GIUSEPPE CALAFIORE AND LAURENT EL GHAOUI

OPTIMIZATION MODELS

SOLUTIONS MANUAL

CAMBRIDGE

Ver. 0.1 - Oct. 2014

DISCLAIMER
This is the first draft of the Solution Manual for exercises in the book "Optimization Models" by Calafiore & El Ghaoui.
This draft is under construction. It is still incomplete and it is very likely to contain errors.
This material is offered "as is," non commercially, for personal use of instructors.
Comments and corrections are very welcome.

Contents

2. Vectors

Exercise 2.1 (Subpaces and dimensions) Consider the set $\mathcal{S}$ of points such that

$$x_1 + 2x_2 + 3x_3 = 0, \ \ 3x_1 + 2x_2 + x_3 = 0.$$

Show that $\mathcal{S}$ is a subspace. Determine its dimension, and find a basis for it.

Solution 2.1 The set $\mathcal{S}$ is a subspace, as can be checked directly: if $x, y \in \mathcal{S}$, then for every $\lambda, \mu \in \mathbb{R}$, we have $\lambda x + \mu y \in \mathcal{S}$. To find the dimension, we solve the equation and find that any solution to the equations is of the form $x_1 = -1/2x_2$, $x_3 = -1/3x_2$, where x_2 is free. Hence the dimension of $\mathcal{S}$ is 1, and a basis for S is the vector $(-1/2, 1, -1/3)$.

Exercise 2.2 (Affine sets and projections) Consider the set in $\mathbb{R}^3$, defined by the equation

$$\mathcal{P} = \left\{ x \in \mathbb{R}^3 \ : \ x_1 + 2x_2 + 3x_3 = 1 \right\}.$$

1. Show that the set $\mathcal{P}$ is an affine set of dimension 2. To this end, express it as $x^{(0)} + \text{span}(x^{(1)}, x^{(2)})$, where $x^{(0)} \in \mathcal{P}$, and $x^{(1)}, x^{(2)}$ are linearly independent vectors.

2. Find the minimum Euclidean distance from 0 to the set $\mathcal{P}$, and a point that achieves the minimum distance.

Solution 2.2

1. We can express any vector $x \in \mathcal{P}$ as $x = (x_1, x_2, 1/3 - x_1/3 - 2x_2/3)$, where x_1, x_2 are arbitrary. Thus

   $$x = x^{(0)} + x_1 x^{(1)} + x_2 x^{(2)},$$

   where

   $$x^{(0)} = \begin{bmatrix} 0 \\ 0 \\ \frac{1}{3} \end{bmatrix}, \ x^{(1)} = \begin{bmatrix} 1 \\ 0 \\ -\frac{1}{3} \end{bmatrix}, \ x^{(2)} = \begin{bmatrix} 0 \\ 1 \\ -\frac{2}{3} \end{bmatrix}.$$

   Since $x^{(1)}$ and $x^{(2)}$ are linearly independent, $\mathcal{P}$ is of dimension 2.

2. The set $\mathcal{P}$ is defined by a single linear equation $a^\top x = b$, with $a^\top = [1 \ 2 \ 3]$ and $b = 1$, i.e., $\mathcal{P}$ is a hyperplane. The minimum Euclidean distance from 0 to $\mathcal{P}$ is the ℓ_2 norm of the projection of 0 onto $\mathcal{P}$, which can be determined as discussed in Section 2.3.2.2. That is, the projection x^* of 0 onto $\mathcal{P}$ is such that $x^* \in \mathcal{P}$ and x^*

is orthogonal to the subspace generating $\mathcal{P}$ (which coincides with the span of a), that is $x^* = \alpha a$. Hence, it must be that $a^\top x^* = 1$, thus $\alpha\|a\|_2^2 = 1$, and $\alpha = 1/\|a\|_2^2$. We thus have that

$$x^* = \frac{a}{\|a\|_2^2},$$

and the distance we are seeking is $\|x^*\|_2 = 1/\|a\|_2 = 1/\sqrt{14}$.

Exercise 2.3 (Angles, lines and projections)

1. Find the projection z of the vector $x = (2,1)$ on the line that passes through $x_0 = (1,2)$ and with direction given by vector $u = (1,1)$.
2. Determine the angle between the following two vectors:

$$x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad y = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}.$$

Are these vectors linearly independent?

Solution 2.3

1. We can observe directly that $u^\top(x - x_0) = 0$, hence the projection of x is the same as that of x_0, which is $z = x_0$ itself.

Alternatively, as seen in Section (2.3.2.1), the projection is

$$z = x_0 + \frac{u^\top(x - x_0)}{u^\top u} u$$

which gives $z = x_0$.

Another method consists in solving

$$\begin{aligned}\min_t \|x_0 + tu - x\|_2^2 &= \min_t t^2 u^T u - 2tu^\top(x - x_0) + \|x - x_0\|_2^2 \\ &= \min_t (u^\top u)(t - t_0)^2 + \text{constant},\end{aligned}$$

where $t_0 = (x - x_0)^\top u/(u^T u)$. This leads to the optimal $t^* = t_0$, and provides the same result as before.

2. The angle cosine is given by

$$\cos\theta = \frac{x^\top y}{\|x\|_2\|y\|_2} = \frac{10}{14},$$

which gives $\theta \approx 41°$.

The vectors are linearly independent, since $\lambda x + \mu y = 0$ for $\lambda, \mu \in \mathbb{R}$ implies that $\lambda = \mu = 0$. Another way to prove this is to observe that the angle is not $0°$ nor $180°$.

