Definiteness of matrices

Let $A \in \mathbb{R}^{n\times n}$. We define $\langle \mathbf{x}, \mathbf{y}\rangle_A = \mathbf{x}^T A\mathbf{y}$ for any vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Proposition: $\langle \mathbf{x}, \mathbf{y}\rangle_A$ is bilinear, and if A is symmetric, then this bilinear form is also symmetric.

Proof: It follows easily from the properties of matrix operations that

$$\begin{aligned}\langle \mathbf{x}+\mathbf{x}', \mathbf{y}\rangle_A =&\langle \mathbf{x}, \mathbf{y}\rangle_A + \langle \mathbf{x}', \mathbf{y}\rangle_A\\ \langle \mathbf{x}, \mathbf{y}+\mathbf{y}'\rangle_A =&\langle \mathbf{x}, \mathbf{y}\rangle_A + \langle \mathbf{x}, \mathbf{y}'\rangle_A\\ \langle \mathbf{x}, \lambda\mathbf{y}\rangle_A =&\lambda\langle \mathbf{x}, \mathbf{y}\rangle_A\\ \langle \lambda\mathbf{x}, \mathbf{y}\rangle_A =&\lambda\langle \mathbf{x}, \mathbf{y}\rangle_A\end{aligned}.$$

If A is symmetric, then $\langle \mathbf{y}, \mathbf{x}\rangle_A = \mathbf{y}^T A\mathbf{x} = (\mathbf{y}^T A\mathbf{x})^T = \mathbf{x}^T A^T\mathbf{y} = \mathbf{x}^T A\mathbf{y} = \langle \mathbf{x}, \mathbf{y}\rangle$, since the transposition does not change the 1×1 matrix $\mathbf{y}^T A\mathbf{x}$.

Def. The symmetric matrix A is

positive semidefinite if $\langle \mathbf{x}, \mathbf{x}\rangle_A \geq 0\ \forall\mathbf{x}$
positive definite if $\langle \mathbf{x}, \mathbf{x}\rangle_A > 0\ \forall\mathbf{x} \neq \mathbf{0}$
negative semidefinite if $\langle \mathbf{x}, \mathbf{x}\rangle_A \leq 0\ \forall\mathbf{x}$
negative definite if $\langle \mathbf{x}, \mathbf{x}\rangle_A < 0\ \forall\mathbf{x} \neq \mathbf{0}$
indefinit if $\exists\mathbf{x}, \mathbf{y}$: $\langle \mathbf{x}, \mathbf{x}\rangle_A > 0$ and $\langle\, \mathbf{y}, \mathbf{y}\,\rangle_A < 0$

Example: The identity matrix I defines the standard scalar product, so it is positive definite: $\mathbf{x}^T I\mathbf{x} = \mathbf{x}^T\mathbf{x} = |\mathbf{x}|^2 > 0$ if $\mathbf{x} \neq \mathbf{0}$.

Proposition: Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Then the following are equivalent:
(i) A is positive semidefinite;
(ii) all the eigenvalues of A are nonnegative;
(iii) $P^T AP$ is positive semidefinite for some/every invertible P.

Proof: $(i) \Rightarrow (ii)$: If $\mathbf{v}$ is an eigenvector with eigenvalue λ, then $0 \leq \langle \mathbf{v}, \mathbf{v}\rangle = \mathbf{v}^T A\mathbf{v} = \mathbf{v}^T\lambda\mathbf{v} = \lambda|\mathbf{v}|^2$, and $\mathbf{v}|^2 > 0$, so $\lambda \geq 0$.
$(ii) \Rightarrow (i)$: Since A is symmetric, all its eigenvalues in $\mathbb{C}$ are real, and there is an orthonormal basis $\mathcal{B} = \{\, \mathbf{b}_1, \dots, \mathbf{b}_n\,\}$ in $\mathbb{R}^n$ consisting of eigenvectors of A. Let $A\mathbf{b}_i = \lambda_i\mathbf{b}_i$. By assumption, $\lambda_i \geq 0$. Then for any $\langle \mathbf{b}_i, \mathbf{b}_j\rangle_A = \mathbf{b}_i^T A\mathbf{b}_j = \mathbf{b}_i\lambda_j\mathbf{b}_j = \lambda_j\langle \mathbf{b}_i, \mathbf{b}_j\rangle$, which is 0 if $i \neq j$, and λ_j otherwise, because $\lfloor$ is orthonormal. Now any $\mathbf{x} \in \mathbb{R}^n$ can be written in the form $\mathbf{x} = \sum_i x_i\mathbf{b}_i$, and then $\langle \mathbf{x}, \mathbf{x}\rangle_A = \sum_{i,j} x_i x_j\langle \mathbf{b}_i, \mathbf{b}_j\rangle_A = \sum_i x_i^2\lambda_i \geq 0$, since $\lambda_i \geq 0$ for every i.
$(i) \Rightarrow (iii)$: $\langle \mathbf{x}, \mathbf{x}\rangle_{P^TAP} = \mathbf{x}^T P^T AP\mathbf{x} = (P\mathbf{x})^T A(P\mathbf{x}) = \langle P\mathbf{x}, P\mathbf{x}\rangle \geq 0$ for every $\mathbf{x}$.
$(iii) \Rightarrow (i)$: We can apply $(i) \Rightarrow (iii)$ with P^{-1}.

