MIT 18.06: Linear Algebra - Notes

Tomasz Bartkowiak
https://github.com/bartkowiaktomasz

April 2019

Contents

1 The geometry of linear equations

No notes.

2 Elimination with matrices

No notes.

3 Multiplication and inverse matrices

If we multiply two matrices $A \cdot B = C$ then columns of C are combinations of columns of A , and rows of C are combinations of rows of B.

4 Factorization into $A = LU$

The cost of elimination of matrix A is $O(n^3)$ because we are eliminating each column of a matrix of size $n, n-1, ..., 1$ and each of the elimination costs n^2 and

$$1 + 2^2 + ... + (n-1)^2 + n^2 = 1/3n^3 + 1/2n^2 + 1/6n$$

If we want to solve the system, then elimination for B in $Ax = B$ is just $O(n^2)$.

The row exchange matrices are called permutation matrices and form a group. For a permutation matrix, a transpose is its inverse, i.e $P^T = P^{-1}$.

5 Transposes, permutations, spaces R^n

For symmetric matrices: $A^T = A$. Also, for any matrix A , $A^T A$ is always symmetric.

Vector space must satisfy:

1. Associativity of addition
2. Commutativity of addition

3. Identity element of addition
4. Inverse element of addition
5. Compatibility of scalar multiplication with field multiplication
6. Identity element of scalar multiplication
7. Distributivity of scalar multiplication with respect to vector addition
8. Distributivity of scalar multiplication with respect to field addition

A line that passes through the zero vector (i.e. origin) is a subspace of a space R^2.

In general, a column space resulting from M columns in R^n is a "flat" $M-$dimensional plane in R^n .

6 Column space and nullspace

1. A plane that passes through the zero vector (i.e. origin) is a subspace of a space R^3.
2. You can solve $Ax = b$ only if b is in the column space of A.
3. Null space is the space of all solutions x to $Ax = 0$.

7 Solving $Ax = 0$: pivot variables, special solutions

1. Rank of a matrix is the number of pivots. It tells us how many independent equations there are. If the number of columns is n and the rank is r, the number of free variables is $n - r$. It tells the dimensionality of solution (line, plane etc.).
2. Echelon form is the "staircase" form (similar to upper-triangular U).
3. Reduced row echelon form ensures that there is a zero above each pivot.

8 Solving $Ax = b$: row reduced form R

NOTE that in this course m means the number of rows and n means the number of rows (opposite to Data Science notation).

For $Ax = b$, the solvability condition on b states: $Ax = b$ is solvable if b is in a column space of A, i.e. $b = C(A)$. Another way to express that is that if there is a combination of rows of A that gives a zero row, then the same combination of entries of b must give zero.

The complete solution to $Ax = b$ is a sum of x_p (a particular solution) and a nullspace of x. The particular solution can be found by setting all free variables to zero (a free variable is a variable in the column that does not have a pivot) and solving $Ax = b$ for pivot variables. Proof:

$Ax_p = b$ and $Ax_n = 0$, where x_n is a nullspace of x. Then $A(x_p + x_n) = b$.

Given a matrix A of size $m \times n$, the full column rank means $r = n$. This means that there are no free variables - the nullspace is only the zero vector. Therefore the solution to $Ax = b$ is just the x_p - there is a unique solution, if it exists (there are either 0 or 1 solutions).

The full row rank means $r = m$. In this case we can solve the equation for every b (not necessarily a unique one). We are left with $n - r$ free variables.

If for a matrix A we have $r = m = n$, then it is called a full rank matrix. This matrix is invertible. The nullspace for this matrix is zero vector only. In case $r < m$ and $r < n$ there are either no solutions or infinitely many.

9 Independence, basis, and dimension

If matrix A consists of columns $v_1, v_2, ..., v_n$, the vectors are independent if the nullspace of A is only a zero vector. If there exists a combination c (of vectors) such that $Ac = 0$, then the nullspace is not only a zero vector (i.e. $c = 0$) and the vectors v are dependent.

Basis for a space s is a set of vectors $v_1, v_2, ..., v_n$ that are independent and span the space. For given space, every basis has the same number of vectors. This number is called a dimension. Therefore: $rank(A) = dim(C(A))$, where $C(A)$ is a column space of A.