Exercise 2.4 (Inner product) Let $x, y \in \mathbb{R}^n$. Under which condition on $\alpha \in \mathbb{R}^n$ does the function

$$f(x,y) = \sum_{k=1}^{n} \alpha_k x_k y_k$$

define an inner product on $\mathbb{R}^n$?

Solution 2.4 The axioms of 2.2 are all satisfied for any $\alpha \in \mathbb{R}^n$, except the conditions

$$f(x,x) \geq 0;$$
$$f(x,x) = 0 \text{ if and only if } x = 0.$$

These properties hold if and only if $\alpha_k > 0$, $k = 1, \ldots, n$. Indeed, if the latter is true, then the above two conditions hold. Conversely, if if there exist k such that $\alpha_k \leq 0$, setting $x = e_k$ (the k-th unit vector in $\mathbb{R}^n$) produces $f(e_k, e_k) \leq 0$; this contradicts one of the two above conditions.

Exercise 2.5 (Orthogonality) Let $x, y \in \mathbb{R}^n$ be two unit-norm vectors, that is, such that $\|x\|_2 = \|y\|_2 = 1$. Show that the vectors $x - y$ and $x + y$ are orthogonal. Use this to find an orthogonal basis for the subspace spanned by x and y.

Solution 2.5 When x, y are both unit-norm, we have

$$(x-y)^\top (x+y) = x^\top x - y^\top y - y^\top x + x^\top y = x^\top x - y^\top y = 0,$$

as claimed.

We can express any vector $z \in \text{span}(x,y)$ as $z = \lambda x + \mu y$, for some $\lambda, \mu \in \mathbb{R}$. We have $z = \alpha u + \beta v$, where

$$\alpha = \frac{\lambda + \mu}{2}, \quad \beta = \frac{\lambda - \mu}{2}.$$

Hence $z \in \text{span}(u,v)$. The converse is also true for similar reasons. Thus, (u,v) is an orthogonal basis for $\text{span}(x,y)$. We finish by normalizing u, v, replacing them with $(u/\|u\|_2, v/\|v\|_2)$. The desired orthogonal basis is thus given by $((x-y)/\|x-y\|_2, (x+y)/\|x+y\|_2)$.

Exercise 2.6 (Norm inequalities)

1. Show that the following inequalities hold for any vector x:

$$\frac{1}{\sqrt{n}}\|x\|_2 \leq \|x\|_\infty \leq \|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\|x\|_2 \leq n\|x\|_\infty.$$

Hint: use the Cauchy-Schwartz inequality.

2. Show that for any non-zero vector x,

$$\mathrm{card}(x) \geq \frac{\|x\|_1^2}{\|x\|_2^2},$$

where $\mathrm{card}(x)$ is the *cardinality* of the vector x, defined as the number of non-zero elements in x. Find vectors x for which the lower bound is attained.

Solution 2.6

1. We have

$$\|x\|_2^2 = \sum_{i=1}^n x_i^2 \leq n \cdot \max_i x_i^2 = n \cdot \|x\|_\infty^2.$$

Also, $\|x\|_\infty \leq \sqrt{x_1^2 + \ldots + x_n^2} = \|x\|_2$.

The inequality $\|x\|_2 \leq \|x\|_1$ is obtained after squaring both sides, and checking that

$$\sum_{i=1}^n x_i^2 \leq \sum_{i=1}^n x_i^2 + \sum_{i \neq j} |x_i x_j| = \left(\sum_{i=1}^n |x_i|\right)^2 = \|x\|_1^2.$$

Finally, the condition $\|x\|_1 \leq \sqrt{n}\|x\|_2$ is due to the Cauchy-Schwartz inequality

$$|z^\top y| \leq \|y\|_2 \cdot \|z\|_2,$$

applied to the two vectors $y = (1, \ldots, 1)$ and $z = |x| = (|x_1|, \ldots, |x_n|)$.

2. Let us apply the Cauchy-Schwartz inequality with $z = |x|$ again, and with y a vector with $y_i = 1$ if $x_i \neq 0$, and $y_i = 0$ otherwise. We have $\|y\|_2 = \sqrt{k}$, with $k = \mathrm{card}(x)$. Hence

$$|z^\top y| = \|x\|_1 \leq \|y\|_2 \cdot \|z\|_2 = \sqrt{k} \cdot \|x\|_2,$$

which proves the result. The bound is attained for vectors with k non-zero elements, all with the same magnitude.

Exercise 2.7 (Hölder inequality) Prove Hölder's inequality (2.4). *Hint:* consider the normalized vectors $u = x/\|x\|_p$, $v = y/\|y\|_q$, and observe that

$$|x^\top y| = \|x\|_p \|y\|_q \cdot |u^\top v| \leq \|x\|_p \|y\|_q \sum_k |u_k v_k|.$$

Then, apply Young's inequality (see Example 8.10) to the products $|u_k v_k| = |u_k||v_k|$.

