Chapter 1

Suggested Solutions

1.1 Overview of Linear Algebra

1. Suppose that $T(\boldsymbol{x}) = A\boldsymbol{x}$ and that $F(y) = B\boldsymbol{y}$, with $A_{m\times n}$ and $B_{k\times m}$.

   (a) Show that $G = F(T(\boldsymbol{x}))$ is also a linear map.

   Solution. $G = F(T(\boldsymbol{x})) = F(A\boldsymbol{x}) = B(A\boldsymbol{x}) = BA\boldsymbol{x} = C\boldsymbol{x}$, where C is a $k \times n$ matrix. Using Lemma **??** because G can be expressed as the multiplication of a constant matrix times a vector, it is a linear map. □

   (b) Show that $||G|| \leq ||F||\ ||T||$. Is the composite of two linear maps continuous?

   Solution. Let $\boldsymbol{x} \in \mathbb{R}^n$. Using the operator norm inequality twice.

$$\begin{aligned} ||F(T(\boldsymbol{x}))|| &\leq ||F||\ ||T(\boldsymbol{x})|| \\ &\leq ||F||\ ||T||\ ||\boldsymbol{x}|| \end{aligned}$$

   Restrict attention to vectors of unit length such that $||\tilde{\boldsymbol{x}}|| = 1$, then $||G(\tilde{\boldsymbol{x}})|| \leq ||F||\ ||T||$. Then right hand side does not depend on the input vector. Therefore we can take the supremum on the left-hand side to show that $||G|| \leq ||F||\ ||T||$. To prove continuity it suffices to use the fact that G is a linear map by using Theorem **??**. □

   (c) Assume that P is a square matrix. Use part (b) to show that for any non-negative integer t, $||P^t|| \leq ||P||^t$.

   Solution. If $t = 1$ then $||P^t|| = ||P||$. Suppose that the statement holds for some $t \geq 1$, then using the previous lemma $||P^{t+1}|| = ||P^t P|| \leq ||P^t||\ ||P||$, where $||P^t|| \leq ||P||^t$ by the induction hypothesis. Therefore, $||P^{t+1}|| \leq ||P||^{t+1}$. By the principle of induction, the result is proved. □

   (d) Show that there exists a number $a > 0$ such that whenever $\boldsymbol{x}$ is a probability vector, it follows that $||\boldsymbol{x}|| \geq a > 0$.

Solution. First, we show that $||\boldsymbol{x}|| > 0$ if $\boldsymbol{x}$ is a probability vector. We proceed using proof by contradiction. Since $||\boldsymbol{x}|| \geq 0$, assume (by contradiction) that $||\boldsymbol{x}|| = 0$ then $\boldsymbol{x} = \mathbf{0}_n$, meaning all its entries are zero. However, the entries of a probability vector must add up to one. Thus we must have $||\boldsymbol{x}|| > 0$. But this is not sufficient for the statement of interest–we want to show $||\boldsymbol{x}||$ is bounded below by a positive constant uniformly across all probability vectors $\boldsymbol{x}$.

Let $\mathcal{P}$ denote the space of probability vectors. Now we show it is compact. Consider a sequence of probability vectors $x_k = (x_{1k}, \ldots, x_{nk})$ with $\sum_i x_{ik} = 1$, $x_k \geq 0$ for all k, and $x_k \to x$ as $k \to \infty$. Since limits preserve equalities and weak inequalities, it follows that $\sum_i x_i = 1$ and $x_i \geq 0$, therefore the limit is still a probability vector and hence the set $\mathcal{P}$ is closed. It is also bounded because all its entries are non-negative and less that or equal to one. Therefore, it is compact.

The function $f(\boldsymbol{x}) = \sqrt{\boldsymbol{x}^t\boldsymbol{x}}$ is continuous because it is a polynomial of the entries of the vector. The extreme-value theorem states that if a function is continuous and the space is compact then it has a maximum and a minimum. Therefore, a minimum exists, and denote $\inf_{\boldsymbol{x}\in\mathcal{P}} \sqrt{\boldsymbol{x}^t\boldsymbol{x}} = \sqrt{\boldsymbol{x}_*^t\boldsymbol{x}_*}$ for some $\boldsymbol{x}_* \in \mathcal{P}$. By our previous result $||\boldsymbol{x}_*|| > 0$ and therefore $||\boldsymbol{x}||$ is bounded away from zero; that is, $||\boldsymbol{x}|| \geq ||\boldsymbol{x}_*|| > 0$ for all $\boldsymbol{x} \in \mathcal{P}$. Letting $a = ||\boldsymbol{x}^*||$ proves the statement.

In fact, $1/\sqrt{n} \leq ||\boldsymbol{x}|| \leq 1$, where n is the dimension of $||\boldsymbol{x}||$. The shortest probability vector has the value $1/n$ as each component of the vector, while the longest probability vector has the value 1 in a single component and 0 in all others. This constitutes an easier proof for the statement:

$$\begin{aligned}
||\boldsymbol{x}|| &= \sqrt{x_1^2 + x_2^2 + ... + x_n^2} = \sqrt{n}\sqrt{\frac{x_1^2 + x_2^2 + ... + x_n^2}{n}} \\
&\geq \sqrt{n}\frac{x_1 + x_2 + ... + x_n}{n} = 1/\sqrt{n},
\end{aligned}$$

where the inequality is due to the AM–GM inequality (the inequality of arithmetic and geometric means) and every element of a probability vector is non-negative with a sum of 1.

□

(e) If P is a stochastic matrix, could it be $||P|| < 1$? What would this imply for our migration example if it were true?

Solution. Let $\mathcal{P}$ denote the space of probability vectors. Then if $\boldsymbol{X} \in \mathcal{P}$ then $P\boldsymbol{x} \in \mathcal{P}$. By applying this argument recursively we know that $P^t\boldsymbol{x} \in \mathcal{P}$. Furthermore, using part (d), $||P^t\boldsymbol{x}|| \geq ||x_*|| \geq a$ where a is a positive constant. Furthermore, $||\boldsymbol{x}|| \leq 1$ because all the entries add-up to one and are non-negative.

Using the operator norm inequality if $||P|| < 1$ then $||P^t\boldsymbol{x}|| \leq ||P||^t \; ||\boldsymbol{x}|| \to 0$. However, this contradicts the fact that $||P^t\boldsymbol{x}|| \geq a > 0$ for all integer t and probability vector $||\boldsymbol{x}|| \in \mathcal{P}$. Therefore, it cannot be that $||P|| < 1$. In our migration example, an implication of $||P|| < 1$ would be that some people go to other states apart from 1 and 2, i.e., the size of population in city 1 and 2 is shrinking (which is why the norm converges to zero). However, the population is not shrinking, but just changing location.