10 The four fundamental subspaces

The four fundamental subspaces:

1. column space $C(A)$
2. nullspace $N(A)$
3. row space $C(A^T)$
4. nullspace $N(A^T)$

For given matrix of size $m \times n$, the dimension of a row space is the same as the dimension of a column space: $dim(A) = dim(A^T) = r(A)$. Tthe dimension of a nullspace is $dim(N(A)) = n - r$ and $dim(N(A^T)) = m - r$.

Row operations preserve row space, not column space. $N(A^T)$ is called a left null space (cokernel), i.e. vectors x such that $x^T A = 0$.

11 Matrix spaces; rank 1; small world graphs

For two subspaces S and U :

$$dim(S + U) = dim(S) + dim(U) - dim(S \cap U)$$

Calculus note: nth order differential equation will have a dimension of a solution space of n.

Rank 1 matrices can be represented in form $A = uv^T$ where v^T is one of the rows and u is the vector column.

12 Graphs, networks, incidence matrices

In mathematics, an incidence matrix is a matrix that shows the relationship between two classes of objects. If the first class is X and the second is Y, the matrix has one row for each element of X and one column for each element of Y. The entry in row x and column y is 1 if x and y are related (called incident in this context) and 0 if they are not.

In a graph (Euler's formula, Topology):

$$\textit{num loops} = \textit{num edges} + (\textit{num nodes} - 1)$$

13 Quiz 1 review

1. $Ax = b$ can be solved if b is in the column space of A
2. If C is invertible then $N(CD) = N(D)$ (N is the nullspace)
3. The nullspace is perpendicular to the rowspace

13.1 Thinking about matrices as transformations of space (Intuition)

Rank means the number of dimensions in the output after transformation. If we squeeze the $3D$ space to a plane using 3×3 matrix, the rank of a matrix is 2.

The set of all possible outputs of Ax is a column space of A. Since each column can be viewed as a transformation to the unit base vectors of given space, if one of the columns is a multiple of the other, then one of those unit base vectors is moved to the other, thus squeezing the dimension. Rank is therefore a dimension of the column space of A.

The nullspace is the set of vectors that, for given transformation, land at the origin. This means that if we transform a $3D$ space into a plane, there exist some line (set of vectors) that got squeezed to one point - origin. This would be a solution to $Ax = 0$.

3×2 matrix means that two columns (i.e. base vectors i, j) land in a 3D world, spanning a 2D plane in 3D. On the other hand, a 2×3 matrix is a transformation from 3D (3 column vectors) to a 2D space.

Some linear transformation to the number line is equivalent to taking a dot product with some dual vector of that transformation (just transpose the row of that transformation to a column).

Orthonormal transformations are the ones that preserve the dot product (the base vectors are still perpendicular to each other) and still unit length (a.k.a rotations).

Expression such as $A^{-1}MA$ represents some transformation M (as seen from our perspective) and A^{-1} and A are the transformations that get some from their coordinate system to ours and back (change of basis), so this is just some linear transformation done in a different coordinate system.

In Diagonal Matrix all of the basis vectors are eigenvectors and their eigenvalues are the entries in the matrix.

A set of basis vectors which are eigenvectors is an eigenbasis (it's a diagonal matrix!). This is useful if you want to compute the nth power of a matrix, transform it to the space at which the basis vectors form an eigenbasis, take it to the nth power there (you just multiply by the diagonal entries to the nth power) and transform back to the original space.

14 Orthogonal vectors and subspaces

Two vectors are orthogonal to each other if $x^T y = 0$. This can be proved using e.g. Pythagorean Theorem, i.e.

$$||x||^2 + ||y||^2 = ||x + y||^2$$

Subspace S is orthogonal to the subspace T means that every vector in S is orthogonal to every vector in T. Those subspaces cannot intersect in any non-zero vector!

Nullspace and row space are always orthogonal components, i.e. nullspace contains all vectors that are perpendicular to the row space.

A^TA matrix is always square and symmetric (proof using the definition of symmetry, i.e. $A^T = A$).

If $Ax = b$ does not have a solution (e.g. the system of equations is noisy and might not have exact solutions), we try to solve for $A^TAx = A^Tb$. Note that the rank of A^TA is equal to the rank of A and their null spaces are also equal.

A^TA is invertible iff columns of A are independent.

