Chapter 2 Matrix Decomposition

At the end of the last chapter we explored two procedures for solving linear equations, namely Gaussian elimination and the LU decomposition. Both of these methods transform the original linear system into a triangular system. The advantage of the LU decomposition is that it needs to be performed once and then one can plug in the values of y to solve two linear triangular system whereas the Gaussian decomposition has to be applied from scratch for each value y. The other observation one needs to make is that the LU decomposition fails if the matrix is singular, whereas the Gaussian elimination might work in this case in the y of interest is in the range of A. Besides the LU decomposition is meant to deal with square matrices while we can always perform the Gaussian elimination. In this chapter we will explore two rather important matrix decompositions that enable us to deal with non-square matrices: The *QR decomposition* and the *singular value decomposition (SVD).*

1 QR decomposition

Before introducing the QR decomposition let us first go back to the linear least-squares problem. We focus our attention on overdetermined linear equations $Ax = y$ where $A \in \mathbb{R}^{m\times n}$, $y \in \mathbb{R}^m$ with $m \geq n$. As there are more constraints then variables, in general, such equations do not have a solution, so we have to settle with an approximate solution. The most common such approximation is the one that minimises the quadratic error

$$||Ax - y||^2 = (Ax - y)^T(Ax - y) .$$

Least-squares problem Let us first consider the existence of such solution. Tot his end we repeat the argument described in the previous Chapter.

Using the fundamental theorem of linear algebra $\mathrm{Ran}(A) \oplus \mathrm{Ker}(A^T) = \mathbb{R}^m$ with $\mathrm{Ran}(A) \perp \mathrm{Ker}(A^T)$ we can decompose the vector as $y = y_R + y_K$ where $y_R \in \mathrm{Ran}(A)$, $y_K \in \mathrm{Ker}(A^T)$ and $y_R \perp y_K$. Let $r = y - Ax$, we have

$$||r||^2 = r^Tr = ||(y_R - Ax) + y_K||^2 = ||y_R - Ax||^2 + ||y_K||^2$$

since $y_R - Ax \perp y_K$.

Therefore minimising $||Ax-y||$ reduces to minimising $||y_R-Ax||$. The optimal solution is the one that makes $||y_R - Ax|| = 0$ and thus $y_R = Ax$ which is possible since $y_R \in \mathrm{Ran}(A)$. Remember that $y_K \in \mathrm{Ker}(A^T)$ and $y_k = y - y_R = y - Ax$. Putting this together we have

$$A^T(y - Ax) = A^Ty_K = \mathbf{0}$$

implying that $A^TAx = A^Ty$ known as the *normal equation.*

Proposition 1 *If A has zero-null space then the solution of the least square problem*

$$\textit{minimise } ||Ax - y||^2$$

is unique and given by

$$x^* = (A^TA)^{-1}A^Ty$$

Proof Note that

$$\begin{aligned} ||Ax - y||^2 &= ||Ax - Ax^* + Ax^* - y||^2 \\ &= ||Ax - Ax^*||^2 + ||Ax^* - y||^2 + 2(Ax - Ax^*)^T(Ax^* - y) \\ &= ||Ax - Ax^*||^2 + ||Ax^* - y||^2 , \end{aligned}$$

since

$$(Ax - Ax^*)^T(Ax^* - y) = (x - x^*)^T(A^TAx^* - A^Ty) = 0$$

Hence $||Ax - y|| > ||Ax^* - y||$ unless $||Ax - Ax^*|| = 0$ which implies $x = x^*$ since A has zero-null space.

QED

The previous proposition enables us to transform the least square problem into solving a linear equation problem, namely solving the *normal equation*,

$$(A^TA)x^* = A^Ty \, .$$

Moreover when A has zero-nullspace, the matrix A^TA is invertible. It is square in $\mathbb{R}^{n\times n}$ and has zero-null space. Since if $A^TAx = 0$ then

$$x^TA^TAx = 0 \quad \Rightarrow \quad ||Ax||^2 = 0 \quad \Rightarrow \quad Ax = 0 \quad \Rightarrow \quad x = 0 \, .$$

In particular the normal equation admits a unique solution.

Let us further examine the consequences of the above proposition. First, note that

$$A^\dagger = (A^TA)^{-1}A^T$$

we have $A^\dagger A = I$. The matrix $A^\dagger$ is known as the *generalised (left) inverse or pseudoinverse of A (a.k.a. Moore-Penrose inverse).*

In addition the vector Ax^* is the orthogonal projection of y on the subspace $\mathrm{Ran}(A)$. More precisely, let

$$P = A(A^TA)^{-1}A^T = A^\dagger A^T$$

It is clear that $P^2 = P$ and that $P^T = P$ so P is an orthogonal projection. Now let us examine the range of P.