□

2. In this section you will expand some of the details of the proof of the Cauchy-Schwarz inequality. Let $\lambda \in \mathbb{R}, \boldsymbol{v}, \boldsymbol{x} \in \mathbb{R}^n$. We know that if $\boldsymbol{z} = \boldsymbol{v} - \lambda\boldsymbol{x}$, $||\boldsymbol{z}|| \geq 0$, then

$$\forall\lambda \in \mathbb{R},\ \boldsymbol{v}^t\boldsymbol{v} - 2\lambda\boldsymbol{v}^t\boldsymbol{x} + \lambda^2\boldsymbol{x}^{\boldsymbol{t}}\boldsymbol{x} \geq 0 \tag{1.1}$$

(a) Show that for all $\boldsymbol{v}, \boldsymbol{x} \in \mathbb{R}^n$, the condition in Equation 1.1 is equivalent to:

$$\inf_{\lambda\in\mathbb{R}}\{\ \boldsymbol{v}^t\boldsymbol{v} - 2\lambda\boldsymbol{v}^t\boldsymbol{x} + \lambda^2\boldsymbol{x}^{\boldsymbol{t}}\boldsymbol{x}\ \} \geq 0$$

Solution. ($\Longrightarrow$) Taking the infimum to the left hand side of Equation 1.1 implies the infimum inequality.

($\Longleftarrow$) Suppose (by contradiction) that there exists a $\lambda \in \mathbb{R}$ such that $\boldsymbol{v}^t\boldsymbol{v} - 2\lambda\boldsymbol{v}^t\boldsymbol{x} + \lambda^2\boldsymbol{x}^{\boldsymbol{t}}\boldsymbol{x} < 0$ for some λ. Then that means that this λ produces a value strictly lower than the infimum, a contradiction. $\square$

(b) Consider the case when $||\boldsymbol{x}|| > 0$. Use the fact that the function is quadratic in λ to show that a minimum exists and that is

$$\frac{\boldsymbol{v}^t\boldsymbol{x}}{\boldsymbol{x}^t\boldsymbol{x}} = \arg\min_{\lambda\in\mathbb{R}^n}\{\ \boldsymbol{v}^t\boldsymbol{v} - 2\lambda\boldsymbol{v}^t\boldsymbol{x} + \lambda^2\boldsymbol{x}^{\boldsymbol{t}}\boldsymbol{x}\ \}$$

Solution. The first order condition with respect to λ yields $-2\boldsymbol{v}^t\boldsymbol{x} + 2\boldsymbol{x}^t\boldsymbol{x}\lambda = 0$ which yields the solution. The second-order condition is $\boldsymbol{x}^t\boldsymbol{x} > 0$ hence it is indeed a minimum. $\square$

(c) Show that if $\boldsymbol{v} = \boldsymbol{x}$, then Cauchy-Schwarz attains equality.

Solution. If $\boldsymbol{v} = \boldsymbol{x}$ then $||\boldsymbol{v}^t\boldsymbol{x}|| = |\boldsymbol{v}^t\boldsymbol{v}| = ||\boldsymbol{v}||^2 = ||\boldsymbol{v}||\ ||\boldsymbol{x}||$. $\square$

1.2 Image and Kernel

1. Suppose that X is a non-zero $m \times n$ rank deficient matrix. Suppose that we partition its columns $X = [X_1, X_2]$ in such a way that $Im(X_1) = Im(X)$ and X_1 is full rank. The block matrices have n_1, n_2 columns, respectively. This is equivalent to dropping redundant variables in a linear regression.

   (a) Show that Equation **??** can be written in block-partitioned form as:

   $$\begin{bmatrix} X_1^t X_1 & X_1^t X_2 \\ X_2^t X_1 & X_2^t X_2 \end{bmatrix} \beta = \begin{bmatrix} X_1^t Y \\ X_2^t Y \end{bmatrix}$$

   Solution. The transpose of X in block-partition form is $\begin{bmatrix} X_1^t \\ X_2^t \end{bmatrix}$. That means that

   $$X^t X = \begin{bmatrix} X_1^t \\ X_2^t \end{bmatrix} \begin{bmatrix} X_1 & X_2 \end{bmatrix} = \begin{bmatrix} X_1^t X_1 & X_1^t X_2 \\ X_2^t X_1 & X_2^t X_2 \end{bmatrix}$$

   Similarly we can show that

   $$X^t Y = \begin{bmatrix} X_1^t \\ X_2^t \end{bmatrix} Y = \begin{bmatrix} X_1^t Y \\ X_2^t Y \end{bmatrix}$$

   □

   (b) Suppose that $\hat{\beta}_1 = (X_1^t X_1)^{-1}(X_1^t Y)$. Construct a vector $\beta^* = \begin{bmatrix} \hat{\beta}_1 \\ \mathbf{0}_{n_2 \times 1} \end{bmatrix}$. Show that β^* is a solution to Equation **??** if and only if $X_2^t X_1 \hat{\beta}_1 = X_2^t Y$.

   Solution. Write the system of equations in block partition form:

   $$\begin{bmatrix} X_1^t X_1 & X_1^t X_2 \\ X_2^t X_1 & X_2^t X_2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \mathbf{0}_{n_2 \times 1} \end{bmatrix} = \begin{bmatrix} X_1^t Y \\ X_2^t Y \end{bmatrix}$$

   We can expand the terms in each block.

   $$\begin{bmatrix} X_1^t X_1 \hat{\beta}_1 + X_1^t X_2 \mathbf{0}_{n_2 \times 1} \\ X_2^t X_1 \hat{\beta}_1 + X_2^t X_2 \mathbf{0}_{n_2 \times 1} \end{bmatrix} = \begin{bmatrix} X_1^t X_1 \hat{\beta}_1 \\ X_2^t X_1 \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} X_1^t Y \\ X_2^t Y \end{bmatrix}$$

   By construction $(X_1^t X_1)\hat{\beta}_1 = (X_1^t Y)$. Therefore the only condition we need to verify is $X_2^t X_1 \hat{\beta}_1 = X_2^t Y$.

   □

   (c) Verify that the columns of X_2 belong in $Im(X_1)$. Use this fact to show that $X_2^t X_1 \hat{\beta}_1 = X_2^t Y$.

   Solution. The columns in X_2 belong in $Im(X)$ which is equal to $Im(X_1)$ by assumption. Let x_{2l} denote the l^{th} column of X_2 which is contained in $Im(X_1)$. Then for $l \in \{1, \ldots, k_2\}$ there exists a vector c_l such that $x_{2l} = X_1 c_l$. We can stack this result in matrix form as $X_2 = X_1 C$. That means that $X_2^t X_1 \hat{\beta}_1 = C^t X_1^t X_1 \hat{\beta}_1$.