15 Projections onto subspaces

Projecting vector b onto a vector a (in 2D) can be viewed as multiplying b by a projection matrix P. Here $P = aa^T/a^Ta$; where a^Ta is a scalar, but aa^T means multiplying column by a vector, which gives a $n \times n$ matrix, so the result, P is a matrix. The rank of P is 1, because the column space is the line that passes through a. The notation P is used to represent the projection matrix, and p as a projected vector, i.e. $p = A\hat{x}$.

Two facts about projection matrices:

1. P is symmetric.
2. A projection matrix applied multiple times, e.g. P^2 will result in the same matrix, i.e. $P^2 = P$ (idempotent).

Projection is used when we want to solve $Ax = b$ but no exact solution can be found (i.e. b is not in the column space of A). In this case we might find the closest $\hat{x}$ that results by solving not for b but for p, i.e. the projection of b onto the column space of A.

If a vector e is in the nullspace of A^T then we know that e is perpendicular to the column space of A.

In many dimensions the solution for $\hat{x}$ (solution for $A\hat{x} = p$) yields

$$\hat{x} = (A^TA)^{-1}A^Tb$$

This is equivalent to solving for Least Squares.

16 Projection matrices and least squares

1. $A^T A$ is supposed to be symmetric, invertible (if A has independent columns) and positive definite.

2. If $Ax = b$ does not exist and we solve for $A^T A\hat{x} = A^T b$ we call these normal equations.

3. Orthonormal = Orthogonal + Unit vector

17 Orthogonal matrices and Gram-Schmidt

Gram-Schmidt is used to make an orthonormal basis out of a given matrix A. For an orthonormal basis $Q : Q^T Q = I$ (view Q as column vectors and if vectors are orthonormal then their product will give zero and a vector v multiplied by itself will give 1).

The convention is to call a matrix orthogonal matrix if it's a square matrix. Here we know that if $Q^T Q = I$ then $Q^T = Q^{-1}$.

Hadamard matrix is a square matrix of -1 and 1s, who's columns (and rows as well) are mutually orthogonal.

Since $Q^T Q = I$, substituting $Q = A$ in $A^T A\hat{x} = A^T b$ we get $\hat{x} = Q^T b$.

In Gram-Schmidt you iteratively, at iteration i, for given vector, remove from it the parts that are projections onto the set of orthogonal vectors from iterations $i - 1$.

The complexity of Gram-Schmidt is $O(nk^2)$ where n is the dimensionality of the vector and k is the number of vectors.

The Gram-Schmidt results in $A = QR$ decomposition, where R is upper triangular matrix.

18 Properties of determinants

Determinant is a number associated with every square matrix.

18.1 Properties of a determinant

1. $det(I) = 1$
2. Exchanging two rows changes the sign of the determinant
3. If you multiply one row of matrix A by t, resulting in a matrix A^T: $det(A^T) = t \times det(A)$
4. Determinant is linear in one row only, i.e. $det(A+B) \neq det(A)+det(B)$, but if we add some row r to one of the rows in our matrix A then the determinant of the resulting matrix would be a sum of determinants of the previous matrix without r and with r alone instead of the original one. But this is only the case if we alter one row only.

As a result:

1. If two rows are equal then $detA = 0$
2. If we subtract a row from any other row, the determinant does not change
3. Row of zeros $\rightarrow$ determinant is zero
4. For a triangle matrix U the determinant is the product of the elements on its diagonal (pivots). That's because we can eliminate all of the off-diagonal elements using elimination without changing the determinant. It's a product because for each row we might factor out the scalar, leaving 1 in the row.
5. $det(AB) = det(A)det(B)$. That's why $det(A^{-1}) = 1/det(A)$
6. $det(kA) = k^n \times det(A)$
7. $det(A^T) = det(A)$

19 Determinant formulas and cofactors

When we calculate determinants of bigger matrices (of size n), we recursively split calculations into n determinants of matrices of size $n-1$. Those smaller determinants are called cofactors (or actually, those determinants with appropriate sign). Without the sign we call them minors.

20 Cramer's rule, inverse matrix, and volume

The general equation for an inverse:

$$A^{-1} = 1/det(A) * C^T$$

Where C^T is the transpose of a matrix of cofactors C.