Remark: A similar theorem can be proved in the same way for positive definite, negative definite and negative semidefinite, and indefinite is just the remaining case.

Example: For $A = \begin{bmatrix} 1 & -1 \\ -1 & 3 \end{bmatrix}$ the characteristic polynomial is $k_A(x) = x^2 - 4x + 2$, the eigenvalues are $2 \pm \sqrt{2} > 0$, so A is positive definite.

The eigenvalues are not always easy to find but we can use the equivalence of (i) and (iii) to find a nicer (actually a diagonal) matrix whose definiteness is the same as that of the original matrix but easier to determine.

If we use elementary row operations to bring the matrix to upper triangular form and we do the same operations on the columns, it means that we multiply the matrix from the left with some invertible matrices and from the right with their transposed matrix. So the new matrix will be of the form P^TAP, which is still symmetric ($(P^TAP)^T = P^TA^TP = P^TAP$), so if it is upper triangular, it must be diagonal. The eigenvalues of a diagonal matrix are its diagonal elements, so the definiteness of the diagonal matrix P^TAP can be determined by the signs of its diagonal elements.

Theorem: Let $A \in \mathbb{R}^{n\times n}$ be symmetric, and $D = P^TAP$. Then the three statements in the rows of the following table are equivalent.

A	the diagonal elements in D	the eigenvalues of A
positive definite, i.e. $\mathbf{x}^TA\mathbf{x} > 0\ \forall \mathbf{x} \neq \mathbf{0}$	all +	all +
negative definite, i.e. $\mathbf{x}^*A\mathbf{x} < 0\ \forall \mathbf{x} \neq \mathbf{0}$	all −	all −
positive semidefinite, i.e. $\mathbf{x}^*A\mathbf{x} \geq 0\ \forall \mathbf{x}$,	+ and possibly 0	+ and possibly 0
negative semidefinite, i.e. $\mathbf{x}^*A\mathbf{x} \leq 0\ \forall \mathbf{x}$,	− and possibly 0	− and possibly 0
indefinite, i.e. $\exists \mathbf{x}:\ \mathbf{x}^*A\mathbf{x} > 0$, and $\exists \mathbf{y}:\ \mathbf{y}^*A\mathbf{y} < 0$	+ and − and possibly 0	+ and − and possibly 0

Pl.: 1.

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 5 & 0 \\ 1 & 0 & 5 \end{bmatrix} \overset{row}{\mapsto} \begin{matrix} \\ r_1 - 2r_1 \\ r_3 - r_1 \end{matrix} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & -2 \\ 0 & -2 & 4 \end{bmatrix} \overset{column}{\mapsto} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & -2 & 4 \end{bmatrix} \overset{row}{\mapsto}$$

$$\begin{matrix} \\ \\ r_3 + 2r_2 \end{matrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{bmatrix} \overset{oszlop}{\mapsto} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

We always performed the same operations on the columns as on the rows: after $r_2 \mapsto r_2 - 2r_1$ and $r_3 \mapsto r_3 - r_1$ comes $c_2 \mapsto c_2 - 2c_1$ and $c_3 \mapsto c_3 - c_1$, and so on.

2.

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \overset{row}{\mapsto} r_1 + r_2 \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \overset{column}{\mapsto} \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} \overset{row}{\mapsto} \begin{bmatrix} 2 & 1 \\ 0 & -\frac{1}{2} \end{bmatrix} \overset{column}{\mapsto} \begin{bmatrix} 2 & 0 \\ 0 & -\frac{1}{2} \end{bmatrix}$$

So this A is indefinite.

Proposition: Let $A \in \mathbb{R}^{m\times n}$. Then $(A^TA)_{n\times n}$ is symmetric, positive semidefinite and $\operatorname{rank} A = \operatorname{rank} A^TA$, which is the number of positive eigenvalues of A^TA (with multiplicities).