On the other hand, substituting the definition of the estimator,

$$X_2^t X_1 \hat{\beta}_1 = C^t X_1^t X_1 (X_1^t X_1)^{-1} X_1^t Y.$$

Some terms cancel out and the expression simplifies to $C^t X_1^t Y = (X_1 C)^t Y = X_2^t Y$. This completes the proof. $\square$

(d) Consider the data matrix,

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \quad Y = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}$$

Construct $X^t X$ and $X^t Y$. Now partition the matrix into X_1, X_2 and compute β^*. Verify that the results that you proved above are true for the following cases:

(i) Construct X_1 using columns 1 and 2.

Solution.

$$X_1 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{bmatrix} \quad X_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad X = [X_1, X_2] \quad X^t X = \begin{bmatrix} 5 & 2 & 3 \\ 2 & 2 & 0 \\ 3 & 0 & 3 \end{bmatrix}, \quad X^t Y = \begin{bmatrix} 15 \\ 3 \\ 12 \end{bmatrix}$$

$$X_1^t X_1 = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}, \quad X_1^t Y = \begin{bmatrix} 15 \\ 3 \end{bmatrix} \quad \beta^* = \begin{bmatrix} 4 \\ -2.5 \\ 0 \end{bmatrix} \quad X^t X \beta^* = \begin{bmatrix} 15 \\ 3 \\ 2 \end{bmatrix}$$

$\square$

(ii) Construct X_1 using columns 1 and 3.

Solution.

$$X_1 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \quad X_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad X = [X_1, X_2] \quad X^t X = \begin{bmatrix} 5 & 3 & 2 \\ 3 & 3 & 0 \\ 2 & 0 & 2 \end{bmatrix}, \quad X^t Y = \begin{bmatrix} 15 \\ 12 \\ 3 \end{bmatrix}$$

$$X_1^t X_1 = \begin{bmatrix} 5 & 3 \\ 3 & 3 \end{bmatrix}, \quad X_1^t Y = \begin{bmatrix} 15 \\ 2 \end{bmatrix} \quad \beta^* = \begin{bmatrix} 1.5 \\ 2.5 \\ 0 \end{bmatrix} \quad X^t X \beta^* = \begin{bmatrix} 15 \\ 3 \\ 2 \end{bmatrix}$$

$\square$

Notice that we follow the convention to write a partition such that $X = [X_1, X_2]$. In this case we select columns 1 and 3, so the matrices X_1, X_2 are different than before in (i).

(e) Is β^* the same in both exercises? How can we interpret the result?

Proof. Typically β^* does not produce the same result. This is an example where there are two mutually exclusive categorical variables and an intercept. For example, column 1 of X presents a constant term, column 2 could represent a binary indicator for whether the individual is female and column 3 could represent a binary indicator for male. The interpretation of the coefficient changes. If we drop the last column, the "reference category" is male. If we drop the second column, the "reference category" is female. However, both models have the same ability to describe the data without loss of information, because their columns have the same image. $\square$

You can use the fact that the inverse of a 2×2 matrix is given by:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad \implies A^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

1.3 Orthogonality

1. In this exercise you will prove a version of the Frisch-Waugh-Lovell Theorem (Greene, 2012) in the detrending example.

(a) Prove that $\tilde{\beta}_1 = (\boldsymbol{X}_1^t M_2 \boldsymbol{X}_1)^{-1}(\boldsymbol{X}_1^t M_2 \boldsymbol{Y})$.

Solution. $\tilde{\beta}_1 = (\hat{U}_1^t \hat{U}_1)^{-1}(\hat{U}_1^t \hat{U}_Y)$. Where $\hat{U}_1 = M_2 X_1$ and $\hat{U}_Y = M_2 Y$. Thus we can rewrite the estimator as

$$\begin{aligned}
\tilde{\beta}_1 &= ((M_2X_1)^t M_2 X_1)^{-1}((M_2X_1)^t(M_2Y) && \text{Plugging-in Expressions } \hat{U}_1 \text{ and } \hat{U}_Y. \\
&= (X_1^t M_2^t M_2 X_1)^{-1}(X_1^t M_2 M_2 Y) && \text{Distributing Transpose} \\
&= (X_1^t M_2 X_1)^{-1}(X_1^t M_2 Y) && \text{Using idempotency and symmetry of } M_2
\end{aligned}$$

$\square$

(b) Show that the system in Equation **??** can be written in block-partition form as:

$$\begin{bmatrix} X_1^t X_1 & X_1^t X_2 \\ X_2^t X_1 & X_2^t X_2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} X_1^t Y \\ X_2^t Y \end{bmatrix}$$

Solution. The transpose of X in block-partition form is $\begin{bmatrix} X_1^t \\ X_2^t \end{bmatrix}$. Therefore

$$X^t X = \begin{bmatrix} X_1^t \\ X_2^t \end{bmatrix} \begin{bmatrix} X_1 & X_2 \end{bmatrix} = \begin{bmatrix} X_1^t X_1 & X_1^t X_2 \\ X_2^t X_1 & X_2^t X_2 \end{bmatrix}$$

Similarly we can show that

$$X^tY = \begin{bmatrix} X_1^t \\ X_2^t \end{bmatrix} Y = \begin{bmatrix} X_1^tY \\ X_2^tY \end{bmatrix}$$

□

(c) Show that second row can be rewritten as $\hat{\beta}_2 = (\boldsymbol{X}_2^t\boldsymbol{X}_2)^{-1}(\boldsymbol{X}_2^t\boldsymbol{Y} - \boldsymbol{X}_2^t\boldsymbol{X}_1\hat{\beta}_1)$.

Solution. The formula for the second row is $(X_2^tX_1)\hat{\beta}_1 + (X_2^tX_2)\hat{\beta}_2 = X_2^tY$. We can solve this equation in terms of the second coefficient as $\hat{\beta}_2 = (X_2^tX_2)^{-1}(X_2^tY - X_2^tX_2\hat{\beta}_1)$. □

(d) Plug the above result into the first row of equations and show that $(\boldsymbol{X}_1^tM_2\boldsymbol{X}_1)\hat{\beta}_1 = (\boldsymbol{X}_1^tM_2\boldsymbol{Y})$. Conclude that $\hat{\beta}_1 = \tilde{\beta}_1$.

