Linear Algebra Fact Sheet I

Amir Kadivar, October 2017

Warning: *This is not intended for learning or for memorization; only use it to review what you have already learned or to find gaps in your knowledge.* **Notation:** *numbers (scalars) are denoted by $a, x, \alpha, \ldots$ and matrices by $A, B, C \ldots$ We write $A_{m\times n}$ to indicate that A has m rows and n columns. Single column matrices (vectors) are denoted by bold letters $\boldsymbol{b}, \boldsymbol{x} \ldots$. An all-zero matrix (or vector) is denoted by $\mathbf{0}$ and its dimensions are to be understood from context.*

I Matrix arithmetic

1. If α is a scalar then $\alpha A_{m\times n}$ is the matrix obtained by multiplying all entries in A by α; i.e αA is an $m \times n$ matrix whose entries are αa_{ij}.
2. The sum $A_{m\times n} + B_{p\times q}$ is defined iff A and B have identical dimensions, i.e $m = p$ and $n = q$. In this case $A + B$ has the same dimensions and has $a_{ij} + b_{ij}$ at row i and column j.
3. We always have: $A + B = B + A, \quad \alpha A + \beta A = (\alpha + \beta)A, \quad \alpha(A + B) = \alpha A + \alpha B$
4. The product $A_{m\times n}B_{p\times q}$ is defined iff $n = p$ in which case the dimensions of the product is $m \times q$. The j-th column of AB is equal to AB_j where B_j is the j-th column of B. The i-th row of AB is equal to A_iB where A_i is the i-th row of A.
5. We always have: $A(BC) = (AB)C, \quad A(B + C) = AB + AC, \quad (\alpha A)B = A(\alpha B) = \alpha(AB)$
6. For any $A_{m\times n}$ we have $AI_m = I_nA = A$. However, it is *not* generally true that $AB = BA$.
7. A matrix $A_{m\times n}$ is *square* if $m = n$. A square matrix $A_{n\times n}$ is *diagonal* if all its off-diagonal entries are zero, i.e $a_{ij} = 0$ for all $i \neq j$. The diagonal entries of a diagonal matrix may or may not be zero. The identity matrix I_n is an $n \times n$ diagonal matrix all whose diagonal entries a_{ii} are 1.
8. The *transpose* of a matrix $A_{m\times n}$, denoted by A^T, is an $n \times m$ matrix whose rows are the columns of A and whose columns are the rows of A. The entry at row i and column j of A^T is a_{ji}.
9. We always have:
$$\left(A^T\right)^T = A, \quad \left(\alpha A\right)^T = \alpha A^T, \quad \left(A + B\right)^T = A^T + B^T, \quad \left(AB\right)^T = B^TA^T$$
10. Multiplying a matrix $A_{m\times n}$ by a diagonal matrix from the left (right) multiplies each of its rows (columns) by the corresponding diagonal entry:
$$\begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_m \end{pmatrix} \begin{pmatrix} A_1 \\ \vdots \\ A_m \end{pmatrix} = \begin{pmatrix} d_1A_1 \\ \vdots \\ d_mA_m \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} A_1 & \cdots & A_n \end{pmatrix} \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix} = \begin{pmatrix} d_1A_1 & \cdots & d_nA_n \end{pmatrix}$$
and as a special case, we always have: $(\alpha I_m)A = A(\alpha I_n) = \alpha A_{m\times n}$.