Solution 2.7 The inequality is trivial if one of the vectors x, y is zero. We henceforth assume that none is, which allows us to define the normalized vectors u, v. We need to show that

$$\sum_k |u_k v_k| \leq 1.$$

Using the hint given, we apply Young's inequality, which states that for any given numbers $a, b \geq 0$ and $p, q > 0$ such that

$$\frac{1}{p} + \frac{1}{q} = 1,$$

it holds that

$$ab \leq \frac{1}{p}a^p + \frac{1}{q}b^q.$$

We thus have, with $a = |u_k|$ and $b = |v_k|$, and summing over k:

$$\begin{aligned} \sum_k |u_k v_k| &\leq \frac{1}{p}\sum_k |u_k|^p + \frac{1}{q}\sum_k |v_k|^q \\ &= \frac{1}{p}\|u\|_p^p + \frac{1}{q}\|v\|_q^q \\ &= \frac{1}{p} + \frac{1}{q} = 1, \end{aligned}$$

where we have used the fact that $\|u\|_p = \|v\|_q = 1$.

Exercise 2.8 (Linear functions)

1. For a n-vector x, with $n = 2m - 1$ odd, we define the median of x as the scalar value x_a such that exactly n of the values in x are $\leq x_a$ and n are $\geq x_a$ (i.e., x_a leaves half of the values in x to its left, and half to its right). Now consider the function $f : \mathbb{R}^n \to \mathbb{R}$, with values $f(x) = x_a - \frac{1}{n}\sum_{i=1}^n x_i$. Express f as a scalar product, that is, find $a \in \mathbb{R}^n$ such that $f(x) = a^\top x$ for every x. Find a basis for the set of points x such that $f(x) = 0$.
2. For $\alpha \in \mathbb{R}^2$, we consider the "power law" function $f : \mathbb{R}^2_{++} \to \mathbb{R}$, with values $f(x) = x_1^{\alpha_1} x_2^{\alpha_2}$. Justify the statement: "the coefficients α_i provide the ratio between the relative error in f to a relative error in x_i".

Solution 2.8 (Linear functions) TBD

Exercise 2.9 (Bound on a polynomial's derivative) In this exercise, you derive a bound on the largest absolute value of the derivative of a polynomial of a given order, in terms of the size of the coefficients[1]. For $w \in \mathbb{R}^{k+1}$, we define the polynomial p_w, with values

[1] See the discussion on regularization in Section 13.2.3 for an application of this result.

$$p_w(x) \doteq w_1 + w_2 x + \ldots + w_{k+1} x^k.$$

Show that, for any $p \geq 1$

$$\forall\, x \in [-1,1] \;:\; \left| \frac{\mathrm{d}p_w(x)}{\mathrm{d}x} \right| \leq C(k,p)\|v\|_p,$$

where $v = (w_2, \ldots, w_{k+1}) \in \mathbb{R}^k$, and

$$C(k,p) = \begin{cases} k & p = 1, \\ k^{3/2} & p = 2, \\ \frac{k(k+1)}{2} & p = \infty. \end{cases}$$

Hint: you may use Hölder's inequality (2.4) or the results from Exercise 2.6.

Solution 2.9 (Bound on a polynomial's derivative) We have, with $z = (1, 2, \ldots, k)$, and using Hölder's inequality:

$$\begin{aligned} \left| \frac{\mathrm{d}p_w(x)}{\mathrm{d}x} \right| &= \left| w_2 + 2w_3 x + \ldots + k w_{k+1} x^{k-1} \right| \\ &\leq |w_2| + 2|w_3| + \ldots + k|w_{k+1}| \\ &= |v^\top z| \\ &\leq \|v\|_p \cdot \|z\|_q. \end{aligned}$$

When $p = 1$, we have

$$\|z\|_q = \|z\|_\infty = k.$$

When $p = 2$, we have

$$\|z\|_q = \|z\|_2 = \sqrt{1 + 4 + \ldots + k^2} \leq \sqrt{k \cdot k^2} = k^{3/2}.$$

When $p = \infty$, we have

$$\|z\|_q = \|z\|_1 = 1 + 2 + \ldots + k = \frac{k(k+1)}{2}.$$

3. Matrices

Exercise 3.1 (Derivatives of composite functions)

1. Let $f : \mathbb{R}^m \to \mathbb{R}^k$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ be two maps. Let $h : \mathbb{R}^n \to \mathbb{R}^k$ be the composite map $h = f \circ g$, with values $h(x) = f(g(x))$ for $x \in \mathbb{R}^n$. Show that the derivatives of h can be expressed via a matrix-matrix product, as $J_h(x) = J_f(g(x)) \cdot J_g(x)$, where $J_h(x)$ is the Jacobian matrix of h at x, i.e., the matrix whose (i, j) element is $\frac{\partial h_i(x)}{\partial x_j}$.

2. Let g be an affine map of the form $g(x) = Ax + b$, for $A \in \mathbb{R}^{m,n}$, $b \in \mathbb{R}^m$. Show that the Jacobian of $h(x) = f(g(x))$ is

$$J_h(x) = J_f(g(x)) \cdot A.$$

3. Let g be an affine map as in the previous point, let $f : \mathbb{R}^n \to \mathbb{R}$ (a scalar-valued function), and let $h(x) = f(g(x))$. Show that

$$\begin{aligned} \nabla_x h(x) &= A^\top \nabla_g f(g(x)) \\ \nabla_x^2 h(x) &= A^\top \nabla_g^2 f(g(x)) A. \end{aligned}$$

Solution 3.1

1. We have, by the composition rule for derivatives:

$$\begin{aligned} [J_h(x)]_{i,j} &= \frac{\partial h_i(x)}{\partial x_j} = \sum_{l=1}^{m} \frac{\partial f_i}{\partial g_l}(x) \frac{\partial g_l}{\partial x_j}(x) \\ &= \sum_{l=1}^{m} [J_f(g(x))]_{i,l} [J_g(x)]_{l,j}, \end{aligned}$$

which proves the result.