- We have that Ran(A) $\subset$ Ran(P). Indeed, if $x \in$ Ran(A) then there exists a z such that $Az = x$ and

  $$Px = A(A^TA)^{-1}A^TAz = Az = x \, ,$$

  so that $x \in$ Ran(P).
- Conversely, we have Ran(P) $\subset$ Ran(A). Since if $x \in$ Ran(P) then $PX = x$ and hence

  $$x = A\left((A^TA)^{-1}A^Tx\right) \in \mathrm{Ran(A)} \, .$$

To solve the least square problem one can perform the LU decomposition of A^TA and then solve the equation $A^TAx = A^Ty$. In fact the matrix A^TA is positive definite if A has zero-null space and one can perform the Cholesky decomposition described in the problem sheet, i.e. write $A^TA = L^TL$ where L is a lower triangular matrix in $\mathbb{R}^{n\times n}$. Subsequently, we do the following: (1) compute $z = A^Ty$ at the cost of $2mn$ operations, (2) perform the product A^TAat the cost of mn^2 operations as it is symmetric, (3) perform Cholesky on A^TA costing $1/3n^3$ operations, (4) solve $Lw = z(= A^Ty)$ and then $L^Tx = w$ each costing n^2 operations.

The total cost of the algorithm is $mn^2 + 1/3n^3 + 2mn + 2n^2$, roughly

$$mn^2 + 1/3n^3 \, .$$

Example 1 *Solve the least squares for* $A = \begin{pmatrix} 3 & -6 & 26 \\ 4 & -8 & -7 \\ 0 & 4 & 4 \\ 0 & -3 & -3 \end{pmatrix}$ *and* $y = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$.

Alternatively one could perform another decomposition known as the QR-decomposition of the matrix A, $A = QR$ where $Q \in \mathbb{R}^{m\times n}$ and $R \in \mathbb{R}^{n\times n}$. The columns of Q form an orthogonal basis of Ran(A). Note that when $m > n$ then Q is not square. The columns of Q are orthonormal so $Q^TQ = I$ however $QQ^T \neq I$.

Gram-Schmidt orthogonalisation procedure Let us first describe the Gram-Schmidt algorithm for computing an orthonormal basis $(q_1, \cdots, q_n)$ of the range of the matrix $A \in \mathbb{R}^{m\times n}$.

Let $a_1, \ldots, a_n \in \mathbb{R}^m$, $m \geq n$ be the column vectors of A. If A has zero-null space these vectors are linearly independent vectors (try to prove it). The Gram-Schmidt algorithm constructs an orthonormal basis for $\text{Span}(a_1, \ldots, a_n) = \text{Ran}(A)$ as follows

$$\begin{aligned} q_1 &= \frac{a_1}{||a_1||} \\ \hat{q}_2 &= (I - q_1 q_1^T) a_2 \\ q_2 &= \frac{\hat{q}_2}{||\hat{q}_2||} \\ \hat{q}_3 &= (I - q_1 q_1^T - q_2 q_2^T) a_3 \\ q_3 &= \frac{\hat{q}_3}{||\hat{q}_3||} \\ &\vdots \end{aligned}$$

Note that the matrices $q_j q_j^T$ defined above are orthogonal projectors and so are the matrices $I - q_1 q_1^T - \cdots q_k q_k^T$. The latter in the projection onto the orthogonal complement of $\text{Span}(q_1, \ldots, q_k)$: at each step, we construct a new vector q_k by projecting a_{k+1} into the orthogonal complement of the previous basis i.e. $q_1, \ldots, q_k$ and then normalising. Also note that for each $k = 1, \ldots, n$, we have (check it yourself)

$$\text{Span}(a_1, \ldots, a_k) = \text{Span}(q_1, \ldots, q_k) .$$

We obtain the matrix R as follows. Let $A = [a_1 \ldots a_n]$ and $Q = [q_1 \ldots q_n]$, then

$$r_{jk} = \begin{cases} q_j^T a_k, & j < k \\ ||\hat{q}_j||, & j = k \\ 0, & j > k. \end{cases}$$

We now describe a recursive procedures for computing such a QR decomposition. The QR decomposition of a matrix $A \in \mathbb{R}^{m\times n}$ with a zero nullspace is of the form $A = QR$ where $Q \in \mathbb{R}^{m\times n}$ is orthogonal and $R \in \mathbb{R}^{n\times n}$ upper triangular with positive diagonal entries.

Let $A = [a_1,\ A_2]$, $Q = [q_1,\ Q_2]$ and $R = \begin{pmatrix} r_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix}$ where $a_1, q_1 \in \mathbb{R}^m$, $A_2,\ Q_2 \in \mathbb{R}^{m\times(n-1)}$, $R_{12} \in \mathbb{R}^{1\times(n-1)}$ and $R_{22} \in \mathbb{R}^{(n-1)\times(n-1)}$.

We want Q to be orthogonal, so that

$$q_1^T q_1 = 1,\ Q_2^T Q_2 = I,\ q_1^T Q_2 = 0 .$$

Moreover we want $r_{11} > 0$ and R_{22} upper triangular.