Solution. The equation in the first row is $X_1^tX_1\hat{\beta}_1 + X_1^tX_2\hat{\beta}_2 = X_1^tY$. Plugging-in the result in part (c) we get that

$$\begin{aligned} X_1^tX_1\hat{\beta}_1 + X_1^tX_2(X_2^tX_2)^{-1}X_2^tY - X_1^tX_2(X_2^tX_2)^{-1}X_2^tX_1\hat{\beta}_1 &= X_1^tY \\ X_1^tX_1\hat{\beta}_1 + X_1^tP_2Y - X_1^tP_2^tX_1\hat{\beta}_1 &= X_1^tY && \text{(Definition } P2\text{)} \\ X_1^t(I - P_2)X_1\hat{\beta}_1 &= X_1^t(I - P_2)Y && \text{(Grouping terms)} \\ X_1^tM_2X_1\hat{\beta}_1 &= X_1^tM_2Y && \text{(Definition } M_2.\text{)} \end{aligned}$$

□

2. In the detrending example:

(a) Show that $\boldsymbol{X}$ full rank implies that $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are full rank. (Hint: Prove by contradiction)

Solution. (By contradiction) suppose that X_1 or X_2 are not full rank. Suppose WLOG that it is X_2. Then by Corollary **??** we can write one of the columns as a linear combination of the other columns in X_2. However, this implies that one of the columns of X can be written as a linear combination of other columns in X, implying that X is not full rank. This is a contradiction. □

(b) Define $B = M_2\boldsymbol{X}_1$. Show that replacing $\boldsymbol{X}_1$ with the matrix B does not change the image, i.e. $Im(\boldsymbol{X}_1, \boldsymbol{X}_2) = Im(B, \boldsymbol{X}_2)$. (Hint: Modify Lemma **??**)

Solution. First rewrite $B = M_2X_1 = X_1 - X_2(X_2^tX_2)^{-1}X_2^tX_1$ and define $\Theta := (X_2^tX_2)^{-1}X_2^tX_1$, which is a $k_2 \times k_1$ vector. Then $B = X_1 - X_2\Theta$.

(i) $Im(B, X_2) \subseteq Im(X_1, X_2)$. Suppose that $z \in Im(B, X_2)$. Then there exists a vector $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$, where $\beta_1 \in \mathbb{R}^{k_1}$ and $\beta_2 \in \mathbb{R}^{k_2}$ such that $z = [B, X_2]\beta$. This can be decomposed as $B\beta_1 + X_2\beta_2$, which is equal to $(X_1 - X_2\Theta)\beta_1 + X_2\beta_2$ and can be written in the form $X_1\beta_1 + X_2(-\Theta\beta_1 + \beta_2)$. Therefore, $x \in Im(X_1, X_2)$.

(ii) $Im(X_1, X_2) \subseteq Im(B, X_2)$. Suppose that $z \in Im(X_1, X_2)$. Then there exists a vector $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$, where $\beta_1 \in \mathbb{R}^{k_1}$ and $\beta_2 \in \mathbb{R}^{k_2}$ such that $z = [X_1, X_2]\beta$. This can be decomposed as $X_1\beta_1 + X_2\beta_2$, which is equal to $(B + X_2\Theta)\beta_1 + X_2\beta_2$ and can be written in the form $B\beta_1 + X_2(\Theta\beta_1 + \beta_2)$. Therefore, $x \in Im(B, X_2)$.

□

(c) Show that if $\boldsymbol{X}$ is full rank then $(\boldsymbol{X}_1^t M_2 \boldsymbol{X}_1)$ is full rank. (Hint: Review Linear Regression Section)

Proof. The matrix can be rewritten as $X_1^t M_2 X_1 = (X_1^t M_2^t M_2 X_1)$ because M_2 idempotent and symmetric implies that $M_2 = M_2^t M_2$. Therefore the equation can be written as $(M_2X_1)^t(M_2X_1)$. By Lemma **??**, the gram matrix is full rank if and only $B = M_2X_1$ is full rank.

Now let us show that $B = M_2X_1$ is indeed full rank. Suppose not. Then there exists some nonzero vector $c_1 \neq 0$ such that

$$(I - X_2(X_2^tX_2)^{-1}X_2^t)X_1c_1 = 0.$$

Define $c_2 = (X_2^tX_2)^{-1}X_2^tX_1c_1$, and hence the above equation could be rewritten as

$$X_1c_1 - X_2c_2 = 0.$$

Since (c_1, c_2) is a nonzero vector, this contradicts the condition that $\boldsymbol{X}$ is full rank. □

1.4 Convex Sets (I): Hyperplanes

1. For any $p \in \mathbb{R}^n \backslash \{0\}$ and $a \in \mathbb{R}$, let

$$h(p,a) \equiv \{x \in \mathbb{R}^n | p^T x \geq a\}$$

be the half space generated by the hyperplane $H(p,a)$. Assume D is a closed subset of $\mathbb{R}^n$. Let E be the intersection of all half spaces that contain D, i.e.

$$E \equiv \bigcap_{h(p,a) \supseteq D} h(p,a).$$

Prove D is convex if and only if $D = E$. This gives another characterization of convexity. (Hint: separating hyperplane theorem.)

Solution. If $D = E$, then D is clearly convex because each half space in the intersection is convex.

Now assume D is convex. Because D is contained in each of the half space in the intersection, $D \subseteq E$. Assume there is $x \in E$ but $x \notin D$. Then because D is convex and closed, there exists a hyperplane $H(p^*, a^*)$ that strictly separates x and D:

$$p^{*T} d > a^* > p^{*t} x \quad \forall d \in D.$$

The first inequality implies $D \subseteq h(p^*, a^*)$, implying $E \subseteq h(p^*, a^*)$. But $x \in E$ implies $p^{*T} x \geq a^*$, a contradiction. Hence $E = D$. □

2. Assume $U \subset \mathbb{R}^n$ is convex. Let $x^* \in U$ be a point. Prove the followings are equivalent:

   (a) there is no $x \in U$ such that $x_i > x_i^*$ for all $i = 1, \cdots, n$,

   (b) there exists $\lambda \in \mathbb{R}^n_+ \backslash \{0\}$ such that x^* solves

$$\max_{x \in U} \lambda^T x.$$

Solution. (a)$\Longrightarrow$(b): Define $W \equiv \{x \in \mathbb{R}^n \big| x_i > x_i^* \ \forall i = 1, \cdots, n\}$. The set W is nonempty and convex, and $W \cap U = \emptyset$ by assumption. Then by supporting hyperplane theorem, there exists $\lambda \in \mathbb{R}^n \backslash \{0\}$ and real number c such that

$$\lambda^T y \geq c \geq \lambda^T x, \quad \forall y \in W, \ x \in U.$$

Because x^* is a limit point of W by construction, clearly $\lambda^T x^* \geq c \geq \lambda^T x$ for all $x \in U$. It remains to show $\lambda \geq 0$. Assume $\lambda_i < 0$ for some i. For any arbitrary $\tilde{y} \in W$, $\lambda^T(\tilde{y} + ne_i)$ tends to $-\infty$ as $n \to \infty$, where e_i is the ith unit vector in $\mathbb{R}^n$. But $\tilde{y} + ne_i \in W$ for all n, contradicting $\lambda^T y \geq c$ for all $y \in W$.