2. Since $g_i(x) = \sum_{k=1}^{n} a_{ik} x_k + b_i$, $i = 1, \ldots, m$, we have that the (i, j)-th element of the Jacobian of g is

$$[J_g(x)]_{ij} = \frac{\partial g_i(x)}{\partial x_j} = a_{ij},$$

hence $J_g(x) = A$, and the desired result follows from applying point 1. of this exercise.

3. For a scalar-valued function, the gradient coincides with the transpose of the Jacobian, hence the expression for the gradient of h w.r.t. x follows by applying the previous point. For the Hessian,

we have instead

$$
\begin{aligned}
[\nabla_x^2 h(x)]_{ij} &= \frac{\partial^2 h(x)}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_j}\frac{\partial h(x)}{\partial x_i} = \frac{\partial}{\partial x_j} a_i^\top \nabla_g f(g(x)) \\
&= \frac{\partial}{\partial x_j}\sum_{k=1}^m a_{ik}\frac{\partial f(g(x))}{\partial g_k} = \sum_{k=1}^m a_{ik}\frac{\partial}{\partial x_j}\frac{\partial f(g(x))}{\partial g_k} \\
&= \sum_{k=1}^m a_{ik}\sum_{p=1}^m \frac{\partial}{\partial g_p}\left(\frac{\partial f(g(x))}{\partial g_k}\right)\frac{\partial g_p(x)}{\partial x_j} \\
&= \sum_{k=1}^m a_{ik}\sum_{p=1}^m \frac{\partial^2 f(g(x))}{\partial g_p \partial g_k}\frac{\partial g_p(x)}{\partial x_j} \\
&= \sum_{k=1}^m a_{ik}\sum_{p=1}^m \frac{\partial^2 f(g(x))}{\partial g_p \partial g_k} a_{pj} \\
&= a_i^\top \nabla_g^2 f(g(x)) a_j,
\end{aligned}
$$

which proves the statement.

Exercise 3.2 (Permutation matrices) A matrix $P \in \mathbb{R}^{n,n}$ is a permutation matrix if its columns are a permutation of the columns of the $n \times n$ identity matrix.

1. For a $n \times n$ matrix A, we consider the products PA and AP. Describe in simple terms what these matrices look like with respect to the original matrix A.

2. Show that P is orthogonal.

Solution 3.2

1. Given the matrix A, the product PA is the matrix obtained by permuting the rows of A; AP corresponds to permuting the columns.

2. Every pair of columns (p_k, p_l) of P is of the form (e_k, e_l), where e_k, e_l are the k-th and the l-th standard basis vectors in $\mathbb{R}^n$. Thus, $\|p_k\|_2 = 1$, and $p_k^\top p_l = 0$ if $k \neq l$, as claimed.

Exercise 3.3 (Linear maps) Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map. Show how to compute the (unique) matrix A such that $f(x) = Ax$ for every $x \in \mathbb{R}^n$, in terms of the values of f at appropriate vectors, which you will determine.

Solution 3.3 For $i = 1, \ldots, n$, let e_i be the i-th unit vector in $\mathbb{R}^n$. We have

$$f(e_i) = Ae_i = a_i,$$

where a_i is the i-th column of A. Hence we can compute the matrix A column-wise, by evaluating f at the points $e_1, \ldots, e_n$.

Exercise 3.4 (Linear dynamical systems) Linear dynamical systems are a common way to (approximately) model the behavior of physical phenomena, via recurrence equations of the form[2]

$$x(t+1) = Ax(t) + Bu(t), \quad y(t) = Cx(t), \quad t = 0, 1, 2, \ldots,$$

where t is the (discrete) time, $x(t) \in \mathbb{R}^n$ describes the state of the system at time t, $u(t) \in \mathbb{R}^p$ is the input vector, and $y(t) \in \mathbb{R}^m$ is the output vector. Here, matrices A, B, C, are given.

[2] Such models are the focus of Chapter 15.

1. Assuming that the system has initial condition $x(0) = 0$, express the output vector at time T as a linear function of $u(0), \ldots, u(T-1)$; that is, determine a matrix H such that $y(T) = HU(T)$, where

$$U(T) \doteq \begin{bmatrix} u(0) \\ \vdots \\ u(T-1) \end{bmatrix}$$

   contains all the inputs up to and including at time $T-1$.

2. What is the interpretation of the range of H?

Solution 3.4

1. We have

$$\begin{aligned} x(1) &= Bu(0) \\ x(2) &= Ax(1) + Bu(1) = ABu(0) + Bu(1) \\ x(3) &= Ax(2) + Bu(2) = A^2Bu(0) + ABu(1) + Bu(2). \end{aligned}$$

   We now prove by induction that, for $T \geq 1$:

$$x(T) = \sum_{k=0}^{T-1} A^k Bu(T-1-k) = \begin{bmatrix} A^{T-1}B & \ldots & AB & B \end{bmatrix} U(T).$$

   The formula is correct for $T = 1$. Let $T \geq 2$. Assume the formula is correct for $T-1$; we have

$$\begin{aligned} x(T) = Ax(T-1) + Bu(T-1) &= A\left(\sum_{k=0}^{T-2} A^k Bu(T-2-k)\right) + Bu(T-1) \\ &= \sum_{k=0}^{T-2} A^{k+1} Bu(T-2-k) + Bu(T-1) \\ &= \sum_{k=1}^{T-1} A^k Bu(T-1-k) + Bu(T-1) \\ &= \sum_{k=0}^{T-1} A^k Bu(T-1-k), \end{aligned}$$

as claimed. Finally, we have $y(T) = HU(T)$, with

$$H = C \cdot \left[\begin{array}{cccc} A^{T-1}B & \dots & AB & B \end{array} \right].$$

2. The range of H is the set of output vectors that are attainable at time T by the system by proper choice of the sequence of inputs, starting from the initial state $x(0) = 0$.