QR factorisation (modified Gram-Schmidt method) Given a $A \in R^{m\times n}$ with zero null space

1. $r_{11} = ||a_1||$
2. $q_1 = \frac{a_1}{||a_1||}$
3. $R_{12} = q_1^T A_2$
4. Compute QR factorisation of $A_2 - q_1 R_{12} = Q_2 R_{22}$.

It suffices to show that first $A_2 - q_1 R_{12}$ has zero null space. Indeed

$$(A_2 - q_1 R_{12})x = \begin{pmatrix} a_1 & A_2 \end{pmatrix} \begin{pmatrix} -1/r_{11} R_{12} x \\ x \end{pmatrix} = A \begin{pmatrix} -1/r_{11} R_{12} x \\ x \end{pmatrix}$$

It can be shown that the total cost is $2mn^2$.

Solving Linear equations using QR To solve a set of linear equations $Ax = y$, when A is nonsingular. We first compute the QR decomposition of $A = QR$. note that

$$Ax = y \iff QRx = y \iff Rx = Q^T y$$

where the final system reduces to solving a linear system with an upper triangular matrix.

Solving least-squares using QR We can solve the least square problem using the QR decomposition. In fact $(A^T A) = R^T Q^T Q R = R^T R$ and $A^T y = R^T Q y$ and the normal equation above becomes

$$R^T R x^* = R^T Q y$$

In particular if we use the QR decomposition based on the modified Gram-Schmidt procedure we have R is invertible if A has zero-null space and the solving the normal equation reduces to solving

$$Rx = Q^T y$$

To sum up, given a $A \in R^{m\times n}$ with zero null space and a vector $y \in \mathbb{R}^m$, the least square solution to $Ax = y$ can be obtained as follows.

1. Compute QR factorisation $A = QR$
2. Compute $v = Q^T y$
3. Solve $Rx = v$.

It can be shown that the total cost is $2mn^2$.

To conclude let us point to the fact that despite being slower than the LU decomposition, the QR decomposition does not require the computation of the product $A^T A$ the result of which, after rounding up the entries of the matrix, might distort the solution of the problem (see examples in accompanying notes).

2 Singular value decomposition

The singular value decomposition is the most popular among the procedures for matrix decomposition as it does not require the matrix in question to be either square or zero-nullspace. Given a matrix $A \in \mathbb{R}^{m\times n}$ the SVD decomposition of A is given by

$$A = U \Sigma V^T$$

where $U \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{n\times n}$ orthogonal and $\Sigma \in \mathbb{R}^{m\times n}$ is diagonal with diagonal elements $\sigma_1 \geq \sigma_2 \geq \sigma_n \geq 0$ and such that the rank of A is given by the number of σ_i that are positive (non-equal to zero).

To construct the SVD of a matrix A we proceed in steps. First let us assume that $m\ geqn$ if not we apply the following to A^T to get the SVD of A by transposing the decomposition of A^T.

Step 1: Eingenvalues and Eingenvectos of A^TA The matrix $A^TA \in \mathbb{R}^{n\times n}$ is symmetric so it is diagonalisable[1], i.e. there exist $\lambda_1, \ldots, \lambda_n$ in $\mathbb{R}$ and $v_1, \ldots, v_n$ in $\mathbb{R}^n$ such that $A^TAv_i = \lambda_i v_i$, without loss of generality we can choose the v_is such that $||v_i|| = 1$ for all i. First, note that since $A^TAv_i = \lambda_i v_i$ then

$$v_i^T A^T A v_i = \lambda_i v_i^T v_i \quad \Rightarrow \quad \lambda_i = ||Av_i||^2 \,.$$

as $||v_i||^2 = 1$. It is not difficult to see that $\lambda_i = ||Av_i||^2 > 0$ unless $v_i \in \mathrm{Ker}(A)$. Let us denote $\sigma_j = \sqrt{\lambda_j}$ for all j.

Moreover if $\lambda_i \neq \lambda_j$ then

$$\lambda_i v_j^T v_i = v_j^T A^T A v_i = v_i^T A^T A v_j = \lambda_j v_i^T v_j$$

since $x^Ty = y^Tx$. Subsequently $v_j^T v_i = 0$ whenever $\lambda_i \neq \lambda_j$.

At the end of this step we have constructed an orthogonal basis of $\mathbb{R}^n$ corresponding to eigenvectors of A^TA. Note that if there are more than one eigenvector per eigenvalue, that are linearly independent, then we can use the Gram-Schmidt procedure to transform these vectors into as many orthonormal vectors.

Step 2: Orthonormal family in $\mathbb{R}^m$ We will define a family of vectors in $\mathbb{R}^m$ as follows.