Cramer's rule says that for $Ax = b$ the solutions $x_1, ..., x_n$ are just ratios $x_1 = B_1/detA$ where B_1 is the matrix A with column 1 exchanged with b. That's because:

$$x = A^{-1}b = 1/det(A) * C^T b$$

and $C^T b$ is just a vector of determinants of $B_1, ..., B_n$, because we can compute determinants of B matrices with respect to column b getting cofactors C.

The complexity of Cramer's rule is $O(n!n)$ using naive determinant computation implementation and $O(n^4)$ using LU decomposition for determinants, where we compute solutions for n unknowns and each LU decomposition is $O(n^3)$.

21 Eigenvalues and eigenvectors

1. Eigenvector is a vector v such that Av is parallel to v
2. A $n \times n$ matrix will have n eigenvalues. Sum of all of those eigenvalues is equal to the trace of the matrix $Tr(A)$
3. $det(A - \lambda I) = 0$ is called a characteristic equation of a matrix A
4. If, for a matrix A you add a multiple of I , e.g. nI, its eigenvalues will be n bigger and its eigenvectors won't change.

5. Product of eigenvalues is equal to the determinant
6. Real matrices can have complex eigenvalues, e.g. rotation matrix rotating by 90°. Those eigenvalues are complex conjugates
7. Symmetric matrices have only real eigenvalues
8. In triangular matrices, the eigenvalues are the entries on the diagonal
9. If there are repeated eigenvalues then the matrix will have less independent eigenvectors, e.g. for 2×2 matrix with one eigenvalue (two repeated) there is just one independent eigenvector (one line).

22 Diagonalization and powers of A

Eigenvectors are used to diagonalise a matrix A , i.e:

$$AS = S\Lambda \Rightarrow S^{-1}AS = \Lambda$$

(if S is invertible), were Λ is a diagonal eigenvalue matrix, and S is the eigenvector matrix.

1. Powers of A have the same eigenvectors, and their eigenvalues are powers of eigenvalues of A.
2. A is diagonalisable if all λ are different. But if some λ are repeated it does not mean that the eigenvectors are not independent!

23 Differential equations and e^{At}

If a system of equations (or system of differential equations) is singular, then one of the eigenvalues it zero. The value of zero will indicate a steady state of the system. The reason why eigenvectors and eigenvalues give solutions to that problem is because they "decouple" the unknowns by diagonalizing matrix A.

1. The system of differential equations is stable, i.e. $u(t) \to 0$ if all eigenvalues are negative (or complex but with real part negative). That's because for a complex number, the length of e.g. $|e^{6it}| = 1$ is one (by Euler's formula).

2. System has a steady state when $\lambda_1 = 0$ and other eigenvalues are negative.

3. System is unstable (i.e. blows up) if at least one eigenvalue is positive or complex with $Re(\lambda) > 0$. The graphical interpretation for stability of differential equations (exponentials) is that the eigenvalues for a system of a matrix are on the left half-plane. On the other hand, in order for the powers of a matrix to $\to 0$, the region of stability is a unit circle, i.e. $|\lambda| < 1$

If a trace of the matrix is positive then the system is unstable. Matrix exponential, e.g. e^{At} (where A is a matrix) should be interpreted as a power series (i.e. Taylor series here) with powers of A. The convergence of different matrix series will depend on the values of eigenvalues, e.g. if any of them is positive, the series will not converge.

The graphical interpretation is that the stability of the systems is when the eigenvalues for a system of a matrix are on the left half-plane.

The nth order differential equation is an equation with n as a highest derivative. Using matrices you can transform nth order differential equation into $n \times n$ 1st order equation. This is because every ODE (Ordinary Differential Equation) can be represented as a vector of derivatives and transformed to a first order derivative of this vector.

24 Markov matrices; Fourier series

Markov matrix has two properties:

1. Every entry ≥ 0

2. Sum of elements of a column is 1

The second property guarantees that the $\lambda = 1$. All other eigenvalues have a magnitude smaller than 1. The steady state then is dictated by the eigenvector with $\lambda = 1$. The rest will disappear with time.

Fourier series is a series

$$1 + a * sin(x) + b * cos(x) + c * sin(2x) + d * cos(2x) + ...$$

which is the same thing as an inner product of infinite dimensional vectors of coefficients and trigonometric functions. The point is that the latter vector is orthogonal, because we can prove that the inner product of each of the functions, e.g. $sinx$ and $cosx$ are orthogonal (integrating their product in the interval $[0, 2\pi]$).