Exercise 3.5 (Nullspace inclusions and range) Let $A, B \in \mathbb{R}^{m,n}$ be two matrices. Show that the fact that the nullspace of B is contained in that of A implies that the range of $B^\top$ contains that of $A^\top$.

Solution 3.5 Assume that the nullspace of B is contained in that of A. This means that

$$Bx = 0 \Longrightarrow Ax = 0.$$

Let $z \in \mathcal{R}(A^\top)$: there exist $y \in \mathbb{R}^m$ such that $z = A^\top y$. We have thus, for any element $x \in \mathcal{N}(A)$, $z^\top x = y^\top Ax = 0$. Hence, z is orthogonal to the nullspace of A, so it is orthogonal to the nullspace of B. We have obtained $\mathcal{R}(A^\top) \subseteq \mathcal{N}(B)^\perp = \mathcal{R}(B^\top)$, as claimed. Here, we have used the fundamental theorem of linear algebra (3.1).

Exercise 3.6 (Rank and nullspace) Consider the image in Figure 3.6, a gray-scale rendering of a painting by Mondrian (1872-1944). We build a 256×256 matrix A of pixels based on this image by ignoring grey zones, assigning $+1$ to horizontal or vertical black lines, $+2$ at the intersections, and zero elsewhere. The horizontal lines occur at row indices 100, 200 and 230, and the vertical ones, at columns indices 50, 230.

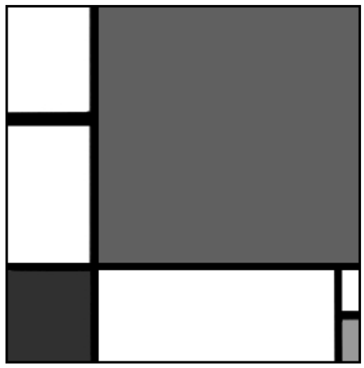

Figure 3.1: A gray-scale rendering of a painting by Mondrian.

1. What is nullspace of the matrix?

2. What is its rank?

Solution 3.6

1. Denote by e_i the i-th unit vector in $\mathbb{R}^{256}$, by $z_1 \in \mathbb{R}^{256}$ the vector with all first 50 components equal to one, by $z_2 \in \mathbb{R}^{256}$ the vector with all last 26 components equal to one, and by $z_3 \in \mathbb{R}^{256}$ the vector with all last 56 components equal to one. Finally, $\mathbf{1}$ denotes the vector of all ones in $\mathbb{R}^{256}$. We can express the matrix as

$$M = e_{100}z_1^\top + e_{200}\mathbf{1}^\top + e_{230}z_3^\top + \mathbf{1}e_{50}^\top + z_3 e_{230}^\top.$$

The condition $Mx = 0$, for some vector $x \in \mathbb{R}^n$, translates as

$$(z_1^\top x)e_{100} + (\mathbf{1}^\top x)e_{200} + (z_3^\top x)e_{230} + (e_{50}^\top x)\mathbf{1} + (e_{230}^\top x)z_3 = 0.$$

Since the vectors $(e_{100}, e_{200}, e_{230}, \mathbf{1}, z_3)$ are linearly independent, we obtain that the five coefficients in the above must be zero:

$$0 = z_1^\top x = \mathbf{1}^\top x = z_3^\top x = e_{50}^\top x = e_{230}^\top x.$$

It is easy to check that the corresponding subspace of $\mathbb{R}^{256}$ is of dimension $256 - 5 = 251$. Indeed, two elements of x are zero ($x_{50} = x_{230} = 0$), the remaining ones satisfy three independent equality constraints. From these we can express (say) x_1, x_{201}, x_{51} from the remaining variables, which then are free of any constraints. We can eliminate a total of five variables from the above five conditions, so the nullspace is of dimension 251.

2. The rank is 5.

Exercise 3.7 (Range and nullspace of $A^\top A$) Prove that, for any matrix $A \in \mathbb{R}^{m,n}$, it holds that

$$\begin{aligned} \mathcal{N}(A^\top A) &= \mathcal{N}(A) \\ \mathcal{R}(A^\top A) &= \mathcal{R}(A^\top). \end{aligned} \tag{3.1}$$

Hint: use the fundamental theorem of linear algebra.

Solution 3.7 First, suppose $x \in \mathcal{N}(A)$, then $Ax = 0$ and obviously $A^\top Ax = 0$. Conversely, suppose $x \in \mathcal{N}(A^\top A)$, we show by contradiction that it must be $x \in \mathcal{N}(A)$, hence proving the first claim. Indeed, suppose $x \in \mathcal{N}(A^\top A)$ but $x \notin \mathcal{N}(A)$. Define then $v = Ax \neq 0$. Such a v is by definition in the range of A, and $A^\top v = A^\top Ax = 0$, so v is also in the nullspace of $A^\top$, which is impossible since, by the fundamental theorem of linear algebra, $\mathcal{R}(A) \perp \mathcal{N}(A^\top)$. Next,

$$\mathcal{R}(A^\top) = \mathcal{N}(A)^\perp = \mathcal{N}(A^\top A)^\perp = \mathcal{R}(A^\top A),$$

which proves (3.1).