- if $\sigma_i \neq 0$ let $u_i = \frac{Av_i}{\sigma_i}$ then
$$||u_i|| = \frac{v_i^T A^T A v_i}{\sigma_i^2} = 1 \,.$$
- if $\sigma_i = 0$, then the corresponding v_i in is $\mathrm{Ker}(A)$. Assume that we have already constructed $u_1, \ldots, u_{j-1}$ then we choose a unit vector u_i, $||u_i|| = 1$ from $(\mathrm{Span}(u_1, \ldots, u_{j-1}))^\perp$. This can be done by taking a vector not in $\mathrm{Span}(u_1, \ldots, u_{j-1})$ and apply the orthogonal projection $I - u_1u_1^T - \cdots - u_{j-1}u_{j-1}^T$ to remove the component in $\mathrm{Span}(u_1, \ldots, u_{j-1})$ and keep its component in $(\mathrm{Span}(u_1, \ldots, u_{j-1}))^\perp$ as in the Gram-Schmidt decomposition. By the above construction if either σ_j or σ_k is equal to zero then $u_j^T u_k = 0$ for $j \neq k$. If $\sigma_j \neq 0$ and $\sigma_k \neq 0$ then
$$u_j^T u_k = \frac{1}{\sigma_j \sigma_k}(Av_j)^T A v_k = \frac{1}{\sigma_j\sigma_k} v_j^T (A^T A v_k) = \frac{\lambda_k}{\sigma_j \sigma_k} v_j^T v_k = 0 \,,$$
since $v_i^T v_k = 0$

In this step we construct an orthonormal family of $\mathbb{R}^m$ which is not necessarily a basis as $m \geq n$.

Step 3: SVD Note that in step 2 we constructed a family of vectors in $\mathbb{R}^m$ such that $Av_i = \sigma_i u_i$. In matrix form this can be rewritten as

$$[Av_1 \; Av_2 \; \ldots \; Av_n] = [\sigma_1 u_1 \; \sigma_2 u_2 \; \ldots \; \sigma_n u_n]$$

or

$$A[v_1 1 \; v_2 \; \ldots \; v_n] = [u_1 \; u_2 \; \ldots \; u_n]\tilde{\Sigma}$$

where $\tilde{\Sigma}$ is in $\mathbb{R}^{n\times n}$ diagonal with diagonal elements $\sigma_1, \ldots, \sigma_n$. In particular we have that $V = [v_1 \, v_2 \, \ldots \, v_n]$ is square and orthogonal since its columns form an orthonormal family. This implies that $V^TV = I = VV^T$ so that

$$AV = \tilde{U}\tilde{\Sigma} \quad \Rightarrow \quad A = AVV^T = \tilde{U}\tilde{\Sigma}V^T \,.$$

This last decomposition looks very much like the SVD described above except that $\tilde{\Sigma}$ is in $\mathbb{R}^{n\times n}$ and $\tilde{U}$ is in $\mathbb{R}^{m\times n}$ with $\tilde{U}^T\tilde{U} = I$ since its column vectors form an orthonormal family.

[1] This is an important property of symmetric matrices that we will not prove here.

To complete the derivation, we first complete the family $u_1, \ldots, u_n$ to construct an orthonormal basis of $\mathbb{R}^m$. To this end, $j = n+1, \ldots, m$, we can add to $\{u_1, \ldots, u_{j-1}\}$ a vector $u_j$from $(\mathrm{Span}(u_1, \ldots, u_{j-1}))^\perp$ as before. We finally obtain an orthonormal basis $\{u_1, \ldots, u_m\}$ and we have that $U = [u_1\, u_2\, \ldots\, u_m]$ is an orthogonal matrix in $\mathbb{R}^{m\times m}$. It remains to define the matrix $\Sigma \in \mathbb{R}^{m\times n}$ by adding a block of zeros to the bottom of $\tilde{\Sigma}$ as follows

$$\Sigma = \begin{pmatrix} & \tilde{\Sigma} & \\ 0 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & 0 \end{pmatrix}$$

Theorem 1 *Let $A \in \mathbb{R}^{m\times n}$ (here m can bigger or smaller than n) then there exist matrices $U \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{n\times n}$ orthogonal and a diagonal matrix $\Sigma \in \mathbb{R}^{m\times n}$ with $\sigma_1 \geq \sigma_2 \geq \sigma_r \geq \sigma_{r+1} = \cdots = \sigma_{\min(m,n)} = 0$ such that $A = U\Sigma V^T$ and r is the rank of the matrix A, where $\min(m,n)$ is the minimum value between n and m.*

In fact, one can easily rewrite $A = U\Sigma V^T$ as

$$A = \sum_{i=1}^{\min(m,n)} \sigma_i u_i v_i^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T \, .$$

In particular let $x \in \mathbb{R}^n$ we have

$$Ax = \sum_{i=1}^{r} \sigma_i (v_i^T x) u_i \, ,$$

and $\mathrm{Ran}(A) = \mathrm{Span}\{u_1, \ldots, u_r\}$, $\mathrm{Ker}(A^T) = (\mathrm{Ran}(A))^T = \mathrm{Span}\{u_{r+1}, \ldots, u_m\}$, $\mathrm{Ran}(A^T) = \mathrm{Span}\{v_1, \ldots, v_r\}$, and $\mathrm{Ker}(A) = \left(\mathrm{Ran}(A^T)\right)^T = \mathrm{Span}\{v_{r+1}, \ldots, v_n\}$.