(b)$\Longrightarrow$(a): Suppose there exists $\lambda \in \mathbb{R}^n_+ \{0\}$ such that $\lambda^T x^* \geq \lambda^T x$ for all $x \in U$. Pick any $x \in \mathbb{R}^n$ satisfying $x_i > x_i^*$ for all i. Because $\lambda_i \geq 0$ and $\lambda \neq 0$, we have $\lambda^T x > \lambda^T x^*$. Thus such x is not in U. □

3. Let D be a nonempty convex subset of $\mathbb{R}^n$. Prove its closure $\overline{D}$ is convex.

Solution. Pick any $x, x' \in \overline{D}$. There must exist sequences $\{x_n\} \subset D$ and $\{x'_n\} \subset D$ such that $x_n \to x$ and $x'_n \to x'$ (if $x \in D$, then let $x_n = x$). So $\lambda x_n + (1-\lambda)x'_n \in D$ for all $\lambda \in [0,1]$. Because $\lambda x_n + (1-\lambda)x'_n$ converges to $\lambda x + (1-\lambda)x'$, $\lambda x + (1-\lambda)x'$ is a point in $\overline{D}$. $\square$

1.5 Convex Sets (II): Cones

1. There are several different characterizations of Farkas' Lemma. For example

   Lemma 1.5.1 (Farkas' Lemma V2)**.** *Let A be an $m \times n$ matrix and $b \in \mathbb{R}^m$. Then one and only one is true:*

   (i) *There exists $x \in \mathbb{R}^n$ such that $Ax \leq b$.*

   (ii) *There exists $y \in \mathbb{R}^m$ such that $y \geq \mathbf{0}_{m\times 1},\ y^tA = \mathbf{0}_{1\times n},\ y^tb < 0$.*

   In this exercise, you will prove the lemma.

   (a) Define $C = [A, -A, I_{m\times m}] \in \mathbb{R}^m \times \mathbb{R}^{2n+m}$. Show that condition (i) is equivalent to $b \in Cone(C)$ (Hint: Use properties of block-partitioned matrices and define a vector $z \in \mathbb{R}_+^{2n+m}$).

   Solution. Before we proceed with the proof we will analyze an object in $Cone(C)$. The vector $z \in Cone(C)$ if there exists a vector $\lambda \in \mathbb{R}_+^{2n+m}$ such that $z = C\lambda$. In block-partition form this means that:

$$z = \begin{bmatrix} A & -A & I \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = A\lambda_1 - A\lambda_2 + \lambda_3 = A(\lambda_1 - \lambda_2) + \lambda_3$$

   ($\Longrightarrow$) We show that condition (i) implies that $b \in Cone(C)$. Suppose that there exists an $x \in \mathbb{R}^n$ such that $Ax \leq b$. Set $\lambda_3 = b - Ax \geq 0$. For each entry of x_j set $\lambda_{1j} = x_j$ if $x_j \geq 0$ and zero otherwise. Similarly, set $\lambda_{2j} = -x_j$ if $x_j < 0$ and zero otherwise. Then $x = \lambda_1 - \lambda_2$ and $\lambda_1, \lambda_2 \in \mathbb{R}_n^+$. Therefore, $b \in Cone(A)$.
   ($\Longleftarrow$) If $b \in Cone(C)$, there exists λ_+^{2n+m} such that $b = C\lambda$. Set $x = \lambda_1 - \lambda_2 \in \mathbb{R}^n$. By definition $b = Ax + \lambda_3 \geq Ax$ since $\lambda_3 \geq 0_{m\times 1}$.

   □

   (b) Show that condition (ii) is equivalent to: There exists $y \in \mathbb{R}^m$ such that $y^tC \geq \mathbf{0}_{1\times(2n+m)}$ and $y^tb < 0$.

   Proof. Before we show the equivalence, let us express the y^tC in block form.

$$y^tC = y^t \begin{bmatrix} A & -A & I \end{bmatrix} \geq 0 \quad \Longleftrightarrow \quad \begin{matrix} y^tA \geq \mathbf{0}_{1\times n} \\ -y^tA \geq \mathbf{0}_{1\times n} \\ y^t \geq \mathbf{0}_{1\times m} \end{matrix}$$

   Combining the two inequalities gives us $y^tA = \mathbf{0}_{1\times n}$ and $y^t \geq \mathbf{0}_{1\times m}$. We can transpose y^t to show that the two conditions are identical.

   □

   (c) Use the original Farkas' Lemma to prove (Version 2).

   Solution. Apply Farkas' Lemma with the matrix C. Then either the statement in question (a) occurs or the statement in question (b). The proof is completed between these statements and those of the lemma that we want to prove. □

2. Consider an alternative restriction on asset prices.

Definition 1.5.1 (Pricing Restrictions)**.** Suppose that there does not exist an $x \in \mathbb{R}^n$ such that ($q^t x \leq 0$ and $Rx > \mathbf{0}_{m\times 1}$) or such that ($q^t x < 0$ and $Rx \geq \mathbf{0}_{m\times 1}$).

(a) Write down an economic interpretation of this condition.

Solution. It says that a market is arbitrage free if an investor cannot purchase a portfolio at (1) zero cost or lower and obtain a positive return in at least one state, or (2) get paid for the assets (negative costs) and receive a non-negative return. □

(b) Suppose that there exists a set of portfolio weights $x \in \mathbb{R}^n$ that yield positive returns in every state ($\Pi x \gg 0$). Show that $Rx > \mathbf{1}_{m\times 1} q^t x$. Give a simple example of a return matrix R, a price vector q and a portfolio x where this holds but the conditions in Definition **??** does not hold.

Proof. By definition, the expected profit from a portfolio is $\Pi x = Rx - \mathbf{1}_{n\times 1} q^t x$. Then $\Pi x \gg 0$ implies that $Rx \gg \mathbf{1}_{n\times 1} q^t x$. Consequently, $Rx > \mathbf{1}_{n\times 1} q^t x$. Let $n = m = 1$, and $q = 1$ and $R = 2$. Then $x = 1$ ensures that $\Pi x \gg 0$. However, $q^t x > 1$ and $Rx > 0$. □

(c) Suppose that there exists a probability vector $\alpha \in \mathbb{R}^m$ with **strictly positive** entries which satisfies $\alpha^t \Pi = \mathbf{0}_{1\times n}$. Show that Definition **??** is satisfied.

Proof. If there exists a vector $\alpha \in \mathbb{R}^m$ with strictly positive probabilities such that $\alpha^t \Pi = \mathbf{0}_{1\times n}$ then $\alpha^t Rx = \alpha^t \mathbf{1}_{m\times 1} q^t x$. Suppose that (i) $Rx > 0$ and $q^t x \leq 0$, then since $\alpha \gg 0$ then $\alpha^t Rx > 0$ and $\alpha^t \mathbf{1} q^t x \leq 0$. On the other hand if (ii) $Rx \geq 0$ and $q^t x < 0$ then $\alpha^t Rx \geq 0$ and $\alpha^t \mathbf{1} q^t x < 0$. This a contradiction because we should have $\alpha^t Rx = \alpha^t \mathbf{1}_{m\times 1} q^t x$. □

1.6 Quadratic Forms

1. Let A be an $n \times n$ square matrix. Assume:

$$x^T A x = 0, \qquad \forall x \in \mathbb{R}^n.$$

   (a) Prove all diagonal components of A are $0 \in \mathbb{R}$.