Exercise 3.8 (Cayley-Hamilton theorem) Let $A \in \mathbb{R}^{n,n}$ and let

$$p(\lambda) \doteq \det(\lambda I_n - A) = \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0$$

be the characteristic polynomial of A.

1. Assume A is diagonalizable. Prove that A annihilates its own characteristic polynomial, that is

$$p(A) = A^n + c_{n-1}A^{n-1} + \cdots + c_1 A + c_0 I_n = 0.$$

Hint: use Lemma 3.3.

2. Prove that $p(A) = 0$ holds in general, i.e., also for non-diagonalizable square matrices. *Hint:* use the facts that polynomials are continuous functions, and that diagonalizable matrices are dense in $\mathbb{R}^{n,n}$, i.e., for any $\epsilon > 0$ there exist $\Delta \in \mathbb{R}^{n,n}$ with $\|\Delta\|_F \le \epsilon$ such that $A + \Delta$ is diagonalizable.

Solution 3.8

1. The result is immediate from Lemma 3.3: if $A = U\Lambda U^{-1}$ is a diagonal factorization of A, then $p(\Lambda) = 0$, since by definition eigenvalues are roots of the characteristic polynomial, hence

$$p(A) = Up(\Lambda)U^{-1} = 0.$$

2. The map $p : \mathbb{R}^{n,n} \to \mathbb{R}^{n,n}$ with values $p(A) = A^n + c_{n-1}A^{n-1} + \cdots + c_1 A + c_0 I_n$ is continuous on $\mathbb{R}^{n,n}$. This map is identically zero on the dense subset of $\mathbb{R}^{n,n}$ formed by diagonalizable matrices (proved in the previous point of the exercise), hence by continuity it must be zero everywhere in $\mathbb{R}^{n,n}$.

Exercise 3.9 (Frobenius norm and random inputs) Let $A \in \mathbb{R}^{m,n}$ be a matrix. Assume that $u \in \mathbb{R}^n$ is a vector-valued random variable, with zero mean and covariance matrix I_n. That is, $\mathbb{E}\{u\} = 0$, and $\mathbb{E}\{uu^\top\} = I_n$.

1. What is the covariance matrix of the output, $y = Au$?

2. Define the total output variance as $\mathbb{E}\{\|y - \hat{y}\|_2^2\}$, where $\hat{y} = \mathbb{E}\{y\}$ is the output's expected value. Compute the total output variance and comment.

Solution 3.9

1. The mean of the output is zero: $\hat{y} = \mathbb{E}y = A\mathbb{E}u = 0$. Hence the covariance matrix is given by

$$\begin{aligned} \mathbb{E}(yy^\top) &= \mathbb{E}(Auu^\top A^\top) \\ &= A\mathbb{E}(uu^\top)A^\top \\ &= AA^\top. \end{aligned}$$

2. The total variance is

$$\begin{aligned} \mathbb{E}(y^\top y) &= \operatorname{trace} \mathbb{E}(yy^\top) \\ &= \operatorname{trace}(AA^\top) \\ &= \|A\|_F^2. \end{aligned}$$

The total output variance is the square of the Frobenius norm of the matrix. Hence the Frobenius norm captures the response of the matrix to a class of random inputs (zero mean, and unit covariance matrix).

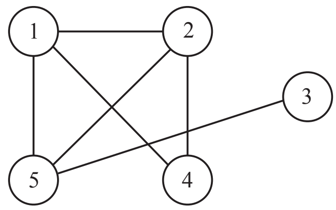

Figure 3.2: An undirected graph with $n = 5$ vertices.

Exercise 3.10 (Adjacency matrices and graphs) For a given undirected graph G with no self-loops and at most one edge between any pair of nodes (i.e., a *simple* graph), as in Figure 3.2, we associate a $n \times n$ matrix A, such that

$$A_{ij} = \begin{cases} 1 & \text{if there is an edge between node } i \text{ and node } j, \\ 0 & \text{otherwise.} \end{cases}$$

This matrix is called the *adjacency* matrix of the graph[3].

[3] The graph in Figure 3.2 has adjacency matrix

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{bmatrix}.$$

1. Prove the following result: for positive integer k, the matrix A^k has an interesting interpretation: the entry in row i and column j gives the number of *walks* of length k (i.e., a collection of k edges) leading from vertex i to vertex j. *Hint:* prove this by induction on k, and look at the matrix-matrix product $A^{k-1}A$.

2. A *triangle* in a graph is defined as a subgraph composed of three vertices, where each vertex is reachable from each other vertex (i.e., a triangle forms a complete subgraph of order 3). In the graph of Figure 3.2, for example, nodes $\{1, 2, 4\}$ form a triangle. Show that the number of triangles in G is equal to the trace of A^3 divided by 6. *Hint:* For each node in a triangle in an undirected graph, there are two walks of length 3 leading from the node to itself, one corresponding to a clockwise walk, and the other to a counter-clockwise walk.