Proof $(A^TA)^T = A^TA^{TT} = A^TA$, so A^TA is symmetric.
$\mathbf{x}^TA^TA\mathbf{x} = (A\mathbf{x})^T(A\mathbf{x}) \geq 0$ for every $\mathbf{x}$, so A^TA is positive semidefinite.
$\operatorname{Ker} A = \operatorname{Ker} A^TA$, since $A\mathbf{x} = \mathbf{0} \Rightarrow A^TA\mathbf{x} = A^T\mathbf{0} = \mathbf{0}$,
and $A^TA\mathbf{x} = \mathbf{0} \Rightarrow 0 = \mathbf{x}^TA^TA\mathbf{x} = (A\mathbf{x})^T(A\mathbf{x}) = |A\mathbf{x}|^2$, so $A\mathbf{x} = \mathbf{0}$. Hence $\operatorname{rank} A = n - \dim(\operatorname{Ker} A) = n - \dim(\operatorname{Ker} A^TA) = \operatorname{rank} A^TA$.
Finally, A^TA is similar to a diagonal matrix with the eigenvalues in the diagonal (the positive eigenvalues coming first). The rank of this diagonal matrix is the same as $\operatorname{rank} A$, and, since it is a row echelon matrix, it is the same as the number of non zero (thus positive) eigenvalues with multiplicities.

Reduced and full singular value decomposition

Def.: The singular values of $A \in \mathbb{R}^{m\times n}$ are $\sigma_1 \geq \cdots \geq \sigma_r > 0$, where $\sigma_1^2 \geq \ldots \geq \sigma_r^2 > 0$ are the positive eigenvalues of A^*A with multiplicities (the number of these is $r = \operatorname{rank} A^*A = \operatorname{rank} A$).

Def.: A matrix $A \in \mathbb{R}^{m\times n}$ is called semiorthogonal if its columns form an orthonormal system.

Theorem (Reduced SVD). *Let A be an $m\times n$ real matrix with rank r. Then there exist semiorthogonal matrices $U \in \mathbb{R}^{m\times r}$ and $V \in \mathbb{R}^{n\times r}$ such that with the diagonal matrix $\Sigma \in \mathbb{R}^{r\times r}$ whose diagonal elements are the singular values $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ in this order, we have $A = U\Sigma V^T$. This is the reduced singular value decomposition of A.*

Proof: The proof also gives an algorithm for calculating the decomposition. Let the positive eigenvalues of A^TA be $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r > 0$, and $\{\mathbf{b}_1, \ldots, \mathbf{b}_r\}$ be an orthonormal system of eigenvectors for the given eigenvalues (such a sytem exists since A^TA is a symmetric matrix). Then $V = [\mathbf{b}_1 \ldots \mathbf{b}_r]$ is a semiorthogonal matrix. Furthermore, $\langle A\mathbf{b}_i, A\mathbf{b}_j\rangle = \mathbf{b}_i^TA^TA\mathbf{b}_j = \lambda_j\mathbf{b}_i^T\mathbf{b}_j = 0$ if $i \neq j$ and λ_i if $i = j$, so columns of the matrix AV are also orthogonal, where the lengthes of the column vectors are $\sigma_1, \ldots, \sigma_r$, so $U = AV\Sigma^{-1}$ is a semiorthogonal matrix. Hence $U\Sigma V^T = AVV^T$.

We only need to prove that $AVV^T = A$. We could complete the orthonormal system $\{\mathbf{b}_1, \ldots \mathbf{b}_r\}$ to a basis of $\mathbb{R}^n$ with eigenvectors $\mathbf{b}_{r+1}, \ldots, b_n$ of A^TA for the eigenvalue 0. Let $M = [\mathbf{b}_1 \ldots \mathbf{b}_n] = [V|T]$. Than M is invertible, and $AVV^TM = [AVV^TV|AVV^TT] = [AV|0] = A[V|T] = AM$, since $\mathbf{b}_{r+1}, \ldots, \mathbf{b}_n \in \operatorname{Ker} A^TA = \operatorname{Ker} A$. We can simplify with M, and we get that $AVV^T = A$.

Example:

Pl.: $A = \begin{bmatrix} 1 & 2 \\ -2 & -4 \end{bmatrix}$, $A^*A = \begin{bmatrix} 5 & 10 \\ 10 & 20 \end{bmatrix}$, $k_{A^*A}(x) = x^2 - 25x = x(x-25)$, $\lambda_1 = 25$, $\lambda_2 = 0$, $\sigma_1 = 5$, $\Sigma = [5]$,
The eigenvector of A^*A for λ_1: $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$, or to have an eigenvecor of length 1, it is $\frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ 2 \end{bmatrix}$.
$V = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $AV = \frac{1}{\sqrt{5}}\begin{bmatrix} 5 \\ -10 \end{bmatrix}$, $U = AV\Sigma^{-1} = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ -2 \end{bmatrix}$.
$A = UAV^T = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ -2 \end{bmatrix} \cdot [5] \cdot \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & 2 \end{bmatrix}$.