25 Symmetric matrices and positive definiteness

Properties of symmetric matrices:

1. Eigenvalues are Real
2. Eigenvectors are perpendicular

In usual case you could decompose any invertible matrix A into $A = S\Lambda S^{-1}$ (diagonal factorisation). For symmetric matrices, since eigenvectors are perpendicular, you can normali'e the columns of S and get $A = Q\Lambda Q^{-1}$ and since $Q^{-1} = Q^T$ we get $A = Q\Lambda Q^T$. This is called a Spectral Theorem. "Spectrum" is a set of eigenvalues of a matrix. In Mechanics it's called a Principal Axis Theorem. Spectral theorem would work for complex matrix A if it was equal to its own conjugate transpose. It's called a Hermitian matrix (or self-adjoint matrix). Those matrices also have real eigenvalues and perpendicular eigenvectors.

Every symmetric matrix is a combination of perpendicular projection matrices.

If we reduced a symmetric matrix using Gaussian Elimination, the signs of pivots would be the same as the signs of eigenvalues.

Positive definite matrix are symmetric with all eigenvalues positive.

26 Complex matrices; Fast Fourier Transform

The length of complex vector is not simply $z^T z$, because we do not get a positive number there. We need to take a conjugate transpose, instead of a simple transpose. The shortcut for that is an uppercase letter H, that stands for Hermitian. Same for inner product, instead of $y^T x$ we do $y^H x$.

The Unitary Matrix means an orthogonal matrix (orthonormal columns) but in complex space.

Fourier Matrix is a square symmetric matrix, where $F_{ij} = w^{ij}$, where $i, j = 0, ..., n-1$ and $w^n = 1$ (w are the roots of unity). Here, $F^H F = I$ if we made columns of F orthonormal (normalize columns since they are already orthogonal). The property behind the Fourier Matrix is that if you square the matrix e.g. 64×64 you get the Fourier Matrix 32×32, because w is the root of unity of $2\times$ smaller angle in 64×64. Thanks to that, the Fast Fourier Transform multiplies two matrices in $O(nlogn)$ time instead of performing $O(n^2)$ multiplications.

27 Positive definite matrices and minima

In Positive semi-definite the eigenvalues are positive or zero. So the determinant can be zero. A quadratic form is a function $x^T Ax$ which will be a quadratic function of $x_1, ...x_n$. For a positive definite matrix, this function will be always positive for any value of $x_1, ...x_n$, i.e. $x^T Ax > 0$. The graph of $f(x) = x^T Ax$ will depend whether it is positive definite or not: it can be an elliptic paraboloid (positive definite) or hyperbolic paraboloid (a.k.a saddle point). The cross section of elliptic paraboloid is an ellipse and a cross section of hyperbolic paraboloid is a hyperbola.

A minimum of a function with n variables has a minimum if the matrix of second derivatives is positive definite.

27.1 Geometric interpretation of the quadratic form

You can think of A as a function that changes one vector x in the space to a new vector Ax in the space. The quadratic form $x^T Ax$ is the inner product of the old vector x with the new vector Ax. A positive definite matrix guarantees that this inner product is positive. But the inner product is the product of the lengths of the two vectors, x and Ax , and the cosine of the angle between them. Ignoring their lengths, the fact that the quadratic form is always positive says that the cosine of the angle between them is always positive, which means that the angle is between $-\pi/2$ and $\pi/2$. Thus, whatever A does to x, the new vector can't be more than 90 degrees away from the old vector. This is a form of stability, not in the length of the new vector (the eigenvalues of A could be large, so vectors could grow very large when A is applied to them), but in the direction of the new vector.

28 Similar matrices and jordan form

For any rectangular matrix A (with all independent columns), $A^T A$ will always be positive semidefinite.

28.1 Similar matrices

Two square matrices A and B are similar if for some $M : B = M^{-1}AM$. Similar matrices have the same eigenvalues, but not the same eigenvectors!

Jordan's theorem: Every square matrix A is similar to a Jordan matrix J (where a matrix J is made of Jordan blocks that have have one eigenvalue each). Those Jordan blocks lie on the diagonal (kind of a generalization of a diagonal matrix)

29 Singular value decomposition

SVD is a factorization of any matrix $A : A = U\Sigma V^T$, where U, V are orthogonal and Σ is diagonal (but does not need to be square, so below diagonal entries there might be zeros). For symmetric, positive definite matrix SVD is equivalent to Spectral Decomposition (or Eigendecomposition).