Solution 3.10

1. We can prove the result by induction on k. For $k = 1$, the result follows from the very definition of A. Let $L_k(i, j)$ denote the number of paths of length k between nodes i and j, and assume that the result we wish to prove is true for some given $h \geq 1$, so that $L_h(i, j) = [A^h]_{i,j}$. We next prove that it must also hold that $L_{h+1}(i, j) = [A^{k+1}]_{i,j}$, thus proving by inductive argument that $L_k(i, j) = [A^k]_{i,j}$ for all $k \geq 1$.

   Indeed, to go from a node i to a node j with a walk of length $h + 1$, one needs first reach, with a walk of length h, a node l linked to j by an edge. Thus:

$$L_{h+1}(i, j) = \sum_{l \in V(j)} L_h(i, l),$$

where $V(j)$ is the neighbor set of j, which is the set of nodes connected to the j-th node, that is, nodes l such that $A_{l,j} \neq 0$. Thus:

$$L_{h+1}(i,j) = \sum_{l=1}^{n} L_h(i,l) A_{l,j}.$$

But we assumed that $L_h(i,j) = [A^h]_{i,j}$, hence the previous equation can be written as

$$L_{h+1}(i,j) = \sum_{l=1}^{n} [A^h]_{i,l} A_{l,j}.$$

In the above we recognize the (i,j)-th element of the product $A^h A = A^{h+1}$, which proves that $L_{h+1}(i,j) = [A^{h+1}]_{i,j}$, and hence concludes the inductive proof.

2. Following the hint, we observe that for each node in a triangle in an undirected graph there are two walks of length 3 leading from the node to itself, one corresponding to a clockwise walk, and the other to a counter-clockwise walk. Therefore, each triangle in the graph produces 6 walks of length 3 (two walks for each vertex composing the triangle). From the previous result, the number of walks of length 3 from node i to itself is given by $[A^3]_{i,i}$, hence the total number of of walks of length 3 from each node to itself is $\sum_{i=1}^{n} [A^3]_{i,i} = \text{trace}(A^3)$, and therefore the number of triangles is $\text{trace}(A^3)/6$.

Exercise 3.11 (Nonnegative and positive matrices) A matrix $A \in \mathbb{R}^{n,n}$ is said to be *nonnegative* (resp. *positive*) if $a_{ij} \geq 0$ (resp. $a_{ij} > 0$) for all $i,j = 1,\ldots,n$. The notation $A \geq 0$ (resp. $A > 0$) is used to denote nonnegative (resp. positive) matrices.

A nonnegative matrix is said to be column (resp. row) *stochastic*, if the sum of the elements along each column (resp. row) is equal to one, that is if $\mathbf{1}^\top A = \mathbf{1}^\top$ (resp. $A\mathbf{1} = \mathbf{1}$). Similarly, a vector $x \in \mathbb{R}^n$ is said to be nonnegative if $x \geq 0$ (element-wise), and it is said to be a *probability vector*, if it is nonnegative and $\mathbf{1}^\top x = 1$. The set of probability vectors in $\mathbb{R}^n$ is thus the set $S = \{x \in \mathbb{R}^n : x \geq 0, \mathbf{1}^\top x = 1\}$, which is called the *probability simplex*. The following points you are requested to prove are part of a body of results known as the Perron-Frobenius theory of nonnegative matrices.

1. Prove that a nonnegative matrix A maps nonnegative vectors into nonnegative vectors (i.e., that $Ax \geq 0$ whenever $x \geq 0$), and that a column stochastic matrix $A \geq 0$ maps probability vectors into probability vectors.

2. Prove that if $A > 0$, then its spectral radius $\rho(A)$ is positive. *Hint:* use the Cayley-Hamilton theorem.

3. Show that it holds for any matrix A and vector x that

$$|Ax| \le |A||x|,$$

where $|A|$ (resp. $|x|$) denotes the matrix (resp. vector) of moduli of the entries of A (resp. x). Then, show that if $A > 0$ and λ_i, v_i is an eigenvalue/eigenvector pair for A, then

$$|\lambda_i||v_i| \le A|v_i|.$$

4. Prove that if $A > 0$ then $\rho(A)$ is actually an eigenvalue of A (i.e., A has a positive real eigenvalue $\lambda = \rho(A)$, and all other eigenvalues of A have modulus no larger than this "dominant" eigenvalue), and that there exist a corresponding eigenvector $v > 0$. Further, the dominant eigenvalue is simple (i.e., it has unit algebraic multiplicity), but you are not requested to prove this latter fact.

   Hint: For proving this claim you may use the following fixed-point theorem due to Brouwer: *if S is a compact and convex set*[4] *in $\mathbb{R}^n$, and $f : S \to S$ is a continuous map, then there exist an $x \in S$ such that $f(x) = x$.* Apply this result to the continuous map $f(x) \doteq \frac{Ax}{\mathbf{1}^\top Ax}$, with S being the probability simplex (which is indeed convex and compact).

5. Prove that if $A > 0$ and it is column or row stochastic, then its dominant eigenvalue is $\lambda = 1$.

[4] See Section 8.1 for definitions of compact and convex sets.

Solution 3.11 (Nonnegative and positive matrices)

1. Let $A \ge 0$, $x \ge 0$, $y = Ax$, and denote with $a_i^\top$ the i-th row of A. Then obviously

$$y_i = a_i^\top x = \sum_{j=1}^{n} a_{ij} x_j \ge 0, \quad i = 1, \ldots, n,$$

which shows that a nonnegative matrix maps nonnegative vectors into nonnegative vectors. Further, if x is a probability vector and A is stochastic, then

$$\mathbf{1}^\top y = \mathbf{1}^\top Ax = \mathbf{1}^\top x = 1,$$

which shows that y is also a probability vector.