Theorem (Full SVD). *Let A be an $m \times n$ real matrix with rank r. Then there exist $M \in \mathbb{R}^{n\times n}$ and $M' \in \mathbb{R}^{m\times m}$ orthogonal matrice such that $A = M'\Sigma' M^T$, where $\Sigma' = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}$ is an $m \times n$-es block matrix, and Σ is the diagonal matrix with the singular values $\sigma_1 \geq \ldots \geq \sigma_r$ in its diagonal. This is the full singular value decomposition of A.*

We can complete the U and V of the reduced SVD to orthogonal matrices: $M = [V|V']$ and $M' = [U|U']$. Then

$$M'\Sigma' M^T = [U \quad U'] \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V^T \\ V \end{bmatrix} = [U\Sigma \quad 0] \begin{bmatrix} V^T \\ V \end{bmatrix} = U\Sigma V^T = A.$$

Example: Find the full SVD of the matrix of the previous example. $M = [V \quad V'] = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}$, and $M' = [U \quad U'] = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}$. Then $A = M'\Sigma' M^T = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 5 & 0 \\ 0 & 0 \end{bmatrix} \cdot \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}$.

Applications of the SVD

Def.: For $A \in \mathbb{R}^{m\times n}$, A^+ is a pseudoinverse of A if $AA^+A = A$, $A^+AA^+ = A^+$, and AA^+ and A^+A are symmetric. The pseudoinverse always exists, and it is unique.

Theorem: If $A = U\Sigma V^T$ a reduced SVD, then $A^+ = V\Sigma^{-1}U^T$

Proof: It is easy to check the four properties of the pseudoinverse, using that $V^TV = I_{r\times r}$ and $U^TU = I_{r\times r}$.

Theorem (Best approximate solution): If $A \in \mathbb{R}^{m\times n}$, and A^+ its pseudoinverse, then the best approximate solution of a (possibly inconsistent) system of equation, $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = A^+\mathbf{b}$. This means that $|A\mathbf{x} - \mathbf{x}|$ is minimal for $\mathbf{x} = A^+\mathbf{b}$.

Proof: The error of the approximate solution $A^+\mathbf{b}$ is $\mathbf{e} := AA^+\mathbf{b} - \mathbf{b}$. We want to prove that $|\mathbf{e}| \leq |A\mathbf{x} - \mathbf{b}|$ for any $\mathbf{x}$. First we show that $\mathbf{e}$ is orthogonal to $\operatorname{Im} A$. For any $\mathbf{x}$,

$$(A\mathbf{x})^TAA^+\mathbf{b} = \mathbf{x}^TA^T(AA^+)\mathbf{b} = \mathbf{x}^TA^T(AA^+)^T\mathbf{b} = \mathbf{x}^T(AA^+A)^T\mathbf{b} = \mathbf{x}^TA^T\mathbf{b} = (A\mathbf{x})^T\mathbf{b},$$

so $\langle A\mathbf{x}, \mathbf{e}\rangle = (A\mathbf{x})^T(A(A^+)b - \mathbf{b}) = 0$. (Here we used that AA^+ is symmetric, and $AA^+A = A$.) This implies that $|A\mathbf{x} - \mathbf{b}|^2 = |(A\mathbf{x} - AA^+) + \mathbf{e}|^2 = \langle A(\mathbf{x} - A^+) + \mathbf{e},\ A(\mathbf{x} - A^+) + \mathbf{e}\rangle = \langle A(\mathbf{x} - A^+),\ A(\mathbf{x} - A^+)\rangle + \langle \mathbf{e}, \mathbf{e}\rangle = |A(\mathbf{x} - A^+)|^2 + |\mathbf{e}|^2 \geq |\mathbf{e}|^2$.

Eckart–Young theorem about low-rank approximation: Let $A = U\Sigma V^T$ be a reduced SVD, and $d < r = \operatorname{rank}(A)$. Then the best approximating matrix of rank at most d for A is

$$A^{(d)} = U^{(d)}\Sigma^{(d)}(V^{(d)})^T,$$

where $U^{(d)}$ and $V^{(d)}$ consist the first d columns of U and V, respectively and $\Sigma^{(d)}$ is the left upper $d \times d$ submatrix of Σ. Here best approximating matrix means that $||A - M||$ is minimal among the matrices M with $\operatorname{rank} M \leq d$ if $M = A^{(d)}$, and for a matrix M, the norm of $||M||$ is $\sqrt{\sum_{i,j} m_{ij}^2}$.