3 Condition number

Now we are going to use the idea of matrix norm to understand the impact of perturbations on the solution of a linear equation.

We define a matrix norm in terms of a given vector norm, we use the vector norm, denoted by $||x||$. Given a matrix $A \in \mathbb{R}^{m\times n}$, we define its norm by

$$||A|| = \max_{x \neq \mathbf{0}} \frac{||Ax||}{||x||} \, .$$

First note that for two matrices A and B,

$$||A\,B|| \leq ||A||\,|||B|| \, .$$

and that $||Ax|| \leq ||A||\,||x||$.

If we consider the Euclidean norm for vectors $||x||_2 = \sum_{i=1}^n x_i^2$, we have the following norm for matrices

$$||A|| = \sigma_1 \, .$$

where σ_1 is the square root of the largest eigenvalue of $A^T A$ defined in the SVD decomposition.

Suppose that we want to solve a linear equation $Ax = y$ where $A \in \mathbb{R}^{n\times n}$ non-singular, i.e. , there is unique solution $x = A^{-1}y$. Now suppose that we have a small perturbation of y, we replace y with $y + \Delta y$. The new solution is $x + \Delta x$ such that

$$x + \Delta x = A^{-1}(y + \Delta y) = A^{-1}y + A^{-1}\Delta y = x + A^{-1}\Delta y \, ,$$

so $\Delta x = A^{-1}\Delta y$. It is not difficult from the definition of the matrix norm

$$||A|| = \max_{\{x \in \mathbb{R}^n,\, x \neq \mathbf{0}\}} \frac{||Ax||}{||x||}$$

to see that

$$||\Delta x|| = ||A^{-1}\Delta y|| \leq |||A^{-1}||| \, ||\Delta y|| \, .$$

Hence if $||A^{-1}||$ is small, then small changes in y result in small changes in x, and vice versa if $||A^{-1}||$ is large. It is however more interesting to look at the relative error

$$\frac{||\Delta x||}{||x||} \leq ||A|| \, ||A^{-1}|| \frac{||\Delta y||}{||y||}$$

The parameter $\kappa(A) = ||A||||A^{-1}||$ is called the *condition number* of A. Using the fact that $||AB|| \leq ||A||||B||$ we easily see that

$$\kappa(A) = ||A|| \, ||A^{-1}|| \geq ||AA^{-1}|| = ||I|| = 1 \, .$$

Note that using the SVD decomposition of A we have $\kappa(A) = \frac{\sigma_1}{\sigma_n}$.

4 Exercises

Exercise 1 Express each of the following problems as a set of linear equations.

(a) Let $f(t) = c_0 + c_1 t + c_2 t^2 + c_3 t^3$ with $f(1) = 0,\ f(-1) = 0,\ f(0) = 1,\ f'(0) = 1$. Find $f(t)$.

(b) Let $g(t) = \frac{a_0 + a_1 t + a_2 t^2}{1 + b_1 t + b_2 t^2}$ with $g(1) = 2.3,\ g(2) = 4.8,\ g(3) = 8.9,\ g(4) = 16.9,\ g(5) = 41$. Find $g(t)$.

Exercise 2 Let A be a nonsingular lower triangular matrix of order n.

(a) What is the cost of computing A^{-1}?

(b) What is the cost of solving $Ax = y$ by first computing A^{-1} and then forming the matrix vector product $A^{-1}y$ to compute x? Compare with the the cost of the method for solving $Ax = y$ with A lower triangular matrix given in the notes.

Exercise 3 Are the following matrices positive definite?

(a) $A = \begin{pmatrix} -1 & 2 & 3 \\ 2 & 5 & -3 \\ 3 & -3 & 2 \end{pmatrix}$

(b) $A = \begin{pmatrix} 1 & a \\ a & 1 \end{pmatrix}$

Exercise 4 Perform the Cholesky decomposition of $A = \begin{pmatrix} 4 & 6 & 2 & -6 \\ 6 & 34 & 3 & -9 \\ 2 & 3 & 2 & -1 \\ -6 & -9 & -12 & 38 \end{pmatrix}$.

Exercise 5 You are given the Cholesky factorisation of $A = LL^T$, $A \in \mathbb{R}^{n\times n}$ positive definite.

(a) What is the Cholesky decomposition of $B = \begin{pmatrix} A & u^T \\ u & 1 \end{pmatrix}$ where $u \in \mathbb{R}^n$, $B \in \mathbb{R}^{(n+1)\times(n+1)}$

(b) What is the cost of computing the factorisation of B when the factorisation of A is given?

Exercise 6 Compute the LU factorisation of the matrix $A = \begin{pmatrix} -3 & 2 & 0 & 3 \\ 6 & -6 & 0 & -12 \\ -3 & 6 & -1 & 16 \\ 12 & -14 & -2 & -25 \end{pmatrix}$.