   Solution. Let $x = e_i$ be the i-th unit vector in $\mathbb{R}^n$. Then $a_{ii} = e_i^T A e_i = 0$. □

   (b) Show by example that condition (**??**) does not imply $A = \mathbf{0}$.

   Solution. For example

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

   works. □

1.7 Determinants

A matrix $B_{n\times n}$ is positive definite if $\forall x \in \mathbb{R}^n, x^T Bx > 0$. An equivalent definition of positive definiteness can be formulated using the determinant:

$$B = \begin{bmatrix} b_{11} & ... & b_{n1} \\ ... & ... & ... \\ b_{n1} & ... & b_{nn} \end{bmatrix}$$

Define the leading principal minor k of B, as the matrix formed by taking the upper left $(k\times k)$ submatrix. In other words:

$$B_1 = \begin{bmatrix} b_{11} \end{bmatrix}, B_2 = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}, ..., B_n = \begin{bmatrix} b_{11} & ... & b_{n1} \\ ... & ... & ... \\ b_{n1} & ... & b_{nn} \end{bmatrix}$$

A matrix is positive definite if and only if $\forall i \in \{1, ..., n\}, \det(B_i) > 0$. (Take this as a given, you do not need to prove it).

1. Define a function $F : \mathcal{M}_{n\times n} \to \mathbb{R}^n$. $F(B) = (\det(B_1), ..., \det(B_n))$. Reformulate the definition of positive definiteness in terms of $F(B)$.

   Solution. The condition is: A matrix is positive definite if and only if $F(B) \in \mathbb{R}^n_{++}$.

   Remark As a reminder, the set $\mathbb{R}^n_{++}$ is the set in $\mathbb{R}^n$ that has strictly positive components for all dimensions.

   □

2. Define a metric for the distance between two matrices, $d(A, B)$. Show that it is a metric: that it is non-negative, symmetric and satisfied the triangle inequality.

   Solution. Let $vec(A), vec(B)$ be the vectorized versions of the matrices (A, B). Then let us define the distance between two matrices as:

   $$d(A, B) = |vec(B) - vec(A)|_{\mathbb{R}^{mn}}$$

   $$d(A, B) = \sqrt{(a_{11} - b_{11})^2 + ...(a_{n1} - b_{n1})^2 + ... + (a_{mn} - b_{mn})^2}$$

   where $|\cdot|_{\mathbb{R}^{mn}}$ is the vector norm in $\mathbb{R}^{mn}$. This metric satisfies the three properties of a metric (because the vector metric is a proper metric):

   (a) It is non-negative and $A = B$ iff $d(A, B) = 0$.

   (b) It is symmetric. $d(A, B) = d(B, A)$.

   (c) It satisties the triangle inequality :

   $$d(A, C) \leq d(A, B) + d(B, C)$$

$\square$

3. Show that the function $F(B)$ is continuous.

   Solution. Let $F_i(B)$ be the i^{th} coordinate of $F(B)$. A vector valued function is continuous if and only if all of its components are continuous functions. Therefore we only need to prove that $F_i(B)$ is continuous $\forall i \in \{1, ..., n\}$.

   $F_i(B) = \det(h_i(B)) = \det(B_i)$, where $h_i(B)$ is a functions that selects the submatrix B_i. We discussed in class that the determinant is a continuous function because it is essentially a polynomial of the components of a matrix, and polynomial functions are always continuous. Furthermore $h_i(B)$ is also a continuous function (it only selects elements from B). Therefore the composite function $F_i(B)$ is also continuous.

   Remark Continuity has to be defined within a metric space. We can choose the metric we selected in part (b).

   $\square$

4. Show that the set of positive definite matrices of size (n) is an open set in $\mathcal{M}_{n\times n}$.

   Remark This shows that under small perturbations in the components of a positive definite matrix, the resulting matrix preserves the property of positive definiteness.

   Solution. One definition of continuity that is very useful in the case is that a function is continuous if and only if the pre-image of an open set is also an open set in the domain. In this cases a matrix is positive definite if $F(B) \in \mathbb{R}^n_{++}$. The set $\mathbb{R}^n_{++}$ is an open set. Therefore, the set of matrices that satisfy the condition is also an open set.

   We can also derive the proof using $\epsilon - \delta$ arguments. Suppose that a matrix is positive definite, then $F(B) \in \mathbb{R}^n_{++}$. There exists an $\epsilon > 0$ such that the all values in an open ball around $F(B)$ also belong to $\mathbb{R}^n_{++}$. By the definition of continuity $\exists \delta > 0$ such that $\forall B' s.t. d(B, B') < \delta \implies d(F(B), d(F(B')) < \epsilon$. This means that a neighborhood around B is also positive definite. Thus the set of positive definite matrices is an open set.

   $\square$

1.8 Eigenvalues and Eigenvectors

1. (25 points) Let P be an $n \times n$ matrix.

   (a) (5 points) Define a *markov matrix* P as an $n \times n$ matrix that has non-negative entries where the entries of each column sum to one. Let π be a non-negative vector whose entries sum to one. Show that π does not belong to the kernel. Further show that $P\pi$ is a vector whose entries sum to one.

   Solution.

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{12} & p_{22} & \cdots & p_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}, \pi = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_n \end{bmatrix}$$

Then $P\pi$ can be written as:

$$P\pi = \pi_1 \begin{bmatrix} p_{11} \\ p_{21} \\ \vdots \\ p_{n1} \end{bmatrix} + \cdots + \pi_n \begin{bmatrix} p_{1n} \\ p_{2n} \\ \vdots \\ p_{nn} \end{bmatrix} = \sum_i \pi_j p_j := \text{ A linear combination of the columns}$$

Therefore, the sum of the entries on $P\pi$ is:

$$\sum_i \sum_j \pi_j p_{ij} = \sum_j \pi_j \sum_i p_{ij} = \sum_j \pi_j = 1$$

The second equality changes the order of the sum. The third equality uses the fact that the elements of each column of P sum to 1. The fourth equality uses the fact that the entries of π_i sum to one.

π belongs to the kernel $\iff$ $P\pi = 0$. However, since its entries of $P\pi$ sum to one, $P\pi \neq 0$. Therefore, π is not part of the kernel.

Using the fact that π is non-negative is not necessary to prove the above properties. However, it implies that $P\pi$ has non negative entries.

$$\begin{aligned} p_{ij} &\geq 0 && \forall i, j \in \{1, ..., n\} \\ \implies \pi_j p_{ij} &\geq 0 && \text{since } \pi_j \geq 0, \forall j \\ \implies \sum_j \pi_j p_{ij} &\geq 0 && \text{which is the } i^{th} \text{ entry of } P\pi \end{aligned}$$

$\square$

   (b) (3 points) Now suppose that $\lim_{m \to \infty} P^m \to P^*$. Show that P^* is also a markov matrix and show that π does not belong to its kernel. (Hint: Show that every P^n is markov).