2. Suppose by contradiction that $A > 0$ and $\rho(A) = 0$. This would imply that A has an eigenvalue of maximum modulus in $\lambda = 0$,

thus, all eigenvalues of A are actually zero. This means that the characteristic polynomial of A is $p_A(s) = s^n$ and, by the Cayley-Hamilton theorem, it must hold that $A^n = 0$, which is impossible since A^n is the n-fold product of positive matrices, hence it must be positive.

3. By the triangle inequality, we have that, for $i = 1, \dots, n$,

$$\begin{aligned} |a_i^\top x| &\leq \sum_{j=1}^{n} |a_{ij} x_j| = \sum_{j=1}^{n} |a_{ij}||x_j| \\ &= |a_i^\top||x|, \end{aligned}$$

whih proves the first part. If $A > 0$ the above relation reads $|Ax| \leq A|x|$ which, for $x = v_i$, becomes

$$A|v_i| \geq |Av_i| = |\lambda_i v_i| = |\lambda_i||v_i|.$$

4. Let S be the probability simplex, and $f(x) \doteq \frac{Ax}{\mathbf{1}^\top Ax}$. From Brouwer's fixed-point theorem there exist $v \in S$ such that $f(v) = v$, that is such that

$$Av = (\mathbf{1}^\top Av)v = \lambda v, \quad \lambda \doteq \mathbf{1}^\top Av.$$

Moreover, since $A > 0$, it holds that $\lambda > 0$ and $v > 0$; thus A has a positive eigenvalue and a corresponding positive eigenvector. We next apply the same result to $A^\top$, obtaining that there exist $w \in S$ such that

$$A^\top w = (\mathbf{1}^\top A^\top w)w = \mu w, \quad \mu \doteq \mathbf{1}^\top A^\top w,$$

where again $\mu > 0$ and $w > 0$. Now, $v^\top w > 0$, and

$$\lambda v^\top w = v^\top A^\top w = \mu v^\top w,$$

which implies that $\mu = \lambda$, whence $A^\top w = \lambda w$.

Next, consider any eigenvalue/eigenvector pair λ_i, v_i for A, and apply the result of point 3. in this exercise, to obtain that

$$|\lambda_i||v_i| \leq A|v_i|, \quad i = 1, \dots, n.$$

Multiply both sides on the left by $w^\top$ to get

$$w^\top |\lambda_i||v_i| \leq w^\top A|v_i| = \lambda w^\top |v_i|,$$

from which we obtain that

$$|\lambda_i| \leq \lambda, \quad i = 1, \dots, n,$$

which proves that λ (which is real and positive, as shown above) is indeed a maximum modulus eigenvalue of A (thus, $\lambda = \rho(A)$), and the corresponding eigenvector v is positive.

5. By definition, if A is column stochastic then $\mathbf{1}^\top A = \mathbf{1}^\top$, which means that $\lambda = 1$ is an eigenvalue of A. Next, recall from Section 3.6.3.1 that the spectral radius of A is no larger than its induced ℓ_1 norm:

$$\rho(A) \leq \|A\|_1 = \max_{j=1,\ldots,n} \sum_{i=1}^{m} |a_{ij}| = 1,$$

hence $\lambda = 1$ is indeed the dominant eigenvalue. An analogous argument applies to row stochastic matrices.

4. Symmetric matrices

Exercise 4.1 (Eigenvectors of a symmetric 2×2 matrix) Let $p, q \in \mathbb{R}^n$ be two linearly independent vectors, with unit norm ($\|p\|_2 = \|q\|_2 = 1$). Define the symmetric matrix $A \doteq pq^\top + qp^\top$. In your derivations, it may be useful to use the notation $c \doteq p^\top q$.

1. Show that $p + q$ and $p - q$ are eigenvectors of A, and determine the corresponding eigenvalues.

2. Determine the nullspace and rank of A.

3. Find an eigenvalue decomposition of A, in terms of p, q. *Hint:* use the previous two parts.

4. What is the answer to the previous part if p, q are not normalized?

Solution 4.1

1. We have

   $$Ap = (cp + q), \quad Aq = p + cq,$$

   from which we obtain

   $$A(p - q) = (c - 1)(p - q), \quad A(p + q) = (c + 1)(p + q).$$

   Thus $u_\pm := p \pm q$ is an (un-normalized) eigenvector of A, with eigenvalue $c \pm 1$.

2. The condition on $x \in \mathbb{R}^n$: $Ax = 0$, holds if and only if

   $$0 = (q^\top x)p + (p^\top x)q = 0.$$

   Since p, q are linearly independent, the above is equivalent to $p^\top x = q^\top x = 0$. The nullspace is the set of vectors orthogonal to p and q. The range is the span of p, q. The rank is thus 2.

3. Since the rank is 2, there is a total of two non-zero eigenvalues. Note that, since p, q are normalized, c is the cosine angle between p, q; $|c| < 1$ since p, q are independent. We have found two linearly independent eigenvectors $u_\pm = p \pm q$ that do not belong to the nullspace (since $|c| < 1$). We can complete this set with eigenvectors corresponding to the eigenvalue zero; simply choose an orthonormal basis for the nullspace.

   Then, the eigenvalue decomposition is

   $$A = (c - 1)v_- v_-^\top + (c + 1)v_+ v_+^\top,$$