II Systems of linear equations

1. A system of m linear equations in n variables (or unknowns) is a collection of linear equations $a_{i1}x_1 + \ldots + a_{in}x_n = b_i$ for $i = 1, \ldots, m$ where a_{ij} and b_i are given. A *solution* to a system is a collection of n numbers $x_1, \ldots, x_n$ that satisfy *all* equations in the system. This can be summarized in matrix notation as $A_{m\times n}\boldsymbol{x} = \boldsymbol{b}$ where A is the *matrix of coefficients* and:
$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}, \quad \boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$$
2. The system $A\boldsymbol{x} = \boldsymbol{b}$ is called *homogeneous* if $\boldsymbol{b} = \mathbf{0}$ and *inhomogeneous* otherwise.
3. A homogeneous system always has a solution ($\boldsymbol{x} = \mathbf{0}$ is always a solution). If $\boldsymbol{x} = \mathbf{0}$ is a solution to a system it is referred to as the *trivial* solution. A homogenous system *always* has a trivial solution and an inhomogenous system *never* has a trivial solution.
4. If the system $A_{m\times n}\boldsymbol{x} = \boldsymbol{b}$ has more than one solution it automatically has infinitely many solutions which can be expressed parametrically. For example, a one parameter solution set has the form $\boldsymbol{x} = \boldsymbol{y}_\circ + s\boldsymbol{y}_1$ where s can be any number (the parameter); a two parameter solution set has the form $\boldsymbol{x} = \boldsymbol{y}_\circ + s\boldsymbol{y}_1 + t\boldsymbol{y}_2$ where the parameters s, t can be any two numbers; and so on. The choice of parameterization is *not* unique; in fact, there are always infinitely many parameterizations.

5. For book-keeping purposes we form an *augmented matrix* by adding $\boldsymbol{b}$ as an additional column of $A_{m\times n}$ and denote it by $(A|\boldsymbol{b})$.
6. We can combine the equations in a system to simplify our calculation and to gain insight into its solution set. *Elementary row operations* are a special class of operations that do not modify the solution set of a system. There are three types of elementary row operations:
   - swapping two rows $R_i \leftrightarrow R_j$,
   - multiplying a row by a *nonzero* scalar $R_i \leftarrow \alpha R_i$,
   - adding a row to another $R_i \leftarrow R_i + R_j$.
7. The *leading nonzero* entry in row i of a matrix $A_{m\times n}$ is the leftmost entry in the i-th row of A that is nonzero. If the leading nonzero is 1 we refer to it as the *leading one*.
8. A matrix $A_{m\times n}$ is in *row echelon form* iff:
   - all-zero rows are all below other rows, and
   - the leading nonzero of any row is to the left of the leading nonzeros of all rows below it.
9. Row echelon form can be obtained by applying elementary row operations as long as there are any two rows with leading nonzeros in the same column. In this case either of the two leading nonzeros can be made zero using the the other.
10. A matrix $A_{m\times n}$ is in *reduced row echelon form* iff:
    - it is in row echelon form, and
    - all leading nonzeros are ones (hence leading ones), and
    - for every leading one all other entries in its column are zero.
11. The reduced row echelon form of a matrix is always unique but the row echelon form is not. However, all important properties of a matrix can be read directly off of either of its row echelon forms (including the unique reduced row echelon form).
12. The *rank* r of a matrix $A_{m\times n}$ is (equivalent definitions):
    - the number of rows that are *not* all-zero in row echelon form (reduced or not).
    - the number of leading nonzeros in row echelon form (reduced or not).
    - the number of leading ones in reduced row echelon form.
13. If the rank of $A_{m\times n}$ is r we always have $r \leq n$ and $r \leq m$, i.e rank can not exceed either of the dimensions of the matrix.
14. A system $A_{m\times n}\boldsymbol{x} = \boldsymbol{b}$ is *consistent* iff in its row echelon form we have (equivalent definitions):
    - for every all-zero row in the coefficients matrix, the corresponding right-hand-side is zero.
    - the rank of the coefficients matrix is the same as the rank of the augmented matrix.
15. A homogenous system is always consistent; an inhomogeneous system may or may not be.
16. Consider the system $A_{m\times n}\boldsymbol{x} = \boldsymbol{b}$ where the rank of $A_{m\times n}$ is r:
    - a solution exists iff the system is consistent.
    - the solution is unique iff the system is consistent and $r = n$.
    - there are infinitely many solutions if the system is consistent and $r < n$ in which case the solutions can be described parametrically using precisely $n - r$ parameters.
17. Among consistency, rank, and number of parameters needed to describe solutions, the right-hand-side $\boldsymbol{b}$ only affects consistency (and hence existence of solutions). The rank and the number of parameters are completely determined by the matrix of coefficients $A_{m\times n}$ regardless of $\boldsymbol{b}$.
18. The effect of each elementary row operation on the system $A_{m\times n}\boldsymbol{x} = \boldsymbol{b}$ can be represented in matrix language by multiplying the coefficients matrix A and $\boldsymbol{b}$ by an *elementary matrix* $E_{m\times m}$ from the left:
    $$\big(A|\boldsymbol{b}\big) \to E\big(A|\boldsymbol{b}\big) = \big(EA|E\boldsymbol{b}\big)$$
19. The elementary matrix corresponding to an elementary operation on $A_{m\times n}$ can be obtained by applying the same row operation on the identity matrix I_m.
20. All elementary matrices are invertible and their inverses correspond to the inverse row operation:
    - The inverse of $R_i \leftrightarrow R_j$ is itself.
    - The inverse of $R_i \leftarrow \alpha R_i$ is $R_i \leftarrow \frac{1}{\alpha} R_i$.
    - The inverse of $R_i \leftarrow R_i + R_j$ is $R_i \leftarrow R_i - R_j$.
21. The effect of applying k elementary row operations, represented by elementary matrices $E_1, \ldots, E_k$, is equivalent to left-multiplication by the product matrix $E_k \ldots E_2 E_1$, namely:
    $$\big(A|\boldsymbol{b}\big) \to E_1\big(A|\boldsymbol{b}\big) \to E_2E_1\big(A|\boldsymbol{b}\big) \to \cdots \to E_k \ldots E_2E_1\big(A|\boldsymbol{b}\big)$$

