilde{\boldsymbol{x}}^{(i)})\left(\frac{p(\tilde{\boldsymbol{x}}^{(i)})}{q(\tilde{\boldsymbol{x}}^{(i)})}\right)$$

Thus

$$w^{(i)} = \frac{p(\tilde{\boldsymbol{x}}^{(i)})}{q(\tilde{\boldsymbol{x}}^{(i)})}$$

[20%]

(b)(ii) For this expression to converge to μ it is necessary for $q(\boldsymbol{x})$ to be non-zero wherever $p(\boldsymbol{x})$ is non-zero. Otherwise there will be a region of the space that cannot contribute to the calculation of μ. [10%]

(b)(iii) The same process is used to compute the variance. Again the expected value of will be μ.

$$\begin{aligned}\text{var}_q(\tilde{\mu}_1) &= \int h(\boldsymbol{x})^2 \left(\frac{p(\boldsymbol{x})}{q(\boldsymbol{x})}\right)^2 q(\boldsymbol{x})d\boldsymbol{x} - \mu^2 \\ &= \int h(\boldsymbol{x})^2 \left(\frac{p(\boldsymbol{x})}{q(\boldsymbol{x})}\right) p(\boldsymbol{x})d\boldsymbol{x} - \mu^2\end{aligned}$$

Thus

$$\text{var}_p(\hat{\mu}_N) - \text{var}_q(\tilde{\mu}_N) = \frac{1}{N}\int h(\boldsymbol{x})^2 \left(1 - \frac{p(\boldsymbol{x})}{q(\boldsymbol{x})}\right) p(\boldsymbol{x})d\boldsymbol{x}$$

What we would like to do is to make this value as large as possible. When $q(\boldsymbol{x}) = p(\boldsymbol{x})$ the differences is zero. However it is possible to make this value positive by setting $p(\boldsymbol{x})/q(\boldsymbol{x})$ to be small when $h(\boldsymbol{x})^2 p(\boldsymbol{x})$ is large. The ideal (usually non-practical) form is to set

$$q(\boldsymbol{x}) = \frac{1}{\mu} p(\boldsymbol{x}) f(\boldsymbol{x})$$

This maximises the differences. [30%]