Exercise 7 For which values of $a_1, \dots a_n$ is the matrix $A \in \mathbb{R}^{n\times n}$ given below nonsingular?

$$A = \begin{pmatrix} a_1 & 1 & 0 & \dots & 0 & 0 \\ a_2 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n-2} & 0 & 0 & \cdots & 1 & 0 \\ a_{n-1} & 0 & 0 & \cdots & 0 & 1 \\ a_n & 0 & 0 & \dots & 0 & 0 \end{pmatrix}.$$

Assuming that A is non singular,

(a) how many flops do we need to solve $Ax = y$?

(b) What is the inverse of A?

Exercise 8 Compute the matrix norm of the following matrices. If A is non-singular, compute the norm of A^{-1} and the condition number $\kappa(A)$

(a) $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$

(b) $A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$

Exercise 9 Compute the QR decomposition of $A = \begin{pmatrix} 2 & 8 & 13 \\ 4 & 7 & -7 \\ 4 & -2 & -137 \end{pmatrix}$

Exercise 10 Formulate the following problems as least-square problems.

(a) Minimise $x_1^2 + 2x_2^2 + 3x_3^2 + (x_1 - x_2 + x_3 - 1)^2 + (-x_1 - 4x_2 + 2)^2$

(b) Minimise $2(-6x_2 + 4)^2 + 3(-4x_1 + 3x_2 - 1)^2 + 4(x_1 + 8x_2 - 3)^2$

Exercise 11 (2010) Let $A \in \mathbb{R}^{n\times n}$ be a non-singular matrix, and consider three vectors $b, c, f \in \mathbb{R}^n$. Given two real numbers α and γ we want to solve the following linear system in $x \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$.

$$\begin{aligned} Ax + b\lambda &= f \\ c^T x + \alpha\lambda &= \gamma\,. \end{aligned} \tag{1}$$

1. Write the system (1) in matrix form, i.e. $My = g$ with $M \in \mathbb{R}^{(n+1)\times(n+1)}$ and $y, g \in \mathbb{R}^{n+1}$.
2. Give a necessary and sufficient condition for the system (1) to admit a unique solution. Justify your answer.
3. In what follows we assume that $\alpha - c^T A^{-1} b \neq 0$.
4. To solve (1), we will use the following algorithm. Let z_0 be the solution of $Az = b$ and h_0 be the solution of $Ah = f$.
$$x = h_0 - \frac{\gamma - c^T h_0}{\alpha - c^T z_0} z_0, \quad \lambda = \frac{\gamma - c^T h_0}{\alpha - c^T z_0}\,.$$
   (a) Show that the above algorithm gives the solution to (1).
   (b) Assuming that we are given the solution of the two linear equations $Az = b$ and $Ah = f$, how many additional operations are required to complete the algorithm.

Exercise 12. Cholesky decompositions Let $A = (a_{ij})_{i,j=1,\ldots,n} \in \mathbb{R}^{n\times n}$ be a symmetric matrix, i.e, $A^T = A$ such that for all $x \in \mathbb{R}^n$ with $x \neq \mathbf{0}$ we have

$$x^T A x > 0\,.$$

Matrices satisfyng the above properties are known as *positive-definite matrices*

1. Let $e_i \in \mathbb{R}^n$ with all its entries equal to 0 except the i-th entry which is equal to 1. Show that, for $i = 1, \ldots, n$
$$a_{ii} = e_i^T A e_i > 0 \,.$$

2. Let C be the Schur complement of a_{11} in A, i.e.
$$C = A_{22} - \frac{1}{a_{11}} A_{21} A_{12} \,,$$
where
$$A = \begin{pmatrix} a_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$
with a_{11} is a scalar, $A_{21} \in \mathbb{R}^{n-1}$, and $A_{22} \in \mathbb{R}^{(n-1)\times(n-1)}$ and $A_{12} \in \mathbb{R}^{1\times(n-1)}$.
Justify the fact that
$$C = A_{22} - \frac{1}{a_{11}} A_{21} A_{21}^T \,.$$

3. Let $v \in \mathbb{R}^{n-1}$ and define $x \in \mathbb{R}^n$ such that
$$x = \begin{pmatrix} -(1/a_{11}) A_{21}^T v \\ v \end{pmatrix} \,.$$
Show that $x^T A x = v^T C v$ and that C is a positive-definite matrix.

4. In what follows we will show that there exists a lower-triangular matrix $L \in \mathbb{R}^{n\times n}$ such that $A = LL^T$. This factorisation is known as the *Cholesky decomposition*.

   (a) Let L be given by
   $$L = \begin{pmatrix} l_{11} & \mathbf{0}^T \\ L_{21} & L_{22} \end{pmatrix}$$
   with l_{11} is a scalar, $L_{21} \in \mathbb{R}^{n-1}$, and $L_{22} \in \mathbb{R}^{(n-1)\times(n-1)}$ and $\mathbf{0} \in \mathbb{R}^{n-1}$. Write the block structure of the matrix LL^T.