The derivation: The goal is to find an orthonormal basis in the rowspace of A, say V and an orthonormal basis of a column space of A, say U, so that the matrix A gets diagonalized, i.e. $AV = U\Sigma$.

In order to obtain U, V:

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

where V are the eigenvectors of A^TA whereas

$$AA^T = U\Sigma\Sigma^T U^T$$

with U as eigenvectors of AA^T. The elements on the diagonal of $\Sigma\Sigma^T$ are the eigenvalues (or squares of singular values). A nice property is that for any A, B the eigenvalues of AB and BA are the same.

30 Linear transformations and their matrices

Rules of linear transformation T:

1. $T(v + w) = T(v) + T(w)$
2. $T(cv) = cT(v)$

Every linear transformation is associated with some matrix A. Derivative is an example of a linear transformation. That is why it is easy to compute derivatives - because we only need to know the derivatives of the basis: $1, x, ...x^n, sin(x)$ etc.

31 Change of basis; image compression

Each image is a vector of $x \in R^{n\times n}$. Compression of images is about figuring out a proper basis (e.g. Fourier basis, Wavelets basis), setting a threshold and removing all of the vectors for which the coefficients are smaller than the threshold value. Then we encode the original image using this subset of basis vectors.

In video compression, each frame is highly correlated with another so we want to encode the difference between the frames, not the whole frames each time.

The original image, vector of pixels P can be denoted as $P = Wc$ where W is a basis matrix. Good basis for compression needs to have a fast inverse, because the vector of coefficients requires inverting W. There exists a Fast Fourier Transform and a Fast Wavelet Transform. Wavelet basis is orthogonal so its inverse is the transform, so its super quick. Also, a good basis should be such that few basis vectors should reproduce the image.

31.1 Change of basis

Given vector x in old basis, and columns of W as new basis vectors, the vector c in new basis is $x = Wc$.

32 Quiz 3 review

1. Anti-symmetric (or skew-symmetric) matrices are always orthogonal. For those matrices: $A^T = -A$.
2. The condition for the orthogonal eigenvectors is $AA^T = A^T A$. This is fulfilled by symmetric, antisymmetric and orthogonal matrices.
3. For Markov matrix, one eigenvalue is always 1, and the other eigenvalues are smaller.
4. Orthogonal matrices have eigenvalues of length 1, because orthogonal matrices do not change lengths (e.g. rotation).

33 Left and right inverses; pseudoinverse

Note m is a number of rows and n is a number of columns. Also recall that Ax will always be in a column space (its a combination of columns).

For a matrix $n \times n$ to be invertible, its rank must be $r = n = m$.

For a left inverse, the matrix must have a full column rank, i.e. $r = n$ (all columns are independent). Here we have either one or more redundant (or contradictory) rows. Since the nullspace is zero, there is no solution to $Ax = b$ or there exists exactly one. Here there exists a left inverse: $(A^T A)^{-1} A^T$ because $(A^T A)^{-1} A^T A = I$.It's dimensions are $n \times m$.

In a right inverse, the matrix has full row rank, i.e. $r = m < n$. There exists a right inverse: $A^T (AA^T)^{-1}$.

If we put a left (right) inverse on the right (left) side instead, that would give us a projection matrix onto the column (row) space of A. It's those nullspaces that are screwing up inverses, cause if a matrix takes a vector to zero, there is no way the inverse can bring it back to life.

In general case, when a matrix has a rank $r < m$ and $r < n$, the matrix maps row space to the column space but kills all the vectors in the nullspace of A. The pseudoinverse A^+ is the mapping back from the column space of A to back to the row space of A.

In order to find a Pseudoinverse:

1. Start from SVD: $A = U\Sigma V^T$. The pseudoinverse of Σ is Σ^+ is just inverted singular values on the diagonal (non-zero singular values). Here $\Sigma\Sigma^+$ and $\Sigma^+\Sigma$ will have 1's on the diagonal with some zeros on the end of the diagonal.

2. Invert the RHS of the SVD: $A^+ = V\Sigma^+U^T$ (since V, U are orthogonal so their transpose is their inverse). It is referred to as Moore–Penrose inverse.