III Invertible matrices and determinants

1. Here we are only concerned with square matrices. The inverse of $A_{n\times n}$ is another matrix $A^{-1}_{n\times n}$ which "undoes" the effect of A, namely: $AA^{-1} = A^{-1}A = I_n$. An inverse, if it exists, is unique. We only need to check one of the above equalities (i.e. if $AA^{-1} = I$ then $A^{-1}A$ is automatically I).
2. We always have:

$$\left(A^T\right)^{-1} = \left(A^{-1}\right)^T, \quad \left(AB\right)^{-1} = B^{-1}A^{-1}, \quad \left(\alpha A\right)^{-1} = \frac{1}{\alpha}A^{-1}$$

   There is no relationship in general between $(A+B)^{-1}$ and A^{-1}, B^{-1}. In fact $A+B$ may not be invertible even if both A and B are.
3. If A is invertible then any system $A\boldsymbol{x} = \boldsymbol{b}$ (homogenous or not) has a unique solution given by $\boldsymbol{x} = A^{-1}\boldsymbol{b}$.
4. If $A_{n\times n}$ is invertible then its reduced row echelon form is necessarily I_n and we have:

$$\left(A_{n\times n}|I_n\right) \to \left(E_1A|E_1\right) \to \cdots \to \left(E_k\dots E_1A|E_k\dots E_1\right) = \left(I_n|A^{-1}\right)$$

5. The product $E_k\dots E_2E_1$ of elementary row operations that bring a matrix A to reduced row echelon form is exactly its inverse $A^{-1} = E_k\dots E_1$.
6. The inverse of an elementary matrix is also an elementary matrix. Therefore if $A^{-1} = E_k\dots E_1$ then A can be written as a product of elementary matrices $A = E_1^{-1}\dots E_k^{-1}$.
7. The *determinant* of a square matrix is a number that satisfies the following properties:
   - the determinant of the identity matrix is 1.
   - multiplying any single row or any single column by α multiplies the determinant by α.
   - swapping any two rows or any two columns flips the sign of the determinant.
8. Adding a multiple of a row (or column) to another does not affect the determinant. Therefore, if a matrix has two rows (or columns) that are scalar multiples of each other, its determinant is zero.
9. The (i,j)-*minor* of a $A_{n\times n}$ is the determinant of the $(n-1)\times(n-1)$ matrix obtained from A by excluding row i and column j of A. The (i,j)-*cofactor* of A denoted by c_{ij} is $(-1)^{i+j}$ times the corresponding minor.
10. The collection of all cofactors c_{ij} forms the *cofactor* matrix whose transpose is called the *adjoint* matrix $\text{adj}(A)$. Both the cofactor matrix and the adjoint matrix have the same dimensions as A. Entry at row i and column j of $\text{adj}(A)$ is the (j,i)-cofactor c_{ji}.
11. Any row or column of A can be used to calculate its determinant in terms of a row or column of its cofactors. For any row i or column j we have:

$$\det(A) = a_{i1}c_{i1} + a_{i2}c_{i2} + \ldots + a_{in}c_{in} = a_{1j}c_{1j} + a_{2j}c_{2j} + \ldots + a_{nj}c_{nj}$$

12. We always have:

$$\det(AB) = \det(A)\det(B), \quad \det\left(A^{-1}\right) = \frac{1}{\det A}$$
$$\det\left(\alpha A_{n\times n}\right) = \alpha^n \det\left(A\right), \quad \det\left(A^T\right) = \det\left(A\right)$$

13. There is no relationship in general between $\det(A+B)$ and $\det(A)$ and $\det(B)$.
14. We always have:

$$A\ \text{adj}(A) = \det(A)I, \quad \det\left(\text{adj}(A_{n\times n})\right) = \det(A)^{n-1}$$

15. If A is invertible the above imply: $A^{-1} = \frac{1}{\det(A)}\text{adj}(A)$.
16. The following statements are equivalent (all pairs are "if and only if"):
   - $A_{n\times n}$ is invertible.
   - The system $A\boldsymbol{x} = \boldsymbol{b}$ has a unique solution for any $\boldsymbol{b}$.
   - The homogenous system $A\boldsymbol{x} = \boldsymbol{0}$ has a unique solution (i.e. the trivial solution).
   - The reduced row echelon form of A is I_n.
   - A is a product of elementary matrices.
   - $\det(A) \neq 0$.

IV Eigenvalues, eigenvectors, and diagonalizable matrices

1. A vector $\boldsymbol{x}$ is an *eigenvector* for *eigenvalue* λ of a matrix $A_{n\times n}$ if we have $A\boldsymbol{x} = \lambda\boldsymbol{x}$. A number λ is an eigenvalue of A if it has a nonzero eigenvector, i.e. $\boldsymbol{x} \neq \boldsymbol{0}$ such that $A\boldsymbol{x} = \lambda\boldsymbol{x}$.