   (b) Let $A = LL^T$. Show that $l_{11} = \sqrt{a_{11}}$, $L_{21} = (1/l_{11})A_{21}$, and $L_{22}L_{22}^T = A_{22} - L_{21}L_{21}^T$.

   (c) Describe a recursive procedure to construct the lower-triangular matrix L such that $A = LL^T$.

   (d) Describe how one would use the above procedure to solve the linear equation $Ax = y$ for $A \in \mathbb{R}^{n\times n}$ positive definite.

   (e) Apply the Cholesky decomposition to the matrix
   $$A = \begin{pmatrix} 25 & 15 & -5 \\ 15 & 18 & 0 \\ -5 & 0 & 11 \end{pmatrix}$$
   and use it to solve the equation $Ax = y$ where $y = \begin{pmatrix} 30 \\ 15 \\ -16 \end{pmatrix}$.

Exercise 13 (2010) Let m and n be two positive integers with $m \leq n$. We consider $A \in \mathbb{R}^{(n+1)\times(m+1)}$ the matrix defined by

$$A = \begin{pmatrix} 1 & x_0 & \dots & x_0^m \\ 1 & x_1 & \dots & x_1^m \\ \vdots & \vdots & & \vdots \\ 1 & x_n & \dots & x_n^m \end{pmatrix},$$

where $x_0, \dots, x_n$ are n distinct real numbers.

Let $\mathbf{0}$ be the vector with all its entries equal to 0 (we will use the same notation for both the zero vector of $\mathbb{R}^{m+1}$ and the one of $\mathbb{R}^{n+1}$).

1. Let $v = \begin{pmatrix} v_0 \\ v_1 \\ \vdots \\ v_m \end{pmatrix} \in \mathbb{R}^{m+1}$.
   (a) Show that if $Av = \mathbf{0}$ then $v = \mathbf{0}$.
   Hint: Use the fact if the polynomial $P(x) = v_0 + v_1 x + \dots + v_m x^m$ has n distinct zeros then $P(x) = 0$.
   (b) Using the previous question, show that if $A^T A v = \mathbf{0}$ then $v = \mathbf{0}$.
   (c) Fix $y \in \mathbb{R}^{n+1}$. Justify the fact that the linear equation $A^T A x = A^T y$ admits a unique solution.
2. In the remainder of this problem, we will denote this solution by w, i.e.
   $$A^T A w = A^T y \,.$$
   For $v \in \mathbb{R}^{m+1}$ and $y \in \mathbb{R}^{n+1}$, define $g(v) = (y - Av)^T (y - Av)$.
   (a) Show that $g(w) = y^T y - y^T A w$, with w defined in 2. a) iii).
   (b) Prove that $g(v) - g(w) = (w - v)^T A^T A (w - v)$.
   (c) Show that for all $v \in \mathbb{R}^{m+1}$, we have $g(v) \geq g(w)$ and that $g(v) = g(w)$ if and only if $v = w$.
3. Let P be a polynomial such that $P(x) = \sum_{k=0}^{m} v_k x^k$. We define the quantity
   $$\Phi_m(P) = \sum_{i=0}^{n} (y_i - P(x_i))^2 \,.$$
   Let $v = \begin{pmatrix} v_0 \\ v_1 \\ \vdots \\ v_m \end{pmatrix} \in \mathbb{R}^{m+1}$ and $y = \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^{n+1}$.
   (a) Show that $\Phi_m(P) = g(v)$.
   (b) Using question 2.b), show that there exisist a polynomial P_w such that $\Phi_m(P) \geq \Phi_m(P_w)$.
4. We now apply the analysis of question 2) c) to a numerical example. Let $n = m = 3$, $x_0 = -1$, $x_1 = 0$, $x_2 = 1$, $x_3 = 2$ and $y_0 = 1$, $y_1 = 2$, , $y_2 = 1$, $y_3 = 0$.
   (a) Solve $A^T A v = A^T y$.
   (b) Derive the expression of the polynomial in $\mathbb{R}_3[X]$ that minimizes Φ_3 and give the minimum value of Φ_3 on $\mathbb{R}_3[X]$. Justify your answer.

Exercise 14 In this problem, we analyse the impact of perturbations on the solutions of linear equations.