Solution. First we will show that P^m is markov. We will do this by induction. For $m = 2$:

$$P^2 = PP = \begin{bmatrix} \uparrow & \cdots & \uparrow \\ Pp_1 & \cdots & Pp_n \\ \downarrow & \cdots & \downarrow \end{bmatrix}$$

Notice that the columns of P are non-negative vectors that sum to one. Therefore, Pp_i is a non-negative vector whose entries sum to one in each column, by the proof in the previous exercise. For $n > 2$:

$$P^m = PP^{m-1}$$

Since P^m is markov, its columns are non-negative entries whose entries sum to one in each column, it follows that P^m is also markov. Let p_{ijm} denote the i, j entry of P^m.

We can summarize the set of conditions that define a markov matrix for a matrix P^m:

$$p_{ijm} \geq 0,\ \forall i, j \in \{1, ..., n\}$$

$$\sum_j p_{ijm} = 1,\ \forall i, j \in \{1, ..., n\}$$

Notice that if $p_{ijm} \to p^*_{ij}$, then it still satisfies the first weak inequality, and the second equality.This means that the set of markov matrices is closed. Therefore, in the limit it still satisfies the restrictions of a markov matrix. This completes the proof of why P^* is markov.

□

(c) (2 points) Show that if P is symmetric, then P^* is symmetric.

Solution. Firs we will show that P^m is symmetric. We will prove this by induction. For $m = 2$:

$$(PP)^t = P^t P^t = P$$

For $n > 2$: $(P^m)^t = (PP^{m-1})^t = (P^{(m-1)})^t P^t = P^{(m-1)}P = P^m$. This shows that P^m is symmetric. Notice that a matrix is symmetric iff:

$$p_{ijm} - p_{jim} = 0\ \forall i, j \in \{1, ..., n\}$$

If $p_{ijm} \to p^*_{ij}$, it will still satisfy this equality b

□

(d) (5 points) Suppose that P^* is such that for every π, $P^*\pi = \pi^*$, for a fixed π^*. Write down what the matrix P^* has to be for $\pi^* = (0.2, 0.3, 0.4, 0.1)$ if P^* is 4×4.

Solution. We will use the elementary basis to construct P^* :

$$P^* = \begin{bmatrix} \uparrow & \cdots & \uparrow \\ Pe_1 & \cdots & Pe_n \\ \downarrow & \cdots & \downarrow \end{bmatrix} = \begin{bmatrix} \uparrow & \cdots & \uparrow \\ \pi^* & \cdots & \pi^* \\ \downarrow & \cdots & \downarrow \end{bmatrix}$$

This is a matrix with identical column vectors π^*. Since $P\pi$ is just a linear combination of the columns, with weights adding to one, then the resulting vector is just π^*, as desired. For the example 4×4 example suggested:

$$P^* = \begin{bmatrix} 0.2 & 0.2 & 0.2 & 0.2 \\ 0.3 & 0.3 & 0.3 & 0.3 \\ 0.4 & 0.4 & 0.4 & 0.4 \\ 0.1 & 0.1 & 0.1 & 0.1 \end{bmatrix}$$

□

(e) (5 points) Under the previous property, for what set of vectors π^* will the implied P^* be symmetric. If it is symmetric, is it idempotent? If so, what is its rank? (Note that if P is symmetric, it implies very special restrictions on what P should converge to).

Solution. In the previous questions we established that under the previous property, $p^*_{ij} = \pi^*_i, \forall i, j \in \{1, ..., n\}$ (all columns are identical to π^*).

On the other hand, symmetry implies that $p^*_{ij} = p^*_{ji}$. Suppose that we take the first column: $\pi^*_i = p^*_{i1} = p^*_{1i} = \pi^*_1, \forall i\{1, ..., n\}$. Therefore all the entries of π^* are identical and equal to $1/n$ because they have to add up to 1. For the 4×4 case this means:

$$P^* = \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix}$$

□

(f) (2 points) Construct an example of a 2×2 symmetric matrix P that doesn't converge. (Hint use zeros and ones only). Compute its eigenvalues. Use the spectral decomposition to give a reason why it doesn't converge.

Solution. An example of a 2×2 matrix that doesn't converge is:

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

The matrix P changes the order of the columns. It can be shown by induction that:

$$P^m = \begin{cases} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} & m \text{ odd} \\ \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} & m \text{ even} \end{cases}$$

To compute its eigenvalues we need to compute the roots of:

$$\det(P - \lambda I) = \det \begin{bmatrix} -\lambda & 1 \\ 1 & -\lambda \end{bmatrix} = \lambda^2 - 1 = 0$$

Then the roots are: $\lambda_1 = 1$, $\lambda_2 = -1$. Now we need to compute the eigenvectors for each eigenvalue, respectively:

$$(P - I)v_1 = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} v_1 = 0 \implies v_1 \in span\{\begin{bmatrix} 1 \\ 1 \end{bmatrix}\}$$

$$(P + I)v_2 = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} v_2 = 0 \implies v_2 \in span\{\begin{bmatrix} 1 \\ -1 \end{bmatrix}\}$$

Therefore we can construct a spectral decomposition of P. Notice that first we have to obtain orthogonal vectors from each span: $\tilde{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}, \tilde{v}_2 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$.

$$\begin{aligned} P &= \begin{bmatrix} \uparrow & \uparrow \\ \tilde{v}_1 & \tilde{v}_2 \\ \downarrow & \downarrow \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \leftarrow & \tilde{v}_1^t & \rightarrow \\ \leftarrow & \tilde{v}_2^t & \rightarrow \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \qquad C\Lambda C^t \end{aligned}$$

We can verify that this decomposition recovers the original matrix P. We can now use this decomposition to compute P^m:

$$\begin{aligned} P^m &= PP \cdots P \\ &= (C\Lambda C^t)(C\Lambda C^t) \cdots (C\Lambda C^t) \\ &= C\Lambda^m C^t \\ &= C \begin{bmatrix} (1)^m & 0 \\ 0 & (-1)^m \end{bmatrix} C^t \end{aligned}$$

The second to third line follow from the fact that $C^t C = I$, since the vectors are orthonormal. This means λ^m oscillates between -1 and 1, depending on m, and thus never converges. The example shows that symmetry does not guarantee convergence.

General remarks for other types of exercises: Notice that if one of the eigenvalues were strictly less than one in absolute value then λ^m would converge to zero. However, at least one of them has to be greater than or equal to zero, otherwise $\Lambda^m \to 0$ and P^* is not a markov matrix (which contradicts what we proved earlier). If $|\lambda| > 1$ then the values would be explosive and diverge to infinity.