2. The equation $A\boldsymbol{x} = \lambda\boldsymbol{x}$ can always be rewritten as a homogenous system $(A - \lambda I_n)\boldsymbol{x} = \mathbf{0}$ or $(\lambda I_n - A)\boldsymbol{x} = \mathbf{0}$. Consequently, λ is an eigenvalue of A if the corresponding system with coefficients $\lambda I_n - A$ has a *nontrivial* solution; hence, any eigenvalue has infinitely many eigenvectors.
3. A matrix $A_{n\times n}$ has at most n and at least 0 eigenvalues. Both extremes are possible (A may have exactly n or 0 eigenvalues).
4. If λ is an eigenvalue of A then the infinitely many solutions to $(\lambda I_n - A)\boldsymbol{x} = \mathbf{0}$, which are the eigenvectors of λ, can be described parametrically.
5. The *geometric multiplicity* of an eigenvalue λ of $A_{n\times n}$ is (equivalent definitions):
   - The number of parameters needed to describe the solutions of $(\lambda I_n - A)\boldsymbol{x} = \mathbf{0}$.
   - the difference $n - r$ where r is the rank of $\lambda I_n - A$.
6. $\det(\lambda I_n - A)$ is a polynomial in λ of degree n and is called the *characteristic polynomial* of A. Its real roots are precisely the eigenvalues of A.
7. The following statements are equivalent (all pairs are "if and only if"):
   - λ is an eigenvalue of A.
   - The homogenous system $(\lambda I_n - A)\boldsymbol{x} = \mathbf{0}$ has a nontrivial solution (and hence infinitely many).
   - $\lambda I_n - A$ is *not* invertible.
   - $\det(\lambda I_n - A) = 0$.
   - λ is a real root of the characteristic polynomial of A.
8. We know that any polynomial of degree n can be decomposed into first order and second order factors. First order factors specify all the real roots. Therefore, the factorization of the characteristic polynomial $c_A(\lambda) = (\lambda - \lambda_1)^{k_1}(\lambda - \lambda_2)^{k_2}\dots(\lambda^2 + a_1\lambda + b_1)\dots$ specifies all eigenvalues: any first order factor $(\lambda - \lambda_i)^{k_i}$ specifies a distinct eigenvalue λ_i whose *algebraic multiplicity* is k_i.
9. The sum of algebraic multiplicities of all eigenvalues is at most n and is strictly less than n iff there are any second order factors (e.g. $\lambda^2 + 1$) in the final factorization of $c_A(\lambda)$.
10. For any eigenvalue both geometric and algebraic multiplicities are at least 1. Furthermore, the geometric multiplicity of any eigenvalue is always at most equal to its algebraic multiplicity. Consequently, if the algebraic multiplicity of an eigenvalue is 1 its geometric multiplicity is also 1.
11. A matrix $A_{n\times n}$ is *diagonalizable* if (equivalent definitions):
    - there exists an invertible matrix $P_{n\times n}$ and a diagonal matrix $D_{n\times n}$ such that $A = PDP^{-1}$.
    - the characteristic polynomial $c_A(\lambda)$ has no complex roots (i.e. first order factors only) *and* the geometric multiplicity of each eigenvalue matches its algebraic multiplicity.
    - the sum of geometric multiplicities of eigenvalues of A equals n.
12. If the characteristic polynomial of $A_{n\times n}$ has n *distinct* real roots then A is diagonalizable. The converse is not true, i.e. A could be diagonalizable without having distinct eigenvalues (e.g. the identity matrix).
13. If $A = PDP^{-1}$ the diagonal entries of D are necessarily eigenvalues of A, the columns of P are corresponding eigenvectors, and the algebraic and geometric multiplicity of each eigenvalue matches the number of times it repeats on the diagonal of D.
14. If $\lambda_1, \lambda_2, \dots$ are eigenvalues of A with algebraic and geometric multiplicities $k_1, k_2, \dots$ such that $k_1 + k_2 + \dots = n$, then we can construct P and D such that $A = PDP^{-1}$ as follows. For each eigenvalue λ_i we find a parametric description of its eigenvectors using k_i parameters and k_i *basic eigenvectors*. If we let $P_{n\times n}$ be the matrix with all basic eigenvectors as its columns and the matrix $D_{n\times n}$ be the diagonal matrix with the corresponding eigenvalues (in correct order) then P is necessarily an invertible matrix and we have $A = PDP^{-1}$.

V Case Study: Diagonal matrices

Let $D_{n\times n}$ be a diagonal matrix with diagonal entries $d_1, \dots, d_n$.

1. For any k in $\mathbb{N}$, the k-th power D^k is another $n \times n$ diagonal matrix with diagonal entries $d_1^k, \dots, d_n^k$.
2. The rank of D is the number of its nonzero diagonal entries.
3. The determinant of D is the product $d_1 \dots d_n$ of its diagonal entries. Consequently, D is invertible iff all d_i are nonzero. In this case, D^{-1} is the diagonal matrix with entries $1/d_1, \dots, 1/d_n$.
4. D is diagonalizable regardless of the values of d_i (it is already diagonal; to check definition set $P = I$). The eigenvalues of D are $d_1, \dots, d_n$ where repeats reflect the multiplicity of each eigenvalue.
5. A diagonalizable matrix $A = PDP^{-1}$ is invertible iff D, the diagonal matrix of its eigenvalues, is invertible, i.e iff all eigenvalues of A are nonzero. In this case, we have $A^{-1} = PD^{-1}P^{-1}$.