1. We will consider the standard Euclidean norm $||x|| = \sqrt{x^T x}$, for $x \in \mathbb{R}^n$ and the associated matrix norm
$$|||A||| = \sup_{x:||x||=1} ||Ax|| \, .$$
   (a) Show that the mapping $A \to |||A|||$ defines a norm on $\mathbb{R}^{n\times n}$.
   (b) Let $x \in \mathbb{R}^n$, and A and B in $\mathbb{R}^{n\times n}$ show that $||Ax|| \leq |||A||| \, ||x||$ and that $|||AB||| \leq |||A||| \, |||B|||$.
2. In this question, we assume that A is a non-singular matrix in $\mathbb{R}^{n\times n}$ and y a non-zero vector in $\mathbb{R}^n$. Let $x_0 \in \mathbb{R}^n$ be the solution of $Ax = y$.
   (a) Let $x_1 \in \mathbb{R}^n$ be the solution of $Ax_1 = y + \delta y$, where $\delta y \in \mathbb{R}^n$. Prove that
$$\frac{||x_0 - x_1||}{||x_0||} \leq |||A||| \, |||A^{-1}||| \frac{||\delta y||}{||y||} \, .$$
   (b) Let $x_2 \in \mathbb{R}^n$ be the solution of $(A + \delta A)x_2 = y$, where $\delta A \in \mathbb{R}^{n\times n}$. Prove that
$$\frac{||x_0 - x_1||}{||x_0||} \leq |||A||| \, |||A^{-1}||| \frac{|||\delta A|||}{|||A|||} \, .$$
   (c) The coefficient $\kappa(A) = |||A||| \, |||A^{-1}|||$ is known a the *condition number* of A.
   Show that $\kappa(A) \geq 1$. Comment on the sensitivity of the solution of the equation $Ax = y$ to perturbations in terms of $\kappa(A)$.
3. Let $A \in \mathbb{R}^{n\times n}$ be a non-singular matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$.
   (a) Derive the eigenvalues of A^{-1}.
   (b) Show that $|||A||| \geq |\lambda_i|$ for all $i = 1, \ldots, n$.
   (c) Derive a lower bound for $\kappa(A)$ in terms of the λ_is.
   (d) Show that if A is (non singular) symmetric then
$$\kappa(A) = \max_{i=1\ldots,n} |\lambda_i| \max_{i=1\ldots,n} \left|\frac{1}{\lambda_i}\right| \, .$$
   Hint: Use the fact that if A is symmetric then there exists an orthonormal basis of eigenvectors of A.

Exercise 15 Consider $A = L+D+R$ with $a_{ii} \neq 0$ for $i = 1, \ldots, n$, where L is a lower triangular matrix with $l_{ii} = 0$, D a diagonal matrix ($d_{ii} \neq 0$) and R an upper triangular matrix with $r_{ii} = 0$ for all $i \in \{1, \ldots, n\}$:
$$A = \begin{pmatrix} 3 & 1 & -\frac{1}{4} & \frac{1}{2} \\ 1 & 5 & \frac{7}{5} & 3 \\ -\frac{2}{3} & \frac{1}{2} & 2 & -\frac{1}{8} \\ 1 & \frac{2}{3} & \frac{1}{6} & 7 \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ -\frac{2}{3} & \frac{1}{2} & 0 & 0 \\ 1 & \frac{2}{3} & \frac{1}{6} & 0 \end{pmatrix}}_{L} + \underbrace{\begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 7 \end{pmatrix}}_{D} + \underbrace{\begin{pmatrix} 0 & 1 & -\frac{1}{4} & \frac{1}{2} \\ 0 & 0 & \frac{7}{5} & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}}_{R}$$

To solve $Ax = b$ we describe the Jacobi Method. It proceeds in steps iterating
$$x^{(k+1)} = x^{(k)} + D^{-1}\left(b - Ax^{(k)}\right)$$
and for $k \to \infty$ the method converges to x^*, the correct solution. This is only possible if $|||D^{-1}(L + R)|||_2 < 1$.

1. Does the Jacobi Method converge for the above matrix A?
   If yes, conduct the method for A and use

$$b = \begin{pmatrix} 15 \\ 12 \\ 8 \\ 23 \end{pmatrix} \text{ and the start vector } x^{(0)} = \begin{pmatrix} 5 \\ -2 \\ 6 \\ 2 \end{pmatrix}.$$

   Iterate until you reach a vector $x^{(k)}$ where $\left|x^{(k)} - x^*\right| \leqslant 10^{-2}$ for all $i \in \{1, \ldots, 4\}$.
   How many steps do you need? You should at least allow an accuracy of 10^{-4} for each iteration.

2. Now let $B \in \mathbb{R}^{n\times n}$ be the the tridiagonal matrix

$$B = \text{tridiag}\,(c, a, b) = \begin{bmatrix} a & b & 0 & 0 & \cdots & 0 \\ c & a & b & 0 & \cdots & 0 \\ 0 & c & a & b & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & c & a & b \\ 0 & 0 & \cdots & 0 & c & a \end{bmatrix} \quad \text{where } bc > 0\ .$$

   Show that the eigenvalues are $\lambda_k = a + 2\sqrt{cb}\cos\left(\frac{k\pi}{n+1}\right)$ and eigenvectors are

$$u_k = \begin{pmatrix} \left(\frac{c}{b}\right)^{1/2} \sin\left(\frac{k\pi}{n+1}\right) \\ \left(\frac{c}{b}\right)^{2/2} \sin\left(\frac{2k\pi}{n+1}\right) \\ \vdots \\ \left(\frac{c}{b}\right)^{n/2} \sin\left(\frac{nk\pi}{n+1}\right) \end{pmatrix} \quad \text{for } B\ .$$

   Give the constraints for a, b and c such that the Jacobi-Method converges for all possible start vectors and all $n \in \mathbb{N}$.