There are many more results for the eigenvalues of markov matrices that are not symmetric, even those that don't have a spectral decomposition. The key thing is to prove which and how many

many eigenvalues are strictly less than one, equal to one and strictly greater than one. I encourage you to keep these concepts in mind in future work involving markov chains.

$\square$

(g) (3 points) Show that the following asymmetric P converges to a P^* such that $P^*\pi = \pi^*$. Compute P^* and π^*.

$$P = \begin{bmatrix} 0.5 & 0 \\ 0.5 & 1 \end{bmatrix}$$

Solution. I will prove that $P^m = \begin{bmatrix} (0.5)^m & 0 \\ 1-(0.5)^m & 1 \end{bmatrix}$ by using induction. The result holds trivially for $m = 1$, then for $m > 1$:

$$\begin{aligned} P^m &= PP^{m-1} \\ &= \begin{bmatrix} 0.5 & 0 \\ 0.5 & 1 \end{bmatrix} \begin{bmatrix} (0.5)^m & 0 \\ 1-(0.5)^m & 1 \end{bmatrix} \\ &= \begin{bmatrix} (0.5)^{(m+1)} & 0 \\ (0.5)(0.5)^m + (1-(0.5)^m) & 1 \end{bmatrix} \\ &= \begin{bmatrix} (0.5)^{(m+1)} & 0 \\ 1-(0.5)^{(m+1)} & 1 \end{bmatrix} \end{aligned}$$

Then $P^* = \lim_{n\to\infty} P^m = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$ and $\pi^* = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. This means that regardless of the initial vector π, $P^m\pi$ will converge to π^*. This highlights that both symmetric and non-symmetric matrices can converge. $\square$

2. This questions asks you to analyze the eigenvalues of stochastic matrices:

(a) (3 points) Let $v \in \mathbb{R}^n$. Show that the entries of the vector Pv add up to $\sum_{j=1}^n v_j$.

Solution. Let $1_n \in \mathbb{R}^n$ be a vector with only 1s in each entry. Then the sum of the entries of Pv can be represented as $1_n^t Pv$. This is a quadratic form that can be represented as a double sum.

$$\begin{aligned} 1_n^t Pv &= \sum_{i=1}^n \sum_{j=1}^n P_{ij} v_j \\ &= \sum_{j=1}^n v_j \sum_{i=1}^n P_{ij} \\ &= \sum_{j=1}^n v_j \end{aligned}$$

$\square$

(b) (9 points) Let $v^* \in \mathbb{R}^n, v \neq 0$ be an eigenvector of P, with corresponding eigenvalue λ. Prove the following statements:

i. (1 point) $P^s v^* = \lambda^s v^*, s \in \mathbb{N}$.

Solution. We can prove this by induction. For $s = 1$, by definition of an eigenvector. $P^1 v^* = Pv^* = \lambda^1 v^*$.
Suppose it holds for s. Then $P^{s+1}v^* = P(P^s v^*) = P(\lambda^n)v^* = \lambda^s P v^* = \lambda^s \lambda v^* = \lambda^{s+1} v^*$.

□

ii. (4 points) Show that if $\sum_{j=1}^n v_j^* \neq 0$, then $\lambda = 1$. [Hint: show that P^s is also markov].

Solution. P is a matrix whose columns $(p_1, ..., p_n)$ sum to one. Let P' be another markov matrix. Then PP' is a matrix with columns $(Pp'_1, ..., Pp'_n)$. By the result in part (a) the columns of PP' must sum to one. Furthermore, since P and P' have non-negative entries, PP' has to have non-negative entries. Now we can show that P^s is markov by induction. For $s = 2$, if $P = P'$ then $PP' = P^2$, which is markov. Now suppose that it holds for s. Then $P' = P^s$. Then $P^{s+1} = PP^s = PP'$ which is also markov.

Therefore, by the result in part (a) the entries of $P^s v^*$ have to sum up to $\sum_{j=1}^n v_j^*$ for all s. From the result in part (b)(i) we know that $P^s v^* = \lambda^s v^*$. Therefore the entries sum up to $\lambda^s \sum_{j=1}^n v_j^*$. This means that $\lambda^s = 1, \forall s \implies \lambda = 1$.

Additional result: It is also possible to show that there exists at least one vector with eigenvalue 1. Since $det(B) = det(B^t), \forall B$ then $det(P - \lambda I) = det(P^t - \lambda I)$. This means that the eigenvalues of P and P^t are the same (although the eigenvectors can be different). Since the rows of P^t sum to one, it can be shown that 1_n is an eigenvector of P^t with eigenvalue $\lambda = 1$. Therefore, P has at least one eigenvector with $\lambda = 1$, which is not necessarily 1_n.

□

iii. (4 points) Show that if $\sum_{j=1}^n v_j^* = 0, v^* \neq 0$, then $|\lambda| \leq 1$. [Hint: show that for any fixed $v \neq 0$ (not necessarily an eigenvector), $\sup_P ||Pv|| \leq M < \infty$, P markov].

Solution. Let $w = Pv$. Notice that $||Pv|| = \sqrt{\sum_{i=1}^n w_i^2} = \sqrt{\sum_{i=1}^n (\sum_{j=1}^n P_{ij} v_j)^2}$. This is a continuous function of the P_{ij}. Suppose that we represent P as $vec(P) \in \mathbb{R}^{n^2}$. If P is markov, then each entry is bounded $P_{ij} \in [0, 1]$ and $\sum_{i=1}^n P_{ij} = 1, \forall j$, which is a closed set. Then for fixed v the norm $||Pv||$ is a continuous function from a compact space in $\mathbb{R}^{n^2}$ (the set of markov matrices) into $\mathbb{R}$. By the maximum theorem, there exists a markov matrix P^* such that $||P^* v|| = \max_P ||Pv|| = \sup_P ||Pv|| = M < \infty$. Consequently $||Pv|| \leq M$, for all P markov.

Notice that P^n is also markov, therefore $||P^n v^*|| \leq M$. By part (b)(i), this implies that $||\lambda^n v^*|| = |\lambda|^n ||v^*|| \leq M, \forall n$. Since $v^* \neq 0$, $||v^*|| > 0$ and $|\lambda|^n \leq \frac{M}{||v^*||}, \forall n$. If $|\lambda| > 1$, there exists an n large enough that $|\lambda|^n > \frac{M}{||v||}$. This is a contradiction, therefore $|\lambda| \leq 1$.

□

Bibliography

Greene, W. H. (2012). *Econometric analysis.* Pearson Education India.

Naiman, D. Q. and E. R. Scheinerman (2017). Arbitrage and geometry. *arXiv preprint arXiv:1709.07446*.

Rudin, W. et al. (1964). *Principles of mathematical analysis*, Volume 3. McGraw-hill New York.